Human estimates of statistical relatedness¹

DWIGHT E. ERLICK

AEROSPACE MEDICAL RESEARCH LABORATORIES, WRIGHT-PATTERSON AIR FORCE BASE

Degree of statistical relatedness between events is an independent variable used by many psychologists investigating concept formation, cue utilization, reinforcement theory and decision theory. Evidence is presented which indicates there is a discrepancy between human estimates and statistical estimates of relatedness.

Knowledge of relatedness among events is of interest to the decision maker attempting to ascertain sources of influence and predict developing trends (Schum, 1965). Manipulation of the degree of relatedness among events is used by psychologists to influence a variety of learning situations such as the formation of concepts, utilization of cues, and the extinction or reinforcement of responses (Byers, 1964; Skinner, 1953). For the philosopher, event relationships are a fundamental aspect in the development of causality (Hume, 1898). The ability of humans to determine how much relatedness actually exists among events has been shown to vary as a function of the spatial and temporal relationships of events (Hammond & Summers, 1965; Carrol, 1963; Erlick, 1961) as well as the expectations of the observer (Michotte, 1963). Studies indicate that man finds it difficult to detect small, but real contingencies in the environment (Smedslund, 1955). On the other hand, there are many instances which demonstrate a readiness by man to exaggerate the importance of certain types of small but chance contingencies and thereby burden himself with superstitious behavior (Frazer, 1959).

If the relationship of two events, each having N different states, is periodically sampled, and we wish to assess the overall degree of relatedness between them, we must in some manner evaluate the occurrence of each of the different combinations of the event states. If $N=2$ (e.g., +, -) for each event, then four different possible combinations can occur. The usual statistical measures of relatedness for $N=2$ (e.g., phi coefficient and tetrachoric r) are a function of comparing the products of occurrences of the similar (++--) to dissimilar (-+ +-) conditions. When human observers perform this task of estimating relatedness between two events ($N=2$), their subjective estimates show little relation to the statistical measures of correlation (Smedslund, 1963; Jenkins & Ward, 1965), but relate significantly only to the frequency of ++ occurrences. This is supported by findings (Wason, 1960) that human observers generally do not search for negative evidence in simple conceptual tasks.

If events have more than two states ($N > 2$), (the more common situation for which humans make estimates) the question arises as to how man goes about integrating the various combinations of these events. Since several

different degrees of agreement (positiveness) and disagreement (negativeness) can occur, an overall assimilation of a series of these combinations would require some type of weighted-frequency analysis. A common statistical correlation method (Pearsonian r) used to measure relatedness takes each of the coordinates for each combination of the events and squares the deviation of the coordinate from a line of regression based on the total series of coordinates being related. This procedure would seem to be quite unmanageable for humans making estimates of relatedness from an extended series of observations. However, man does make subjective estimates of relatedness between events which have several states of occurrence. The present study was an attempt to determine what methods he uses to derive these estimates. In this investigation, five different states of occurrence ($N=5$) for each of two events were displayed to observers over an extended period of time. Correlation (Pearsonian r) was used to statistically quantify the degree of relationship presented to the observers. Results of subjective estimates plotted against this statistical measure are shown in Fig. 1. The apparatus used was two semi-circular dials (x and y) each having five easily discriminable equidistant indices. Observers were presented a random sequence of 20 x, y values such that the overall series resulted in one of 21 dif-

Fig. 1. Mean estimated correlation as a function of statistical correlations (Pearsonian r) presented. Each point represents 250 estimates based on 10 observers.

ferent r 's ranging from -1.00 to $+1.00$ in increments of $.10$. Each r was represented in two different ways: *small variability*—in which most of the x, y coordinates were close to the theoretical line of regression (TLR), and *large variability*—in which most x, y coordinates were on the TLR and a few deviated quite markedly.

If the method of deriving subjective estimates of r were based on the statistical methods used for estimating r , i.e., squared deviations from the TLR, there should be no difference between estimates of r for the two variability conditions. However, if x, y discrepancies were weighted simply in terms of their actual discrepancies, large variability conditions would be overestimated toward positive relationships, because for the same r , the sum of a few large discrepancies would be less than that of many small discrepancies (if all discrepancies are zero, then $r = +1.00$).

Ten Ss, naive in terms of statistical assumptions and procedures, served as observers. They estimated the degree of positive or negative relatedness on a scale from -10 through $+10$ with 0 indicating no relationship. Thus, a 21-point scale was used for recording responses. Twenty-five estimates were made by each observer for each of the small and large variability conditions.

An analysis of variance on the mean estimation error showed a significantly greater error for negative than for positive r relationships. Also, there was a significant bias toward the use of positive responses. Estimates of large variability r 's showed a positive r bias relative to small variability r 's (see Fig. 1).

Subjective estimates were compared to several other statistical measures of relatedness, e.g., contingent uncertainty (Garner, 1962), eta, phi and coefficient of contingency. None of these measures proved any more effective than r in reducing the discrepancy between subjective estimates of small and large variation r 's.

However, if the different combinations of event-states are weighted in terms of how closely there is agreement between the two events at each occurrence and if we then multiply these weights by their frequency in the overall series, the resultant sums of this weighted-frequency model predicts both the subjective estimates of small and large variability r 's along the same general line. Five different weights were used based on the five

different degrees of discrepancy possible for the two events. The weights used increased in an approximately linear fashion as a function of the degree of discrepancy present between the event states.

It thus appears that subjective estimates of relatedness between two events each having several different states of occurrence, closely approximates a summing or averaging of the discrepancies between the events over time. This results in subjective estimates varying from statistical estimates (r). The degree of difference between the two appears to be a function of the variability that occurs between the states of the two events—large variability producing more bias toward positive estimates of relatedness.

References

- Byers, J. L. Rate of information acquisition in concept attainment. *Psychol. Rep.*, 1964, 15, 111-117.
- Carroll, J. D. Functional learning: The learning of continuous functional mappings relating stimulus and response continua. *Educ. Testing Service Res. Bull.*, RB-63-26, Sept. 1963.
- Erlick, D. E. Judgments of the relative frequency of a sequential series of two events. *J. exp. Psychol.*, 1961, 62, 105-112.
- Frazer, J. G. *The new golden bough*. New York: Criterion Books, 1959.
- Garner, W. R. *Uncertainty and structure as psychological concepts*. New York: Wiley, 1962.
- Hammond, K. R., & Summers, D. A. Cognitive dependence on linear and nonlinear cues. *Psychol. Rev.*, 1965, 72, 215-224.
- Hume, D. *Essays moral, political and literary*. London: Longmans & Green, 1898.
- Jenkins, H. M., & Ward, W. C. Judgment of contingency between responses and outcomes. *Psychol. Monogr.*, 1965, 79, 1-17.
- Michotte, A. *The perception of causality*. New York: Basic Books, 1963.
- Schum, D. A. Inferences on the basis of conditionally nonindependent data. *USAF AMRL TDR No.* 65-161, 1965.
- Skinner, B. F. *Science and human behavior*. New York: Macmillan, 1953.
- Smedslund, J. The concept of correlation in adults. *Scan. J. Psychol.*, 1963, 4, 165-173.
- Smedslund, J. *Multiple probability learning*. Oslo: Akademisk Forlag, 1955.
- Wason, P. C. On the failure to eliminate hypotheses in a conceptual task. *Quart. J. exp. Psychol.*, 1960, 12, 129-140.

Note

1. The research reported in this paper was sponsored by the Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio. This paper has been identified by Aerospace Medical Research Laboratories as AMRL-TR-66-109. Further reproduction is authorized to satisfy needs of the United States Government.