Do We Know a Vector From a Scalar?
Why Measures of Association (Not Their Squares) Are Appropriate Indices of Effect

MICHAEL J. BEATTY
University of Missouri-St. Louis

"Variance accounted for"—calculated by squaring one of the various measures of association—is the most common estimate of experimental effect or strength of association reported in communication studies. However, methodologists in other social science disciplines have made compelling cases that the statistic itself, not its square, is the appropriate index of shared variation. The basic principles and arguments for interpreting unsquared measures of effect are presented and the implications for the practice of communication theory and research are discussed.

The importance of measures of association and experimental effect relative to significance testing was underscored by Cohen's (1990) observation that "the primary product of research inquiry is one or more measures of effect size, not *p* values" (p. 1310). Certainly, theoretical advancement is enhanced when statistical significance is de-emphasized and accuracy of predications as evidenced by obtained measures of effect is relied upon as the primary criterion for evaluating a theoretical model. "Variance explained" calculated by squaring one of the various measures of association (e.g., *r*, *eta*, *omega*, *phi*) is the most common estimate of experimental effect or magnitude of association reported in the communication literature. In fact, estimates of such effects are required for quantitative studies by the editorial policies of major journals such as *Communication Monographs* and *Human Communication Research*. For some time now, however, methodologists in other social science disciplines have recognized that the statistic itself (e.g., *r*) and not its square is the appropriate index of variation explained (D'Andrade & Dart, 1990; Hunter & Schmidt, 1990; Jensen, 1971; Ozer, 1985; Rosenthal & Rubin, 1982; Tryon, 1929). Ozer (1985) demonstrated that "squaring the correlation coefficient to yield a percent of variance accounted for is often unjustified by an underlying model. . . . By contrast, the correlation is a directly interpretable effect size

Michael J. Beatty (Ph.D., Ohio State University, 1976) is Professor and Chair, Department of Communication, University of Missouri-St. Louis, St. Louis, MO 63121-4499

Human Communication Research, Vol. 28 No. 4, October 2002 605–611

indicator and may in circumscribed conditions be used as a coefficient of determination" (p. 134). Over a decade ago, in his treatment of effect sizes, Cohen (1988, p. 533) alerted social scientists to Ozer's essay.[1]

A review of the works just cited reveals agreement that r and not r^2, for example, is the appropriate estimate of shared variance when correlations between two variables are spurious, which is to say due to a common underlying latent variable. Ozer (1985) drew an illustration from personality research, noting that "most trait models suggest that some latent variable underlies scores on both measures; and that the latent variable is responsible for the covariance between the measured variables. In this model, the correlation coefficient itself provides the appropriate coefficient of determination" (p. 312). Likewise, correlations between twins' scores are not squared in formulae designed to estimate genetic inheritance of traits (Falconer, 1989; Jensen, 1971): Common genetic structures are responsible for the covariance between twins' scores on the measured variables.

A decision to interpret unsquared measures of association as indices of shared variance in spurious correlation scenarios alone would have dramatic implications for reinterpreting much of the communication research literature. However, D'Andrade and Dart (1990) have made a strong case that unsquared measures of association are the appropriate estimates of effect or association in all cases. Despite the lack of published counterarguments, we continue to square measures of association to determine effects in communication studies. In most, but not all cases, this practice yields gross underestimates of obtained effects. This brief essay is focused on the basic principles and arguments underlying the use of unsquared measures of association, and the implications of this practice for communication theory and research.

BASIC PRINCIPLES AND ARGUMENTS

D'Andrade and Dart (1990) point out that measures of association have a direct meaning with respect to the data whereas their squares have at best an abstract meaning. As an illustration, consider a two-by-two contingency table in which 20 couples were separated into four groups based on median splits of communication competence and relational commitment scores. Suppose 9 couples were high in both competence and commitment, 1 couple was high in competence but low in commitment, 1 couple was low in competence but high in commitment, and 9 couples were low in both competence and commitment. The phi coefficient for those data is .80 (phi-square = .64). If the researchers expected a positive relationship between competence and relational commitment, two misclassified cases (or 10% of the sample) are observed (i.e., the low competence/high commitment and the high competence/low commitment couples). D'Andrade and Dart's preference for phi over phi-square in the two-by-two case resides in the fact that the proportion of misclassified cases can be calculated from phi (Levy, 1967) by solving the simple equation, $(1 - r) / 2$. Applied to the hypothetical example, $(1 - .80) / 2$ indicates that 10% of the sample or 2 couples were misclassified, which of course corresponds exactly to the frequencies reported above. To emphasize D'Andrade and Dart's point, compared to the abstract, albeit familiar, construction "communication competence explained 64% of the variance in relational satisfaction," the phi coefficient (.80) more clearly represents the pattern of the data and more accurately describes the magnitude of the relationship.

D'Andrade and Dart (1990) demonstrated that unsquared statistical estimates of effects or association also have direct and meaningful interpretations when data are interval. Correlation coefficients provide an "accurate estimate of the proportion of variation (not variance) accounted for" (D'Andrade & Dart, p. 50). Although the terms "variation" and "variance" are often used interchangeably, variation refers to the *absolute deviations* from the mean whereas variance refers to *squared deviations* about the mean. D'Andrade and Dart contended that our goal is to account for variation, not variance, in human behavior. As they put it:

> Variance is a square measure, like acres or square miles. It tells one the expected value of the squared differences between any score in a group and the group mean. When dealing with variance round the mean it is normal practice to use the square root of variance, or the standard deviation, to talk about how "far" some score is from the mean. One does not say that some score is 9 "variances" away from the mean. Why not? Because it is like saying that somebody is 9 square miles from San Diego. (p. 53)

Accordingly, "the square root of the variance is taken in calculating the standard deviation in order to return to the same dimensionality as the original measurements, thereby making that measure directly comparable to the original measurements" (D'Andrade & Dart, p. 53). D'Andrade and Dart's position was that "just as in the case of variance versus standard deviation, in the case of r^2 versus r, it is r which corresponds most directly to our intuitive notion of strength of association" (p. 53).

D'Andrade and Dart (1990) attributed our loyalty to the concept of variance and our reluctance to embrace variation to two factors. First, variance has the "nice quality" of being additive. However, D'Andrade and Dart reminded us that "while it is true that variances combine additivity . . . *this does not mean that variance is a good measure of the strength of* *relationship between X and Y*" (p. 54). Second, most social science researchers do not understand "that correlational and ANOVA operations on variables are operations on vectors, not on scalars" (D'Andrade & Dart, p. 54). D'Andrade and Dart's description of variables as vectors is perhaps their most compelling observation.

D'Andrade and Dart (1990) pointed out that "A variable consists of a vector of scores . . . represented geometrically as a directional line through *n* space, where *n* equals the number of scores" (p. 55). The length of such a vector equals the standard deviations of the individual scores. Assuming the correlation between variables is zero, the Pythagorean theorem can be used to compute the length of the vector that represents the sum of the two variables (i.e., the length of that vector is equal to the square root of the sum of the squared lengths of the individual vectors). Thus, the sum of the two variables equals the standard deviation of the summed scores. The relevance to the correlation coefficient is that if (a) the "first variable" is the explained mean deviation (i.e., $\hat{Y} - \overline{Y}$ scores), and (b) the "second variable" is unexplained mean deviation (i.e., $Y - \hat{Y}$ scores), the sum of the two variables, computed by applying the Pythagorean theorem, equals the total mean deviation (i.e., $Y - \overline{Y}$ scores). D'Andrade and Dart concluded that

> the issue is what number best represents the $\hat{Y} - \overline{Y}$ scores—the scores which make up "explained deviation." The argument here is that vector length has a direct interpretation in terms of "forces" or "cause," or "explanatory power," often used in physics, while "variance" has no known direct interpretation. (p. 55)

Furthermore, it can be shown that from any data set the ratio $\Sigma(\hat{Y} - \overline{Y}) / \Sigma(Y - \overline{Y})$, which D'Andrade and Dart use to calculate shared variation, is equal to the correlation coefficient.[2]

IMPLICATIONS OF UNSQUARED ESTIMATES FOR COMMUNICATION THEORY AND RESEARCH

Although this essay is far from comprehensive, it should be sufficient to stimulate reflection regarding the calculation of effect sizes in communication research. Accurate estimation of effects is fundamental to the theory-building enterprise because observed effects or associations between variables are directly related to the predictive power of the models derived from theories. Over a decade ago, Cohen (1990) reminded us that "error variance in our observations should challenge us to efforts to reduce it and not simply to thoughtlessly tuck it into the denominator of an

F and *t* test" (p. 1310). Certainly, factors such as the sensitivity of measurement and the strength of experimental inductions must be considered when evaluating the size of an effect (Cohen, 1988). However, the calculation of effect sizes is also dependent on several factors. As Ozer (1985) pointed out, "it is not only the interpretation of an effect size that is partially theory-dependent; the actual estimate of a coefficient of determination requires theory as well" (p. 313). A common theme across the works cited in this essay is that, in general, squaring correlations and other measures of association has been unjustified in the social sciences. These incorrect calculations produced gross underestimates of effects, followed by misleading evaluations of research and theory.

Take, for example, Mischel's (1968) critique of the personality literature which Ozer (1985) as well as D'Andrade and Dart (1990) point to as a prime example of the consequences of incorrectly estimating the magnitude of association, even in the simple bivariate case. Mischel concluded that personality had only a negligible effect on behavior because most correlations between personality measures and behavioral measures were approximately .30, yielding an r^2 of less than 10%. D'Andrade and Dart noted, however, that

> using r rather than r^2 as an estimate of the amount of variation accounted for leads to a different conclusion; namely, that some personality measures predict behavioral measures *moderately* well, often making it possible to select cases 30% as effectively as direct selection on the behavioral measures themselves, and accounting for approximately 30% of the variation in such behavioral measures. (1990, p. 56)

Although articles explaining the inappropriateness of squaring correlations when determining effects in latent variable scenarios (such as those posited by trait models) had already appeared in mainstream psychology journals (e.g., *Psychological Bulletin*, *Psychological Review*), many communication theorists bought into Mischel's "cross-situational consistency" folly, and railed against trait conceptualizations of communication. Had Mischel known a vector from a scalar, the cross-situational argument would never have been hatched, and had communication theorists appreciated the difference, the cross-situational consistency issue probably would not have appeared in our scholarly journals.

Overall, it appears that we have underestimated the predictive power of our communication theories. Although social scientists, including communication scholars, often accept the premise that models of human behavior will never attain the precision of models in the physical sciences, it may be that the differences in predictive accuracy have more to do with our inadequate determination of effect sizes than the subject matter we study. After all, physicists, as D'Andrade and Dart (1990) seem to recognize, would never square their obtained effects. Our inaccurate calculation of effects has stimulated considerable effort to convince ourselves that small effects can actually be important. Certainly, cases exist in which small effects have practical implications but had r rather than r^2 been the interpretive standard, many of the obtained effects would never have been considered small in the first place. It seems clear that much of the unexplained variation in human behavior—which opponents of the scientific study of behavior point to as indices of the randomness, inherent unpredictability, and the indefinable complexity of humans—is actually due to a self-inflicted calculation error by social scientists. Ironically, the situation indicates a need for more rather than less emphasis on quantitative methods, especially among our theorists. Although it is true that substantial correspondence between Y and $\hat{Y}$ requires coefficients of .80 and larger[3], such predictive precision is more likely if we become more attentive to psychometric properties of measures and characteristics of sample distributions than if we rationalize our small observed effects. Refinement of methodological practices should be the principal priority for communication scholars in the next decade. It would be unfortunate indeed if we settled for low power models of communication or, worse yet, abandoned our commitment to scientific explanation merely because we did not know a vector from a scalar.

NOTES

1. Although Ozer's full thesis and set of proofs is beyond the scope of this essay, interested readers should note that Ozer argued that the determination regarding whether r or r^2 should be interpreted depends on the underlying theoretical model. Ozer concurred with D'Andrade and Dart concerning latent variable cases. However, unlike D'Andrade and Dart, Ozer maintained the r^2 is the appropriate index of variance explained if the underlying model posited X as a causal factor and if X and Y do not share variance in common with an underlying latent variable. While I urge readers to compare both positions for themselves, I found D'Andrade and Dart's evidence for treating all measures of association in their unsquared form the more compelling case.

One complication flowing from the discussion of interpreting correlation coefficients, which none of the writers cited attempted to tackle, concerns decomposition procedures when independent variables in multiple predictor models are correlated. If r rather than r^2 is indicated as the appropriate measure of association between predictor variables, then the overlap between the two correlated predictors equals r not r^2. In cases involving considerable multicollinearity and relatively small multiple correlations, results are likely to be overestimated if r^2 is used to estimate overlap among predictors.

2. D'Andrade and Dart illustrated the parallel between vectors and physical forces by presenting a hypothetical scenario involving a westbound plane traveling at 500 mph which encounters a southerly wind equal to 500 mph. The velocities combine to drive the plane to the northwest. The total velocity the plane will travel to the NW can be calculated by taking

the distance of the hypotenuse between the point of origin and the resultant velocity. D'Andrade and Dart asked that we assume the maximum velocity of both the plane and wind equals 1000 mph. Applying the Pythagorean theorem it is possible to calculate the total velocity that could be imparted on the plane (i.e., total possible velocity equals the square root of the sum of the squared maximum velocities of the plane and the wind or 1414 mph). D'Andrade and Dart made the point that:

> one can see that the independent velocities of 500 miles per hour have imparted exactly *one-half* of the total velocity possible. However, these velocities have accounted for only 25% of the total area through which resultant velocities could occur given various combinations of airplane speeds and wind speeds. Wouldn't it be confusing to say that because independent velocities combine according to the sum of squares, that the strength of the two velocities covers an area that includes only 25% of the total possible combinations of velocities? While correct, such statement would be misleading because, in fact, the plane has attained exactly one-half of its total possible velocity, and it is its velocity, not *area* of possible velocities which it might have attained, which is usually of interest to us. . . . The parallel between physical forces and the addition of variances in statistics is not just an analogy. (1990, p. 55)

3. If the reduction in error, expressed in standard deviation units (Guilford & Fruchter, 1978) is plotted along the y axis and correlation coefficients are arranged along the x axis, a rather severe power curve is observed. The curve hugs the floor of the graph until $r = .80$, at which point it sweeps upward. In other words, predictive accuracy for correlations under .80 is not much better than merely knowing the standard deviation of Y.

REFERENCES

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*, 1304–1312.

D'Andrade, R., & Dart, J. (1990). The interpretation of r versus r^2 or why percent of variance accounted for is a poor measure of size of effect. *Journal of Quantitative Anthropology, 2*, 47–59.

Falconer, D. S. (1989). *Introduction to quantitative genetics*. New York: Wiley.

Guilford, J. P., & Fruchter, B. (1978). *Fundamental statistics in psychology and education* (6th ed.). New York: McGraw-Hill.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

Jensen, A. R. (1971). Note on why genetic correlations are not squared. *Psychological Bulletin, 75*, 223–224.

Levy, P. (1967). Substantive significance of significant differences between two groups. *Psychological Bulletin, 67*, 37–40.

Mischel, W. (1968). *Personality and assessment*. New York: Wiley.

Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin, 97*, 307–315.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effects. *Journal of Educational Psychology, 74*, 166–169.

Tryon, R. C. (1929). The interpretation of the correlation coefficient. *Psychological Review, 36*, 419–445.