Psychological Bulletin
1982, Vol. 91, No. 2, 393-403

0033-2909/82/9102-0393$00.75

Analyzing Sequential Categorical Data on Dyadic Interaction: A Comment on Gottman

Paul D. Allison
Department of Sociology
University of Pennsylvania

Jeffrey K. Liker
Department of Human Development
and Family Studies, Cornell University

Recent theoretical developments emphasize that social interactions are dynamic and reciprocal, and this has led to widespread use of time-series data on behavior in two-person systems. In principle, such data allow one to separate the influences of two actors on each other. Statistical methods currently being used, however, are deficient in several respects. In this article, we show that a statistic proposed by Sackett and later "proved" by Gottman is incorrect. We also show that the failure to control for autodependence can produce misleading results. Finally, we introduce new procedures that are based on both traditional and more recently developed methods for the analysis of contingency tables. Using these procedures, we show how to test for dependency, how to measure dependency, how to test for differences in dependency across subgroups, and how to test for dominance in reciprocal behavior sequences.

Recent theoretical developments have shifted research attention from static, cross-sectional snapshots of social relationships to dynamic, reciprocal models of interaction sequences. For example, early work on childhood socialization viewed parent–child relationships as static and unidirectional; parents were thought to influence but not be influenced by their children. An emerging life-span perspective on child rearing recognizes that parents affect, and are affected by, their children in an evolving, reciprocal relationship (Bell & Harper, 1977; Lerner & Spanier, 1978). Similarly, marital relationships have been viewed primarily in static terms that rely on concepts like relative power, role, and marital satisfaction. John Gottman's pioneering work on *Marital Interaction* (1980b) is the first systematic attempt to measure and characterize patterns of interaction sequences in distressed and nondistressed couples. Gottman developed dynamic concepts to characterize common marital patterns such as positive and negative reciprocity and marital dominance.

This theoretical shift from a static to a dynamic framework has been accompanied by a shift in the kinds of data collected. In particular, there has been increased reliance on sequential observation data (see Burgess & Huston, 1979; Clarke-Stewart, 1973; Gottman, 1980b; Mishler & Waxler, 1966; Patterson, 1980; Schenkein, 1978; Wright, 1967) and on sophisticated methods for coding these data to capture both verbal and nonverbal behavior. Unfortunately, statistical expertise in analyzing these rich, carefully coded interaction sequences has not kept pace with the development of the conceptual apparatus and coding schemes. To analyze his data, Gottman made extensive use of a statistic proposed by Sackett (1979) and later derived rigorously by Gottman (1979) himself. Unfortunately, both Gottman's proof and the original statistic are in error. Even if the statistic were correct, moreover, many of Gottman's applications would still be problematic.

In this article we describe the problems with Gottman's statistical methods, and we propose alternative procedures based on recently developed methods for the analysis of categorical data. These methods can be used to estimate and test the degree of lagged dependence of the behavior of one actor on

Support from National Institute of Mental Health Grant MH34172 (Glen H. Elder, Principal Investigator) is gratefully acknowledged.

Requests for reprints should be sent to Paul D. Allison, Department of Sociology, 3718 Locust Walk CR, University of Pennsylvania, Philadelphia, Pennsylvania 19104.

the behavior of another. Tests are also proposed for determining whether one actor dominates the other. Wherever possible, we illustrate the methods with data reported by Gottman (1980b), and we compare our results with his.

A Test For Lagged Dependence

A basic aim in the analysis of sequential interaction data is to estimate and test for lagged dependence in the behavior of the two actors. Let us suppose that we observe the behavior of a married couple at a sequence of n points in time ($t = 1, \ldots, n$), and we wish to test for the dependence of the husband's behavior at time $t + k$ on the wife's behavior at time t. We define the variables H_t and W_t to indicate the husband's and the wife's behavior, respectively, and, for the moment, we consider only the case in which both variables are dichotomous, taking on values of 1 or 0 at each point in time. Due to the lag of k time units, the data consist of a set of pairs (W_t, H_{t+k}) for $t = 1, \ldots, n - k$. Thus, the last k values of W and the first k values of H are not used in the analysis.

Gottman (1979, 1980a, 1980b) argued that the hypothesis of *no* dependence corresponds to the statement that the conditional probability of the husband's behavior given the wife's behavior is the same as the unconditional probability, that is,

$$Pr[H_{t+k} = 1|W_t = 1] = Pr[H_{t+k} = 1].$$

To test this hypothesis, Gottman used a statistic proposed by Sackett (1979). To discuss this statistic, we need some additional notation. Let $p_{H|W}$ and p_H be the observed proportions corresponding to the conditional and unconditional probabilities in the null hypothesis. That is, $p_{H|W}$ is the proportion of times that the husband's behavior is coded 1 *given* that the wife's behavior k units earlier is coded 1. Let p_W be the proportion of times that the wife's behavior is coded 1.

Sackett (1979) took the usual approach of constructing a z statistic by dividing the difference between the observed proportions $p_{H|W} - p_H$ by the estimated standard error of that difference, under the assumption that there is no autodependence for H_t and W_t (i.e., that H_t and W_t do not depend on their own values at previous points in time). The statistic he obtained is

$$z_s = \frac{p_{H|W} - p_H}{\sqrt{\dfrac{p_H(1 - p_H)}{(n - k)p_W}}},$$

which should have an asymptotically normal distribution with a mean of zero and a variance of one (a standard normal distribution). Unfortunately, the denominator of this statistic is not the correct standard error of the numerator. It would be correct only if p_H were the true probability and not merely an observed proportion subject to sampling error. It can be shown that a correct statistic is given by

$$z_1 = \frac{p_{H|W} - p_H}{\sqrt{\dfrac{p_H(1 - p_H)(1 - p_W)}{(n - k)p_W}}},$$

which does have, asymptotically, a standard normal distribution. In more recent work, Gottman (1980a) himself advocated the use of z_1 rather than z_s.

It is apparent that this statistic is always greater than Sackett's (1979) statistic, specifically,

$$z_1 = \frac{z_s}{\sqrt{1 - p_W}}.$$

The difference can be substantial. In one analysis Gottman (1980b, p. 228) reported $p_H = .70$, $p_{H|W} = .76$, and $z_s = 1.57$, which is not significant at the .05 level (all significance levels mentioned here assume a two-tailed test). Because $p_W = .70$, however, the correct statistic is $z_1 = 2.87$, which is significant at beyond the .01 level.

The correct statistic z_1 also happens to be identically equal to two other statistics that are much more familiar. One is the test of a difference between two proportions estimated from two independent samples (Blalock, 1979, pp. 232–234). Suppose we divide the pairs (W_t, H_{t+k}) into two groups: those in which $W_t = 1$ and those in which $W_t = 0$. We let m_1 be the number of times that $W_t = 1$ and m_0 be the number of times that $W_t = 0$. Within each group we calculate the proportion of times that $H_{t+k} = 1$. These two proportions are denoted by p_{H11} and p_{H10}. The test statistic is then

$$z_1 = \frac{p_{H11} - p_{H10}}{\sqrt{p_H(1 - p_H)\left(\frac{1}{m_1} + \frac{1}{m_0}\right)}}.$$

The equality of this statistic and the one based on conditional and unconditional proportions is proved in the Appendix. Intuitively, the reason they are equal is that they are testing exactly the same null hypothesis. If the conditional probability is equal to the unconditional probability, it necessarily follows that the two conditional probabilities are equal to each other, that is,

$$Pr[H_{t+k} = 1|W_t = 1] = Pr[H_{t+k} = 1|W_t = 0].$$

Conversely, if the two conditional probabilities are equal, each of them must equal the unconditional probability.

A third statistic is obtained by viewing the data as a 2 × 2 contingency table. For example, for Gottman's data described above, it is possible to reconstruct approximately the 2 × 2 table:

	H_{t+1}	
W_t	1	0
1	109	35
0	35	27

A natural way to test for dependence in such a table is to apply the usual Pearson's chi-square statistic after first calculating expected values under the hypothesis of independence. It is well known (e.g., Fienberg, 1980; Kendall & Stuart, 1961) that Pearson's chi-square test for a 2 × 2 table is identically equal to the square of the z test for the difference between two proportions. For these data $X^2 = 7.63$ with a square root of 2.76 (the difference between this and the previously calculated value of $z_1 = 2.87$ is due to rounding error incurred by reconstructing the table from the given proportions). As we shall see, the advantage of the contingency table approach is that it can be easily generalized to more complicated situations.

A Measure of the Strength of Lagged Dependence

We now have appropriate statistics for testing the null hypothesis that there is no relationship between the wife's behavior at time t and the husband's behavior at time $t + k$. Suppose that the null hypothesis is false, that the husband's behavior *does* in some way depend on the wife's past behavior. One then needs some measure of the strength of that dependence in order to make comparisons with other kinds of couples, other experimental treatments, and so forth. In essence, what is needed is a measure of association for a 2 × 2 table. Bishop, Fienberg, & Holland (1975) reviewed the extensive literature on this topic and concluded that a key distinction is between those measures that are sensitive to the marginal totals and those that are not. An example of a measure that is sensitive to the marginal totals is the difference between the conditional and unconditional probabilities

$$Pr[H_{t+k} = 1|W_t = 1] - Pr[H_{t+k} = 1],$$

which was proposed by Gottman (1980b, p. 31) to measure reduction in uncertainty. Clearly, if the marginal probability $Pr[H_{t+k} = 1]$ is very high or very low, the range of this difference is greatly restricted. As a consequence, comparisons such as those between distressed couples with moderate rates of negative behavior and nondistressed couples with relatively low rates of negative behavior may be questionable. The same criticism applies to the difference between the two conditional proportions:

$$Pr[H_{t+k} = 1|W_t = 1] - Pr[H_{t+k} = 1|W_t = 0],$$

which is commonly used under the name "percentage difference" (Davis, 1971).

A measure that is invariant to the marginal totals can be obtained by first taking the logit transformation (Berkson, 1944; Fienberg, 1980) of each of the two conditional probabilities and then taking the difference. The logit transformation of a probability p is just $\text{logit}(p) = \log_e [p/(1 - p)]$. Thus, the measure we shall use here is defined to be

$$\beta = \text{logit}[Pr(H_{t+k} = 1|W_t = 1)] - \text{logit}[Pr(H_{t+k} = 1|W_t = 0)].$$

It has a range of from minus infinity to plus infinity and is zero when the two variables are independent.

The measure β is better known in the statistical literature as the logarithm of the cross-product ratio (or odds ratio). The reason is that if the expected frequencies in a 2×2 table are given by

$$\begin{matrix} F_1 & F_2 \\ F_3 & F_4 \end{matrix},$$

then $\beta = \log_e (F_1F_4)/(F_2F_3)$. Moreover, if f_i are the *observed* frequencies in the 2×2 table, the maximum likelihood estimator of β is just $\log_e (f_1f_4)/(f_2f_3)$. An alternative way of expressing the definition of β is

$$\text{logit}[Pr(H_{t+k} = 1|W_t)] = \alpha + \beta W_t,$$

which may be thought of as a "logit-linear" model. As we shall see, this form is convenient for generalizing to additional variables.[1]

Comparing Dependency Levels Across Groups

We turn now to the problem of comparing the degree of dependency across two or more groups. Consider the contingency table in Table 1, which is reconstructed (approximately) from data in Gottman's (1980b) Table 12.5. The two groups are distressed and nondistressed couples, and the aim is to determine whether the dependence of the wife's behavior on the husband's behavior is stronger for one group than for the other. Gottman's approach is to compute Sackett's z statistic within each group and then compare the statistics across groups. Using the correct z statistic, we get a highly significant 3.26 for distressed couples and a nonsignificant .51 for nondistressed couples, suggesting that dependency is much stronger for the former group.

There are two possible flaws in this procedure. One is that test statistics are highly sensitive to sample size. As a result, differences in sample size could produce major differences in the test statistics for different groups even when the degree of dependence is identical. A second problem is that Gottman has no statistical test for the *difference* between the test statistics. It is well known that one statistic can be highly significant and another nonsignificant without there being a significant difference between the two.

Table 1
Frequency Counts for Husband's Behavior × Wife's Behavior for Distressed and Nondistressed Couples

	Distressed W_{t+1}		Nondistressed W_{t+1}	
H_t	1	0	1	0
1	76	100	80	63
0	79	200	43	39

These difficulties can be avoided by testing the null hypothesis that our measure of association β is the same in each group. When there are only two groups, the easiest procedure is to estimate β in each group, take the difference between these estimates, and divide by the standard error of the difference. This yields a z statistic given by Fienberg (1980, p. 37):

$$z_2 = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{\sum_i \frac{1}{f_i}}}.$$

[1] Those who are unfamiliar with β may find its values somewhat difficult to interpret, especially because it ranges from minus infinity to plus infinity. An intuitive interpretation is facilitated by focusing on exp (β), which is the odds ratio. In this context, the odds ratio is interpreted as the odds that $H_{t+k} = 1$, given $W_t = 1$, divided by the odds that $H_{t+k} = 1$, given $W_t = 0$. Thus, if exp (β) = 2, we can say that the odds that $H_{t+k} = 1$ are doubled by W_t shifting from 0 to 1. Alternatively, for those who prefer a measure that varies between -1 and 1, a simple transformation of β produces Yule's Q, a familiar measure of association first proposed in 1900. The transformation is

$$Q = \frac{e^\beta - 1}{e^\beta + 1}.$$

The f_i are simply the eight observed cell frequencies in the two 2×2 tables. As explained in the previous section, each $\hat{\beta}$ is computed by taking the logarithm of the observed cross-product ratio in each table. Under the null hypothesis, this statistic has an asymptotic standard normal distribution. For these data, we have

$$z_2 = \frac{.651 - .141}{\sqrt{\frac{1}{76} + \frac{1}{79} + \frac{1}{100} + \frac{1}{200} + \frac{1}{80} + \frac{1}{43} + \frac{1}{63} + \frac{1}{39}}} = 1.493,$$

which is not significant at the .05 level. Thus, the evidence is insufficient to conclude that there is a difference between the two groups.

An alternative approach is especially useful when there are more than two groups to be compared. Instead of viewing the data as two two-way tables, we can treat them as a single three-way table with the additional variable X, which equals 1 for distressed and 0 for nondistressed couples. The null hypothesis can then be expressed as a logit-linear model:

$$\text{logit}[Pr(W_{t+1} = 1 | H_t, X)] = \alpha + \beta H_t + \gamma X.$$

The key feature of this equation is that there is no interaction term for X and H_t, expressing the fact that β is the same for both groups. This model may be fit to the data by the method of maximum likelihood using programs such as GLIM (Baker & Nelder, 1978). If a logit program is unavailable, one can make use of the fact that all logit models for contingency tables are special cases of log-linear models for which estimation programs are widely available (e.g., ECTA, MULTIQUAL, FREQ). For this logit model, the equivalent log-linear model is one that specifies no three-way interaction.

Once the model is fit, the hypothesis is tested by comparing observed cell frequencies to predicted frequencies using either Pearson's chi-square statistic or the likelihood-ratio chi-square statistic. For the data in Table 1, Pearson's chi-square is 2.235, and the likelihood-ratio chi-square is 2.230, both with one degree of freedom. These values may be compared with the z test simply by taking the square root, which yields 1.495 and 1.493, respectively. Although all three tests give virtually identical results in this case, they are not identically equal. For one thing, the z test cannot be computed when any of the cell frequencies is zero, but this does not necessarily prevent the logit-linear model from being fit.

To compare more than two groups, the logit-linear model is modified to include a dummy variable for each of the groups less one. Thus, if there are four groups, the dummy variables could be $X_1 = 1$ for Group 1 and zero otherwise, $X_2 = 1$ for Group 2 and zero otherwise, and $X_3 = 1$ for Group 3 and zero otherwise. The model is then fit by the method of maximum likelihood and the null hypothesis tested with either of the two chi-square statistics. The number of degrees of freedom for these tests will be one less than the number of groups. Again the logit-linear model is equivalent to a log-linear model with no three-way interaction.

This same approach offers the possibility of estimating β, the effect of the husband's behavior on the wife's behavior, while statistically controlling for other variables. Indeed, the β in the logit-linear equation above is a partial β controlling for group membership. If one has sequential data on many different couples differing on such characteristics as race, religious preference, or social class, one could construct a multiway contingency table and fit logit-linear models (or the equivalent log-linear models), which include additional dummy variables to represent the other characteristics.

Using a somewhat different fitting procedure (Hanushek & Jackson, 1977; Nerlove & Press, 1973), one can even include control variables such as age, income, or years of schooling, which are measured as continua. The resulting analysis is very much like the analysis of covariance. Finally, one can adjust for all differences *between* couples by pooling all the data for all couples but including a dummy variable for each couple (less one), a procedure that Gottman (1980b) recommended in a somewhat different context.

The Problem of Autodependence

All of the tests discussed so far assume that an actor's behavior at any time does not depend on the same actor's behavior at any previous time (i.e., there is no autodependence). This assumption makes it possible to treat each pair of observations (H_t, W_{t+k}) as though it were independent of all other pairs from the same couple. Although this greatly simplifies the construction of statistical tests, it is obviously not very realistic. Moreover, the violation of this assumption can produce very misleading results.

Consider, for example, the data in Table 2, which gives frequencies for the three-way cross-classification of H_t, W_t, and W_{t+1}.[2] Suppose we wish to test for the dependence of W_{t+1} on H_t. Following the procedures given earlier, we can test for dependence by first collapsing the table over W_t to get the $H_t \times W_{t+1}$ table and then computing a chi-square test for independence. The Pearson chi-square for this collapsed table is a highly significant 27.64 with one degree of freedom, and the estimated value of β is .404. This test fails to take account of the fact that both H_t and W_{t+1} are strongly associated with W_t, which may be seen by looking at the $W_t \times W_{t+1}$ table and the $W_t \times H_t$ table. The former collapsed table has an enormous chi-square of 1,184.7, whereas the latter has a chi-square of 62.35 (both on one degree of freedom). These interrelationships suggest that the apparent dependence of W_{t+1} on H_t may be a consequence of the fact that both of these variables depend on W_t. To rule out this possibility, we can estimate and test the effect of H_t on W_{t+1} *controlling* for W_t.

There are several equivalent ways to do this, but all require analysis of the full three-way table. The null hypothesis can be expressed as

$$Pr[W_{t+1} = 1|W_t, H_t] = Pr[W_{t+1} = 1|W_t]$$

for all values of W_t and H_t. In essence, this says that W_{t+1} may depend on W_t but not on H_t. The easiest way to test this hypothesis is to treat Table 2 as two 2×2 tables, one table for each value of W_t. The test is performed by computing a chi-square test for independence in each of these subtables and then summing the two values. For these data, the resulting Pearson chi-square is 5.91 with two degrees of freedom—not quite significant at the .05 level. Thus, when the autodependence is partialed out, the cross-dependence is greatly attenuated.[3]

Again, an alternative approach is to fit a logit-linear model by maximum likelihood methods. In this case, the null hypothesis is expressed by the model

$$\text{logit}[Pr(W_{t+1} = 1|W_t, H_t)] = \alpha + \beta W_t\,.$$

If Pearson's chi-square is used to evaluate the fit of this model, the result will be iden-

Table 2
Frequency Counts for a Three-Way Cross-Classification

		W_{t+1}	
W_t	H_t	1	0
1	1	577	139
1	0	222	76
0	1	169	1089
0	0	149	839

[2] These data are taken from Coleman (1964, p. 168). For consistency with the rest of the article, the variables have been labeled H_t and W_t, but actually these refer to variables measuring "membership in the leading crowd" and "attitude toward membership in the leading crowd" at two different times.

[3] This test has been criticized for lack of power (Cochran, 1954). A more powerful test having a single degree of freedom was proposed by Mantel and Haenszel (1959) and is described by Bishop et al. (1975). An alternative test with one degree of freedom is described in Footnote 4. Both of these tests assume that H_t and W_t do not interact in their effects on W_{t+1} (i.e., the effect of H_t on W_{t+1} is the same at both levels of W_t). Before applying either of these tests, one should test for interaction using the methods described above for comparing dependency levels across groups (here the two groups correspond to $W_t = 1$ and $W_t = 0$). The z_2 statistic for these data is 2.39, suggesting that there *is* an interaction, and, hence, the single degree of freedom tests are inappropriate.

tical to the test described above.[4] The advantage of this approach is that it can handle much more complicated forms of autodependence. For example, we could allow for autodependence up to lag 3 by fitting the model

$$\text{logit}[Pr(W_{t+1} = 1|H_t, W_t, W_{t-1}, W_{t-2})] = \alpha + \beta_1 H_t + \beta_2 W_t + \beta_3 W_{t-1} + \beta_4 W_{t-2}$$

to the five-way table for W_{t+1}, W_t, W_{t-1}, W_{t-2}, and H_t. The hypothesis that W_{t+1} does not depend on H_t is then equivalent to the null hypothesis that $\beta_1 = 0$. Gottman (1980a) suggested even more complicated models for interval-level data; completely analogous models for categorical data can be estimated by the methods outlined here. For futher details of the fitting and testing procedure, see Goodman (1970), Bishop et al. (1975), Fienberg (1980), or Haberman (1978).[5]

Testing for Dominance

Gottman (1980b) gave the following definition of dominance: "In a dyad, if B's future behavior is more predictable from A's past behavior than conversely, then A is said to be dominant" (p. 71). To test for dominance, Gottman used a procedure that was based on spectral analysis, which unfortunately did not control for autodependence. He has, however, since proposed a method suitable for interval-level data that does control for autocorrelation (Gottman & Ringland, 1981). We now describe an analogous method for categorical data.

A naive approach would be to compute a z test for the dependence of H_{t+k} on W_t and another z test for the dependence of W_{t+k} on H_t. If one test statistic is substantially larger than the other, it could be taken as evidence for dominance. As we have seen earlier, however, the fact that two test statistics differ does not imply that the strength of the dependence is significantly stronger in one direction than in the other. What is needed is a single statistical test for the difference between the measures of association for each relationship. Of course, it is also necessary to introduce appropriate controls for the lagged dependence of each spouse's behavior on his or her own prior behavior. As we shall see, proper controls can actually reverse the direction of dominance.

Table 3
Frequency Counts for a Hypothetical Four-way Table

		W_2	H_2	W_2	H_2	W_2	H_2	W_2	H_2
W_1	H_1	1	1	1	0	0	1	0	0
1	1	300		79		5		16	
1	0	70		11		5		14	
0	1	14		5		11		70	
0	0	16		5		79		300	

Suppose that for a sample of 100 couples we observe behavior at two times. Thus, we observe H_1, H_2, W_1, and W_2. As shown in Table 3, the data can be displayed as a four-way contingency table. These data were constructed for illustrative purposes, but there is no reason why similar data could not arise in real life.

Ignoring any autodependence, we first test for cross-dependence of H_2 on W_1 and of W_2 on H_1. To do this, we collapse the four-way table into two two-way tables and then compute Pearson's chi-square test for independence in each table. For H_2 on W_1 the chi-square is 270.4; for W_2 on H_1 the chi-square is 350.5, suggesting that the husband is dominant. This is corroborated by the fact that $\hat{\beta} = 2.305$ for H_2 on W_1 as compared with $\hat{\beta} = 2.723$ for W_2 on H_1. The difference between these statistics may be tested using the same test we employed to compare distressed and nondistressed couples:[6]

---

[4] A more powerful test is obtained by first fitting this model and then fitting the model

$$\text{logit}[Pr(W_{t+1} = 1|W_t, H_t)] = \alpha + \beta_1 H_t + \beta_2 W_t.$$

The test statistic is then the difference between the likelihood-ratio chi-squares for these two models, with a single degree of freedom. As explained in Footnote 3, this test assumes that there is no interaction between H_t and W_t in their effects on W_{t+1}.

[5] The full justification of these methods requires that the sequential data be generated by a stationary Markov chain. The underlying statistical theory is discussed by Anderson and Goodman (1957) or in greater detail by Billingsley (1961).

[6] In this context, z_2 is not strictly appropriate because it presumes that the two estimates of β are uncorrelated. This is unlikely to be the case because both are estimated from the same sample. If the correlation is positive (which is the most plausible case), z_2 will be lower than the "true" z statistic. Hence, it may be regarded as a conservative test.

$$z_2 = \frac{2.723 - 2.305}{\sqrt{\frac{1}{380} + \frac{1}{120} + \frac{1}{120} + \frac{1}{380} + \frac{1}{398} + \frac{1}{102} + \frac{1}{102} + \frac{1}{398}}} = 1.937.$$

Because this is not quite significant at the .05 level, the evidence is only marginal for husband dominance.

This analysis ignores the fact that there is substantial autodependence for the wife's behavior and even stronger autodependence for the husband's behavior. To partial out this autodependence, we apply the techniques described in the preceding section. We first analyze the three-way table $H_1 \times W_1 \times H_2$ to estimate the logit-linear model:

$$\text{logit}[Pr(H_2 = 1|H_1, W_1)] = \alpha + \beta_1 W_1 + \beta_2 H_1.$$

This can be interpreted as a structural model for the dependence of the husband's behavior on the wife's and on the husband's own previous behavior. Next, we estimate the corresponding table for the $H_1 \times W_1 \times W_2$ table:

$$\text{logit}[\Pr(W_2 = 1|H_1, W_1)] = \alpha^* + \beta_1^* H_1 + \beta_2^* W_1.$$

The appropriate comparison is between β_1 and β_1^*, which are the two measures of cross-dependence, controlling for autodependence.

These equations were estimated using the program GLIM (Baker & Nelder, 1978), and the estimates are displayed as a path diagram (Goodman, 1973) in Figure 1. With the introduction of controls the direction of dominance is reversed, with estimates for β of 2.265 for wives on husbands and 1.443 for husbands on wives. But is this difference significant? Gillespie (Note 1) presents one method of testing for the significance of the difference, but a simple method is available if the computer program provides standard errors for each of the estimated parameters.[7] For this example, the estimated standard error of $\hat{\beta}_1$ is $s = .183$. For $\hat{\beta}_1^*$ we have $s^* = .240$. The test statistic is

$$z_3 = \frac{\hat{\beta}_1 - \hat{\beta}_1^*}{\sqrt{s^2 + s^{*2}}} = \frac{2.265 - 1.443}{\sqrt{(.183)^2 + (.240)^2}} = 2.72,$$

which is significant at beyond the .01 level. We conclude that wives are dominant over husbands for these data.

Polytomous Data

So far we have assumed that sequential behavior is coded as a dichotomy. There will often, however, be more than two categories, and Gottman (1980b) typically dealt with three: positive (1), neutral (0), and negative (−1). His general strategy was to do two separate analyses, one focusing on "positive reciprocity" $Pr[W_{t+k} = 1|H_t = 1]$ and the other focusing on "negative reciprocity" $Pr[W_{t+k} = -1|H_t = -1]$. Of course, these two analyses are not independent, and if the neutral category is small they will be almost mirror images.

We believe that a better strategy is to do a simultaneous analysis of all three categories after which one can test more refined hypotheses. This is easily accomplished, often by quite conventional methods for analyzing contingency tables. Suppose, for example, that both the husband's and the wife's behavior is coded by the three-category scheme just mentioned, and the aim is to test for dependence of the husband's behavior at time $t + k$ on the wife's behavior at time t. If we ignore autodependence, the null hypothesis may be expressed as

$$Pr[H_{t+k} = i|W_t = j] = Pr[H_{t+k} = i]$$

for $i = -1, 0, 1$ and for $j = -1, 0, 1$. This hypothesis can be tested by forming the 3×3 contingency table and applying the standard chi-square test for independence. To determine whether the degree of dependence *differs* across different groups (e.g., distressed and nondistressed couples), the best approach is to fit the appropriate log-linear model to the three-way contingency table and then test the goodness of fit. The log-

[7] Both of these methods suffer from the difficulty discussed in Footnote 6: They ignore the fact that $\hat{\beta}_1$ and $\hat{\beta}_1^*$ are likely to be positively correlated. As a consequence, the test statistics will be biased downward.

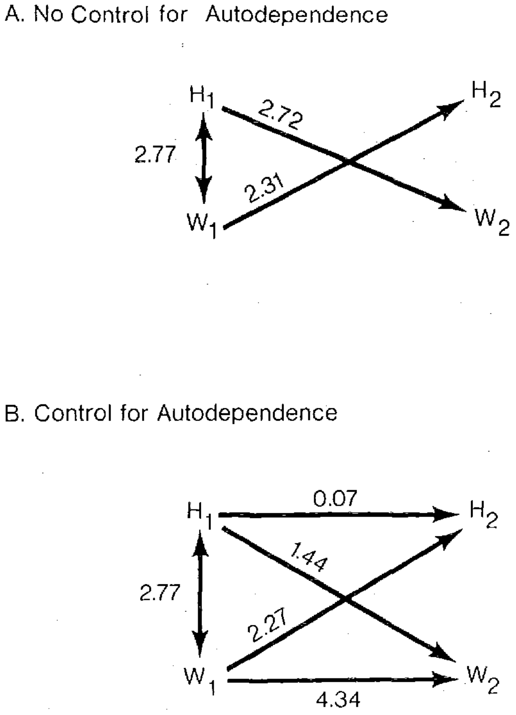

Figure 1. Path diagrams for alternative models for a four-way table.

linear model should exclude the three-way interaction term, and, as usual, the fit may be evaluated using either Pearson's chi-square or the likelihood-ratio chi-square.

These tests do not, however, tell anything about the direction and form of the dependency of H_{t+k} on W_t. For that, it is useful to convert the log-linear model for the contingency table into its equivalent logit-linear form. There are several ways to "parameterize" the polytomous logit-linear model, but for the three-category case we favor the following expression. Let $W(+) = 1$ if $W_t = 1$ and zero otherwise, and let $W(-) = -1$ if $W_t = -1$ and zero otherwise. The model then consists of two equations:

$$\log_e \left\{ \frac{Pr[H_{t+k} = 1 | W_t]}{Pr[H_{t+k} = 0 | W_t]} \right\} = \alpha_1 + \beta_{11} W(+) + \beta_{12} W(-)$$

$$\log_e \left\{ \frac{Pr[H_{t+k} = 0 | W_t]}{Pr[H_{t+k} = -1 | W_t]} \right\} = \alpha_2 + \beta_{21} W(+) + \beta_{22} W(-).$$

The first equation describes the effects of W on whether H is positive or neutral, and the second equation describes the effects of W on whether H is neutral or negative. These effects are represented by four β coefficients. Each of these coefficients corresponds to the logarithm of the odds ratio in one of the 2×2 tables formed by deleting one extreme row and one extreme column from the 3×3 table.

Within this framework, one can test hypotheses of the following sort. With respect to the wife's behavior, one might hypothesize that the effect of going from negative to neutral is the same as going from neutral to positive. This hypothesis is equivalent to assuming that $\beta_{11} = \beta_{12}$ in the first equation and $\beta_{21} = \beta_{22}$ in the second. Similarly, with respect to the husband's behavior, one might hypothesize that the difference between negative and neutral is the same as the difference between neutral and positive. This amounts to hypothesizing that $\beta_{11} = \beta_{21}$ and $\beta_{12} = \beta_{22}$. If both hypotheses are true, one has the uniform association model of Duncan (1979) for ordered categorical data. Haberman (1979) described methods for testing these hypotheses. Just as in the dichotomous case, the models for polytomous variables can be easily extended to allow for more complicated lag structures and for additional control variables. Finally, if the sequential behavior is coded on a continuous or nearly continuous scale, one can use standard procedures for the analysis of multivariate time-series data that have been developed by econometricians (see, e.g., Maddala, 1977) and that have been recently advocated by Gottman and Ringland (1981).

Conclusion

The statistical procedures described here can be used to test the kinds of propositions advanced by Gottman (1980b) as well as many other common propositions in the study of social interaction. They can be applied to data on a small number of couples followed over many time points or many couples followed through as few as two time points. The advantages of these procedures over those used by Gottman are essentially threefold: (a) They explicitly take into account the categorical nature of the data; (b) they allow for multivariate as opposed to merely bivariate analysis; and (c) they conform to conventional practices for statistical inference.

The implications for Gottman's (1980b) substantive conclusions are unclear. In a few instances where Gottman presented sufficient data for reanalysis, we were able to show that the use of more appropriate methods could lead to different conclusions. In most cases, however, reanalysis was impossible, so we have no way of knowing how serious the biases are. On the one hand, the use of Sackett's (1979) incorrect statistic would tend to result in acceptance of the null hypothesis when, in fact, it could be rejected. On the other hand, the failure to test explicitly for the significance of differences between groups is likely to result in concluding that a difference exists when there is actually insufficient evidence for a statistically reliable difference.

Given the importance of the hypotheses and the apparent high quality of the data, we hope that further analyses will be undertaken to answer these questions. Without appropriate statistical procedures, even the most elaborately coded data and carefully controlled collection procedures will not further our understanding of social interaction.

Reference Note

1. Gillespie, M. W. *Another look at the sixteenfold two-wave two-variable table: The analysis of partially recursive causal models of categorical data.* Unpublished manuscript, University of Alberta, 1981.

References

Anderson, T. W., & Goodman, L. A. Statistical inference about Markov chains. *Annals of Mathematical Statistics*, 1957, *28*, 89–109.

Baker, R. J., & Nelder, J. A. *The GLIM system.* Oxford, England: Numerical Algorithms Group, 1978.

Bell, R. Q., & Harper, L. V. *Child effects on adults.* New York: Wiley, 1977.

Berkson, J. Application of the logistic function to bioassay. *Journal of the American Statistical Association*, 1944, *39*, 357–365.

Billingsley, P. *Statistical inference for Markov processes.* Chicago: University of Chicago Press, 1961.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. *Discrete multivariate analysis.* Cambridge, Mass.: MIT Press, 1975.

Blalock, H. M., Jr. *Social statistics* (Rev. 2nd ed.). New York: McGraw-Hill, 1979.

Burgess, R. L., & Huston, T. L. (Eds.). *Social exchange in developing relationships.* New York: Academic Press, 1979.

Clarke-Stewart, K. A. Interactions between mothers and their young children: Characteristics and consequences. *Monographs of the Society for Research in Child Development*, 1973, *38*, (6 & 7, Serial No. 153).

Cochran, W. Some methods for strengthening the common χ^2 tests. *Biometrics*, 1954, *10*, 417–451.

Coleman, J. S. *Introduction to mathematical sociology.* New York: Free Press, 1964.

Davis, J. A. *Elementary survey analysis.* Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Duncan, O. D. How destination depends on origin in the occupational mobility table. *American Journal of Sociology*, 1979, *84*, 793–803.

Fienberg, S. E. *The analysis of cross-classified categorical data* (2nd ed.). Cambridge, Mass.: MIT Press, 1980.

Goodman, L. A. The multivariate analysis of qualitative data: Interactions among multiple classifications. *Journal of the American Statistical Association*, 1970, *65*, 226–256.

Goodman, L. A. The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. *Biometrika*, 1973, *60*, 179–192.

Gottman, J. M. Detecting cyclicity in social interaction. *Psychological Bulletin*, 1979, *86*, 338–348.

Gottman, J. M. Analyzing for sequential connection and assessing interobserver reliability for the sequential analysis of observational data. *Behavioral Assessment*, 1980, *2*, 361–368. (a)

Gottman, J. M. *Marital interaction: Experimental investigations.* New York: Academic Press, 1980. (b)

Gottman, J. M., & Ringland, J. T. The analysis of dominance and bi-directionality in social development. *Child Development*, 1981, *52*, 393–412.

Haberman, S. J. *Analysis of qualitative data* (Vol. 1). New York: Academic Press, 1978.

Haberman, S. J. *Analysis of qualitative data* (Vol. 2). New York: Academic Press, 1979.

Hanushek, E. A., & Jackson, J. E. *Statistical methods for social scientists.* New York: Academic Press, 1977.

Kendall, M. G., & Stuart, A. *The advanced theory of statistics* (Vol. 2). London: Charles Griffin, 1961.

Lerner, R. M., & Spanier, G. B. *Child influence on marital and family interaction.* New York: Academic Press, 1978.

Maddala, G. S. *Econometrics.* New York: McGraw-Hill, 1977.

Mantel, N., & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 1959, *22*, 719–748.

Mishler, E. G., & Waxler, W. E. *Interaction in families: An experimental study of family process in schizophrenia.* New York: Wiley, 1966.

Nerlove, M., & Press, S. J. *Univariate and multivariate log-linear and logistic models.* Santa Monica, Calif.: Rand, 1973.

Patterson, G. R. Mothers: The unacknowledged victims. *Monographs of the Society for Research in Child Development*, 1980, *45*(5, Serial No. 186).
Sackett, G. P. The lag sequential analysis of contingency and cyclicity in behavioral interaction research. In J. Osofsky (Ed.), *Handbook of infant development.* New York: Wiley, 1979.
Schenkein, J. (Ed.). *Studies in the organization of conversational interaction.* New York: Academic Press, 1978.
Wright, H. F. *Recording and analyzing child behavior.* New York: Harper & Row, 1967.

Appendix

Proof of the Equivalence of Two Statistics for Testing Lagged Dependence

We show here that the statistic z_1 is identically equal to the standard test for the difference between two binomial probabilities estimated from independent samples. We have

$$z_1 = \frac{p_{H|W} - p_H}{\sqrt{\dfrac{p_H(1 - p_H)}{(n - k)p_W}}}.$$

Using additional notation developed above, we can rewrite this as

$$z_1 = \frac{p_{H|1} - [p_{H|1}p_W + p_{H|0}(1 - p_W)]}{\sqrt{\dfrac{p_H(1 - p_H)}{(n - k)p_W}}}$$

$$= \frac{(1 - p_W)(p_{H|1} - p_{H|0})}{\sqrt{\dfrac{p_H(1 - p_H)}{(n - k)p_W}}}$$

$$= \frac{p_{H|1} - p_{H|0}}{\sqrt{\dfrac{p_H(1 - p_H)}{(n - k)p_W(1 - p_W)}}}.$$

This, in turn, can be written as

$$z_1 = \frac{p_{H|1} - p_{H|0}}{\sqrt{p_H(1 - p_H)\left(\dfrac{1}{m_1} + \dfrac{1}{m_0}\right)}},$$

which is the desired result.

Received May 1, 1981 ■