Reprinted from HUMAN ORGANIZATION, Volume 35, Number 3, Fall, 1976.

Informant Accuracy in Social Network Data

PETER D. KILLWORTH
H. RUSSELL BERNARD

Abstract

This paper examines the problem of informant accuracy in the production of social network data, through the use of a self-monitoring network. This allows a comparison between cognitive network data and informants' interactive behavior. Against expectations, it turns out that informants are extremely inaccurate. In other words, informants' reports of their behavior bear little resemblance to their behavior. If an informant claimed to have communicated with some person "the most frequently" then, in fact, he communicated with that person between first and fourth most frequently only 52% of the time. The implications of our findings for sociometric and network analysis are: (1) Attempts to filter out noise in a sociometric network matrix by using sophisticated software are likely to be unproductive. This is because such manipulations assume a much lower level of noise than actually occurs. (2) Due to the low level of informant accuracy, theories of social structure built upon presently available network data are suspect.

Exactitude informatrice
dans les indications de réseau social

Cet article examine le probleme de l'exactitude informatrice dans la production d'indications de réseau social, par l'emploi d'un service d'écoute automatique. Ceci permet de comparer entre les données du réseau cognitif et le comportement mutuel des informateurs. Contre toutes previsions, il s'avere que les informateurs sont extrêmement inexacts. C'est à dire que les rapports des informateurs sur leur propre comportement ont peu de ressemblance avec leur comportement. Si un informateur prétend avoir communiqué avec une certaine personne "tres fréquemment," en realité il a communiqué avec cette personne environ une à quatre fois plus seulement dans 52% du temps. Les implications de nos découvertes pour l'analyse sociométrique et de réseau sont: (1) Les essais de filtrer le bruit dans une matrice sociométrique ou de réseau en utilisant des objets sophistiquès et nous sont apparemment improductifs. Il en est ainsi parce que de telles manipulations présument un niveau de bruit beaucoup plus bas que ce que se produit en réalité. (2) En raison du bas niveau de l'exactitude informatrice, les théories de strúcture sociale construites sur les indications du réseau actuellement disponibles sont suspectes.

Objetividad de informantes
en información de agrupaciones sociales

El presente informe explora el problema de la objetividad de los informantes involucrados en investiga-

H. Russell Bernard and Peter D. Killworth are a naturalist-theoretician research team. Bernard is Associate Professor of anthropology at West Virginia University; Killworth is Research Fellow, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, U.K. They have produced a series of papers on numerical aspects of social networks, and generally take turns as senior author. Both authors maintain more traditional career lines: Killworth specializes in modeling the Antarctic and Mediterranean oceans; Bernard specializes in Otomi Indian sociolinguistics and the culture of the Greek islands. *The team's work in network analysis is funded by the Office of Naval Research, Code 452, Contract #N00014-73-A-0417-0001. The opinions expressed in this article are those of the authors and do not necessarily reflect the position of the supporting agency. Reproduction in whole or in part is permitted for any purpose of the United States government. The data were collected by Dr. William Stokoe, Gallaudet College, under NSF grant #SOC-7414724.*

ción de agrupaciones sociales por medio de un sistema autocorrectivo. Ello permite que se establezca una comparación entre la información cognitiva y la conducta interactuante del informante. Contrario a lo que pueda esperarse, resulta que los informantes son exageradamente inexactos. Es decir, lo que ellos dicen sobre lo que hacen y lo que hacen son dos cosas distintas. Si un informante se jactaba de haberse comunicado con una persona "con màs frecuencia," ello significaba que se había mantenido en contacto con ella sólo un 52% del tiempo a lo máximo. Las implicaciones de este hallazgo para análisis sociométrico y de interacción matriz son: (1) Los intentos de filtrar interferencias de ruido en una matriz sociométrica o de interacción resultan vanos aún si se usan instrumentos avanzados. Ello se debe a que tales manipulaciones presuponen un nivel de ruido inferior al que en realidad existe. (2) Debido al bajo nivel de objetividad del informante, las teorías sobre estructura social que se han usado de cimientos hasta el presente carecen de total veracidad y se prestan a sospechas.

I. Introduction

SOCIAL NETWORK DATA are much like sociometric data; they are collected by asking people to make their best guess about who they interact with. For the most part, sociometric data are "preferential" or affective (Bjerstedt 1956), while network data are effective. Both are riddled with measurement error, as is well known (Holland and Leinhardt 1973; Hallinan 1972; Killworth 1974). The papers mentioned have examined the problem only as regards the effect of error on interpretations of the data. They make the assumption that sociometric data will always be riddled with error and that we shall never know the nature of the error.

Assessment of the error within the instruments used to collect sociometric or network data involves four considerations: validity, reliability, veridicality, and accuracy. The problem of validity is well known: no matter how carefully one phrases the question, one never really knows whether people are responding in the same way. For example, if we make the classic request for people to name their "three best friends" we do not know whether the word "friend" connotes the same thing to everyone we ask.

The problem of reliability is similarly well known. It is usually tested by asking the same question after a suitable length of time. Two weeks seems to crop up with some regularity in the literature. The assumption seems to be that two weeks is long enough for people to have forgotten what they said the first time and short enough for the situation not to have changed. Evidence presented elsewhere (Bernard and Killworth 1975), however, demonstrates a gross change in sociometric relations over a two-week period.

In addition to validity and reliability, we have the annoying problem of veridicality. If we ask people to tell us how racially prejudiced they are, we get predictably useless results. Loaded questions get loaded answers. (A review of the general untrustworthiness of attitudes as predictors of behavior may be found in Deutscher 1966.)

Assuming that straightforward sociometric queries receive honest answers, there remains an unanswered question: Is cognition a reasonable proxy for behavior? Clearly, the validity of much social science depends on the answer to this question. Sociometry (i.e., affective measurement) has universally avoided the issue, and with good reason.[1] Network analysis cannot.

How, then, does one measure the "accuracy of informant cognition about their interactive behavior" (hereafter "accuracy")? The requirements are rather stringent. Reiss (1971) has posed them succinctly:

> The means of observation and recording, whether a person or some form of technology, must be independent of that which is observed and the effects of observing and measuring must be measureable.

Clearly, the best way to achieve the idyllic state described by Reiss is to have no observer; this yields no bias and, unfortunately, no data. The next best state of affairs is unobtrusive observers. However, as Kandel (1974) has noted, "monitoring a communications network, especially one of any size and spatial distribution is almost impossible for one or even a few observers." Moreover, assigning many observers to collect data would be so obtrusive as to modify the network being studied.

The problem boils down then, to "collection and recording of ongoing network data *accurately* with *minimum interference*" (Kandel 1974, emphasis added). Kandel suggested mounting minitransducers on the members of an interaction set and recording their output via an analog computer. However, it is not known how knowledge of "being bugged" might affect behavior. A better approximation to the ideal state is a system of interaction which self-monitors as a matter of course.

The data described in this paper were collected by such a system. The key issue which arises from such data, and which occupies the remainder of the paper is:

> Given network data obtained in the normal way from a member of a group, what can we say (and with what reliability) about whom he talked to and how much talking he did?

II. Method of Data Collection

The data considered here are part of a larger study of the social and communications structure of the deaf community in the Washington, D.C., area. The study is being conducted by the Linguistics Research Laboratory at Gallaudet College, under the direction of William Stokoe. An intrinsically interesting segment of the community is the "elite." Few ethnographies deal with elites, for obvious reasons. The elite of the deaf community, however, have been genuinely cooperative in the overall effort to study their activities and roles. The "group" under study was selected from among a much larger initial population by an expert panel of deaf persons. The final group consisted of 33 members in government, commerce, and academic life.

What concerns us here is the range of communication techniques employed by the group. In face-to-face interaction, they sign; for long range communication they rely on teletypes (TTY's) or their oscilloscope equivalents (VDU's) in lieu of telephones. A characteristic of the group is that they can afford such units at home as well as at work.

One may legitimately inquire whether TTY communication is "natural." Chapanis (1975) has shown that, indeed, communications via TTY are quite different from face-to-face voice interaction. For example, in experiments under controlled conditions, problem solving took twice as long with TTY's as with voice mode. The time difference could not be accounted for by lack of typing experience.[2] In addition, those using the TTY were terse compared with persons using voice mode. On the other hand, Chapanis found that sentence fragmentation, false starting, and other features of voice communication were duplicated on TTY's. Regardless of whether such laboratory communication can be considered "natural," there is no doubt that the interaction of the elite deaf by TTY *is* natural. It is, quite simply, their accustomed means of long-range communication.

A feature of the TTY, which makes it ideal for our purposes, is that a hard copy of a conversation is produced which can be compared with a person's guess about whom he talks to.

With this in mind, we first asked (during December 1974) each member of the group to describe his communications on TTY with the rest of the group. This was done by giving each person a deck of 31 cards containing the names of the others in the group (one member had moved to Chicago in the interval between planning and implementation of the study). Each person ranked the deck from "most" to "least" in response to the question, "How much do you communicate with each of these people on your TTY?" Predictably, informants asked if we wished them to distinguish between home and office, and among time spent interacting, importance of content, and frequency of contact. (Later experiments are planned to compare the results of asking such questions.) Since we had no a priori way of deciding the relative merits of such distinctions we conveniently instructed informants to combine home and office communications, and to "do the best overall job" they could concerning the ranking.

Of the 32 potential informants, 31 did the ranking (one was in the hospital and remained so during the rest of the data collection). Not all the group felt capable of ranking all 31 others. Eleven persons ranked fewer members and demurred that they could not make meaningful judgments about the rest.

Up to this time the group was unaware that anything more was required from them. We now asked them to retain the hard copy of their TTY's for a period of 21 days (15 work days plus six weekend days). We provided them with log sheets and requested that they fill in, for each TTY communication, the following information: date, whether the call was dialed or answered, the name of the other party, and (as a measure of communication length) the number of TTY lines in the conversation. We were already aware that individuals differ as regards style of communication via TTY. Some prefer to fill up each line, while others prefer to use short lines; some persons are also more terse than others. However, we felt that these differences would average themselves out across a group. From the practical point of view, counting words would be an intolerable task. We would not ask for the hard copy itself, for the obvious reason of privacy.

Another problem we encountered was the use of VDU's versus TTY's. Since VDU's produce no hard copy, we asked people with these machines to log minutes rather than lines. Two informants (28, 25) conveniently logged both lines and minutes of TTY. Making the assumption that typing speed on TTY is the same as on a VDU, these data should enable us to convert minutes of VDU to lines of TTY. Unfortunately, the comparison proved less easy than we had hoped. The ratio lines/minutes for person 25 was 3.56 (S.D. 1.5), while for person 28 it was 1.43 (S.D. 0.34). Student's t between these two means is 7.49, with 59 degrees of freedom; this value is highly significant.

Having thus established that teletype habits differ violently between users, there was no obvious reason to select one user's mean rather than another's. This presented a problem. Fortunately, the amount of VDU usage in our sample was much smaller than the amount of TTY usage. (Only 21% of the total contacts reported were on VDU's.) This meant that the bias introduced by selecting the "wrong" lines/minutes ratio, should not affect the data too strongly, and we settled on the lower ratio. In fact, we used a value of 1.49 which was computed before all the data were in.

This was by no means the end of our problems. Of the 31 original rankers, only 21 returned log sheets. (Actually, 25 returned logs. However, four of these had no communication with the group during their period of logging, and we did not consider them further.) They did not all complete the task during a single 21-day period. Some spent as much as three months out of a four-month period (January-April 1975) logging the requisite number of TTY use-days. Several informants had not managed to log 21 days when we finally stopped the experiment. There was no way to overcome these difficulties, or to know what effect they might have on our conclusions from the data. Since these are the only data of their kind, we felt that, despite their imperfections, their analysis is very worthwhile.

III. Overview of the Data

This section gives a brief description of the data. In the following sections we will introduce various measures of informant accuracy, and we will discuss the results of applying these measures to our data.

Tables 1, 2, and 3 summarize the raw data. We have introduced three terms, "to," "from," and "undistinguished." By these we mean, respectively, communications *to* others, *from* others, and both combined. (In what follows, "undistinguished" is assumed, unless otherwise specified.) We have also indicated the number of lines logged of communication to the group (mean 173, S.D. 157 for undistinguished communication) and the total lines logged (mean 819, S.D. 553). In fact, on average, 75% of each informant's TTY interaction was with persons outside the group. This, of course, demonstrates the difficulty of defining a closed social group.

Communications to and from the group were split almost equally. The mean number of TTY lines to the group was 84 (S.D. 107), i.e., 49% of the total undistinguished TTY lines logged to the group. The mean number of TTY lines logged from the group was 89 (S.D. 90), i.e., 51% of the total undistinguished lines to the group. Three people had no communication to the group, and two had none from the group.

The number of contacts reported also differed radically between informants. The mean number of undistinguished contacts reported was 37 (S.D. 15.6). Of these, 7.9 on average (S.D. 7), were to members of the group. In other words, 79% of a person's contacts lay outside the group. This compares well with the 75% fraction noted above for number of TTY lines. Both globally and just to the group, there were more contacts *from* others than *to* others, although the difference is not statistically significant. (Mean number of contacts *to* others globally = 15.6, S.D. = 10.2; mean number of contacts *from* others globally = 21.7, S.D. = 18.4; mean number of contacts *to* others in the group = 3.4, S.D. = 2.7; mean number of contacts *from* others in group = 4.6, S.D. = 5.2.)

The number of different *people* communicated with

TABLE 1. PROPORTIONAL AMOUNT OF COMMUNICATION WITH THE GROUP WHICH A PERSON HAD WITH

Person	No. Ranked	Communications with the Group	Total Communication	RANKINGS 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	31	262	1569	0.160	0.585		0.061							0.068			
2	31	92	1017	0.185	0.174		0.272		0.043	0.326							
3	31	39	83	0.346			0.192	0.192								0.269	
4	16	137	936				0.095	0.905									
5	12	298	1264	0.158	0.047					0.124	0.158	0.262					
6	8	9	266			0.556				0.444							
7	31	29	213										0.483				
9	31	179	430	0.430	0.184	0.039					0.061	0.179		0.022			
10	25	515	1605	0.045	0.017		0.014	0.017		0.014	0.217	0.099	0.179	0.021	0.064		
12	20	223	853	0.367		0.067		0.100			0.067		0.067		0.267		
14	6	66	99	0.667					0.152								
16	31	103	636	0.632	0.136		0.072	0.072									
18	31	359	1303		0.008	0.964		0.964					0.028				
19	31	392	1391	0.196	0.543					0.041						0.102	
20	31	100	881	0.500				0.370		0.130							
21	31	24	189				0.500		0.500								
24	31	165	813			0.170		0.133			0.073		0.158				0.327
25	31	525	1119	0.455					0.248								
26	31	38	769							0.632						0.368	
27	31	25	660							1.000							
32	31	63	107	0.524	0.127												
Means				0.222	0.087	0.040	0.103	0.085	0.045	0.136	0.029	0.028	0.048	0.006	0.017	0.041	0.018
Standard Deviation				0.229	0.166	0.122	0.227	0.204	0.118	0.262	0.059	0.071	0.115	0.016	0.060	0.102	0.075

by any individual was quite small: mean 4.6, S.D. 3.2. One individual spoke only to one person. This presents some problems, as we shall see later.

We did not check to see if a group contact reported by an informant was reciprocated on the appropriate log sheet. The time scale (four months) over which the data were collected meant that many persons who communicated with each other were not logging simultaneously.

IV. Definitions of Accuracy

We return now to the key question of accuracy as raised in Section I. There are two distinct approaches. The first is to answer the question directly; given that a person has ranked the group, what predictions can be made concerning his actual communicants? The second is to define some measure of an informant's accuracy, based on the degree of difficulty which one believes the informant is capable of handling. One may then compare informants' accuracy scores between themselves and against various parameters. We tried both of these approaches.

The first approach may be tested in at least three successively more rigorous ways:

(a) *Section V*. We can test to see if there are significant differences between the rankers and the amounts of communication they had to the people they ranked first, second, third, and so on. If the difference between ranks is not significant, then we cannot conclude that people's guesses (or rankings) are better than chance.

(b) *Section VI*. We can ask, "i *says* he talked to j; did he? Does this depend on where he ranked j?" If i's talking to j does not depend on his ranking of j, then we cannot conclude that i's ranking is meaningful.

(c) *Section VII*. We can ask, "how many of the people that i ranked must be included in order to account for a given percentage of his total communication with varying reliability?" If most or all of i's rankings are required, then i's rankings are again meaningless.

The second approach required the manufacture of ad hoc definitions of accuracy along successively more stringent lines, as follows:

(a) *Section VIII*. We may attempt to correlate i's ranking of any j with the amount of communication i had with that j. This correlation may serve as a crude (hopefully negative) measure of i's accuracy. If the correlation does not differ significantly from zero, then we must conclude that i is unable to make meaningful rankings.

(b) *Section IX*. We may assume that people "do

EACH OTHER PERSON; "UNDISTINGUISHED" COMMUNICATION

RANKINGS

15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0.023			0.031	0.053					0.019							
									0.517							
										0.084						
	0.050				0.076	0.064	0.076									
			0.033	0.033												
				0.087												
			0.041													
				0.048			0.091									
	0.297															
											0.349					
0.001	0.019		0.006	0.013	0.004	0.004	0.010		0.034	0.005	0.023					
0.005	0.068		0.013	0.025	0.018	0.016	0.028		0.125	0.020	0.087					

better" at estimating their communication with higher ranked individuals than with lower ranked persons. We may then define suitably modified rank coefficients A and Â. It will turn out that the statistical significance of A and/or Â is difficult to test. However, we may still examine whether these coefficients vary with certain parameters.

(c) *Section X.* We may ask, "if i ranked j nth, say, was j really the nth most frequently communicated with?" How much leeway would we have to allow (say nth plus or minus 1, or 2, or whatever ranks) in order to improve i's accuracy significantly? If i's accuracy is *not* significantly better than chance, at any stage of leeway, then i cannot be said to be accurate.

We may now examine the data to see if informants' cognition about their behavior is accurate. We should make it clear that in all our tests we are examining the informants' ability to report accurately their behavior over a given, specific period of time. It could be argued that informants will tend to give rankings more closely corresponding with a "typical week" or an "average week" than with any specific week. This would naturally make any correlation between reported and actual communication rather worse than it perhaps ought to be. In a similar vein, one might seek to confirm this by examining the correlations between rankings on repeated samples over time.

We have discounted this for two reasons. First, it is unlikely that the group is in any way—save statistically—steady. Since measurement is the only way to confirm or deny this, repeated measurements to compare reliabilities—which rely on steadiness—give somewhat circular reasoning. Similarly, only measurement can give a picture of a "typical week"; yet it is the very accuracy of those measurements we are examining. There appears to be no way to finesse the problem; we cannot define a "typical week" without including some time-averaging scale which is itself undefined.

For these reasons we limit the discussion specifically to the accuracy of the data as it stands; later experiements will examine these other issues.

V. "Rankers Versus Ranked"

Are there significant differences between the rankers and the amount of communication they had to the people they ranked first, second, third, and so on? We addressed this problem as follows:

We took the data presented in Tables 1, 2, and 3, and established a cut-off at rank K where K is an integer taking successively the values 2, 3, 4. . . . A two-way analysis of variance was performed for each value of K comparing the effects of those who ranked and the

TABLE 2. PROPORTIONAL AMOUNT OF COMMUNICATION WITH THE GROUP WHICH A PERSON HAD WITH

Person	No. Ranked	Communications with the Group	Total Communication	RANKINGS 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	31	129	563		0.800									0.138			
2	31	50	476		0.320				0.080	0.600							
3	31	7	52				1.000										
4	16	137	552				0.095	0.905									
5	12	162	535	0.142	0.086						0.290	0.481					
6	8	5	76			1.000											
7	31	15	52														
9	31	62	140	0.581		0.113								0.065			
10	25	113	136		0.080		0.062	0.080				0.106	0.230		0.142		
12	20	119	476	0.500		0.125									0.313		
14	6	10	18						1.000								
18	31	152	535		0.020		0.980										
19	31	376	1298	0.205	0.566					0.043						0.106	
20	31	13	536							1.000							
21	31	3	78				1.000										
24	31	29	498					0.483									
25	31	361	511	0.476					0.091								
27	31	25	370							1.000							
Means				0.106	0.104	0.069	0.174	0.082	0.065	0.155	0.017	0.037	0.014	0.013	0.028	0.007	
Standard Deviation				0.194	0.221	0.229	0.367	0.228	0.228	0.339	0.068	0.118	0.056	0.036	0.081	0.027	

effects of the ranks into which they placed the rest of the group (up to rank K only) on the "scores." ("Scores" are the fractional amount of total communication with the group which each person had to each other member of the group. These are, of course, the very entries in Tables 1, 2, and 3.)

In no case was there any significant difference between rankers. Thus, there is no reason to suppose that any one informant does "better" than any other. However, there were significant differences between the ranks both for contacts with the group which were "undistinguished" and "from" others. For undistinguished contacts, and for K increasing from 2 onward, the value of the F ratio was: significant at the 1 or 5% level for all but $K = 5$, 7, or 8, and significant at the 1% level for $K \geqslant 11$. For "from" contacts the F ratio was significant at the 1% level for all values of K, except 7 where it was only significant at the 5% level.

In direct contrast to the above, there was *no* significant difference between ranks on contacts "to" others in the group. This leads to the uncomfortable possibility that people "know" more about who will call them than the reverse. Of course, there is no "accuracy" or direct knowledge implied by significant differences between the ranks.

Having found no significant difference between rankers, we could pool them into various subgroups with impunity, in order to test the effect of other parameters. We limited our attention to undistinguished contacts only, and ran two experiments. In the first, we created two subgroups, the 16 people who chose to rank all 31 members of the group, and the five people who chose to rank fewer than 31. An analysis of variance between these two subgroups and ranks 1-25 (this cut-off was chosen to avoid empty cells) showed no difference between the two subgroups. Once again (not surprisingly) there were significant differences between the ranks.

In the second experiment three subgroups were created consisting of persons with low, medium, and high communication with the group at large. These categories were manufactured by taking the maximum number of TTY lines produced by any informant with the group (M, say) and allotting an informant to the "low" category if he had less than $\frac{M}{3}$ lines; to the "medium" category if he had more than $\frac{M}{3}$ but less than $\frac{2M}{3}$ lines; and to the "high" category otherwise. Again, no significant differences were obtained between categories of rankers; and, again, there were significant differences between the ranks.

We also performed a variety of other experiments on these lines which are worth mentioning briefly. The group was split into two or three subgroups, based again on TTY communication with the group. Analysis of variance between these subgroups and between ranks

EACH OTHER PERSON; "TO" COMMUNICATION

RANKINGS

15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
			0.062													
									1.000							
										0.242						
						0.071	0.080									
			0.063													
							0.517									
	0.432															
	0.029		0.009			0.005	0.046		0.077	0.019						
	0.108		0.022			0.019	0.138		0.266	0.064						

1-K (K = 2, 3, . . ., 6) was performed for "undistinguished," "to," and "from" contacts. No new effects were found for "undistinguished" or "to" communications; "from" contacts, however, showed significant variations between quantities other than ranks. For K = 6, and either two or three subgroups, the interaction (subgroups X ranks) was significant; for K = 3, and three subgroups, there were also significant variations between subgroups. Again note the apparently counterintuitive result that people have more understanding about who will call them than who they will call.

From this it would appear that neither (a) amount of communication, nor (b) number of people chosen to be ranked have any strong relationship to a person's ability to rank his communication with others in a group.

VI. "i Says He Talked to j; Did He?"

A slightly more stringent question to ask about the data is, "Given that i ranked j nth (implying, as far as this section is concerned, simply that i believes he talks to j), then what is the chance that i did talk to j? Is this probability a function of n or of the other parameters in the system?" Figure 1 shows the probability of i actually talking to j as a function of the rank of j. There is a strong relationship between rank and probability

FIGURE 1. PROBABILITY OF i TALKING TO j AS A FUNCTION OF THE RANK i ALLOTTED TO j

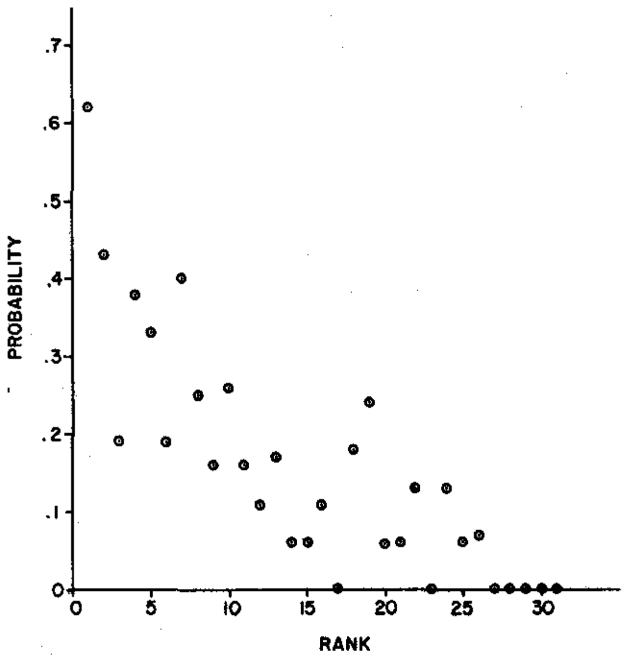

TABLE 3. PROPORTIONAL AMOUNT OF COMMUNICATION WITH THE GROUP WHICH A PERSON HAD WITH

Person	No. Ranked	Communication with the Group	Total Communication	RANKINGS 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	31	133	1006	0.316	0.376		0.120										
2	31	42	541	0.405			0.595										
3	31	31	31	0.429				0.238								0.333	
5	12	136	729	0.176						0.272							
6	8	4	190							1.000							
7	31	14	161										1.000				
9	31	117	290	0.350	0.282						0.094	0.274					
10	25	402	1469	0.057						0.017	0.279	0.097	0.164	0.027	0.042		
12	20	104	377	0.214				0.214			0.143		0.143		0.214		
14	6	56	81	0.786													
16	31	103	377	0.632	0.136		0.072	0.072									
18	31	207	768				0.952						0.048				
19	31	16	1093														
20	31	87	345	0.575				0.425									
21	31	21	111				0.429		0.571								
24	31	136	315			0.206		0.059			0.088		0.191				0.397
25	31	164	608	0.409					0.591								
26	31	38	387							0.632						0.368	
32	31	63	85	0.524	0.127												
Means				0.256	0.048	0.011	0.114	0.053	0.061	0.107	0.034	0.022	0.091	0.002	0.015	0.044	0.025
Standard Deviation				0.250	0.105	0.046	0.253	0.112	0.178	0.265	0.072	0.067	0.236	0.006	0.051	0.116	0.096

(r = −.81, highly significant). As expected, then, there is a strong tendency for there to be communication with those who are ranked low; however, since the a priori probability of talking to anyone is 4.61/30 = 0.15, probabilities below this are not meaningful. So, given the high probabilities for low ranks, it follows that the occasional appearances of high probability at a high rank is indicative of serious inaccuracy on the part of the informant. In fact, rankings beyond 8 are generally no better than chance.

In addition, we examined the difference between the probabilities obtained by limiting attention to those who ranked fewer than 31 and those who ranked all 31. Analysis of variance on ranks 1-6 (beyond 6 there are very few people in the less-than-31 category, which makes the probabilities suspect) reveals no significant difference between the two subgroups.

Similarly, we divided the rankers into two subgroups based successively on (a) less than or greater than 1,000 lines of TTY output to the world; (b) whether or not the TTY output to the group as a fraction of the individual's TTY output to the world was less than 0.4 (chosen to keep the numbers in the higher category from getting too small); (c) whether or not the number of TTY lines to the group was less than half the maximum number registered (575). In no case was there a systematic effect of the subgrouping.

The conclusion would seem to be that none of these parameters affect people's "accuracy" in any way, as before.

VII. *"How Many People Need to Be Ranked?"*

Perhaps the strongest question which might be asked of the data is, "If we wish to account for s% of a person's total communication with the group, how far down the list of those ranked by an individual must one go, in order to be accurate t% of the time?" In other words, we hope that people tend to rank the people they talk with fairly high, so that the first few rankings should contain quite a high percentage of the total communication. Is this true? Table 4 shows that, although on average very few ranks are necessary to account for as much as 60% of a person's total communication, the spread of these values is unpleasantly high.

From Table 4 we generate Table 5. Patently, the numbers of total rankings needed to account for s% of a person's total communication, apply only to this set of data. The percentages of total rankings, however, may apply to groups in general, up to, say, order 50 persons. Beyond that we cannot ignore the limitations of human information processing ability. Obviously, in a group of

EACH OTHER PERSON; "FROM" COMMUNICATION

RANKINGS																
15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0.045				0.105					0.038							
	0.065				0.097	0.062	0.075									
				0.071												
				0.087												
			1.000													
				0.059												
											0.349					
0.003	0.004		0.063	0.020	0.006	0.004	0.005		0.003		0.025					
0.011	0.016		0.242	0.036	0.023	0.016	0.019		0.009		0.090					

10,000 people (a community, for instance) it should not be necessary for an informant to rank his first 320 people in order for a researcher to have a 50% chance of accounting for only 10% of that informant's communication.

The percentages in Table 5 for 95% reliability appear somewhat large, even for a group of this size. For example, requiring an informant's first 24 ranked communicants in order to obtain 70% of his total communication would seem rather pointless; why ask an informant to rank in the first place?

How can we judge if the numbers in Table 5 are significantly high or low? Although there seems to be no easy answer to this, we can at least establish two null hypotheses against which to test Table 5. The first is that everyone communicates with everyone else an equal amount. Thus, to obtain s% of the communication entails choosing s% of the population. Examination of Table 5 shows that our communicants fared poorly at high reliability and small s, but did better than chance elsewhere.

The second, slightly more realistic null hypothesis takes into account the fact that people actually communicated with 4.61 others in the group on average. Thus the probability of choosing one such person in the group at random is 4.61/30. If we make the further assumption that communication is spread evenly between all the 4.61 communicants a person has, then we can set up a crude binomial distribution against which to test Table 5. Because of the absurd assumptions involved, only a very limited sub-set of Table 5 can, in fact, be tested. The results are as follows:

(a) To account for roughly 20% of the total communication, 5 people are necessary to achieve 50% reliability (cf. 2 in Table 5).

(b) To account for roughly 20% of the total communication, 8 people are necessary to achieve 75% reliability (cf. 5 in Table 5).

(c) To account for roughly 40% of the total communication, 12 people are necessary to achieve 50% reliability (cf. 4 in Table 5).

(d) To account for roughly 40% of the total communication, 17 people are necessary to achieve 75% reliability (cf. 7 in Table 5).

TABLE 4. THE NUMBERS OF THOSE RANKED WHICH MUST BE RETAINED TO ACHIEVE S% OF A PERSON'S TOTAL COMMUNICATION

no. of guesses for		10%	20%	30%	40%	50%	60%	70%	80%	90% of total communication
person	1	1	2	2	2	2	2	2	4	18
	2	1	2	2	4	4	4	7	7	7
	3	1	1	1	4	4	5	5	13	13
	4	5	5	5	5	5	5	5	5	5
	5	1	2	7	8	9	9	9	–	–
	6	3	3	3	3	3	7	7	7	7
	7	10	10	10	10	24	24	24	24	24
	9	1	1	1	1	2	2	8	9	11
	10	7	8	8	9	10	10	16	20	22
	12	1	1	1	3	5	8	12	12	12
	14	1	1	1	1	1	1	6	6	–
	16	1	1	1	1	1	1	2	4	5
	18	4	4	4	4	4	4	4	4	4
	19	1	2	2	2	2	2	2	13	18
	20	1	1	1	1	1	5	5	5	7
	21	4	4	4	4	4	6	6	6	6
	24	3	5	5	10	10	14	14	14	19
	25	1	1	1	1	6	6	6	16	16
	26	7	7	7	7	7	7	13	13	13
	27	7	7	7	7	7	7	7	7	7
	32	1	1	1	1	1	2	26	26	26
mean		2.95	3.29	3.52	4.19	5.33	6.24	8.86	10.75	12.63
S.D.		3.59	3.83	4.07	4.76	6.96	7.65	10.54	11.81	13.19
mode		1	1	1	1	1/4	2	2/5/7/6	4/5	7
median		1	2	2	4	4	5	7	9	16

TABLE 5. NUMBER/% OF TOTAL RANKINGS NEEDED TO OBTAIN A GIVEN PERCENTAGE OF TOTAL COMMUNICATION

Percentage of communication	10%	20%	30%	40%	50%	60%	70%	80%	90%
Required Reliability									
50%	1 3.2	2 6.5	2 6.5	4 12.9	4 12.9	5 16.1	7 22.6	7 22.6	11 35.5
75%	4 12.9	5 16.1	5 16.1	7 22.6	7 22.6	7 22.6	12 38.7	13 41.9	18 58.1
90%	7 22.6	7 22.6	7 22.6	9 29.0	10 32.3	10 32.3	16 51.6	20 64.5	24 77.4
95%	7 22.6	8 25.8	8 25.8	10 32.3	10 32.3	14 45.2	24 77.4	24 77.4	26 83.9

Hence, people generally rank better than might be expected by chance but reliable data still requires an inordinate number of rankings.

VIII. Correlations of Rankings with Scores

We now turn to direct methods of scoring an informant's accuracy. As the first of these we used the correlation between where i ranked j and what proportion of i's communication with the group was actually *with* j. The ranks were subdivided into 1, . . . , K (where K is some integer greater than 0) and "the rest," which was given the numerical value of (K + 1) but contained all the other proportions of communication with persons ranked more than K. (In fact, the proportions of communications with unranked persons can be added to "the rest" without altering the conclusions.) For comparison, E_{yx} (eta) was also calculated.

For undistinguished contacts, r hovers around −0.3 for all K, its modulus reaching a maximum of 0.33 at K = 10 and 11. Such values are highly significant, although of course no systematic trend can be seen if the data are plotted. E_{yx} naturally takes higher values and increases monotonically from 0.3 at K = 1 to 0.412 at K = 30. The difference between E_{yx} and r is highly significant except for K = 2 and 3.

For "to" contacts, r is still negative, but its modulus only attains 0.24 around K = 11. These values are also highly significant. E_{yx}, lying around 0.3, differs significantly from r only for K less than 9.

For "from" contacts, r is of order −0.28 and E_{yx} about 0.4. These values continue to be highly significant, and E_{yx} is significantly greater than r.

There are several points worth making. First, correlations make dubious measures of accuracy; they need to be about 0.9 before any relationship becomes obvious. However, such negative correlations do confirm the rough hypothesis that people tend to rank higher the people with whom they communicate more. Second, the nonlinear relationship between rank and communication implied by the difference between E_{yx} and r is a little surprising. One might have expected little net nonlinear effect despite presumable nonlinearities for small rankings. Third, small violent extrema can significantly increase E_{yx} and |r| without indicating any connection between the variables being correlated. Examination of the data shows many examples of this.

Thus, the correlation, if used as a measure of accuracy, gives a two-valued result: on the one hand, this data yields significant correlations between ranks and proportional communication; on the other hand, the values of the correlations are *not* large, indicating that the relationship between ranks and proportional communication is not a strong one.

IX. The Â and A scores

We now consider two more stringent ways of scoring an informant's accuracy. The first of these is the Â coefficient. The correlation used in the last section suffers as a scoring device because it mitigates equally against any form of error. However, the assumption among sociometricians has always been that people, in some undefined sense, "know" better who their first few friends or contacts are than their less frequent contacts. This assumption has obvious intuitive appeal. Hence, a better scoring technique should take this assumption into account.

The simplest way to do this is to define a "chunking number" M, say. Call ranks 1-M group 1, and ranks (M + 1)-31 group 2. To obtain $\hat{A}$ we simply count the errors between the guessed rankings, and the correct rankings obtained from the interaction data. An error is defined to be when a person whose true rank lies in group two is incorrectly guessed to be in group one. Thus, the maximum number of possible errors is M when the first M guesses are of people whose true ranks are all greater than M. Similarly, the minimum number of possible errors is zero, when the first M guesses are of people whose true ranks are the integers $1, 2, \ldots, M$ in some order.

$\hat{A}$ is then defined for each informant, and for each M, by

$$\hat{A} = 1 - \frac{\text{Number of errors}}{\text{Maximum number of errors}}$$

and lies between 1 (zero errors) and 0 (maximum errors). Modifications must be made for those who ranked few people and for those who communicated with few people by replacing M by

$$M' = \min \left\{ \begin{array}{l} \text{M, number ranked, number} \\ \text{communicated with.} \end{array} \right\}$$

The advantage of $\hat{A}$ as a scoring device is that it does not demand strict accuracy in ranking the informant's first M choices; merely that he rank them *in some order* from 1 to M. Thus, $\hat{A}$ for $M = 1$ merely indicates whether the informant ranked his most frequent interactor correctly ($\hat{A} = 1$) or not ($\hat{A} = 0$). The disadvantage of $\hat{A}$, for our data at least, is the necessity to replace M by M' as above. Since the mean number communicated with was 4.61, chunking much beyond 5 yields no change in most informants' scores.

We thus chose to limit attention to $M = 5$. The scores were not encouraging: mean 0.31, S.D. 0.25. The maximum score was 0.8 (i.e., four out of the first five ranked in the first five), achieved only once. What these scores indicate is that people do *not* "know" their first few contacts very well (in fact, they know less than one-third of them). With such bad scores, it is perhaps not surprising that we could find no significant differences between subgroups based on any of the following criteria: number of people ranked, "to" or "from," amount of communication to the group, amount of communication to the world at large, or the proportion of total communication which was with the group.

The $\hat{A}$ coefficient, although revealing how poor is the informants' cognition about their communicants, is nonetheless perhaps too biased toward the highly ranked contacts. Clearly, interchanging one's 24th and 25th ranked communicants is not such a critical mistake as interchanging one's 1st and 2nd communicants, but it is still a mistake. However, interchanging 1st and 3rd is more of a mistake than interchanging 1st and 2nd. What is required, then, is a measure of accuracy which weights the errors according to their magnitude and to where they appear. A measure such as the Spearman rank coefficient weights errors only according to their magnitude (although, of course, its primary function is a correlation).

The A coefficient, then, is defined as a suitably weighted rank coefficient as follows: define for each i, two vectors g_j, r_j which are the guessed and actual ranks of person j as perceived by person i. Then A is defined by

$$A = 1 - \frac{\sum_j (g_j - r_j)^2 / r_j^s}{\sum_j (\bar{g}_j - r_j)^2 / r_j^s}$$

where s is a (so far undefined) non-negative power, and $\bar{g}_j$ are the set of "worst possible" guesses which i could have made. Then A, like $\hat{A}$, lies between 0 ($g_j \equiv \bar{g}_j$) and 1 ($g_j \equiv r_j$) as desired.

Let us examine the rationale of the A coefficient. It consists of summing the errors $(g_j - r_j)^2$ made about each j, just as in the Spearman coefficient, except that these errors are weighted by r_j^s, so that errors for which r_j is large are less important than those for which r_j is small. The *degree* of the weighting depends on the power s. If $s = 0$, A is linearly related to Spearman's R (except that ties appear in both guesses and ranks), and there is no weighting by rank. If $s = \infty$, only the j for which $r_j = 1$ is summed, and the accuracy reduces to a test of how far away from 1st i ranked that j.

The only problem is to define a suitable s. After experimentation with the data, we found that $s \geqslant 1$ yielded, for most informants, too high a score to be intuitively acceptable; in addition, there was little discrimination between what we felt to be reasonably accurate and highly inaccurate informants. A final value of $s = 0.5$ was settled upon.

Finally, the worst possible guesses, $\bar{g}_j$, are easily obtained. They consist of reversing the rank order of the r_j and then removing ties (since a decision to rank individuals who are really tied is itself an error).

The mean score obtained was 0.74, with a S.D. of 0.11. For comparison, other values are: for $s = 0$, the mean was 0.55; for $s = 1$, the mean was 0.86; higher values of s yield higher mean scores. Correlations (single, multiple, and partial) were calculated between all combinations of scores, number ranked, and communication with the group. The only significant correlation was between number ranked, and scores of −0.465; it rose to −0.494 with the effects of amount of communication partialled out. (This correlation is also significant for $s = 0$, but not for $s \geqslant 1$.) This indicates that there is a tendency for persons who rank fewer people to score higher. Part of this effect is because ties in the guesses (i.e., an informant could rank no further) coinciding with zero communication (i.e., tied ranks) produce less error than fully ranked guesses. However, this reflects some extra knowledge on the part of such informants, who, thus, deserve to score higher.

The significance of the above correlation was reinforced when we examined the effects of the same group parameters on the scores as were conducted on Â. The only significant differences in scores were obtained between those ranking less than, and those ranking equal to 31. The same was true of merely "to" communication, but there were *no* significant differences for "from" communication. This suggests that people better understand their "to" communication, than their "from," in contradiction to the results of Section V. No complete solution can be provided for this with only one set of data, of course. At best, we can say that the conclusion here is probably more relevant to the general problem of informant accuracy; stronger conclusions must await more data.

X. *"How Stringent Does One Have to Be?"*

Undoubtedly the most stringent question to ask of our data is simply: "i states that he ranks j nth. Did he?" If it should turn out that this question yields disastrous results, we may then ask, "How much leeway should we give i in his rankings before we define him to be in error?" In other words, instead of insisting that j should genuinely be nth, we might allow j to be ranked (n ± k)th, where k is some hopefully small integer.

We asked these questions of our data for a variety of k, and the results are summarized in Table 6, from which the overriding indication is that people only seem capable of ranking their most frequent communicator with any accuracy—and then only half the time! An indication of how badly people fare can be obtained from the last row of the table. Only 24% of the time will

TABLE 6. PROBABILITIES THAT THE PERSON RANKED NTH HAS AN ACTUAL RANK LYING IN THE RANGE (N ± K), FOR N ⩽ 6

k \ n	1	2	3	4	5	6
0	0.29	0.10	0	0.05	0	0
1	0.52	0.19	0.05	0.10	0.10	0
2	0.52	0.19	0.05	0.14	0.14	0
3	0.52	0.19	0.05	0.24	0.24	0

an informant, who ranks j fifth, actually talk to that person 2nd, or 3rd, or . . . , or 8th most frequently; 76% of the time one cannot count on that.

If we examine the significance of these probabilities, compared with the null hypothesis that there is a probability of $\frac{1}{31}$ of ranking anyone correctly (the interaction terms are small and are neglected for our purposes), then all the probabilities in Table 6 could easily be attained by chance save for those in column 1. Recent literature cited earlier has dealt with the effects of measurement error in normal sociometric questions. The argument has been that restriction of choices to three, or seven, or whatever, biases the results away from revealing any potential group structure. As convincing as these arguments have been, we reluctantly conclude that choices beyond one are totally untrustworthy; group structures based on one-choice sociograms are unlikely ever to yield significant results.

Despite this result, we performed an analysis of variance between the probabilities evaluated on the data of those who ranked less than 31, and on those who ranked 31. Not only was there a significant difference between ranks, as expected, but also the probabilities calculated on those who ranked all 31 were significantly greater than the others, for leeways of ± 1, ± 2, and ± 3. Again, this conclusion apparently contradicts earlier results, showing how important it is to define precisely what one means by informant accuracy.

XI. *Effects of Inaccuracy on Structural Models.*

The previous sections showed the degree of inaccuracy within informants' cognition about their behavior. We now inquire as to the effects of this error on structural interpretations of the data. To do this we use the routines detailed by Holland and Leinhardt (1975) for the enumeration of triads and calculation of the appropriate τ measures. τ measures whether a given structure (such as transitivity, for example) occurs more or less than chance. A τ of zero indicates chance; positive τ above chance; negative τ below chance. Conveniently, values of τ are in units of standard deviation, so that $|\tau| > 1.96$ is significant.

The τ calculation was designed by its inventors for use, in a purely statistical sense, on many (~800) data sets in an attempt to locate structural signals among the noise of measurement error. Holland and Leinhardt have shown, furthermore, that certain types of measurement error are well filtered by their method.

Despite both the above remarks, and the fact that there is no reason to expect affective-type structure in effective data, we ran Holland and Leinhardt's program[3] to seek various structures (describable by triad censi) in both cognitive and behavioral data. The results are summarized in Table 7. The left column names and describes the ten potential structures we examined. Across the top is shown the cut-off K (defined as before) used to construct a sociogram (whose entries are 1 if the appropriate ranking is $\leqslant$ K, 0 if $>$ K). Each pair of entries shows the τ measures for the appropriate structure: cognitive above, behavioral below.

The results are quite startling. Systematically, the actual communication shows no structure whatsoever; equally systematically, the cognitive pattern shows strong structure on all plausible structures, and significant lack of structure on all implausible structures (e.g., intransitivity).

In short, the cognitive and actual communications not only differ from each other–which we have interpreted as informant inaccuracy–but also in that their statistical structural analysis yields quite different answers. In other words, informant inaccuracy has here *produced* structural signals which did not exist in the actual communications pattern, in conflict with the findings of Holland and Leinhardt (1973). This suggests, then, that methods and theory may have to deal with the problem that cognitive and behavioral structures are quite different entities. Moreover, since we are unable to locate plausible structures in the behavioral data, we have to assume that either (a) behavioral data are chaotic; or (b) the structural signals are far more global than those found in cognitive data. From these limited data, it would appear that people impose a cognitive structure on their interactive behavior with others when such a structure either does not exist or is too large (global) to comprehend.

XII. *Discussion*

Before discussing our conclusions, the limitations of our data must be acknowledged. In the first place, this is, to our knowledge, the only set of such data extant;

TABLE 7. OCCURRENCES OF VARIOUS POSSIBLE STRUCTURES IN THE DATA

Structure	Cutoff K 1	2	3	4	5	6	7	8	9	10
1. Transitivity	–	+	–	–	+	+	+	+	+	+
i→j, j→k⇒i→k	–	–	+	+	+	+	+	2.08	2.16	3.49
2. Intransitivity	–	–	+	–	–	–	+	+	+	+
i→j, j→k⇒i↛k	+	+	–	–	–	–2.16	–	–2.09	–2.18	–3.26
3. Positive balance I	–	–	–	–	+	+	+	+	+	+
i↔j, i→k⇒j→k	–	–	+	+	+	+	+	2.05	+	3.26
4. Positive balance II	+	+	–	–	–	–	–	–	–	–
i↔j, i↛k⇒j↛k	–	–	–	–2.60	–2.11	–2.07	–2.40	–2.03	–2.07	–2.74
5. Positive imbalance	–	+	+	+	–	–	+	+	+	+
i↔j, i→k⇒j↛k	+	+	–	2.12	+	+	+	+	+	+
6. Anti-transitivity I	–	–	–	–	–	–	–	–	–	–
i→j, j↛k⇒i↛k	–	–	–	–4.17	–3.62	–4.00	–3.68	–3.78	–3.52	–5.14
7. Anti-transitivity II	2.32	+	+	+	+	–	–	–	–	–2.21
i↛j, j→k⇒i↛k	+	+	2.07	+	+	2.99	2.92	2.92	2.99	3.52
8. Anti-intransitivity I	+	+	+	+	+	–	+	–	+	–
i→j, j↛k⇒i→k	2.20	+	2.56	5.65	5.16	6.05	4.78	4.57	4.00	5.23
9. Anti-intransitivity II	–2.44	–1.98	–2.25	–	–	–	+	+	+	+
i↛j, j→k⇒i→k	–2.28	–2.54	–2.33	–2.40	–2.37	–3.47	–3.88	–4.22	–4.31	–5.46
10. Cyclic behavior	–	–	–	–	+	+	+	+	+	+
i→j, j→k⇒k→i	–	+	–	–	+	+	+	+	1.98	2.19

The table shows the τ statistic (Holland and Leinhardt 1975) computed on the data discussed here. The left-most column shows the potential structure being examined (the notation i→j means that person i chooses person j in the derived sociomatrix; a slash indicates negation). The other 10 columns contain the τ values for cutoffs 1-10. The *upper* value is based on the *actual* structure; the *lower* value is based on the *cognitive* structure. Only τ values greater in modulus than 1.96 (the 5% significance level) are given numerically; the others are indicated by sign only. On only 5 occasions out of 100 is significance attained by the actual data.

therefore, generalizations are premature. Second, these are network (i.e., effective) rather than sociometric (i.e., affective) data. We have no way of examining the "accuracy" of affective data; nor do we know how to define such a concept. Third, a period of three weeks (disjointed, at that) is probably too short a time to catch the full range of an informant's communications. Perhaps it isn't; we simply do not know.

Fourth, we asked people to give us rankings based on the situation as it was (and, therefore, implicitly as it had been). However, the logging would relate to the situation as it was going to be over the period *after* we obtained the rankings. Cognition about past behavior can only be an adequate predictor of future behavior providing (a) the situation remains unchanged over time; and (b) the cognition was correct in the first place. Social groups are time dependent, and this places obvious limitations on our results; much of the discrepancy between cognition and reality can be ascribed to time dependence. One way to remove much of this effect would be to ask for rankings at the end of the logging period as well as at the beginning. There are some serious logistic problems in doing this; for example, if informants know they will be asked to make a post hoc ranking, they will be subject to a strong temptation to peek at the logs before doing it.

This raises further problems which are not inherent to our data, but which are part of the data collection technique. (a) The process of ranking forces informants to think about their interaction with others in a manner which is probably "unnatural." Moreover, the more rankings an informant makes (especially if he is keeping logs) the more accurate he is likely to become.

(b) It is not clear whether people should rank others in a group or provide a scaling (0-10, say) of their interaction with others. It may be easier for people to be accurate (in the ways we have defined here) if they are given the opportunity to assess their interaction with others on an individual basis. Other definitions of accuracy might then be necessary.

(c) The criterion for ranking which we used was "total communication." Other, equally plausible criteria include frequency of contact, importance of communication, and perhaps "content" of communication. It is clear only that much testing is required before any or all of these criteria can be selected as the "best" for the collection of network data.

In spite of all these limitations, one conclusion stands out from our data: people simply do not know, with any degree of accuracy, with whom they communicate. A graphic representation of this may be seen in Figures 2 and 3. (We have omitted several persons because they did not return any logs. However, some of these persons were quite accurately ranked by those who did return logs.) Only 29% of those guessed as being the top communicant were, in fact, communicated with the most. Similarly, only 34% of those guessed as being in the top five communicants were, in fact, among the top five communicated with.

Consider the implications of even the more lenient case in Figure 3. Of all the guesses made, 66% are erroneous. This is, in itself, rather worrying for network analysis. More to the point, however, we have no way to know which the erroneous guesses are. The previous sections reveal no *systematic* effect on accuracy of any of the parameters we examined, implying that informant error is probably independent of these parameters. This leaves the unpleasant possibility that error is produced by one or more psychological or sociological factors which we will have to uncover before we can know the accuracy of any data.

Even if it can be determined, by analysis of these factors, whether a large amount of error exists in a given set of data, it is not clear what to do about it. Our evidence suggests that error is fairly random; this is rather different from the errors considered in the theoretical papers cited earlier. The results of section XI suggest that informant error produces rather strange answers when compared with the well-known structural models of balance, transitivity, and the like. Specifically, informants' *cognition* yields systematic significant results at all levels; informants' *actual* communication yields no structure at all. Furthermore, recall that the structural indicators used are independent of the problem of lack of fit between the cognitive and behavioral data. It is not at all clear how to interpret this.

One may reasonably question these rather negative observations about sociometry and network analysis. After all, thousands of practical users have found sociometric descriptions both plausible and useful over the years. Similarly, in an effort to check the accuracy of our own network techniques (Bernard and Killworth 1973; Killworth and Bernard 1974), we have taken the results back to the group concerned. We have been encouraged by apparently knowledgeable members of these groups to believe that our descriptions are accurate. We assume that sociometricians also believe their descriptions to be accurate.

Consider, however, this one last thought: comparison of sociogram or network data with a group's cognition is bound to be successful since the data represent the group's cognition in the first place.

FIGURE 2. SOCIOGRAM AND REALITY FOR 1ST RANKED COMMUNICANT ONLY. FIGURE 2a SHOWS THE ACTUAL LINKS BETWEEN 1ST COMMUNICANTS TAKEN FROM THE LOG SHEETS. FIGURE 2b SHOWS THE GUESSES FOR 1ST COMMUNICANTS MADE BY THE INFORMANTS. FIGURE 2c SHOWS THE OVERLAP BETWEEN FIGURES 2a AND 2b. NON-LOG-KEEPERS ARE OMITTED.

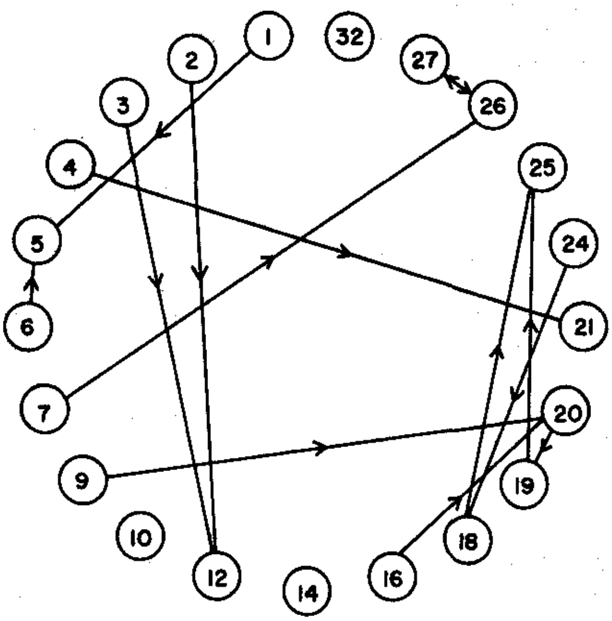

Figure 2(a)

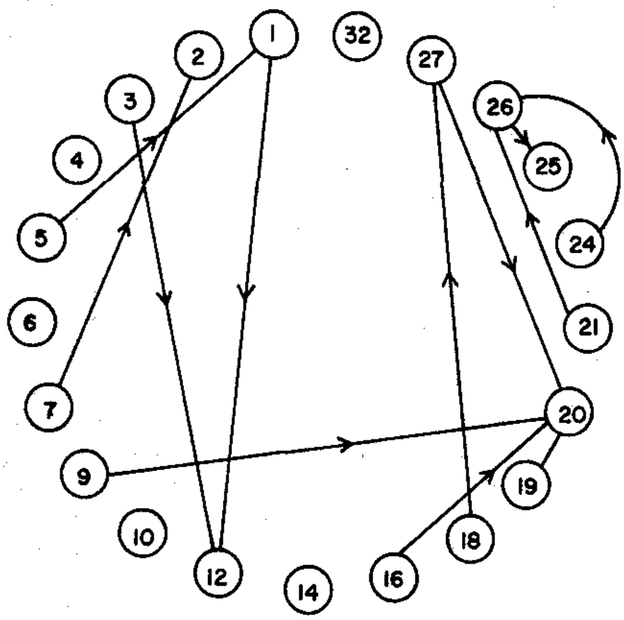

Figure 2(b)

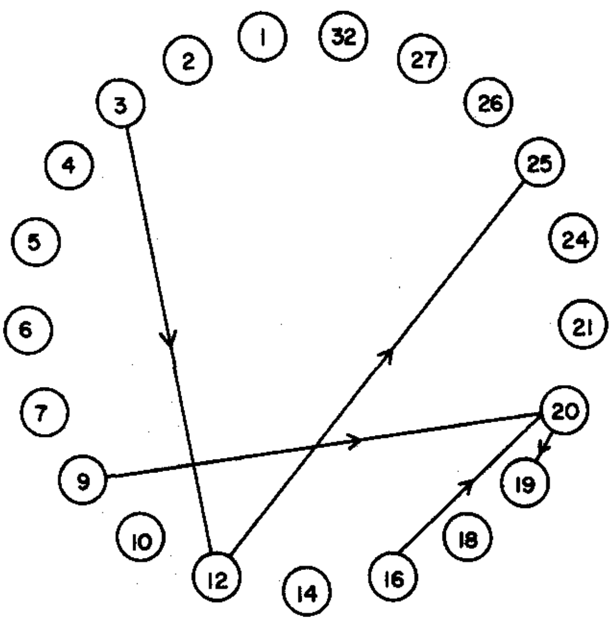

Figure 2(c)

FIGURE 3. SOCIOGRAM AND REALITY FOR FIRST 5 RANKED COMMUNICANTS. FIGURE 3a DRAWS LINKS FROM i TO j IF i ACTUALLY TALKED TO j ANYWHERE FROM FIRST TO FIFTH MOST FREQUENTLY. FIGURE 3b DRAWS LINKS FROM i TO j IF i ESTIMATED THAT HE TALKED TO j ANYWHERE FROM FIRST TO FIFTH MOST FREQUENTLY. FIGURE 3c SHOWS THE OVERLAP BETWEEN FIGURES 3a AND 3b. NON-LOG-KEEPERS ARE OMITTED.

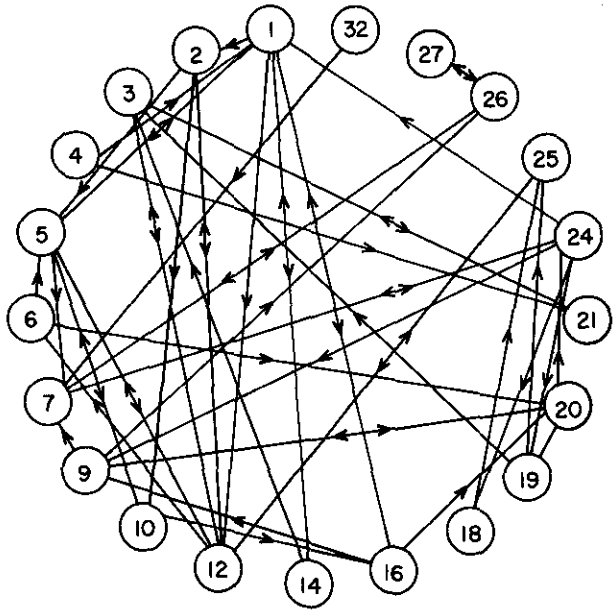

Figure 3(a)

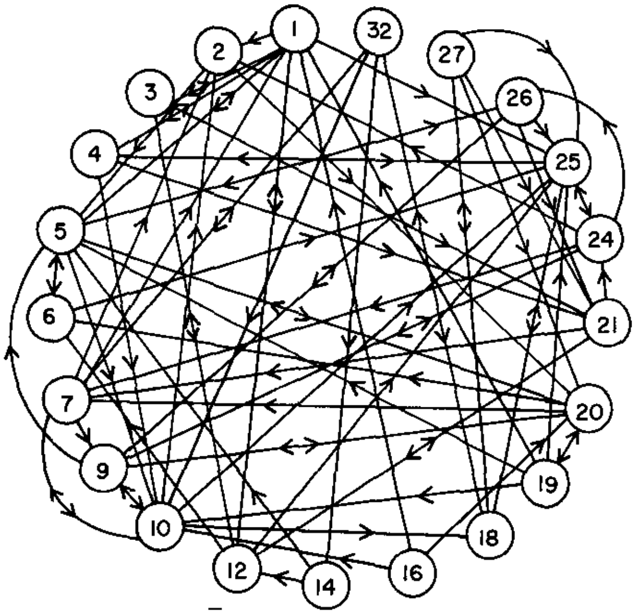

Figure 3(b)

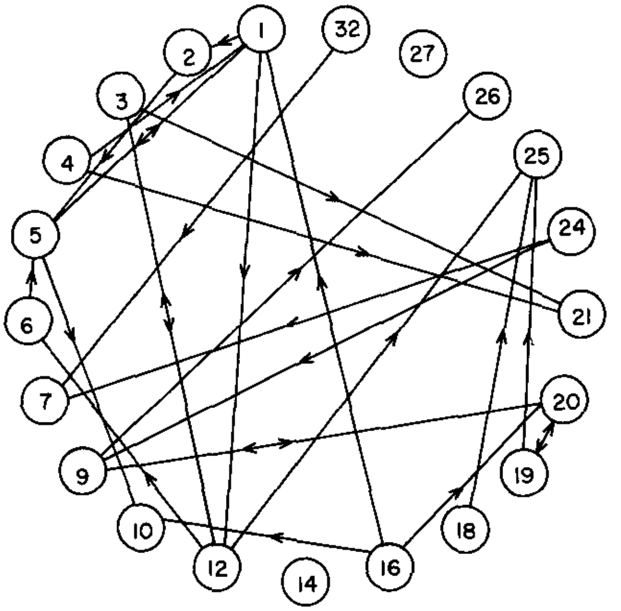

Figure 3(c)

NOTES

1. It is unknown whether avoidance of this question or the affective bias came first in sociometry. To the best of our knowledge, Tagiuri, Blake, and Bruner (1953) are the only workers ever to examine affective accuracy. At that, they were forced to "measure" accuracy of an informant's guesses about how others would feel about him, in terms of those others' statements. Unfortunately, no indication of *level* of accuracy was given. The problem of accuracy in affective data continues to defy solution.
2. In a problem solving laboratory situation, Chapanis (1975) has shown that typing speed and ability is irrelevant to the total flow of information. (It turns out that all TTY users–even highly skilled ones–spend a lot of time thinking rather than typing.)
3. Our thanks to Sam Leinhardt and Paul Holland for use of their computing facilities in running the statistics shown in Table 7. Their facilities are supported by NSF Grant SOC-73-05489 to Carnegie Mellon University. Our thanks also to Richard Dietrich for help in accessing these facilities and in running the data.

REFERENCES CITED

BERNARD, H. R., and P. D. KILLWORTH

1973 On the social structure of an ocean-going research vessel and other important things. Social Science Research 2:145-84.

1975 On the structure of affective and effective sociometric relations in a closed group over time. Technical Report #BK-107-75, Office of Naval Research, Washington, DC.

BJERSTEDT, A.

1956 The Methodology of Preferential Sociometry. Lund: University of Lund Press.

CHAPANIS, A.

1975 Interactive human communication. Scientific American 232(3):36-50.

DEUTSCHER, I.

1966 Words and deeds: social science and social policy. Social Problems 13:235-54

HALLINAN, M.

1972 Comment on Holland and Leinhardt. American Journal of Sociology 77:1201-05.

HOLLAND, P., and S. LEINHARDT

1973 The structural implications of measurement error in sociometry. Journal of Mathematical Sociology 3:85-112.

1975 The statistical analysis of local structure in social networks. *In* Sociological Methodology, D. F. Heise, ed. San Francisco: Jossey-Bass.

KANDEL, J.

1974 The quantification of networks. Paper presented at the 73rd Annual Meeting of the American Anthropological Association, Mexico City.

KILLWORTH, P. D.

1974 Intransitivity in the structure of small closed groups. Social Science Research 3:1-23.

KILLWORTH, P. D., and H. R. BERNARD

1974 Catij: a new sociometric and its application to a prison living unit. Human Organization 33:335-50.

REISS, A. J., JR.

1971 Systematic observation of natural social phenomena. *In* Sociological Methodology, H. Costner, ed. Washington: Jossey-Bass.

TAGIURI, R., R. R. BLAKE, and J. S. BRUNER

1953 Some determinants of the perception of positive and negative feelings in others. Journal of Abnormal and Social Psychology 48:585-92.