University of Pennsylvania
ScholarlyCommons

Departmental Papers (ASC)

Annenberg School for Communication

9-1970

On Generating Data in Communication Research

Klaus Krippendorff

University of Pennsylvania, kkrippendorff@asc.upenn.edu

Follow this and additional works at: https://repository.upenn.edu/asc_papers

Part of the [Communication Commons](#)

Recommended Citation

Krippendorff, K. (1970). On Generating Data in Communication Research. *Journal of Communication*, 20 (3), 241-269. <https://doi.org/10.1111/j.1460-2466.1970.tb00883.x>

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/asc_papers/273
For more information, please contact repository@pobox.upenn.edu.

On Generating Data in Communication Research

Abstract

Communication research requires data that are rich enough to contain explicit evidence about processes of communication. The paper formally distinguishes among and discusses three kinds of data:

aggregational data (such as accumulated in much of psychological inquiries, small group experiments and survey research) and *network data* (representing observable patterns of interest to relational biology and sociology, for example) are found to be not rich enough to qualify as *communication data* proper.

The paper describes an elementary form of communication data that would contain explicit evidence about the process in question. It exposes some conceptual degenerations in communication research as a consequence of heavy reliance on inadequate data. Pointing to the fact that the advancement of knowledge is critically linked to the ability to process a certain kind of data, the paper concludes that communication research must develop new analytical techniques that are compatible with its most basic concept: communication.

Disciplines

Communication | Social and Behavioral Sciences

On Generating Data in Communication Research¹

KLAUS KRIPPENDORFF

Abstract

Communication research requires data that are rich enough to contain explicit evidence about *processes of communication*.

The paper formally distinguishes among and discusses three kinds of data: *aggregational data* (such as accumulated in much of psychological inquiries, small group experiments and survey research) and *network data* (representing observable patterns of interest to relational biology and sociology, for example) are found to be not rich enough to qualify as *communication data* proper.

The paper describes an *elementary form* of communication data that would contain explicit evidence about the process in question. It exposes some conceptual degenerations in communication research as a consequence of heavy reliance on inadequate data. Pointing to the fact that the advancement of knowledge is critically linked to the ability to process a certain kind of data, the paper concludes that communication research must develop new analytical techniques that are compatible with its most basic concept: communication.

In the field of communication research there exists considerable confusion regarding appropriate investigative methods. Assertions about the importance of communication processes for the functioning of society and individual well being cannot provide a logical basis for inquiries into the phenomena, nor can they define disciplinary boundaries. As part of a larger methodological concern, this paper is concerned with the nature of data relevant or irrelevant to inquiries into communication processes. It will be argued that the form of communication data is fundamentally different from much of traditional behavioral science data. This presents a unique analytical challenge which needs an appropriate response.

Klaus Krippendorff (Ph.D., University of Illinois) is associate Professor at The Annenberg School of Communications, University of Pennsylvania. His current interest lies in the methodology of communication research, particularly of content analysis and of applying cybernetic models to social processes.

Figure 1

The role of data in scientific inquiries

The Role of Data in Scientific Inquiries

The concept of "data" as used in this paper may be elucidated with the help of the diagram in Figure 1. The upper part of the diagram is adapted from Coombs' *A Theory of Data* [11:4]. Here it is suggested that, as a first step, the behavioral scientist must decide on the observations to which attention is to be paid. These observations are not yet data, however.

In a second step, observations have to be formalized, i.e., they have to be identified, labeled, classified, measured or "translated" into terms amenable to subsequent analysis. This step is often referred to as the making of fundamental measurements [20], as data-making [32] or as recording. It should be pointed out that the outcome of this step—data—cannot be considered objective in the usual sense of the word. In deciding what is relevant to an investigative problem and how observations are structured and assigned to the formal terms of an analysis, the nature of data reflects the scientist's orientation and values.

The third step in the diagram involves the detection of relationships, order and structure in data and produces explicit evidence for inferential classifications. Evidence of this kind follows as a logical consequence of the data and of the assumptions that are implicit in the analytical procedure. Usually, there are many ways of analyzing data and the choice among techniques for transforming data into evidence must be justified in the light of the scientist's analytical objectives.

Scientific inquiries often involve a fourth step at which deductively derived evidence is inductively related to potential

observations. Analytical procedures, it should be pointed out, operate exclusively on what might be called syntactical structures which are manifest in the formal terms of data. Evidence should represent such structures in a theoretically meaningful way. In contrast, this fourth step might be said to provide a semantic interpretation of evidence in the sense that the terms in which the evidence appears are assigned significations and denotations. The distinction between analysis and interpretation is important because only the former can be conclusive in the strict logical sense. The latter always involves some uncertainty, nonverifiable assumptions, and guesswork.

As Coombs has noted, the behavioral scientist enters each of these steps in a creative way in the sense that alternatives are exhausted by him and his decisions will significantly determine the results obtained from the analysis. Thus the universe of available events can be the same for a large number of disciplines, but what is observed and what is extracted from it may be quite different depending on (a) the researcher's analytical objectives, (b) his theoretical commitments, and (c) the investigative tools available to him. For example, a psychologist may study society as a statistical aggregate of individual properties; a sociologist might consider the same as a network of roles, while a biologist would perhaps analyze his observations in organismic terms.

The lower part of the diagram in Figure 1 suggests a bypass of the explicit analysis. Intuitive conceptualizations that may have been formed directly from observations can control the research results to a more or less significant degree. Judgments regarding "face validity" exemplify such a case. However, truly critical situations arise where the research results are claimed to be supported by evidence that cannot be derived from data by explicit techniques. In such cases uncontrolled information must have entered the semantic interpretation surreptitiously. This need not result from dishonest intentions. Rather, such situations may stem from a confusion of explicit evidence with a semantic interpretation thereof. And this confusion may stem from a lack of understanding of the formal nature of data, and particularly of the kind of evidence that can or cannot be deduced from available data.

A comparison of the informational content of a string of symbols, such as the words in a natural language with a frequency characterization of the symbols in this string, provides an obvious example. A frequency distribution can be derived from this string in a logically determinate way. However, the operation cannot be reversed without supplying information about the order relationships among those symbols. If someone claims to have studied the grammar of a language from data that consisted of nothing but word counts, the internal validity of the analytical process is in question on the ground that the data base is insufficient and hidden information must have entered the process somewhere.

It is possible to ascertain, on purely formal grounds, whether data can answer given questions, i.e., whether or not the data base of an inquiry is rich enough to contain relevant information. Decisions of this kind presuppose that the concepts for which evidence is sought are describable formally or syntactically. If the form of data is such that it cannot be mapped (in the mathematical sense) onto the form of the desired evidence (corresponding to the concepts under consideration) then the structure of the data is inadequate, and can be rejected as irrelevant to the desired evidence.

The Notion of "Communication Process"

Several books and a large number of articles devote at least part of their discussion to how communication can be apprehended intuitively. A review of such concepts goes beyond the scope of this paper. Let it merely be stated here that some authors consider it a process by means of which one mind affects another [23]; some writers assign crucial importance to "messages" as mediators between communication agents [16]; and others identify it as the fabric of society or as the web that holds any kind of organization together [3, 12, 36]. When a group of people talk to each other, it is usually assumed that communication is taking place. But scholars disagree considerably as to whether a mutual understanding is a necessary prerequisite for deciding whether communication had occurred. The audible exchanges among animals that facilitate coordination of their social interaction can also be regarded as communication processes. Here the objections come from scholars with anthropocentric attitudes who wish to

limit such processes to inter-human activities. Similar controversies exist regarding the criterion of intentionality. The transmission of values and behavior patterns from generation to generation is known to be essential to the maintenance of a culture's identity beyond the biological life of its carriers. Formally similar is the transmission of genetic information from parents to offspring. It is difficult to see why the former process is often regarded as one of communication while the latter is rarely so interpreted in the social sciences. In both cases, ultimate receivers are difficult to determine and the imputation of purpose is doubtful.

Most of these debates are irrelevant to a theory of data. As stated above, in order to recognize communication processes in data, it is imperative that the intuitive concept of communication be explicated and stated in an appropriate technical language. Intuitively meaningful references to man, animals or machines or to the channels among them may guide the behavioral scientist's choice of observations. But such references cannot influence the evidence that data can provide unless they are themselves formalized and take their place in the form of data.

Elsewhere I have argued for a syntactical definition of communication. It seems that the only reasonable delineation of the empirical domain of the social communication sciences stems from a theoretical commitment, i.e., a commitment regarding the form of objectifications (theories, models, etc.) of real world events. Communication sciences do not characteristically view their objects as things or as combinations of properties, but as patterns of dynamic interdependencies. The theoretical commitment underlying the various approaches to communication research seems reflected in the understanding of "*communication*" as *a process of transmission of structure among the parts of a system which are identifiable in time and space* [22].

Such a formulation can readily be put into a mathematical form which can in turn be searched for in adequate data. It can also be given a variety of semantic interpretations thus showing it to be sufficiently general. The "systems" may be biological, cognitive or social in nature. The "structures" transmitted may involve genetic, linguistic or cultural patterns of meaning. The notion of "process" subsumes that of behavior or changes over time. "Trans-

Figure 2
Evidence provided by minimum and rich communication data

mission" implies information exchange and control over geographic distances. The emphasis on transmission "among the parts" directs attention to the complex fabric that integrates individual behaviors into organized wholes rather than to the nature of a system's individualized components.

Communication Data

The task of communication research is presumably one of providing conclusive evidence about the parameters according to which the process may be explained, predicted and/or controlled. However, there are virtually no limitations as to the number and kind of parameters that might be considered relevant for this purpose. Consequently, there is no upper limit as to the informational richness that communication data may exhibit. But there is a lower limit below which data remain meaningless as far as communication constructs are concerned. This seems to provide a more reasonable definitional criterion of communication data: communication data must provide explicit evidence at least about the existence or non-existence of communication processes. Figure 2 depicts the evidence that minimum communication data and, what might be called, "rich" communication data respectively provide. The analysis of minimum communication data merely involves a decision with two outcomes. The analysis of rich communication data may involve a large number of decisions regarding which of the possible parameters provides the best account of the process manifested in data. Evidently, making choices among the parameters of communication presupposes decisions regarding the presence of the process. Thus rich communication

data fully include the evidence that minimum communication data contain.

With this very rough framework in mind it is now possible to consider three kinds of data and ask in each case whether they can provide evidence for communication processes as defined.

AGGREGATIONAL DATA

The Form of Aggregational Data

In the sciences dealing with man, by far the most elaborate analytical methods are used in psychology. Here the basic theoretical commitment lies in the assumption that human behavior can be understood as a response to both environmental stimuli and characteristics of the individual, his personality for example. In addition, psychological schemes of explanation exhibit a minimum concern for the intervening physiological mechanisms [2]. Data that are gathered with these commitments in mind usually exhibit the following form:

- (1) There is either one set A or two distinct sets A and B of observations.
- (2) Each member of the set is described in terms of (classified or scaled along) one or more dimensions x , i.e., $A \subset \Pi x_r$ and/or $B \subset \Pi x_s$.
- (3) One kind of binary relation is specified between a collection of pairs of observations, i.e., $d(a_i, a_j)$ or $d(a_i, b_j)$ where d is either an order relation or a distance, $a \in A$, $b \in B$.

This is essentially Coombs' [11] analysis, which he developed in much greater detail than is required here.

Among the many semantic interpretations that can be given to aggregational data are the following examples: a set of individuals are asked to judge the appearance of a set of mass circulation magazines along semantic differential scales. Such scales constitute what Osgood *et al.* [27] call a semantic space, and distances between pairs of points are presumed to be indicative of effective similarities.

In experimental settings in which subjects are asked to indicate their preference between television shows, consumer products or works of art, individual responses specify an order relation be-

tween pairs of stimuli. Similarly, answers to such questions as the extent to which a subject desires to continue watching a violent movie or the extent of disapproval of a political candidate specify proximity relations between pairs of observations, concepts or individuals. Distances of this sort define a psychological space in which data may be represented as points.

It is important to note that observations are recorded and treated as collections (involving no inherent order) that define statistical distributions in point spaces. The analysis of such data invariably involves an aggregation, a statistical summation of individual cases. The recorded responses of a number of individuals to a set of political speeches may be used to classify the individuals involved or to group the speeches to which each was exposed. They may yield a simpler description of stimulus material involving fewer factors, or they may produce variables that best predict how individuals react to speeches of a certain kind. In either case the result is an aggregation of many responses by many individuals to many speeches.

The Scope of Aggregational Data

A few quite different semantic interpretations of aggregational data exhibit an extremely wide scope: the correlations between a variety of daily habits and the occurrence of cancer are based on aggregational data. The changes of stock market prices in response to political crises, the co-occurrence of riots with warm weather and other environmental factors, the effects of fertilizers on the growth of grain under a variety of climatic conditions, etc.—all are based on the aggregation of a large number of observations that are scaled or categorized along many dimensions and between which not more than binary relations are specified. Evidently, when aggregational data are stripped of their particular semantic interpretation, there is nothing peculiarly human or peculiarly psychological about it. It is in this sense that data are regarded here in syntactical terms and as objects of mathematics. Similarly the mode of analysis is independent of what the data may represent.

While aggregational data exhibit a wide scope they also have severe limitations: not all phenomena can be captured when data

are viewed as collections of the characteristics of individuals, stimuli, responses, etc. As Barton put it with reference to survey research in sociology:

"... using random sampling of individuals, the survey is a sociological meat grinder, tearing the individual from his social context and guaranteeing that nobody in the study interacts with anyone else in it. It is a little like a biologist putting his experimental animals through a hamburger machine and looking at every hundredth cell through a microscope; anatomy and physiology get lost; structure and function disappear, and one is left with cell biology [7].

The same metaphor applies when designing psychological experiments or examining statistical records of large-scale phenomena. Whether individuals, diseases, telephone calls or riots are aggregated, they must be regarded as independent of each other. Where the generation of data can to some extent be controlled, behavioral scientists take care to make sure that their units of enumeration are independent in fact.

In his paper, Barton reviews attempts to bring social structure "back in" survey data. Among the more prominent devices is to record individuals' perception of interpersonal environments, to request that sociometric choices be revealed, or to sample institutional settings and inter-institutional relationships. The intuitive link between social structure and the evidence that aggregational data may provide lies in the postulated effect of such a structure on its component parts: if such a structure has some degree of reality then it would presumably be reflected in the way individuals perceive each other or in the way inter-institutional relationships are distributed. For example, if there is something to a social group as opposed to a collection of strangers, sociometric choices among members should be expected to exhibit a non-random distribution. This is presumably the result of how members work together, talk with each other, etc. It is on the basis of such assumptions that aggregational data may become *indicative* of possible dependencies among the aggregated units. But these dependencies are not represented in such data.

Use and Limitations in Communication Research

Suppose a group of students is observed debating how to stage a sit-in or whether to support a political candidate. No doubt

we experience by any definition that members of this group communicate with each other and we may observe the results of such communication in the collective activities that follow. The most common way of mapping such a situation into aggregational data is to record what each individual receives and how he responds to it. The data that would be generated in this case come close to those of Bales [6].

It is not difficult to see that an aggregation of such data over the individuals concerned is effectively equivalent to the generation of data from experiments in which isolated individuals respond to the stimuli presented. From such data alone it is impossible to ascertain whether individuals were in fact communicating with each other during the observation period or whether they behaved in isolation. The behavior of isolated individuals or, more specifically, data that record the behavior of individuals as isolated regardless of the social experimental setting (which may or may not include communication activities), cannot provide explicit evidence for the existence of communication processes.

The situation would not change very much if attention were to shift from the individuals that are connected through a communication net to their verbal exchanges. The collection of "who says what to whom" or Newcomb's [26] co-orientation is only another interpretation of the collection of "what causes X's response" or, loosely, of stimulus-response connections. Both conform to the formal characteristics of aggregational data. Both define point distribution in multidimensional space in which distances may become psychologically meaningful. Both permit a variety of inferential classifications but neither contains evidence of the dynamics of the process or of the network characteristics of these exchanges. Aggregational data are simply inadequate to contain evidence of communication processes. They do not satisfy the minimum criteria for communication data.

What designates studies using such data as "communication research" is not the kind of evidence it provides; rather, it is how the results of such an analysis are interpreted semantically. Often, it is the observer's assumed authority as a communication researcher which leads him to decide whether the group members are in fact communicating with each other and whether the stimulus is a communication. The uncertain observer may assure

himself by asking the subjects if they felt that they did communicate. But such answers beg the question by not exhibiting what goes on among the individuals rather than within each.

Yet, "communication research" flourishes despite its heavy reliance on aggregational data and on analytical techniques that are so conveniently geared to this kind of data. This is evident in the literature on persuasion initiated by Hovland [19] and recently reviewed by Sears and Abeles [29]; in the work done on information diffusion [28]; in research into small groups phenomena [15] including communication nets [30]; and in inquiries into psycholinguistic processes [13], as well as in almost all the studies in mass communication reviewed by Tannenbaum and Greenberg [33].

The heavy use of aggregational data puts communication research on extremely uncertain grounds. The fact that evidence of communication processes cannot be deduced from such data has caused another unfortunate development: severe conceptual impoverishment. If the intuitive and powerful notion of communication process is stripped of its relational qualities and processual characteristics (as it is when aggregational data are generated), then communication reduces to nothing but another variable. Just as anxiety is thought to vary from low to high, attitudes are scaled along dimensions such as strong-weak and violence is measured in degrees, so is communication reduced to the more or less of one or the other quality. This is indeed the common denominator of most of the studies that are otherwise methodologically impeccable. The explanatory power of the notion of "communication process" is then virtually eliminated. Blumer [8] too observes inadequacies of sociological evidence relying on aggregational data. He argues from a slightly different perspective that "modes of interpretation" cannot be described by "variables" and that researchers who force their observations into such forms might not be able to tab the critical process intervening between overtly obvious phenomena. In other words, communication can then not be regarded as a mediational phenomenon.

NETWORK DATA

The Form of Network Data

In contrast to the psychological focus on the behavior of indi-

viduals taken as a whole, there is another research tradition which is concerned with organizations involving individuals. Theoretical commitments in sociology, political science and anthropology suggest explaining the social behavior of groups and their members in terms of the way *individuals are related to each other*, i.e., in terms of social structure.

Such concepts as "level of integration of political units," "alienation of a class of individuals," or "pattern-maintenance functions of an institution" are clearly *understood* as correlates of particular relations among individuals although they are rarely *defined* on those relations. This is presumably because analytical techniques are not as highly developed in these disciplines. The situation would be different if such properties are deduced from the recorded patterns of interaction, from organizational charts, from flow-diagrams, or from a graph of possible telephone connections from networks of interpersonal communication. To explain social behavior in terms of the relations among individuals or among societal units presupposes that the relations of interest are explicitly identifiable in relevant data, and not merely inferred from their presumed behavioral consequences.

The kind of data that would contain evidence about the way individuals, groups and organizations are related satisfies the following requirements:

- (1) There are many distinct sets $A, B, \dots, Z$ of observations.
- (2) Each observation may be described in terms of (classified or scaled along) one or more dimensions, i.e., $A \subset \Pi x_e$, $B \subset \Pi x_f, \dots, Z \subset \Pi x_m$.
- (3) One or more kinds of many-valued relations R are specified among the observations, e.g., $R_1(a, b, c)$, $R_2(a, b, c, \dots, z)$.
- (4) Some relations imply other relations, e.g., $R_1(a, b, c, d) \Rightarrow R_3(a, c, d)$.

For lack of a better term, I call these data *network data*.

A kinship system provides a good example of what network data may contain. Individuals may be described in terms of sex, age, maturity, etc. But of crucial importance are such relational attributes as lineal, colineal and ablineal descendent, and generational differences. It is the relationships which usually determine

the kind of interpersonal interaction that may emerge and how kinsmen perceive themselves as related. The way A is related to B usually implies how B is related to A and delineates, to some extent, how A and B behave toward each other. Adequate data on kinship would have to exhibit which relations are equivalent and which relations between A and B and between A and C imply which relations between B and C.

The way an organization analyst may chart lines of authority in a bureaucratic institution is another example of the use of network data. Observations regarding the scope of each office's supervisory functions and the formal and informal channels of influence connecting the individuals involved yield relational nets, typically hierarchies, from which, among other things, the possible paths of a command may be deduced. That A's direct authority over B and B's direct authority over C often implies A's indirect authority over C again exemplifies the implications inherent in network data. Recently Friedell [14] described organizations as semilat-ices using data from *Street Corner Society* as examples. He thus made use of a way to represent many valued dependencies as a hierarchy of relations.

In the study of communication processes, the mapping of networks is basic. Lasswell's question "Who says what to whom with what effect?" requires at least a relational answer. The practice of answering it as if it consisted of several independent questions seems linguistically legitimate but eliminates the crucial relational aspect of communication. When it is said that A communicates with B who in turn communicates with C, a transmission chain is identified. When, in addition, C communicates with A, a transmission loop is described. The use of digraphs in diagramming complex communication nets is common [18] even though the mathematical implications of this approach are not always utilized.

I should like to add that graphical representations of networks, e.g., using digraphs, are often misleading. Simply because it is difficult to connect more than two modes with one arrow, such descriptions readily suggest that relations between sources and receivers, for example, are exclusively binary in character. If a number of communicators coordinate their activity in a certain

way, the resulting relation among them may not be describable in binary terms. Tertiary relations (involving three individuals), quaternary relations, . . . , n -ary relations are difficult to visualize and present even more serious problems of analysis.

The important analytical task of manipulating network data is therefore a decomposition of many-valued relations into relations of a lower order [5]. Systems analysts talk about hierarchical decomposition [31]. Quantitative notions of the relative independence and density of a network, but more important, such mathematical properties of relations as symmetry, transitivity and reflexivity provide powerful analytical devices for the decomposition of network data.

Suppose we observe a large group of individuals at a social gathering. Unless highly institutionalized forms of behavior exist—such as at a public speech—it is very unlikely that interaction patterns involving, say, 1000 individuals need to be recorded. Perhaps the order of the manifest relations does not exceed twenty at any one point in time and relations that are persistent during such a gathering may not involve more than five individuals. Without any doubt such a finding would constitute a tremendous saving of subsequent analytical and conceptual efforts. Similarly does the web of a city's telephone connections look unmanageably complex to an alien. However, if the calls between any two stations are found not to be conditional on a third, then the apparently complex network can at once be decomposed into a set of binary connections. It is precisely when communication is nearly absent, such as among the drivers of cars, among people being interviewed in isolation from each other, or among the members of large mass media audiences, that mere aggregational properties appear significant.

Relational biology, on the other hand, has shown that organic functions often constitute an irreducible complex of mutual conditionality. If the connection between a pair of organs is interrupted, the whole complex may collapse. Mutual conditionality is also the essence of organization [3]. To treat social organizations as if they were fully decomposable—for example by interviewing each individual member—may miss their unique properties. Whether and the extent to which a many-valued

relation can be decomposed, i.e., whether such a relation can be considered as an aggregate of lower order relations, should not be a matter of *a priori* decisions. Data themselves must provide the answers.

Relation to Aggregational Data

The possibility of decomposition links aggregational data with network data. Aggregational data can always be interpreted as a collection of distance—or ordering—relations between pairs of points, i.e., as a collection of either one or another kind of binary relation. Thus, if an elaborate network can be fully decomposed into a set of binary relations with a metric interpretation and if the values (individuals, stimuli, etc.) of these relations are drawn from not more than two distinct sets, then, in this special case, network data and aggregational data are equivalent. Formally, aggregational data are a special case of network data, the latter being informationally richer and, as far as their explanatory potential is concerned, more powerful. This can already be seen as implied by their formal definition.

Often the underlying network of communication cannot enter the data because of (a) experimental constraints, (b) limited observational schemes, and (c) methodological biases. In gathering sociometric choices, for example, it is the experimental setting which precludes network data from emerging. While filling out the forms, communication among individuals is not permitted. Assurances to subjects that individual choices will not be disclosed to members of the group is another way of suppressing the emergence of higher order "choices." The fact that sociometric choices among known friends tend to correlate, shows that the method has considerable face validity. But it should be pointed out that one essential property of social structure is that choices of friends, or perceptions of social proximities and the like tend to be mutually conditioned. Such a conditionality is reflected in the possibility that A's choice of B—if communicated to C—will affect C's choice of D. This is already a quaternary relation which simply cannot emerge because the experimental setting systematically suppresses communication.

Another factor which prevents higher-order relations from ap-

pearing in data are powerless observation schemes. Bales [6] again demonstrates this point. The constraints which members of a freely interacting problem solving group mutually impose upon each other are simply ignored except where they affect binary relations of the type "A asserts C to B." Subsequent analysis often reduces these binary relations, of which Bales differentiates 16 different kinds, to properties associated with individuals. Thus even the minimal relational content of aggregational data is eliminated.

Shaw's [30] review of inquiries into communication networks within small groups reveals how methodological biases can obscure evidence about the effects of higher-order communication patterns. In this research tradition communication links among individuals are manipulated so that the effects of a variety of channel configurations on group behavior can be studied. However, since traditional methods do not readily provide descriptive devices for such networks (other than mere verbal designations in such terms as "chain," "wheel," and "Y") researchers seek to quantify and regard as a variable the properties of the nodes of such networks. Experimental evidence tends then to be cast in terms of correlations among the individual's centrality, peripherality, saturation, and independence and many-valued channel configurations which the experiments utilize are discarded to conform to established methods of analysis. (It is interesting to observe that the experimental design of communication networks has by and large restricted itself to communication among five but not more than six individuals. This is most probably due to the fact that the number of possible communication patterns is an exponential function of the number of individuals involved. As the number of individuals increases, computational requirements quickly spiral beyond the scope of current analytical techniques.)

Although analytical techniques for network data are not as well known as those for aggregational data, they are not altogether absent. In general, such techniques tend to be qualitative, i.e., algebraic and logical, rather than statistical. Qualitative techniques can be as conclusive as statistical ones. Recent advances in the algebra of relations and in the computability of hierarchical representations, e.g., in terms of list structures, permit symbolic manipulation of complex networks without too much loss of detail.

The Use and Limitations of Network Data in Communications Research

As I argued above, the notion of communication always implies a relational component: a message is sent from A via B to C. Interaction among individuals is understandable largely in terms of the stable relationships that have developed among them [35]. The question must therefore be raised as to whether or not communication processes can be detected in network data. Unfortunately, the answer is negative.

If communication patterns are recorded as networks, there is nothing peculiar in the properties of communication relations that could not be found in entirely different domains. For example, the binary relations "A talks to B," "A likes B," "A signifies B," and "A is smaller than B" are all formally identical as far as their anti-symmetry is concerned. What makes the first of these relations one of communication is not its formal property, but its semantic interpretation. There is very little in network data that would permit the analyst to differentiate a communication pattern from a family tree, a work schedule or a graph of functional requirements as used in architectural design [1]. While being informationally richer than aggregational data, network data are still not powerful enough to represent the formal properties of communication processes.

This inadequacy of network data is quite obvious. It stems from their static nature. Relations either hold or do not hold and, when a number of "time slices" are summed, relations may be said to occur with a certain frequency. Clearly, a representation of this kind does not provide implications along the time dimension as the notion of a process requires.

Relations such as "A dominates B," "A supports B provided that C supports A" or "A, B, C, and D trade with each other cyclically" often are the reflections of underlying communication processes. As in the case of aggregational data, network data may represent correlates of communication processes (the outcome of interpersonal communication or the contractual result of bargaining) but not the process in which such relations may have evolved. Although it cannot be ignored that such data may aid intuition about communication or serve as a particular abstraction, there is

no way to explicate and analyze *processes* of communication on the basis of network data alone.

ELEMENTARY COMMUNICATION DATA

Formal Requirements on Content

After having shown that aggregational data and network data obliterate the very properties which constitute the target of inquiries into communication processes, I will now turn to a more positive argument and point out at least the minimal formal requirements which communication data must satisfy. This restriction to "minimal formal requirements" is dictated by the well known fact that analytical and conceptual efforts quickly spiral into unmanageable proportions once all conceivable intricacies of communication processes are considered.

An abstract statement of these formal requirements is not too difficult:

- (1) All requirements of network data must be satisfied.
- (2) Three-or-higher-valued transformations, T , involving time, are specifiable over at least two distinct sets of configurations, R_n (describable in terms of network data), e.g.,

$$T(R_1 \times R_2)_{t-1} = (R_2)_t.$$

First, there must be identifiable in time and space two or more component parts of a system. The individual members of a small group, the organs of a complex organism, but also the components of a modern electronic computer might be interpreted in these terms. Each component may be described geographically (as occupying a certain space) and chronologically (along a time dimension) by the set of states it can take or by the possible configurations it can receive, memorize or emit. A description of each of these configurations, pattern, structures, etc., remains essentially within the power of network data as discussed above.

Second, the component parts and consequently the system as a whole should possess dynamic properties jointly and of their own. Relevant data must have the potential to exhibit an orderly variation of states, a regularity or some predictability along the time dimension. Therefore, one important analytical task is to identify some constraint over the succession of states and describe

it as a transformation or as a set of possible transitions of one configuration at one point in time into another configuration at some later point in time. Just as we say that an individual goes through a sequence of behavioral routines that are in part culturally prescribed or psychologically motivated, so may the behavior of a social organization be described as a changing pattern of more or less coordinated activities involving many individuals. The delivery of a speech clearly involves a sequence of bodily gestures and linguistic pattern and the visible behavior of a newspaper are the configurations of print together with their semantic interpretations that are turned out in regular intervals.

However, a behavior that could be depicted by transformations of the form:

$$R_{t-1} \rightarrow R_t$$

is of little interest to communication researchers. It would be equivalent to describing the behavior of closed and whole systems without reference to their components. If such a transformation could be isolated successfully, communication is irrelevant as a construct. Communication processes are conceptually linked to a conditionality of behavior across the components of a system and communication data must exhibit at least three-valued transformations. More specifically, if R is the repertoire of possible inputs or the set of receivable message configurations, M is the set of internal states or memory configurations, and E is the set of outputs or emittable message configurations, then the kind of three-valued transformations that may be deduced from data could have the form:

$$(R \times M)_{t-1} \rightarrow M_t$$

and

$$(R \times M)_{t-1} \rightarrow E_t$$

The former characterizes a change in the internal state of an individual regardless of what he emits. This may correspond to attitude change or, more generally, learning. The latter describes his emission as a consequence of both his previous receptions and his state of memory or cognition. Individuals often talk without external stimulation and change their minds without overt changes

in behavior. There may not be any statistical correlation between input and output, or between stimulus and response. Transformations of the above form can account for such facts. However, they still have little to do with communication processes. They merely characterize behavior as an open system as it has been studied extensively under the name of automata theory [17, 24, 25, 34]. But automata theory is heavily geared toward *synthesis* while its analytical component is not as highly developed. Progress can be noted, however, in its application to linguistic theory [9, 10].

According to our definition of communication data we require, thirdly, that such data provide the basis for deciding whether or not structure is transmitted among the component parts of a system. Communication always involves a coding of patterns of one form in one medium into another form of another medium. The most common technical example is the translation of visual images into electrical impulses, radiowaves, movements of an electrode, and ultimately back into visual forms. This is quite analogous to the process of transforming mental configurations into verbal expressions and back into cognitive structures. Such a transmission does not require exact replication: the translation of the script for a play into a performance may be regarded as a transmission process just as the transfer of technology from one domain of application into another constitutes such a process regardless of the media involved. Communication has to do with the transfer of behavior patterns, decoding and encoding of linguistic structures, and reproduction of cognitive configurations, but communication also involves time and may involve delays, autocorrelation, "chunking," and conditioning, i.e., communication channels may exhibit memory as one of several other characteristics.

We can thus regard the input to the communication process as the emission, and its output as the reception, of each component part of a system. If the memory or channel characteristics of the process are denoted by C , the transformations for which evidence is to be sought then has the following form:

$$(E_1 \times E_2 \times \dots E_n \times C)_{t-1} \rightarrow C_t$$

and

$$(E_1 \times E_2 \times \dots E_n \times C)_{t-1} \rightarrow (R_1 \times R_2 \times \dots \times R_n)_t$$

Figure 3

Flow diagram of a system with n components and the communication among them

where E denotes emissions, R denotes receptions and subscripts designate the communicators involved.

A formal comparison of the transformations accounting for the behavior of each of a system's parts with the transformations representing possible communication processes reveals them to be formally equivalent. Both pairs describe processes of information transmission. The former describes such process within, the latter across, the communicators. Both can be treated by the same analytical techniques.

This surprising isomorphism, visualized in Figure 3, suggests that from a mathematical point of view the distinction between

Figure 4

Identification of behavior and transmission in communication data

individual behavior and the communication process is quite arbitrary. Just as individuals are said to be linked by a communication net, so can it be said that the communication net is closed by the individuals concerned. The distinction is merely conceptual and often leads to chicken-egg-type problems. Together both constitute one system the understanding of which seems to me to be the aim of communication research.

The Form and Analysis of Communication Data

In order to provide evidence about the existence and nature of the transformations which I have delineated, data must be recorded so that (a) the sequence of states of the system as a whole can be examined in their chronological order. In addition, (b) a diachronical differentiation must be made among at least two sets of configurations corresponding to the states of at least two components. While communication data may have several different forms, the easiest way to understand such data is as a protocol [4], or as the tape of a Turing Machine with many channels, each of which contains information about the states of the communication agents involved, their emissions and their receptions.

The task of analyzing minimum communication data then becomes one of ascertaining whether or not some constraint on the

possible *succession of states* can be demonstrated to exist *across the components* of the system *over time*. This is the obverse to saying that the task is one of testing for whether or not interaction among the individuals exhibits some regularity over time or is rule-like. If such a constraint can be shown to be present in data, then some transformation, some explanatory device, might be formulated to account for it.

Klir [21] visualizes the domains and ranges of the possible transformations by means of a mask with windows defined over the protocol entries as in Figure 4. Such a device covers those states or configurations which the analyst does not wish to consider and makes relevant configurations available in its openings for inspection. As the protocol is moved to the left, those configurations over which orderly variations are expected or transformations are sought become successively available.

Klir's work centers around the problem of describing the behavior of the system as a whole. The domain and range of the transformation to be sought is then defined over any possible set of windows of the mask. However, when evidence about the existence or nonexistence of communication is to be extracted from the protocol, the configurations of interest are those that might be transmitted across and processed within each component.

Figure 4 depicts two masks. The one labeled "behavior" exposes those configurations that are relevant to describe the behavior of one component of the system (e.g., a hypothesized communicator). It makes available for analysis the messages he receives, his particular memory configurations and the subsequent configurations of emission. The arrows correspond to those of the three-valued transformations discussed above. Unless some constraint on the co-occurrence of states in E_1 , R_1 and M_1 can be demonstrated to persist over time, there is little sense in trying to describe the process by a transformation or as information processing.

The mask labeled "transmission" exposes those configurations that are relevant to describe processes of transmission among the components suspected to communicate. This includes the messages that each individual emits at time $t-1$ (the input into the communication net), the memory states of the channels, and the

configurations of each individual's receptions (the output of the communication net). The arrows which connect the windows' openings of this mask likewise correspond to the arrows used above to define the many-valued transformations. If a constraint on the possible co-occurrence of states on the emission side and on the reception side can be shown to persist over time, then evidence for the existence of communication is provided, interaction is present, and information is transmitted. To account symbolically for such constraints across communication agents, to describe their dynamic conditionality as interactions, or to specify the transmission process as a transformation requires additional analytical efforts with which I cannot deal within the limits of this paper.

A realistic example requires at least three-dimensional transition tables that are too large to make the constraint we are seeking easily recognizable and might still be an oversimplification. But oversimplification is necessary here to make the point. Let me therefore merely use a few more words in place of the symbols.

Suppose we observe a two-person telephone conversation and we have a protocol consisting of state descriptions of

- E_1 = his emissions
- E_2 = her emissions
- R_1 = his interpretation of her emissions
- R_2 = her interpretation of his emissions
- M_1 = his internal states
- M_2 = her internal states
- C = the "memory" characteristics of the channel.

The emissions may consist of linguistic expressions, gestures, and sound not all of which are transmitted through the channel. These emissions are interpreted by the receiving individual (and monitored by the source). The internal states of the individuals may be described in terms of emotions, attitudes, conceptions of self and the environment or as cognitive states of mind.

If communication between the two individuals is present, then one should expect that his emissions are in some systematic way related to her interpretation of his emissions and vice versa. Here one does not need to require a perfect reproduction of the message configuration sent; any distortion or misunderstanding will do as long as it is systematic. Often though, the receiver's interpreta-

tions are inaccessible, in which case we have to identify communication in the conditionality that his verbal behavior imposes on her verbal behavior at some later step in time and vice versa. Communication must also be regarded as evident when conventions and inter-individual relations are established in the course of the interaction according to which subsequent interpretations and emissions are regulated.

Communication cannot be considered evident, for example, either when he gives everything that she says the single interpretation: "nagging" and responds to his singular interpretation regardless of the variation her verbal behavior exhibits, or when his interpretations of her speech are entirely erratic, random and in no way constrained by what she says. In such a situation we would expect that all triples in

$$(E_2 \times C)_{t-1} \times C_t$$

and in

$$(E_2 \times C)_{t-1} \times (R_1)_t$$

and/or all quadruples in

$$(E_2 \times C)_{t-2} \times (M_1)_{t-1} \times (E_1)_t$$

would be observed freely. Any constraint on the occurrence of such m-tuples across individuals must be regarded as evidence for the existence of a process of communication from her to him. Communication in the other direction can be considered by analogy.

This example should be interpreted with the intended generality in mind. A protocol may describe many more such variables for suspected communicators which in turn may represent biological organs, people, social institutions, or machines. The reliance on an abstract mathematical formulation prevents such generalities from becoming vacuous.

The importance of decomposing the transmission processes involved must be emphasized. If not all messages reach each of the possible destinations, or if information circulates only within subgroups of individuals, a considerable simplification may be achieved. Similarly, when interactions are only two-way (as in an idealized telephone conversation) the system may be decomposed into independent systems involving pairs.

By regarding communication processes as an open and dynamic system of messages, analysis of data can determine whether or the extent to which such processes exhibit "a life of their own" or possess "super-individual characteristics." This is often said to be a property of communication among men and a defining criterion for social organization but it has to my knowledge never been demonstrated. Inadequate data have so far prevented presentation of evidence to support such contentions.

I have restricted my argument to the form of minimal communication data. While more sophisticated data need to be considered in order to provide interesting insights about communication processes, they cannot exclude the basic form I have discussed.

CONCLUSIONS

1. This classification of three types of data is not exhaustive. Finer differentiations along more dimensions are necessary. However, in order to examine their methodological implications for communications research such a division is sufficient.

2. Neither aggregational nor network data are powerful enough to provide explicit evidence about the existence and nature of communication processes. This does not mean that such data are totally useless. Their use in communication research may supplement existing intuitions about communication and can provide evidence about certain marginal phenomena and correlates of the process. But the defining characteristics of communication can neither be detected in such data nor can the analysis of parameters of communication be based on them.

3. This fundamental insufficiency of data in communication research accounts for several current conceptual problems associated with communication. On one hand, social problems that press toward solution seem to nourish questionable ontological claims and dogmatic assertions. Questions regarding what is communication are often settled by philosophical arguments (e.g., a thing is what it is) or they are answered by reference to an authority in communication (he knows what communication is and what it is not). On the other hand, exclusive reliance on inadequate methods of investigation leads to the impoverishment of the fruitful concept of "communication process" to the kind of concepts for which such methods can provide explicit

evidence. Thus, in the case of aggregational data, communication is often studied as a variable, or at best as statistical association between variables. In the case of network data, communication reduces to a static pattern of relationships or to a fixed symbolic structure. Consequently, the concept of communication is stripped of much of its potential explanatory power.

4. One must realize that the major stumbling block in advancing knowledge about social communication is not the small quantity of research that is being done in the field. In fact, if data are not powerful enough to represent relevant parameters of the process, and/or analytical techniques are inadequate to the phenomenon under observation, no increase in the number of studies can possibly be expected to produce a breakthrough. What is needed is an extensive development of computational techniques for processing communication data. Fostering such developments should be a prime target of the social communication sciences.

5. The fact is pointed out that more and more data archives are emerging in the social sciences. The way data are recorded and made available to future users significantly determines what will one day be studied and what may be discovered. Unless meaningful proposals for the storing and manipulation of communication data are presented, communication research may severely constrain its potential contribution to knowledge.

6. We have, I think, to admit the discomfoting fact that our current knowledge of communication processes hardly stems from the explicit findings of research. It is our intuitive participation in human interpersonal communication that has provided the primary source of current insight. Intuitive knowledge of this kind tends to enter the semantic interpretation of research results surreptitiously. We know very little about how to describe and reproduce communication processes symbolically or simulate them on a computer. Methods that would provide explicit evidence are scarce and techniques to control such processes in the real world are almost completely unreliable.

It is hoped that this article was able at least to point to some formal obstacles to progress in communication research. It depends on many creative minds to overcome them once they are identified as such.

REFERENCES

1. Alexander, Christopher. *Notes on the Synthesis of Form*. Cambridge, Mass.: Harvard University Press, 1964.
2. Allport, Gordon W. "The Open System in Personality Theory." In *Modern Systems Research for the Behavioral Scientist* (Edited by Walter Buckley). Chicago: Aldine Publishing Co., 1968, 343-50, p. 345.
3. Ashby, W. Ross. "Principles of the Self-Organizing System." In *Principles of Self-Organization* (Edited by Heinz von Foerster and George W. Zopf). New York: Pergamon Press, 1962, p. 255-78.
4. Ashby, W. Ross. *An Introduction to Cybernetics*. New York: John Wiley, 1963.
5. ———. "Constraint Analysis of Many-Dimensional Relations." *General Systems* 9:99-105, 1964.
6. Bales, Robert F. *Interaction Process Analysis: A Method for the Study of Small Groups*. Cambridge, Mass.: Addison-Wesley, 1950.
7. Barton, Allen H. "Bringing Society Back In: Survey Research and Macro-Methodology." *American Behavioral Scientist* 12(2):1-9, 1968, p. 1.
8. Blumer, Herbert. "Sociological Analysis and the 'Variable.'" *American Sociological Review* 21:683-90, 1956.
9. Chomsky, Noam. *Syntactic Structures*. The Hague: Mouton, 1957.
10. Chomsky, Noam and George A. Miller. "Introduction to the Formal Analysis of Natural Languages." In *Handbook of Mathematical Psychology*. Vol. II (Edited by R. Duncan Luce, Robert B. Bush and Eugene Galanter). New York: John Wiley, 1963, p. 269-321.
11. Coombs, Clyde H. *A Theory of Data*. New York: John Wiley, 1964.
12. Deutsch, Karl W. *The Nerves of Government: Models of Political Communication and Control*. New York: Free Press, 1963.
13. Ervin-Tripp, Susan M. and Dan I. Slobin. "Psycholinguistics." *Annual Review of Psychology* 17:435-74, 1966.
14. Friedell, Morris F. "Organizations as Semilattices." *American Sociological Review* 32:46-54, 1967.
15. Gerard, Harold B. and Norman Miller. "Group Dynamics." *Annual Review of Psychology* 18:287-332, 1967.
16. Gerbner, George. "Toward a General Model of Communication." *Audio-Visual Communication Review* 4(3):171-99, 1956.
17. Glushko, W. M. *Theorie der Abstrakten Automaten*. Berlin: VEB Deutscher Verlag der Wissenschaften, 1963.
18. Harary, Frank, Robert Z. Norman, and Dorwin Cartwright. *Structural Models*. New York: John Wiley, 1965.
19. Hovland, Carl. "Effects of Mass Media of Communication." In *Handbook of Social Psychology* (Edited by Garner Lindzey). Boston: Addison-Wesley, 1954, p. 1062-104.
20. Kaplan, Abraham. *The Conduct of Inquiry: Methodology for Behavioral Science*. San Francisco: Chandler, 1964, p. 187.

21. Klir, Jiri. "Processing of General System Activity." *General Systems* 12:193-98, 1967.
22. Krippendorff, Klaus. "Values, Modes, and Domains of Inquiries into Communication." *Journal of Communication* 19:105-33, 1969.
23. Mead, George Herbert. *Mind, Self and Society from the Standpoint of a Social Behaviorist*. Chicago, Ill.: University of Chicago Press, 1934.
24. Mealy, G. H. "A Method for Synthesizing Sequential Circuits." *Bell System Technical Journal* 34:1045-79, 1955.
25. Moore, E. F. "Gedanken-Experiments on Sequential Machines." In *Automata Studies* (Edited by Claude E. Shannon). Princeton, N. J.: Princeton University Press, 1956, p. 129-53.
26. Newcomb, Theodore M. "An Approach to the Study of Communicative Acts." *Psychological Review* 60:393-404, 1953.
27. Osgood, Charles E., George J. Suci and Percy H. Tannenbaum. *The Measurement of Meaning*. Urbana: University of Illinois Press, 1957.
28. Rogers, Everett M. *Diffusion of Innovations*. New York: Free Press, 1962.
29. Sears, David O. and Ronald P. Abeles. "Attitudes and Opinions." *Annual Review of Psychology* 21:253-88, 1969.
30. Shaw, Marvin E. "Communication Networks." *Advances in Experimental Psychology*. Vol. 1 (Edited by Leonard Berkowitz). New York: Academic Press, 1964, p. 111-47.
31. Simon, Herbert A. "The Architecture of Complexity." *General Systems* 10:63-76, 1965.
32. Singer, J. David. "Data-Making in International Relations." *Behavioral Science* 10:68-80, 1965.
33. Tannenbaum, Percy H. and Bradley S. Greenberg. "Mass Communication." *Annual Review of Psychology* 19:351-86, 1968.
34. Turing, A. M. "On Computable Numbers, with an Application to the Entscheidungs Problem." *Proceedings of the London Mathematics Society* 42:230-65, 1936; 43:544, 1937.
35. Watzlawick, Paul, Janet Helmick Beavin, and Don D. Jackson. *Pragmatics of Human Communication*. New York: Norton, 1967.
36. Wiener, Norbert. *The Human Use of Human Beings, Cybernetics and Society*. Boston, Mass.: Houghton Mifflin Co., 1950.