George A. Borden

INSTITUTE OF COMMUNICATIONS RESEARCH

SOME TERMS AND ASSOCIATED MEASURES FOR TALKING ABOUT HUMAN COMMUNICATION

CHARLES E. OSGOOD
and
KELLOGG V. WILSON

THE UNIVERSITY OF ILLINOIS, URBANA

SOME TERMS AND ASSOCIATED MEASURES

FOR TALKING ABOUT HUMAN COMMUNICATION

Charles E. Osgood
University of Illinois
and
Kellogg V. Wilson
Duke University

The terms employed in information theory, as developed by Shannon, Wiener and others (cf., 6, pp. 47-49, for a bibliography), provide at least the beginnings of an arbitrary metalanguage for talking about communication. They have the advantage of extreme generality and, although the lexicon has to be extended a bit to cover human communication situations, the mechanistic nature of the language serves as a partial safeguard against unjustified implicit assumptions. This is particularly useful in a field that deals with human activities, like human communications, where everyone has a ready explanation of any phenomenon, but in terms from the lay language that are usually loaded with diverse and sometimes contradictory connotations. While it is not our intention to try to impose a new language, we have tried to select terms which avoid identification with any particular human communcation situation (and thus pressure toward conviction by analogy). Also, we have tried to define these terms as precisely as possible, and we have associated these terms with a set of measurement operations.

We have concluded that a relatively small number of basic measures can be used to describe a relatively large variety of communication situations, thus increasing the comparability and economy of such descriptions. This also means that the lexicon is considerably larger than the number of measurement operations. It is important to point out at the outset, however, that what follows in this paper is _not_ a predictive or explanatory theory of the human communication process but rather a descriptive model with measurement implications. This is for the most part true of information theory as it has been applied to human communication situations. However, in developing any theory of human communication, such arbitrarily defined terms as used here can become the constructs whose interrelationships are specified by the principles of the theory, and the measurement operations associated with such terms can facilitate tests of the theory.

As is the case in any definitional scheme, one must start with certain undefined notions for which agreement in meaning is simply assumed. We shall assume the meanings of terms like number, state, event, same, different, maximum, minimum, exist, occur, connect, time, infinite, finite, indeterminate, and a few others are self-evident.

(1) System: a system is anything capable of existing in one or more states (or anything in which one or more events can occur).

Different systems vary in the types of states in which they can exist or the types of events that can occur. Some display states definable in terms of physical continua (e.g., radiant energy, sound-wave energy, etc.) others have states defined in terms of neurophysiological variables (e.g., impulse frequencies), and yet others display states definable only in terms of psychological continua (e.g., meanings, loudnesses). This is a qualitative but necessary characterization.

One of the terms we will use is information. According to Shannon, information is measured by the same formula as entropy, as that notion is employed is statistical mechanics to indicate the degree of randomness or indeterminacy in a system. This is confusing in the context of human communication, being more equivalent to potential information in a system than to the actual information utilized. For our present purposes, we shall substitute the term uncertainty for entropy in its original sense of indeterminacy and reserve the term information to refer to uncertainty measures computed where we can say information is transmitted in the usual sense of this term.

(2) Uncertainty: Uncertainty is a characteristic of a system which increases with the number of its states and the degree to which these states occur with equal frequency or probability.

In some instances we shall be concerned with the properties of a single system, in others with the properties of sequences of states in a system, and in yet others with relations between systems.

(3) Maximum uncertainty: the maximum uncertainty of a system is the maximum possible indeterminacy of its different states,

$$H_{max} = \log_2 \bar{n}$$

where $\bar{n}$ equals the number of different states in which a system

---

* References (3) and (6) contain fuller discussions of information theory conceptions in fairly non-technical terms.

can exist.

When (a) all states of a system are equally probable and (b) all states are sequentially independent, this maximum degree of uncertainty is displayed. In other words, maximum entropy is the logarithm to the base 2 of the number of equally probable events. The reasons for use of this function are demonstrated in the following examples.

First, consider a biased coin which always comes up heads when tossed so that $\bar{n}=1$. In this case $H_{max} = \log_2 1=0$. (This is true of a logarithm of any base). Since uncertainty equals 0 when $\bar{n}$ equals 1 -- an invariant signal tells us nothing, we have demonstrated one reason for using a logarithmic function. Now let us take the system in which a single fair coin is tossed. Two events, H or T, can occur, and both are equally probable. The occurrence of either one of these events reduces uncertainty to zero and the system is therefore said to contain 1 bit of 'information.' In this case, $\log_2 = 1$ bit. The choice of a base 2 logarithm in the function thus defines the unit of measurement, which is the amount of uncertainty associated with a two state system where each state occurs equally often. It is obvious that a two state system is a natural choice for the definition of our measurement unit, since there can be no uncertainty associated with a one event system. Now take the system in which two unbiased and independent coins are tossed simultaneously. Four events -- HH, HT, TH, or TT -- can occur and all are equally probable. The occurrence of any single H or T on a coin reduces the uncertainty by 1/2 and the occurrence of any single H or T on the other coin reduces the uncertainty to zero -- there are 2 bits of 'information' in this system (e.g., $H= \log_2$, 4= two bits). It should be noted that the uncertainty of this new system, 2 bits, is the sum of the uncertainties of the single coin systems when considered separately. This is another result of our use of a logarithmic measure. In general, the entropy of a system equals the sum of the entropies of its independent sub-systems.

Since $\bar{n}$ is the number of states in which a system can exist, estimation of this valu implies extended observation or sampling of the system's states. Presumably systems of the same type (e.g., capable of assuming corresponding states) may vary in their $\bar{n}$'s, and in any communication situation involving such systems, the possible amount of communication is limited by the system displaying the smallest $\bar{n}$. This is because the

H_{max} of the intercommunicating systems can be no greater than the smallest H_{max}, which is a function of the smallest $\bar{n}$ (equation 3). In training one's pet dog, for exam the possible complexity and diversity of the "tricks" that can be mastered depends not upon the ingenuity of the master but upon the ingenuity of the dog.

So far we have been dealing with the situation in which the alternatives are equally probable. What about the more general case in which this is not so? It can b appreciated intuitively that where the alternatives are equally probable the degree of ur certainty in the system will be maximal and that where the probability of a single alternative is 1.00, and that of all others .00, the degree of uncertainty in the system will be its minimum of zero.

(4) Absolute uncertainty: the absolute uncertainty of a system is the average degree of indeterminacy of its states displayed throu infinite time:

$H(I) = -\Sigma p(i) \log_2 p(i)$ where $p(i)$ is th probability of the ith state, and where I the class of all states of the system.

This formula has the intuited properties above and gives the number of bits of potential information in a system displaying any degree of deviation from equal probability of alternative events. $H(I)$ is the expected value, or mean, of the amounts of uncertainty associated with each of the system's states. It can be seen that the degree of uncertainty in a system will vary with both the number of alternatives and the balance of their several probabilities. However, we are sometimes more interested in the degree to which the uncertainty in a system approaches its maximum.

(5) Relative uncertainty: the relative uncertainty of a system is the ratio of its absolute uncertainty to its maximum uncertainty.

$$H_{rel}(I) = \frac{H(I)}{H_{max}}$$

This measure varies from a minimum value of 0 when only one state occurs, and $H(I)$ is therefore 0, to a maximum of 1.00 when all states are equally probable, and $H(I)$ therefore equals H_{max}.

In situations where sequences or sets of several events are being considered another formula is used, that for joint uncertainty:

(6) Joint uncertainty: the joint uncertainty of a system is the average degree of indeterminacy displayed by sets of its concurrently or sequentially occurring states.

$$H(I,J) = -\sum_{i,j} p(i,j) \log_2 p(i,j)$$

where $p(i,j)$ is the probability of states i and j occurring together.

If two sets of events are being considered, such as sets of input and output events for a system, i represents any of one set of events, I, (say, the input) and j represents any of the other set of events, J, (say, the output). If sequences of events of r length are being considered, i represents any one of the possible combinations of r-1 antecedent events in the sequence while j represents any one of the possible rth events. If the events designated by any i and j are completely dependent on each other, it is intuitively clear that there is as much uncertainty when the events are considered together as when the i of the j are considered separately, and the following relation holds:

$$H(I,J) = H(I) = H(J)$$

If the events designated by i and j are completely independent of each other, the uncertainty of the sets of events considered together is equal to the sum of the uncertainties of the i and the j, and the following relation holds:

$$H(I,J) = H(I) + H(J)$$

As the above equations imply, joint uncertainty is maximal when the events considered are completely independent and is minimal when they are completely dependent on each other. Any factors increasing or decreasing $H(I)$ or $H(J)$ separately will have the same effects upon $H(I, J)$.

A final notion we will require from information theory is that of conditional uncertainty, the degree to which events in a subsequent system (or the subsequent states of the same system) are predictable from events in an antecedent system (or antecedent states of the same system).

(7) Conditional uncertainty: conditional uncertainty is the average degree of uncertainty of subsequent states of a system (or states in a subsequent system) given the occurrence of

antecedent states of the system (or states in an antecedent system).

$$H_I (J) = - \sum_{i\,j} p\,(i,j) \log_2 p_i\,(j)$$

where the p (i,j) is the probability of $\underline{i}$ and $\underline{j}$ occurring and p i (j) is the probability of j occurring when $\underline{i}$ has occurred.

If $\underline{i}$ and $\underline{j}$ are completely dependent on each other, then $\underline{H}_I$ (J) is equal to $\underline{H}_J$ (I) and both are equal to 0, which is the minimal value of conditional uncertainty. If $\underline{i}$ and $\underline{j}$ are completely independent of each other, then $\underline{H}_I$ (J) is equal to $\underline{H}$ (J) and $\underline{H}_J$ (I) is equal to $\underline{H}$ (I), and conditional uncertainty is maximal. As can be seen, the maximal indeterminacy of the events in a subsequent system (J) given knowledge of the events in the antecedent system (I) is equal to the degree of indeterminacy in the subsequent system, $\underline{H}$ (J), considered separately.*

TYPES OF COMMUNICATION SITUATIONS

We assume that any communication situation can be described as a set of interrelated informational systems. We further assume, that each system involved in any communication situation must be either directly or mediately (via intervening system coupled with every other system. Figure 1 gives the schematic diagram of a generalized communication situation as presented by Shannon. Here we have communication whenev one system, a <u>source</u>, influences another, the destination, by manipulation of the alternative signals which can be transmitted over the <u>channel</u> connecting them.

---

* These relations of joint and conditional entropy correspond to certain more familiar statements in probability theory. The identity below holds for all sets of events, wheth dependent or independent:

$$p\,(a,b) = p\,(a)\, p_a\,(b) = p\,(b)\, p_b\,(a)$$

where $\underline{p}$ (a,b) is the probability of both $\underline{a}$ and $\underline{b}$ occurring, $\underline{p}$ (a) and $\underline{p}$ (b) are the probabilities of $\underline{a}$ and $\underline{b}$, $\underline{p}_a$ (b) is the probability of $\underline{b}$ occurring given $\underline{a}$, and $\underline{p}_b$ (a) is the probability of $\underline{a}$ occurring given $\underline{b}$. If $\underline{a}$ and $\underline{b}$ are independent, then $\underline{p}_a$ (b) = $\underline{p}$ (b) and $\underline{p}_b$ (a) = $\underline{p}$ (a) and we have the familiar identity: p (a,b) = p (a) p (b). Corresponding to this identity is the condition in which maximal joint entropy holds:

$$H\,(I,J) = H\,(I) + H\,(J)$$

Corresponding to the more general identity, $\underline{p}$ (a,b) = $\underline{p}$ (a) p_a (b) = p (b) p_b (a), is the following: $H(I,J) = H\,(I) + H_I\,(J) = H(J) + H_J\,(I)$ where $\underline{H}_I$ (J) and $\underline{H}_J$(I) are measures of conditional entropy.

The information source is conceived as producing one or more messages which must be transformed by a transmitter system into signals which the channel can carry; these signals must then be transformed by a receiver system back into messages which can be utilized by the destination system. In telephone communication, for example, the messages produced by a speaker are in the form of variable sound pressures and frequencies which must be

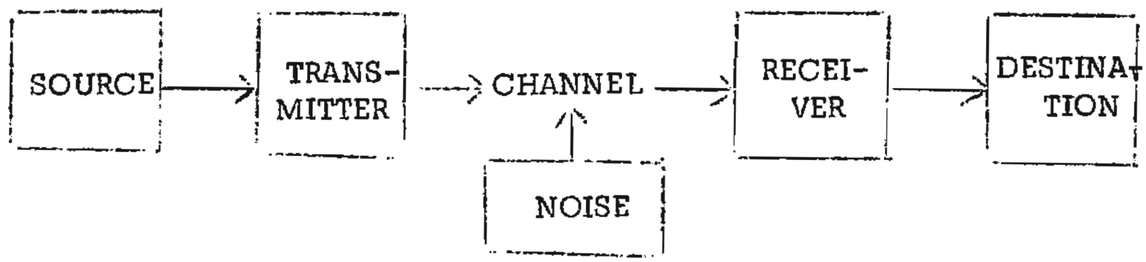
FIGURE 1 ABOUT HERE

transformed into proportional electrical currents by the transmitter; these energies are carried in the wires (channel) to a receiver which must transform them back into variable sound pressures and frequencies, which constitute the message to be utilized by the destination. The activity of the transmitter is usually referred to as encoding and that of the receiver as decoding. Anything that produces unpredictable variation in the signal, either in the channel or at its termini, is called noise. For the sake of logical consistency -- and to meet the assumption of unbroken coupling of informational systems- we should treat the channel in Shannon's diagram as another system (i.e., should enclose it in a box and, in fact, should make all system 'boxes' contiguous). For the most part, the mathematical aspects of Information Theory and its applications to communication problems have dealt with relations between transmitter and receiver via signals in noiseless or noisy channels, rather than with relations between sources and destinations.

The Shannon model is not entirely satisfactory as a description of the essential processes in human communication. For one thing, it implies a separation of source and destination, of transmitter and receiver, which is typical of mechanical communication situations but not of human ones. One solution, often employed by psychologists studying learning phenomena with the Information Theory model, has been to treat the human subject as the channel through which information is transmitted between input (stimulus) and output (response) systems (cf., Quastler, 9). This model corresponds to, and is translatable into, the single-stage S-R learning theory model. Such a single-stage or "empty organism" psychological model is adequate for most associative learning

FIGURE I

The Shannon Model of Communication



phenomena, but it proves insufficient for many human communication situations (in the ordinary sense), where the emphasis is upon meaningful phenomena such as those involving intentions in sources and significances in destinations.

Another solution is to treat the human individual as a communicating unit which includes all of the parts of the Shannon model as sub-systems. The individual human being as a communication unit functions more or less simultaneously as a source (selector) and a destination (interpreter) of messages and as both a transmitter (encoder) and receiver (decoder) of messages -- indeed, he is typically a decoder-interpreter of the messages he himself selects-encodes via various feed-back mechanisms. Figure 2 describes in Information Theory language (for the most part) a complete human communica tion event participated

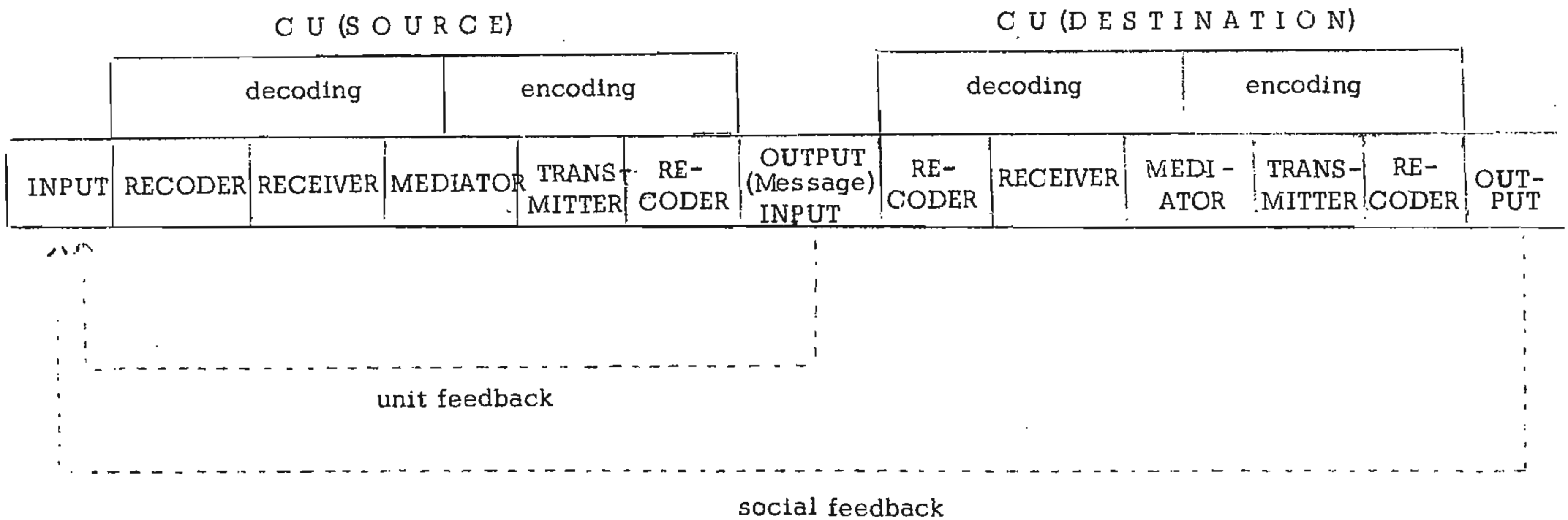
FIGURE 2 ABOUT HERE

in by two communication units (CU), one the source of a message and the other its destir tion. The source CU decodes certain input information (e.g. perceives a dollar bill on the sidewalk while walking downtown with his friend), the source CU then encodes this cognitive information into the signals of a shared linguistic code, a message (e.g. says "Look, there's a dollar bill on the sidewalk!") the destination CU decodes and interprets this linguistic message input (e.g., hears and understands the significance of what his friend says), and then the destination CU may encode in turn, either linguistically (e.g. saying, "So there is -- we're in luck!") or behaviorally (e.g., stopping, stooping, and picking up the bill), the destination CU thereby becoming the source CU in another communication event.

There are several points about the model in Figure 2 which need elaboration. (1) For the source and destination systems of the Shannon model we have substituted the single mediator system (and the reader may substitute the term cognition here if he wishe When the human communicating unit is functioning as a decoder, the alternatives select in this system constitute the significances of what is heard, seen, etc., i.e., the destination system; when the same communicating unit is functioning as an encoder, the alternatives in this system constitute the intentions for what is said, done, etc., i.e., the source system. (2) Receiver and transmitter systems in this model are assumed to correspond to what psychologists ordinarily treat as sensory integration (perception) and motor integration (skill), while the two recoder systems in the model are assumed to correspond to transformations between nervous and physical forms of energy that must

FIGURE 2

Scheme of Communication Involving Two Human Communication Units



take place both on the input side (sensory recoding) in receptors and the output side (motor recoding) at motor nerve endplates. Whereas sensory and motor recoding follow 'wired-in' structural principles, the recoding that takes place in receiver, transmitter and mediator systems follows learning principles. A two-(or three-) stage model such as this corresponds to, and is translatable into, a learning theory model that has been described by one of the authors (Osgood, 8). (3) To the extent that communication occurs, the events in the message system are at once the output of the source unit and the input to the destination unit, e.g., the vocal-auditory pattern of "Look, there's a dollar bill on the sidewalk!" as a physical event is equally a response of the source CU and a stimulus to the destination CU. (4) Finally, it may be noted that the sequence of systems between the source mediator and the destination mediator in Figure 2 essentially duplicates the Shannon model in Figure 1. Noise, rather than being a system, is a variable attribute of any system.

It is probably obvious that the number of sub-systems into which any communication situation is divided is arbitrary and hence dependent upon the purposes and interests of the investigator. In Figure 2, for example, we indicate sensory recoding as a single system. Yet a specialist in the process of auditory reception might find it fruitful to divide it into a sound-pressure system coupled to a tympanic membrane vibration system coupled to a mechanical bone transmission system coupled to fluid hydraulic system coupled to a mechanical hair-cell pressure system coupled to a piezoelectric system coupled to a nervous impulse system -- all these systems displaying a variety of states and conditional dependencies of these states upon antecedent systems, hence being informational. Going in the other direction, each human communicating unit, despite its own complexity, can be treated as one of the alternatives in the exchange of messages within a larger social system -- and in fact we shall discuss just such system in a final section on communication networks. What is required for the operations to be described here is that the states of any system be discriminable from one another, and divided into two or more non-overlapping classes. It is not required, incidentally, that the states of any system be capable of ordering along any similarity continua -- although as we shall see in one of our quantitative illustrations, this places a limitation on the Information Theory type of description.

For the sections which follow, we have tried to isolate some basic types of communication situations to which generalizable terms and measures can be applied. These are: (1) the inherent characteristics of single systems (e.g., the informational characteristics of the set of phonemes employed in English); (2) information transfer between coupled systems (e.g., the dependencies existing between variations in sound pressures and vibrations of the tympanic membrane); (3) transmission characteristics of single systems (e.g., of the human subject in a learning experiment as determined from dependency relations between the stimulus input and his response output); (4) information transfer between non-coupled but corresponding systems (e.g., communication of meanings between two human individuals via gestures); and (5) the inherent characteristics of communication networks (e.g., the distribution of messages among and between the individuals making up a business organization). With the exception of (1) and (5) - the latter being essentially an elaboration into the social situation from the former - these situations in which we might be interested in assessing informational characteristics are assumed to be independent of the size or nature of the systems involved or their states. The concepts and measures of the Shannon-Weiner theory of information will be used as the basis of a vocabulary for talking about communication in general in these types of situations, but occasionally we shall find it necessary to go beyond these concepts and measures. The number of different terms in the lexicon is much greater than the number of different measures; this is because essentially the same measures, computationally and logically, are found to apply to quite different communication situations.

INHERENT CHARACTERISTICS OF SINGLE SYSTEMS

Here we deal with probabilistic distribution of states or events within a single system, and the sequential dependencies of these states or events upon each other, within some finite observation period, i.e., with the information in that system without reference to any other systems to which it may be coupled. Reference to Figure 3, which provides a general summary of the measures described, will be helpful in following the analysis. The rows represent antecedent states or events and the columns subsequent states or events in the system. If event i is followed 22 times by event j, 22/N (N being the total number of observed events in the period of measurement) is the entry in cell ij; if event j is followed by event i only 3 times during this period, then 3/N is the entry in cell ji, and so forth. At a later point in this section, the sequences of pairs of phonemes in English are analysed as a concrete illustration of this type of single system measurement.

Applications of entropy measures to sequences of events in single systems are actually fairly common. Newman (5) has analysed sequences of consonants and vowels in the printed orthography of various languages; Miller and Frick (4) have used such measures to analyse the sequences of responses by animal subjects in learning situations.

Measures of the <u>maximum uncertainty</u> (3), <u>absolute uncertainty</u> (4), and <u>relative uncertainty</u> (5) of the system are as given in the previous section on basic Information Theory notions:

(3) $H_{max} = \log_2 \bar{n}$

(4) $H(J) = -\sum_{i=1} p(j) \log_2 p(j)$

(5) $H_{rel} = \frac{H(J)}{H_{max}}$

where $\bar{n}$ is the number of different states or events that occur during the period of measurement and $\underline{p}$ (j) is the probability of the $\underline{j}$th event.

If the probabilities of the various events in the system tend toward equality, then $\underline{H}$ (J) will approximate $\underline{H}_{max}$ and $\underline{H}_{rel}$ will approximate 1.00 (the system will be characterized by a high degree of uncertainty as to what event will occur at any given moment). Conversely, if H (J) approximates zero and H_{rel} likewise, the system is characterized by a low degree of uncertainty and hence is not utilizing the information potential.

(8) <u>Relative transitional uncertainty</u>: the relative transitional uncertainty of a system is the ratio of its absolute transitional uncertainty to its maximum, H (J), assuming H (J) to be fixed at some constant value.

$$H_{rel\ I}(J) = \frac{H\ I\ (J)}{H(J)}$$

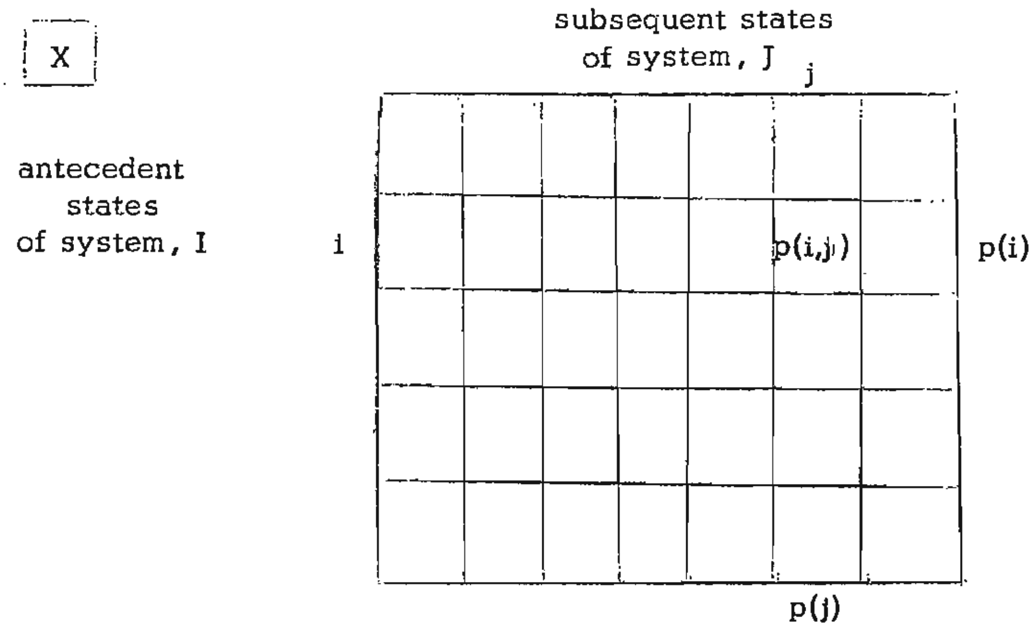
In this case, absolute transitional uncertainty is operationally identical with <u>conditional uncertainty</u> (7), when considering sequences of events in which the $\underline{i}$ are any $\underline{r}$ - 1 events in the sequence and the $\underline{j}$ are any $\underline{r}$th events. It will be recalled that at its maximum value $\underline{H}_I$ (J) is equal to $\underline{H}$ (J). Therefore $H_{rel\ I}$ (J) will tend toward 1.00 when subsequent events in the system are relatively unpredictable from knowledge of antecedent events and will tend toward 0 when subsequent events are relatively predictable from knowledge of antecedent events. The value of H_I (J) is determined from entries in

FIGURE 3 ABOUT HERE

the body of the matrix in Figure 3.

FIGURE 3

Inherent Characteristics of Single Systems



(3) $H_{max} = \log_2 \bar{n}$

(4) $H(I) = \sum_i p(i) \log_2 p(i)$; $H(J) = -\sum_j p(j) \log_2 p(j)$ $\quad [0, \log_2 \bar{n}]$ *

(5) $H_{rel}(I) = \frac{H(I)}{H_{max}}$; $H_{rel}(J) = \frac{H(J)}{H_{max}}$ $\quad [0, 1]$

(6) $H(I,J) = -\sum_{i,j} p(i,j) \log_2 p(i,j)$ $\quad [0, 2 \log_2 \bar{n}]$

(7) $H_I(J) = -\sum_{i,j} p(i,j) \log_2 p_i(j)$ $\quad [0, H(J)]$

(8) $H_{rel\,I}(J) = \frac{H_I(J)}{H(J)}$ $\quad [0, 1]$

(9) $R = H(J) - H_I(J)$ $\quad [0, H(J)]$

(10) $R_{rel} = \frac{R}{H(J)} = 1 - H_{rel\,I}(J)$ $\quad [0, 1]$

(11) $r(i,j) = p_j(j) - p(j)$ $\quad [-p(j), 1 - p(j)]$

(12) $r_{rel}(i,j) = \frac{r(i,j)}{1 - p(j)}$, if $r(i,j) > 0$; $= \frac{r(i,j)}{p(j)}$, if $r(i,j) < 0$

* The figures enclosed in the brackets are the minimum and maximum values of the measure.

It is generally more meaningful to deal with values which represent degrees of predictability (or certainty) than with entropy measures themselves. The measurement of redundancy is a case in point:

(9) Absolute redundancy: the absolute redundancy of a system is the degree to which its states are dependent upon antecedent states.

$$R = H\ (J) - H_I(J)$$

As can be seen from this formulation, the amount that the uncertainty in the sequential states of a system is reduced by knowledge of antecedent states is a measure of the transitional predictability involved. If the occurrence of antecedent states is not predictive of subsequent states (and H_I (J) equals H (J), its maximum), then the amount of redundancy in the system is zero. If the occurrence of antecedent events is completely predictive of subsequent states (and H_I (J) is zero), then the amount of redundancy in the system is maximum (i.e., equals the value of H (J)). However, the value of H (J) varies with the total number of possible states or events. It is sometimes more convenient to have a redundancy measure which is independent of the total number of events and which will therefore render various systems comparable.

(10) Relative redundancy: the relative redundancy of a system is the ratio of its absolute redundancy to its maximum redundancy.

$$R_{rel} = \frac{R}{H\ (J)} = \frac{H\ (J) - H_I\ (J)}{H\ (J)}$$

If the rth state in all sequences of r states is completely independent of the previous $r - 1$ states, then R_{rel} takes on its minimal value of 0. If the rth state is completely determined by the previous $r - 1$ states, then R_{rel} equals its maximum of 1.00. This is an over-all measure of the degree to which states of any system are predictable from antecedent states of the system, i.e., a measure of the rigidity of lawfulness of sequencing within the system.

It may often be of interest to compare particular states of a system with respect to their dependency upon preceding states. In the case of phoneme sequences, for example, it might be of interest to know which phonemes in English carry the most independent information in Shannon's sense; i.e., which phonemes are least predictable from antecedent phonemes.

(11) Specific redundancy: the specific redundancy of a given state of a system,

$\underline{j}$, is the degree to which its occurrence or non-occurrence is dependent on the prior occurrence of some other state or sequence of states, $\underline{i}$,

$$r(i,j) = p_i(j) - p(j)$$

where $p(j)$ is the overall probability of state $\underline{j}$ and $p_i(j)$ is the conditional probability of $\underline{j}$ when $\underline{i}$ has occurred.

If $\underline{j}$ occurs more frequently after $\underline{i}$ than in general, i.e., if $p_i(j) > p(j)$, then we will obtain a positive value of $r(i,j)$. The maximum value of $r(i,j)$ is obtained when $\underline{j}$ always occurs after $\underline{i}$, i.e. if $p_i(j) = 1$ and $r(i,j)$ equals $1 - p(j)$. If $\underline{j}$ occurs less frequently after $\underline{i}$ than in general, i.e. if $p_i(j) < p(j)$, then we will obtain a negative value of $r(i,j)$. The minimum value of $r(i,j)$ is obtained when $\underline{j}$ never occurs after $\underline{i}$ (i.e. if $p_i(j) = 0$) and $r(i,j)$ equals $- p(j)$. If the frequency of occurrence of $\underline{j}$ is not influenced by the prior occurrence of $\underline{i}$, i.e. if $\underline{i}$ and $\underline{j}$ are independent and $p_i(j) = p(j)$, then $r(i,j)$ will equal zero.

Again, in order to make specific redundancy measures comparable for differing values of $p(j)$, it is convenient to express these redundancies in a relative form:

(12) Relative specific redundancy: the relative specific redundancy of a given state is the ratio of its specific redundancy to the maximum value for its sign.

$$r_{rel}(i,j) = \frac{r(i,j)}{1 - p(j)} \qquad \text{if } r(i,j) > 0$$

$$= \frac{r(i,j)}{p(j)} \qquad \text{if } r(i,j) < 0$$

Thus, $r_{rel}(i,j)$ will vary from 0 to 1 if $\underline{j}$ occurs more often after $\underline{i}$ than in general (i.e., if $p_i(j) > p(j)$ and $r(i,j) > 0$). If $\underline{j}$ occurs less often after $\underline{i}$ than in general (i.e. if $p_i(j) < p(j)$ and $r(i,j) < 0$), then $r_{rel}(i,j)$ will vary from -1.00 to 0. If $\underline{i}$ and $\underline{j}$ are independent (i.e. if $p_i(j) = p(j)$ and $r(i,j) = 0$), then $r_{rel}(i,j)$ will equal zero.

Informational Properties of English Phonemes

The measures described in this section have been applied to a sequence of 20,028 phonemes in a sample of spoken American English, tabulated in sequences of two as in Figure 3. This tabulation was directed by John B. Carroll in connection with the 1952 Summer Seminary on Psychology and Linguistics.* Since there are 34 phonemes in the tabulation, the value of maximum uncertainty (eq. 3) is $H_{max} = \log_2 34 = 5.09$.

* This tabulation may be obtained by writing to Professor John B. Carroll, Graduate School of Education, Harvard University.

Since the phonemes were tabulated as pairs of antecedent (I) and subsequent (J) events, the obtained absolute uncertainty values (eq. 4) for I and J differed only in the third decimal figure:

$$H\,(I) = H(J) = 4.45$$

The relative uncertainty values (eq. 5) are therefore:

$$H_{rel}\,(I) = H_{rel}\,(J) = 0.88.$$

Thus, English phonemes are 88% as "random" as they might be. This reflects the not too surprising fact that some phonemes are more frequently used than others.

The most interesting aspect of this analysis is the examination of redundancy in phoneme sequences. The first step in this analysis was to obtain a joint entropy measure (eq. 6) for phoneme pairs.

$$H(I,J) = 7.81$$

The amount of transitional uncertainty (eq. 7) is

$$H_I\,(J) = 3.36,$$

and the value of relative transitional uncertainty (eq. 8) is therefore

$$H_{rel\ I}\,(J) = 0.75.$$

Subsequent phonemes are only about 75% as "unpredictable" as they could possibly be. In other words antecedent and subsequent phonemes are not entirely independent in a statistical sense. This is also indicated by the absolute redundancy value (eq. 9) of

$$R = 1.09$$

and the relative redundancy value (eq. 10) of

$$R_{rel} = 0.25;$$

25% "redundancy" is equivalent to 75% "unpredictability".

By definition, the phoneme must be a fairly non-redundant unit of language encoding -- at least for short sequences of speech sounds (say, two, three or four). If two speech sounds occurred together with very high frequency in a language, it is quite likely that they would be included in the same phoneme class according to the rules of linguistic analysis. Also, if one pair of sounds occurred only in particular sequences of sounds (phonetic environment) and another "similar" pair occurred only in different sequences, both pairs would very likely be classed as allophones of the same phoneme class. However, we should not regard the value of 25% "redundancy" as anything but an "average lower bound" on the redundancy of English phonemes. If longer phoneme sequences -- say, three, four or five -- had been tabulated, certainly no lower value would have been obtained and it is likely that the value would have been higher. We should also realize that such measures, which are essentially "averages" over a complexly organized series of events, cannot be especially sensitive to specialized types of redundancy -- e.g. redundancy in the structure of English nouns. (A fuller discussion of the application of Information Theory measures to sequences of language events is contained in 7, pp. 103-110.)

The redundancy measures above only indicate the degree of redundancy -- not its nature. The measures of specific and relative specific redundancy (eq. 10,11) provide some very interesting clues regarding the nature of this redundancy, however. Since there are $34^2 = 1{,}156$ of each of these two kinds of measures, it is not possible to reproduce them in this paper, but the grouped frequency distribution for the measures of relative specific redundancy in Table 1 indicate the most interesting aspects of these measures.

TABLE 1 ABOUT HERE

The distribution in Table 1 is obviously bimodal with the largest mode for values near -1.00 and a smaller mode near zero. A large negative value (near -1.00) or $r_{rel}(i,j)$ indicates that a particular phoneme pair tends <u>not</u> to occur together. A value near zero indicates that the phonemes occur together just about as often as would be expected by chance, assuming independence. A large positive (near +1.00) value would indicate that a particular phoneme pair tends to always occur together. Table 1 quite clearly indicates that a large proportion (over one-half) of the possible contiguous phoneme pairs rarely or never occur, that a sizeable proportion (about one-fifth) are nearly independent, and that very few tend to occur together.* (No $r_{rel}(i,j)$ values greater than +0.78 were obtained.) Thus, the main source of redundancy for phoneme pairs is that many particular pairs rarely or never occur together.

We may now ask why such a large proportion of phoneme pairs do not occur together. Saporta (10) has indicated at least a partial explanation of this phenomena

* One such phoneme pair which tends to occur together is /st/. Also, /ž/ followed by a vowel, such as /žɔ/ tended to occur in pairs.

TABLE 1

Grouped Frequency Distribution of Relative Specific Redundancy r_{rel} (i,j) Measures. (Interval Size = 0.15)

Interval Midpoint	Frequency	Percentage
-0.93	609	52.68
-0.78	67	5.80
-0.63	55	4.76
-0.48	44	3.81
-0.33	29	2.51
-0.18	32	2.77
-0.03	224	19.38
0.12	69	5.97
0.27	17	1.47
0.42	2	0.17
0.57	5	0.43
0.72	3	0.26
	1,156	100.01

is his analysis of clusters of consonant phonemes. He measured the "similarity" of consonant phoneme pairs in terms of the number of distinctive features by which the phoneme classes differ. * Two consonant phonemes which differ in few distinctive features (say, 1 or 2) would be easy for a speaker to encode in immediate succession since only minor changes in the position of the articulatory organs would be required. However, such sequences would be difficult to decode because of the similarity of the sounds involved. By the same token, two consonant phonemes which differ in many distinctive features (say 9 or 10) would be easy to decode because of their dissimilarity but hard to encode because of the major changes in position of the articulatory organs required. In analysing the frequencies of consonant pairs in Carroll's sample, Saporta found that consonant pairs differing in few or many distinctive features are quite rare, while differences in a moderate number (4 or 5) of features was common in consonant pairings. Thus, the non-occurrance of at least certain consonant phoneme pairs can be explained on the basis of an apparent compromise between the demands of encoding and decoding systems. The disjunction between consonants and vowels also presumably contributes to the high degree of less-than-chance pairing -- in terms of linguistic analysis of the spoken language (if not the orthography), VV or CC sequences are very infrequent as compared with CV or VC sequences.

One of the most useful properties of the measures being described in this paper is that of comparison between systems. It would be both feasible and informative to apply the same measures to the phoneme systems of several additional languages. Would the relative uncertainty of Chinese differ significantly from that of English? Would the sequential redundancies of Hungarian differ from English? Or could it be shown that there is an extremely high degree of stability across languages in the properties isolated by these information measures? Unfortunately, to make such comparisons it is necessary to have both comparable phonemic analysis and then adequate samples of comparable sequential utterances recorded, transcribed into phoneme sequences, and analysed statistically. Although some comparative work of this sort is underway, such data were not available to us at the time of this writing.

---

* See (6, pp. 11-12) and (2) for a fuller discussion of the distinctive feature concept.

INFORMATION TRANSFER BETWEEN COUPLED SYSTEMS

At each point in any communication process information is being transferred from one system to some adjacent and subsequent system. It seems intuitively reasonable that immediately adjacent systems in any communfcation series should be non-corresponding systems, i. e., systems incapable of assuming identical states. It is this non-correspondence of states which makes it possible for us to identify the existence of two different systems rather than one-and -the-same system. If the nervous systems of two individuals were completely in immediate contact with one another, they would necessarily function as a single system. In actuality of course, the corresponding neural systems of two individuals are separated by <u>non</u>-corresponding systems, e.g., the physical environment, and this is the <u>raison</u> <u>d'etre</u> for human communication problems.

The upper portion of Figure 4 describes two adjacent systems, X and Y, where the direction of information flow is from X to Y. In this case, we refer to Y as the output system with respect to X and to X as the input system with respect to Y.

(13) <u>Coupling:</u> any two systems are coupled when information in one is dependent to any degree upon information in the other, and no other systems intervene between them.

(14) <u>Input</u>: the input to any system is the information in any other antecedent system or systems to which it is coupled.

(15) <u>Output:</u> the output of any system is the information in any other subsequent system or systems to which it is coupled.

In dealing with relations between two coupled systems, the basic measure required is the degree of dependency of states of the subsequent system upon states of the antecedent one.

(16) <u>Relative conditional uncertainty</u>: the relative conditional uncertainty between two systems is the ratio of the absolute conditional uncertainty to the maximum conditional uncertainty.

$$H_{rel\ K}\ (L) = \frac{H_K\ (L)}{H\ (L)}$$

where H (L) is assumed to be fixed at some constant value,

$$H_K(L) = -\sum_{k,l} p(k,l) \log_2 p_k(l)$$

$$H(L) = -\sum_{l} p(l) \log_2 p(l)$$

and k and l are states in the antecedent and subsequent systems respectively and K and L are the sets of all such states.

It will be noted that (16) is formally equivalent to the definition of relative transitional uncertainty (8) and that both are special cases of general absolute and conditional entropy measures (5,7). Whereas we were dealing with antecedent and subsequent states of the same system in the preceding section, here we are dealing with states of antecedent and subsequent systems. Figure 4 provides an illustrative computational matrix:

FIGURE 4 ABOUT HERE

the rows represent states of the antecedent and the columns states of the subsequent of two coupled systems. $H.(L)$ is computed from the column sums, $H_K(L)$ from the body of the matrix, and $H_{rel\,K}(L)$ as shown above. If the states of the subsequent system (the set L) are completely determined by the states of the antecedent system (the set K), then $H_K(L)$ and $H_{rel\,K}(L)$ will both take on their minimal values of 0. If all l are completely independent of all k, then $H_{rel}\,K(L)$ takes on its maximum value of 1.00.

Thinking in terms of degrees of predictability rather than degrees of uncertainty, the following equation measures <u>dependency</u> between systems as a function of entropy measures:

(17) <u>Absolute dependency</u>: The absolute dependency of a system on an antecedent system to which it is coupled is the degree to which its states are predictable from states of that antecedent system.

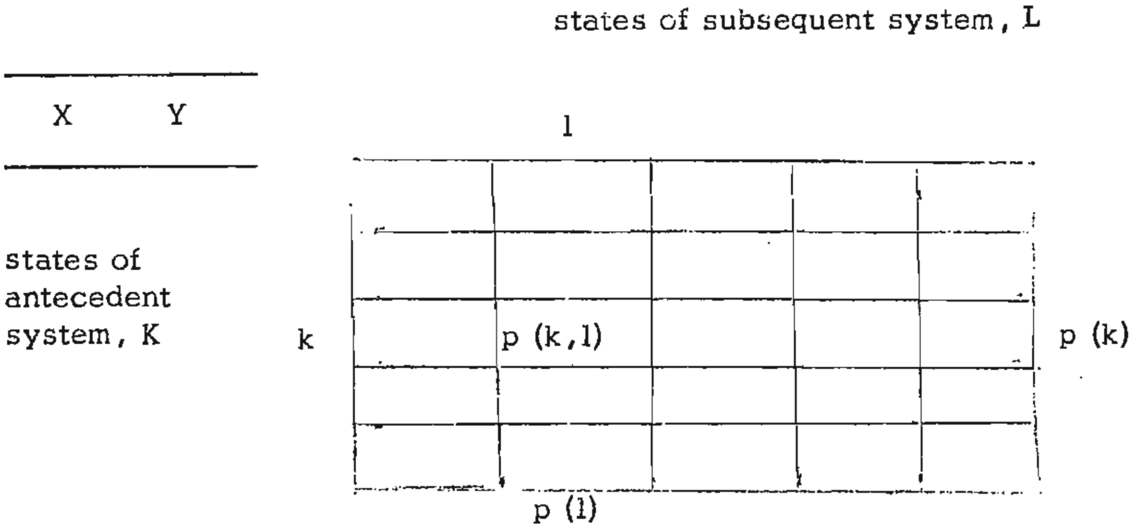
$$D = H(L) - H_K(L)$$

where K and L are defined as in (16).

If all l are completely independent of all k, D has its minimum value of zero. If all l are completely dependent upon k (i.e. if all $p_k(l)$ are 0 except one for each k which equals 1), then D has its maximum value of H (L). The value of $H(L)$ will vary with the number of states which the subsequent system can assume, and since we often wish

FIGURE 4

Information Transfer Between Coupled System



(16) $H_{rel\,K}(L) = \frac{H_K(L)}{H(L)} = \frac{-\sum_{k,l} p(k,l) \log_2 p_k(l)}{-\sum_{l} p(l) \log_2 p(l)}$ $[0,1]$

(17) $D = H(L) - H_K(L)$ $[O, H(L)]$

(18) $D_{rel} = \frac{D}{H(L)}$ $[O,1]$

(19) $d(k,l) = p_k(l) - p(k)$ $[-p(l), 1 - p(l)]$

(20) $d_{rel}(k,l) = \frac{d(k,l)}{1 - p(l)}$, if $d(k,l) > O$; $= \frac{d(k,l)}{p(l)}$ if $d(k,l) < O$

(21) $A_{k,l} = [-\sum_{k,l} \log_2 p(l) - \log_2 (l)]\ f(k,l)$

$[O, \{-\sum_{k,l} \log_2 p(l)\}\ f(k,l)]$

(22) $E_{k,l} = \frac{A_{k,l}}{A_{max}} = \frac{A_{k,l}}{\{-\sum_{k,l} (\log_2 p(l)\}\ f(k,l)}$ $[O,1]$

to compare different couplings with respect to information transfer, we usually will transform $\underline{D}$ into $\underline{D}_{rel}$.

(18) Relative dependency: the relative dependency of a system upon any antecedent system to which it is coupled is the ratio of its absolute dependency to its maximum dependency.

$$D_{rel} = \frac{D}{H\ (L)}$$

where $\underline{H}$ (L) is assumed to be fixed at some constant value.

Suppose that X is a system producing sound waves of varying frequency and Y is the basilar membrane in the inner ear: To the degree that states of the membrane system (e.g., the loci of its maximum vibration) are predictable from the states of the oscillator (e.g., sound wave frequencies), $\underline{D}_{rel}$ will approach 1.00; to the degree that action of this membrane is independent of sound frequency, $\underline{D}_{rel}$ will approach 0. Note that there is no question here of correspondence between states of X and Y - loci of points along a continuum are not qualitatively comparable to frequencies of vibration.

Again in analogy to redundancy within single systems, we may wish to indicate the degree to which some specific state of Y is dependent upon some specific state, or set of states, of X.

(19) Specific dependency: the specific dependency of a given state, l, of a subsequent system upon a given state, or set of states, k, of any antecedent system to which it is coupled is the degree to which its occurrence or non-occurrence is predictable from the prior occurrence of the given state, or set of states, of the antecedent system.

$$d(k,l) = p_k(l) - p(l)$$

where $\underline{k}$ and $\underline{l}$ are defined as in (16)

The statements made respecting specific redundancy (11) are equally applicable here, except that they apply to specific dependencies of states in one system upon states in another. Again, it is useful for purposes of comparability to deal with relative specific dependency.

(20) Relative specific dependency: the relative specific dependency of a given state, l, in a subsequent system upon a given state, or set of states, k, in an antecedent system to which it is coupled is the ratio of the specific dependency to the maximum value for its sign.

$$d_{rel}(k,l) = \frac{d(k,l)}{1 - p(k)} \qquad \text{if } d(k,l) > 0$$

$$= \frac{d\ (k,l)}{p(k)} \qquad \text{if } d(k,l) < 0$$

It seems intuitively reasonable that the greater the dependency of one system upon another, the greater the amount of information that can be transmitted during any period of time. Whether dealing with directly coupled, non-corresponding systems (as here) or with non-coupled, but corresponding systems connected by a channel (see below), it may often be useful to indicate by a single number the amount of information exchanged or transmitted during some particular period of observation or measurement.

(21) Amount of information transmitted between coupled systems: the amount of information transmitted from one system to another is a multiplicative function of the dependency between these systems and the frequency with which states occur in them, during some finite time interval $(t_o\ ,\ t_1)$.

$$A_{k,l} = -\sum_{k,l} \left[\log_2 p(l) - \log_2 p_k\ (l)\ \right] f(k,l)$$

$$= -\sum_{k,l} \left[\log_2 \frac{p(l)}{p\ k(l)}\ \right]\ f\ (k,l)$$

where $\underline{k}$ and $\underline{l}$ are as defined in (16) and $\underline{f}$ (k,l) is the frequency of $\underline{k}$ and $\underline{l}$ occurring together during the time interval from $\underline{t}_o$ to $\underline{t}_1$.

The quantity $-\left[\log_2 p(l) - \log_2 p\ _k(l)\ \right]$ gives the number of bits of information transmitted to the subsequent system when $\underline{k}$ and $\underline{l}$ occur. These quantities are weighted by the frequencies of the $\underline{k}$ and the $\underline{l}$ in a finite interval, $\underline{t}_o$, $\underline{t}_1$, to give the total amount of information transmitted. If the states of a subsequent system are completely independent of the states of an antecedent system, then $\underline{A}$ must take on its minimum value of 0, regardless of the number of states that occur in either $\underline{k}$ or $\underline{l}$ during the period of observation. If the states of the subsequent system are completely dependent on the states of the antecedent system, then $\underline{A}$ will take on its maximum value, $\underline{A}_{max} = -\sum_{k,l} \left[\log_2 p(l)\ \right] f(k,l)$. In other words, the possible amount of information transfer between systems is limited by the absolute entropy of the system receiving the information. These statements apply only to the sub-sets of states displayed during time $\underline{t}_o$ to $\underline{t}_1$. To compare systems having varying maxima, it is useful to transform this amount measure to what we may call an efficiency measure of information transmission.

(22) Efficiency of information transmission between coupled systems: the efficiency of information transmission between two systems is the ratio of the actual amount of information transmitted ($\underline{A}$) to the maximum possible amount ($\underline{A}_{max}$), given that $\underline{H}$ (L) of the subsequent system is fixed.

$$E_{k,l} = \frac{A_{k,l}}{A_{max}} = \frac{A_{k,l}}{-\sum_{k,l} \left[\log_2 p(l)\right] f(k,l)}$$

This measure serves to indicate the degree to which two systems utilize the available transmission facilities. If the states of the subsequent system are completely independent of the states of the antecedent system, then $\underline{E}$ is 0. If there is complete dependency between the systems, then $\underline{E}$ will have its maximum value of 1.00. It should be noted that the maximum here is relative only to the condition where the subsequent system is completely dependent upon the antecedent system -- changes in $\underline{H}$ (K) or in $\underline{H}$ (L), or in the form of coding, which might change $\underline{A}_{max}$ are not considered.

Information theory measures have rarely been applied to this type of situation, in psychology or the study of human communication. Because of the similarity of the measures used in this section to those described in the section concerning non-coupled corresponding systems, the analysis of communication by facial expressions following the later section can serve as an example of both types of applications.

TRANSMISSION CHARACTERISTICS OF SINGLE SYSTEMS

Each system in a series linked together to accomplish any communication process is involved in transmitting information from one point to another in the chain. Therefore it is often necessary to inquire into the transmission characteristics of a given system, quite apart from its inherent characteristics. In this inquiry, as shown in Figure 5, we treat the particular system under study (X) as a channel connecting the two other systems to which it is coupled, its antecedent or input (I) system and its subsequent or output (O) system. In other words, the transmission characteristics of any given system are estimated from relations between the states of its input and output.

The first transmission characteristic of a system in which we are interested is the reliability with which it transforms input states into predictable output states, or what we shall call its stability.

(23) Absolute stability: the absolute stability of a system is the degree to which the states of its output system are predictable from the states of its input system.

$$S = H\,(N) - H_M\,(N)$$

$$= - \sum_m p\,(m) \log_2 p\,(m) + \sum_{m,n} p\,(m,n) \log_2 p_m\,(n)$$

where $H\,(N)$ is the absolute uncertainty of the output system, $H_M\,(N)$ is the absolute conditional uncertainty of the output given the input, m is any of the set of events M displayed during the period of measurement, and n is any of the set of output events N displayed during the period of measurement.

If the occurrence of the output events is independent of the occurrence of the input events (i.e., if all $p_m(n) = p\,(n)$), then S takes on its minimal value of O. If the output events are completely dependent upon the input events (i.e., if all $p_m(n)$ are O or 1), then S takes on its maximum value of $H\,(N)$. It is to be noted that this measure is identical in form to those for absolute redundancy (9) and absolute dependency (17). * In the illustrative computational matrix given in Figure 5 the rows correspond to states of the input system of any given test system and the columns correspond to the states of the output system from the same test system. The value in each cell therefore re-

FIGURE 5 ABOUT HERE

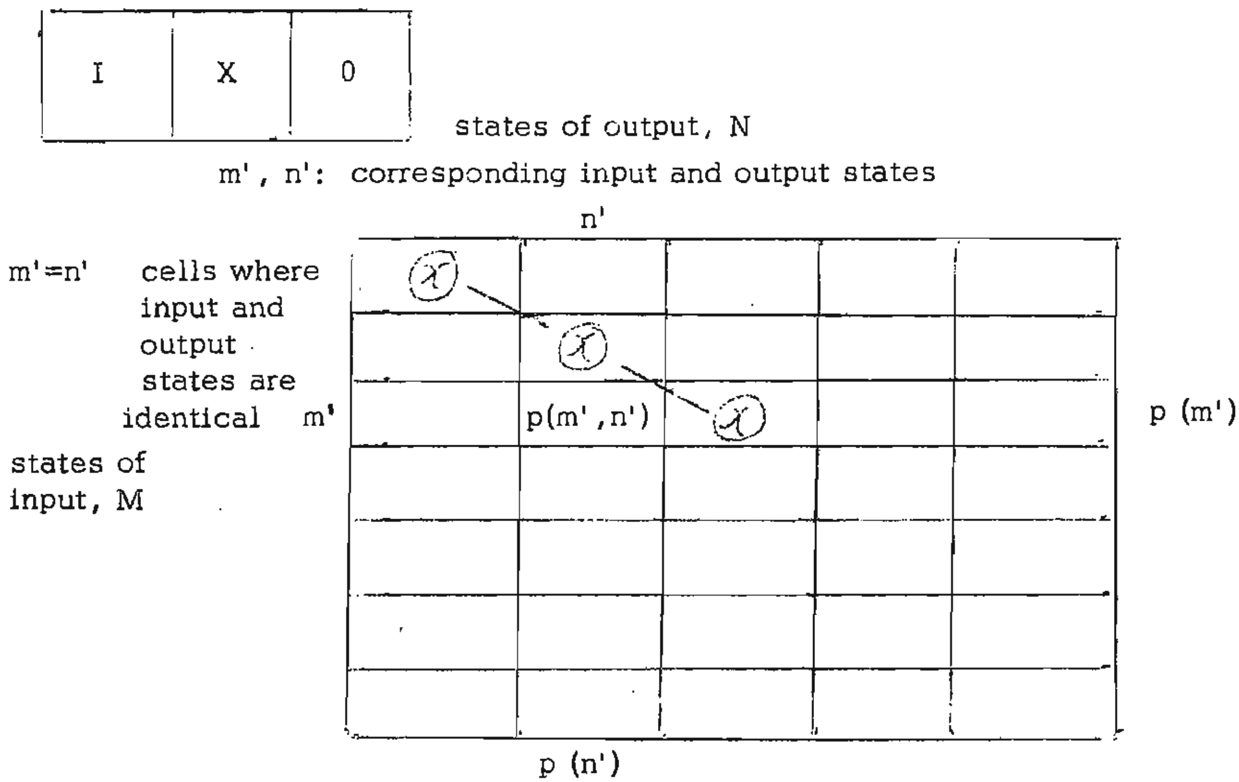
presents that frequency with which a particular input state is accompanied or followed by a particular output state. To the extent that there is concentration of entries to this matrix in specific cells, the stability of the test system is high, i.e., the test system operates in such a way that specific states of its input result in predictable specific states in its output. $H\,(N)$ is estimated from the column totals and $H_M\,(N)$ from the body of the matrix. Since the values of $H\,(N)$ will vary from system to system, and since we will often wish to compare different systems in terms of their stabilities, it is useful to transform this measure into relative terms.

---

* S is also equivalent to the usual measure of information transmission through X as a channel.

FIGURE 5

Transmission Characteristics of Single Systems



(23) $S = H(N) - H_M(N)$ $[0, H(N)]$

(24) $S_{rel} = \frac{S}{H(N)}$ $[0, 1]$

(25) $CC = \sum_{m', n'} p(m', n')$ $[0, 1]$

(26) $F = \sum_{m', n'} p(m', n' = m')$ $[0, CC]$

(27) $F_{rel} = \frac{F}{CC}$ $[0, 1]$

(28) $A_{m,n} = \sum_{m,n} \left\{ \log p(n) - \log_2 p(n) \right\} f(m,n)$

$\left[0, \left\{ \sum_{m,n} \log_2 p(n) \right\} f(m,n) \right]$

(29) $E_{m,n} = \frac{A_{m,n}}{A_{max.}}$ $[0, 1]$

(24) Relative stability: the relative stability of a system is the ratio of its absolute stability to its maximum stability, assuming that $\underline{H}$ (N) is fixed at some constant value.

$$S_{rel} = \frac{S}{H\ (N)}$$

When output is independent of input, $\underline{S}_{rel}$ equals 0. When output is completely dependent upon input, then $\underline{S}_{rel}$ equals 1.00. This index is formally equivalent to relative redundancy (10) and relative dependency (18). One could also devise measures for specific stability and relative specific stability (analogous to specific and relative specific redundancy and dependency (11, 12 and 19, 20) should need for them arise.

Measurement of the stability of a system does not require that the states of its input system and output system be in any way comparable or corresponding. It merely indicates the predictability of some state of the output system given some state of the input system. In certain communication problems, however, it is important to ask to what extent states of the input are translated into corresponding states of the output. Take the problem of machine translation systems, for example: A machine translator could exhibit perfect stability, as we have defined it, if a particular vocabulary item in language A (input) resulted in the printing of some particular vocabulary item in language B (output) -- and it wouldn't make any difference whether input Pferd resulted in consistent output horse or man or happy, or anything else. But we would like to measure what we call the fidelity of this translator, e.g., the extent to which particular vocabulary items in language A result in the printing of corresponding items in language B, that is, the predictability of getting horse when Pferd is fed in. Analysis of this sort assumes some external criterion of correspondence.

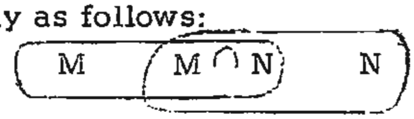
(25) Correspondence: the correspondence between any two systems is the proportion of the states in one that are identical with states of the other.

$$C\,C = \sum_{m',\,n'} p\ (m',n')$$

where m' and n' are any members of the set of common input and output events, $M \cap N$.*

---

* $M \cap N$ is defined as the intersection of the sets M and N. $M \cap N$ may be indicated graphically as follows:



If all the states of one system are displayed by another system, then their correspondence is 1.00; if none of the states of one system are displayed by another system, then their correspondence is 0. It is apparent that the possible fidelity of any test system (e.g., our mechanical translator above) will be limited by the degree of correspondence between its input and output systems. To the extent that vocabulary items in one language do not have semantic equivalents in another language, perhaps as functions of the alien nature of the cultures involved, no translator (machine or otherwise) can manifest fidelity.

(26) Fidelity: the fidelity of any system is the degree to which the states of its output depend upon corresponding states of its input.

$$F = \sum_{m', n'} p\ (m', n' = m')$$

$\underline{F}$ is simply the sum of the probabilities of corresponding input and output events. In terms of the computational matrix in Figure 5, it is the sum of the $\underline{p}$'s for the cells of the main diagonal within the sub-matrix defined by corresponding rows and columns. If the input and output systems were in perfect correspondence, then this measure would be the sum of all probabilities in the main diagonal and the value could approach 1.00 to the extent that the test system translated input events into corresponding output events without error. The smaller the sub-matrix of corresponding states in relation to the total matrix, the lower is the maximum fidelity of the system concerned.

(27) Relative fidelity: the relative fidelity of a system is the ratio of its fidelity to its maximum possible fidelity.

$$F_{rel} = \frac{\Sigma\ p\ (m', n' = m')}{\Sigma\ p\ (m', n')}$$

where $\Sigma\ p(m', n')$ is the sum of all $\underline{p}$'s falling within the sub-matrix defined by corresponding rows and columns.

If a system operates in such a way that all of its output states which have corresponding input states are completely dependent upon these corresponding input states, its relative fidelity will be 1.00. If none of its output states depend upon corresponding input states, or if there is no correspondence between input and output states, its relative fidelity will be 0. It should be noted that these fidelity measures are not entropy measures, although they have a logical function within the model as a whole. It should also be noted that the maximum possible fidelity of a system is limited by its stability. A system can have high fidelity only if it has high stability, but a system can have high stability and zero fidelity. A device which substituted z for a, y for b, etc. in printed texts would be such a system - i.e. a simple recoder. The close analogy between stability and fidelity, as used here, and reliability and validity, as used in psychometric testing theory, will be apparent.

The amount of information handled by a given test system and the efficiency of its operation in this respect can be estimated in the same fashion as described in the preceding section.

(28) Amount of information transmitted by a system:

$$A_{m,n} = - \sum_{m,n} \left[\log_2 p(n) - \log_2 p_m(n) \right] f(m,n)$$

(29) Efficiency of information transmission:

$$E_{m,n} = \frac{A_{m,n}}{A_{max}}$$

where m and n are states of input and output systems of a given test system respectively.

In other words, in the present case we are estimating the efficiency of a particular system in transmitting information between two points rather than the efficiency of transmission between two systems over a coupling.

Most applications of information theory in psychology have been of this type. A set of stimuli is the set of inputs and a set of responses is the set of outputs, and one or more human subjects constitute the system whose transmission characteristics are being studied. Garner and Hake's (1) proposed application of information measures to the psychophysical method of absolute judgment is one example. Senders and Cohen's study (9, pp. 282-289) concerning the effect of sequential dependence on instrument reading performance is another example of this type of application. The measure of the amount of information transmitted, T, used in these applications is equivalent to our measure of stability, S.

INFORMATION TRANSFER BETWEEN NON-COUPLED BUT CORRESPONDING SYSTEMS

When one individual talks to another in their common language, we have a communication situation in which information is being transferred between two more-or-less corresponding systems (the semantic or mediator systems of the two individuals) via a channel which includes non-corresponding systems (recoders, message system, and so forth). Similarly, when the artist expresses his moods and meanings via oils on canvas and other, less talented, mortals observe and appreciate his work, information is being transferred between more-or-less corresponding systems via non-corresponding ones. Or if we analyse the content of messages produced by one communicating unit (e.g., radio programs) and relate them to the content of messages produced by other communicating units (e.g., members of the audience) who have received them, we are again dealing with transfer between more-or-less corresponding systems via a channel including non-corresponding ones. These are all typical situations that are of concern to those working with human communication problems.

(30) Channel: a channel is a series of one or more non-corresponding systems connecting two corresponding systems.

(31) Source: the source (X) is that one of two corresponding systems connected by a channel whose information is independent of the information in the other during the period of observation.

(32) Destination: The destination (Y) is that one of two corresponding systems connected by a channel whose information is dependent upon information in the other during the period of observation.

These relations are shown in the upper portion of Figure 6. It is understood, of course, that a given system may alternately assume source and destination functions (e.g., when two people talk back and forth). It should also be pointed out that "non-corresponding" and "corresponding" are matters of degree, as is the dependency relation which defines source and destination.

The first measures we require here are estimates of the degree of over-all and specific dependencies between the states in one system (source) and the states in the other (destination). These measures will be strictly analogous to those for redundancy of sequential states in one system (9, 10, 11, 12), dependency of states in coupled systems (17, 18, 19, 20), and stability of input/output for a given test system (23, 24). When dealing with dependencies between corresponding systems connected by a channel, we shall refer to the contact between source and destination.

(33) Absolute contact:

$$C = H(Y) - H_X(Y)$$

the absolute contact between two systems is the degree to which the states of the destination, Y, are predictable from the states of the source, X.

(34) Relative contact: the relative contact between two systems is the ratio of their absolute contact to the maximum contact, assuming that H (Y) is fixed at some constant value.

$$C_{rel} = \frac{C}{H\,(Y)}$$

(35) Specific contact: the specific contact between a given state, y, of a destination and a given state, or set of states, x, of the source is the degree to which its occurrence is predictable from the occurrence of the given state, or set of states, of the source.

$$C\,(x,y) = p_x\,(y) - p\,(y)$$

(36) Relative specific contact: the relative specific contact between a destination state, y, and a source state, x, is the ratio of the specific contact to the maximum value for its sign.

$$c'_{rel}\,(x,y) = \frac{c(x,y)}{1 - p(y)} \quad \text{if } c(x,y) > 0$$

$$= \frac{c(x,y)}{p\,(y)} \quad \text{if } c(x,y) < 0$$

In Figure 6 the rows represent states of the source system and the columns states of the destination system, with corresponding states denoted by identical letters; H (Y) is

FIGURE 6 ABOUT HERE

computed from the column totals, H_X (Y) from the body of the matrix, and c (x,y) from the values in particular cells.

In the previous section distinction was made between the stability and fidelity of a single system; a similar distinction needs to be drawn here, between contact and what we shall term communication. Take for example the case of the abstract artist: he may, as a source, exhibit a high degree of contact with his audience, in the sense that whenever he intends some mood his viewers consistently get some specific reaction; but yet he may exhibit a low degree of communication, in the sense that when he's intending "quiet pleasure" his audience experiences "abandon", when he's intending "hope" his audience experiences "dizziness", and so on. When people use the same language (e.g., the same sign vehicles) but have different meanings for the terms, the same result is obtained -- a high degree of contact but a low degree of communication. Experiences

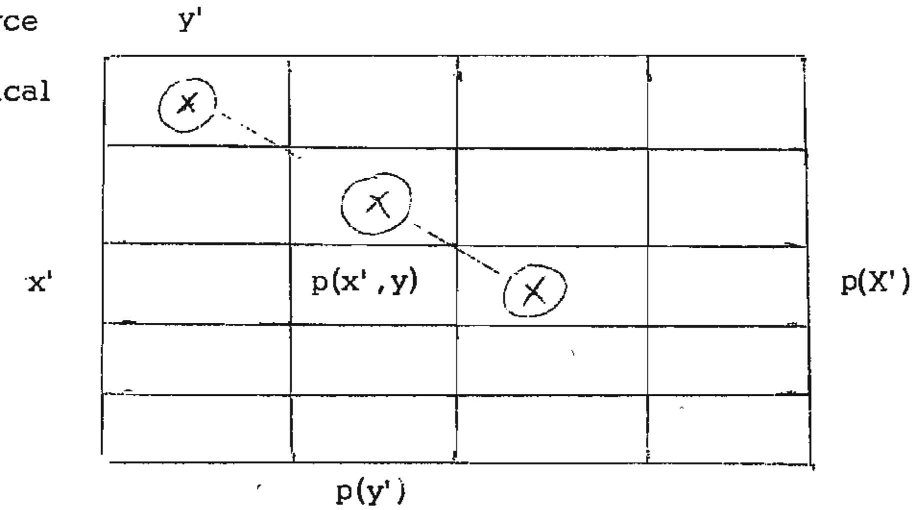
FIGURE 6

Information Transfer Between Non-Coupled but Corresponding Systems

X	Channel	Y

states of destination, Y

cells where source and destination states are identical



(33) $C = H(Y) - H_x(Y)$ $\quad [O, H(Y)]$

(34) $C_{rel} = \dfrac{C}{H(Y)}$ $\quad [O, 1]$

(35) $c(x,y) = p_x(y) - p(y)$ $\quad [-p(y), 1 - p(y)]$

(36) $c_{rel}(x,y) = \dfrac{c(x,y)}{1 - p(y)}$ if $c(x,y) > 0$, $\dfrac{c(x,y)}{p(y)}$ if $C(x,y) < 0$

(37) $U = \sum_{x',y'} p(x', y' = x')$ $\quad [O, \sum_{x',y'} p(x', y')]$

(38) $U_{rel} = \dfrac{U}{\sum_{x',y'} p(x', y')}$ $\quad [O, 1]$

(39) $A_{x,y} = -\sum_{x,y} \left\{ \log_2 p(y) - \log_2 p_x(y) \right\} f(x,y)$

$[O, \sum_{x',y'} \left\{ \log_2 p(y) \right\} f(x,y)]$

(40) $E_{x,y} = \dfrac{A_{x,y}}{\sum_{x',y'} \left\{ \log_2 p(y) \right\} f(x,y)}$ $\quad [O, 1]$

of American soldiers with English girls during the last war provided some humorous examples of this situation.

(37) Communication: the communication between two corresponding systems connected by a channel is the degree to which the states of the destination system depend upon corresponding states of the source system during a period of measurement.

$$U = \sum_{x',y'} p\,(x', y' = x')$$

where x' and y' are any events in the set $X \cap Y$.

U is simply the sum of the probabilities of corresponding source and destination states, i.e., in Figure 6, the sum of the p's for the main diagonal cells within the sub-matrix defined by corresponding rows and columns. Perfect communication between source and destination occurs when entries are restricted to the main diagonal and the two systems are in perfect correspondence. Again, as was the case with fidelity measurement, it is useful to have an index of relative communication -- different source and destination pairs will vary in over-all correspondence and hence in the absolute communication possible, e.g., speakers of different languages depending on gestures.

(38) Relative communication: the relative communication between a source and destination is the ratio of their communication to the maximum communication possible.

$$U_{rel} = \frac{\Sigma\, p\,(x', y' = x')}{\Sigma\, p\,(x', y')}$$

where $\Sigma\, p\,(x', y')$ is the sum of all p's falling within the sub-matrix defined by corresponding rows and columns and $\Sigma p\,(x', y' = x')$ is the sum of all p's falling within the diagonal of this sub-matrix.

Thus, two speakers of different languages might display little communication when estimated in terms of the total semantic states of which each was capable, yet within the sub-system of gestures there might be a fairly high degree of relative communication.

Measures of the amount of information transmitted (A) and the efficiency of information transmission (E), described in connection with information transfer between coupled systems, are equally applicable in the present situation.

(39) Amount of information transmitted between non-coupled, corresponding systems:

$$A_{x,y} = -\sum_{x,y} (\log_2 p\,(y) - \log_2 p_x\,(y))\, f\,(x,y)$$

(40) Efficiency of information transfer between non-coupled, corresponding systems:

$$E_{x,y} = \frac{A_{x,y}}{A_{max}}$$

where $\underline{x}$ and $\underline{y}$ are states of source and destination systems respectively.

A complete analysis of activity in any channel -- including a source and a destination connected via a series of non-corresponding systems -- would involve measurement of the inherent characteristics of each system (maximum and relative uncertainty, relative redundancy, and specific redundancies), information transfer between each system and the succeeding system to which it is coupled (relative dependency, specific dependencies, amount and efficiency of information flow), the transmission characteristics of each system (its relative stability, relative fidelity, and amount and efficiency of handling information), and finally information transfer between the corresponding systems in the channel, particularly the source and destination (relative contact, relative communication, and amount and efficiency of information transfer). In practice, however, as our illustrations have shown, each particular problem will generally demand certain of these measures rather than others, e.g., the inherent characteristics of a system when studying phoneme patterning and the transfer of information between corresponding systems when studying art as a form of communication.

Informational Properties of Judgments of Facial Expressions.

An experiment was conducted where student actors portrayed emotional states designated by a set of 40 verbal labels (e.g., expectancy, awe, dread, loathing).* The student observers selected one label from the same set to designate the emotion they thought was being portrayed. The communication situation may be diagrammed as below:

I	X	Channel	Y	O
Input (40 labels)	Semantic (Intentional) System	Facial Expressions	Semantic (Significance) System	Output (40 labels)

* This experiment was carried out by the first author. A report, which duplicates much of the material in this section is contained in (9, pp. 374-384). This report also contains a fuller account of the "dimensional" analysis of facial expressions.

It seems reasonable to assume that for actors and observers, who all use highly similar language systems, the semantic states (i.e. "meanings") generated by the 40 labels will be highly similar; e.g. the label <u>sullen</u> <u>anger</u> will have about the same "meaning" for all actors and observers. If this assumption is correct, we can study the information transmission between the non-coupled, but corresponding, semantic systems of the actors and the observers through the channel of facial expressions by an informational analysis of the relations between input and output labels as outlined in Figure 6.

X is the set of labels the actors intended to portray and Y is the set of choices made by the subjects. Since events in X and Y correspond, H_{max} (eq. 3) for both X and Y is

$$H_{max} = \log_2 40 = 5.32$$

The obtained values of H (X) and H (Y) (eq. 4) are

$$H(X) = 5.19 \text{ and } H(Y) = 5.20$$

while the corresponding values of relative uncertainty (eq. 5) are

$$H_{rel}(X) = 0.97 \text{ and } H_{rel}(Y) = 0.98$$

The values for X are not surprising since the states the actors were asked to portray were randomly selected. The results for Y indicate that the observers were equally "unbiased" in distributing their judgments, presumably because they were told each label would be portrayed once.

The value of $H_X(Y)$ (eq. 7) was 4.24, so the measure of <u>contact</u> (eq. 33) is

$$C = 0.93$$

and relative contact (eq. 34) is

$$C_{rel} = 0.17$$

The measure of communication (eq. 37) is

$$U = 0.129$$

U_{rel} has the same value since the correspondence of X and Y makes

$$\sum_{x',y'} \underline{p}(s',y') = 0.129$$

Both $\underline{C}$ and $\underline{U}$ seem to indicate that there is little accuracy in judgments of facial expressions. However, an inspection of the original data indicated that, while observers made a considerable proportion of "errors" (about 87%) in interpreting the actors' intentions, there was a marked tendency to choose labels "similar" to those intended by the actor. For example, fear expressions produced sizable proportions of anxiety and horror judgments but very few judgments of glee. This suggests that the insensitivity of $\underline{C}$ and $\underline{U}$ to "near misses" does not reflect an appreciable amount of communication that actually occurs in this situation. The value of the $\underline{U}$ measure is not altered by any sort of change in the values of the off-diagonal cells of Figure 6, regardless of the effect of these changes on the "accuracy" of judgments. $\underline{C}$ is not quite so insensitive, but it would be unaltered by any rearrangement (i.e., permutation) these off-diagonal values provided that the p (x') and p(y') values were not altered.

The preceding comments are based on an implicit assumption that "similarity" of events in X or Y is a meaningful concept. If these events constitute discrete, unordered categories (i.e., a nominal scale) like the dots, dashes or spaces of Morse code telegraphy, so that the events are merely the "same" or "different", then the concept of similarity is meaningless. If these events are ordered along a single continuous dimension (i.e., an interval or ratio scale) like sound intensity or, presumably, absolute loudness judgments, then similarity is a meaningful concept and can be defined in terms of separation of scale units. For such continuous dimensions it would be possible to measure degree of correspondence between X and Y events -- i.e., degree of communication -- by one of the various correlational measures.

The ordering of facial expressions seems to fall somewhere between these two extremes. Intuitively, we would agree that glee and joy are more similar than glee and rage. However, there is no standard scale of facial expressions and we do not even know the nature or number of dimensions necessary for a complete system of measurement. An attempt was made to scale the facial expressions by computing a measure of distance between all possible pairs of expressions. This distance measure was the measure of profile similarity devised by Osgood and Suci (7) which, in this application, was obtained from the formula below:

$$D^2_{1,2} = \Sigma [p\ (x,y_1) - p\ (x,y_2)\ ]^2$$

where $p\ (x,y_1)$ and $p\ (x,y_2)$ are the probabilities of an actor's "intention" of x being judged as y_1, or y_2 respectively. The 40 x 40 matrix of distances obtained in this manner could be quite accurately reproduced in three independent dimensions. One dimension was labeled Intensity with complacency and quiet pleasure at the low end and dismay and contempt at the high end. A second dimension was labeled Pleasantness with sulkiness and sullen anger at the low end and glee and joy at the high end. A third dimension was labeled Control with awe and horror at the low end and annoyance and loathing at the high end.* This analysis provided an ordering of the facial expressions along three dimensions and made possible the measurement of the degree of agreement between the intentions of the actors and the judgments of the observers by means of product-moment correlations. The correlation was +0.32 for the Intensity dimension, +0.38 for the Pleasantness dimension and +0.50 for the Control dimension. The multiple correlation for all three dimensions was +0.71, indicating that about half of the variance in significances is associated with the intentions of the actors. Thus, there appears to be a considerable degree of communication by means of facial expressions.

The discussion above should not be taken to mean that information measures should not be applied to data where the events in X or Y are not ordered along a nominal scale. Rather, it merely indicates that our interpretations should be cautious in such cases and that other measures may be needed to clarify the results.**

INHERENT CHARACTERISTICS OF COMMUNICATION NETWORKS

Many of the most interesting problems in communication analysis arise in the consideration of information exchange within networks of communicating units. Any group of individuals capable of displaying corresponding states and having available channels may constitute a network -- various types of formal and informal groups, co-workers in a department of some organization, the managerial staff of the organization, and the teacher and students in a course would be examples. If our underlying assumption is valid -- that communication processes serve to organize and determine the natures of groups -- then network analysis becomes a basic tool of the social scientist. The communicating units which comprise a network need not be individuals: A group of manufacturing concerns which together handle products of a certain kind may be treated as a network for purposes of study; the "messages" exchanged in a network may be material goods or services or money as well as words (which suggests the possibility

---

* This analysis is a sort of "rough-and-ready" multidimensional stimulus scaling. Since this type of scaling is in a very early stage of development, no definite conclusions can be made concerning the validity of these results.

** See Quastler (9, p. 384-386) for some suggestions concerning analysis of this kind of data using information measures.

of applying some of these methods to problems in economics); a collection of psychological journals may be treated as a network, with each journal conceived as receiving information from other journals (citing references) and delivering information to other journals (being cited) -- this will be our computational illustration at the end of the presentation.

Each unit, whether individual or institutional, must be conceived as simultaneously functioning in all stages of the communication process: as receiving information from other units, as interpreting this received information, as selecting information to be communicated, and as transmitting this information as messages to other units. Different types of communicating units in a society (such as a "man on the street", a "teacher", a "civic leader", and "eighth grade gang of boys", or a "local newspaper") will vary widely in what aspects of the total communication process are emphasized and how these various functions are interrelated in their communicative acts -- indeed, one of our tasks will be to devise common scales against which various types of units can be compared and distinguished.

(41) Communicating unit: any aggregate of systems including a receiver, a mediator, and a transmitter -- coupled in that order -- constitutes a communicating unit and may be treated as a single system.

For most problems in mass communication we are only indirectly concerned with the essentially psychological processes of decoding and encoding by individuals, in the analysis of which we need to treat receivers, mediators and transmitters as separate systems. It becomes convenient, therefore, to treat the individual as a whole as a single communicating unit. Similarly, it is possible to analyze separately the receiving, mediating (e.g., policy), and transmitting functions of an organization, and this is a problem for management analysts, efficiency experts, and the like -- but for many problems in mass communications it is more convenient to treat the organization as a whole as a single communicating unit.

(42) Communication network: any collection of communicating units capable of displaying corresponding states and having available channels each to the other may constitute a communication network.

Since the communicating units included in a network are, in theory, at least, completely interchangeable as sources and destinations of messages, the network can be treated as a single system in which each CU can be either an antecedent or subsequent state.

In other words, the inherent characteristics of networks can be studied in a fashion strictly analogous to that developed for studying the inherent characteristics of single systems, e.g., a system of phonemes which precede and follow one another. Rather than saying "state a is followed by state b", we will now say "unit A, as a source, sends a message to B, as a destination", but the same operations can be made, along with some others that have not been described previously.

To have sufficient generality to cover the types of communication situations suggested as examples at the beginning of this section, our definition of "message" must be a broad one.

(43) Message: a message is any event occurring in that non-corresponding system in the channel connecting two corresponding systems which is equivalently the output from the source and the input to the destination.

This definition requires some clarification. Its intention is to specify as the "message" that particular system in the channel which is (a) equally a part of the output from a source and part of the input to a destination and (b) independent of either source or destination in the sense that it can be studied without reference to either of them. In ordinary face-to-face communication, for example, neither the vocal muscle action of the source nor the auditory nerve action of the destination satisfy the definition, but the oscillating sound wave frequencies and pressures in air do -- they are at once part of the input to the destination and part of the output from the source, and they can be recorded and studied quite independent of either source or destination. In telephone communication only the patterns of electrical pulses in the wires meet these criteria. In the mass media, the printed newspaper itself, images on the television screen, and sounds coming from the radio speaker satisfy these criteria -- they are at once dependent upon their respective sources and part of the input to their respective audiences, as well as being available for independent study (e.g., the newsprint can be studied by a content analyst, the face of the television tube can be photographed, and the output of the radio can be recorded). Defined in this broad fashion, money, material goods, and even women (cf., a paper by Levi-Strauss in which kinships systems are analyzed in information theory terms), as well as language, may serve as the media of exchange, i.e., as messages.

Networks vary considerably in what might be called their "closure". If all of the units constituting a network only received messages originating from other units within the network, and only sent messages to other units within the network, we would be dealing with a completely closed system. This, of course, is an ideal rather than typical situation. Ordinarily, only a small portion of the total activity of a given communicating unit is limited to intercourse within any particular network -- only a portion of the total activity of the business manager in a plant is concerned with exchange of messages to and from other people in the plant. Ideally, it would be useful to have a measure of the degree of closure characteristic of a network: Imagine a matrix in which the rows represent all of the sources from which units in a network receive messages and the columns represent all of the destinations to which messages are sent by units in the network; the units included in the network will necessarily constitute a sub-matrix defined by corresponding rows and columns, and the degree of closure will be the ratio of the number of messages tallied within this sub-matrix to the total number of messages exchanged.

(44) Closure: the degree of closure characteristic of a network is the number of messages exchanged between communicating units within the defined network as a proportion of total number of messages exchanged between these units and all sources and destinations during the period of observation.

$$(C) = \sum_{a,b} p\,(a,b)$$

where a and b are any of the complete set of sources or destinations within the network and where $p\,(a,b)$ is the probability of a, as a source, sending a message to b, as a destination (and where $p\,(b,a)$ is the probability of b sending a message to a).

This measure of closure, it will be noted, is formally identical with that for the degree of correspondence between two systems (25). If the network being studied were completely closed, and all messages exchanged were between units specified as being in the network, then the value of (C) would be 1.00. If this value were zero, of course, we would be dealing with a network in a hypothetical sense only. In practical terms, it would seldom be possible to get the information necessary to make this estimate, since it would involve cataloguing every message exchange between each included unit and all other units communicated with during the entire period of observation --

we would have to follow our business manager about from morning to night as well as all the other people included in the network being studied.

This limitation is not a restrictive one, however. For all of the inherent characteristics of networks we shall describe below, *they will be treated as perfectly closed systems*, i.e., we shall deal only with a matrix defined by the corresponding rows and columns representing units within the network. Such a computational matrix is shown in Figure 7: the rows represent the communicating units participating in a

FIGURE 7 ABOUT HERE

network as they function as sources and the columns these same units as they function as destinations. The value in each cell would represent the relative frequency or probability of messages exchanged between each source and destination pair -- the value in cell *ab* represents the relative frequency of messages sent from *a* (as a source) to *b* (as a destination). Values in the main diagonal represent the degree of self-communication, assuming this is definable in some manner (e.g., a journal citing itself). Thus, the value in cell *aa* is the relative frequency of messages sent by unit *a* to itself. Any messages that may be received by these units from sources outside the network or may be sent by these units to destinations outside the network are not considered.

(45) *Traffic (of a unit)*: the traffic of any unit in a network is the number of messages it handles during some finite time, coded both as to source and destination.

(46) *Traffic (of a network)*: the traffic of any network is the total number of messages exchanged during some finite time, coded as to both source and destination.

Defined in this manner, traffic in a network is equivalent to information in a system. For computational purposes, the total traffic of a unit is the sum of all messages which it handles, either as source of destination, i.e.,

$$T d = \sum_a f(a,d) + \sum_b f(d,b) - f(d,d)$$

where $\sum_a f(a,d)$ is the total frequency of messages received by unit *d* as a destination, $\sum_b f(d,b)$ is the total frequency of messages sent from *d* as a source and $f(d,d)$ is the frequency of messages sent by *d* to itself which is subtracted to correct for double addition.

FIGURE 7

Inherent Characteristics of Networks

$\boxed{C_N}$

destination units, B

source units, A

		a	b		
a		p(a,a)	p(a,b)		p(.a)
b		p(b,a)	p(b,b)		p(b.)
		p(.a)	p(.b)		

(44) $(C) = \sum_{a,b} p\,(a,b)$ $[0,1]$

(45) $Td = \sum_{a} f\,(a,d) + \sum_{b} f\,(d,b) - f\,(d,d)$ $[0,\infty]$

(46) $Tn = \sum_{a}\sum_{b} f\,(a,b)$ $[0,\infty]$

(47) $H_{rel}\,(A) = \dfrac{H\,(A)}{H_{max}\,(A)}$ $[0,1]$

(48) $H_{rel}\,(B) = \dfrac{H\,(B)}{H_{max}\,(B)}$ $[0,1]$

(49) $Cn = rp\,(.a)\,p\,(a\,.)$ $[-1,1]$

(50) $F/C = \dfrac{H_{rel}\,(B)}{H_{rel}\,(A)}$ $[0,\infty]$

(51) $O\,R\,G = H\,(B) - H_A\,(B)$ $[0, H\,(B)]$

(52) $O\,R\,G_{rel} = \dfrac{O\,R\,G}{H\,(B)}$ $[0,1]$

(53) $S\,F = \sum_{a,b} p\,(a,b=a)$ $[0,1]$

(54) $S\,F_d = p\,(d,d)\,/\,p\,(.d)$ $[0,1]$

(55) $I/O_d = \dfrac{p\,(.d)}{p\,(d.)} = \dfrac{\sum_a p(a,d)}{\sum_b p(d,b)}$ $[0,\infty]$

(56) $NS_d = Td \times 0/Id$ $[0,\infty]$

The total traffic of a network is simply the sum of all messages exchanged within it during the period of observation, i. e.

$$T_N = \sum_a \sum_b f\ (a,b)$$

This value is that obtained by adding the frequencies of exchange of messages between all pairs of units in the network, i.e. the sum of all cells.

We may first briefly indicate the entropy measures that can be made here in exactly the same fashion as described for inherent characteristics of a single system. The maximum uncertainty of the network can be estimated from either row or column totals and is necessarily the same for sources and destinations:

$$H_{max}\ (A) = H_{max}\ (B) - \log_2 \bar{n}$$

where $\bar{n}$ is the number of communicating units in the network. The absolute uncertaint of both sources and destinations is computed separately from the row and column totals respectively.

$$H\ (A) = - \sum_a p\ (a) \log_2 p\ (a)$$

$$H\ (B) = - \sum_b p\ (b) \log_2 p\ (b)$$

Measures of relative uncertainty of both sources and destinations are themselves indice of important characteristics of networks, which we shall term their balance.

(47) Source balance: the source balance of a network is the degree to which the units it includes produce messages to various destinations with equal frequency.

$$H_{rel}\ (A) = \frac{H\ (A)}{H_{max}\ (A)} = \frac{H\ (A)}{\log_2 \bar{n}}$$

(48) Destination balance: the destination balance of a network is the degree to which the units it includes receive messages from various sources with equal frequency.

$$H_{rel}\ (B) = \frac{H\ (B)}{H_{max}\ (B)} = \frac{H\ (B)}{\log_2 \bar{n}}$$

When H_{rel} (A) has its maximum value of 1.00, all of the communicating units within the network are producing messages in the system with equal frequency, i.e., the input to the network is maximally balanced. When $\underline{H}_{rel}$ (A) has its minimal value of zero, a single communicating unit is producing all the messages that are handled within the system. When $\underline{H}_{rel}$ (B) has its maximal value of 1.00, all units within a network are receiving messages with equal frequency, i.e., the output of the network is maximally balanced. When $\underline{H}_{rel}$ (B) has its minimal value of zero, all messages are being received by a single unit within the network.*

If a network exhibits a high degree of what we shall call congruence, there will be high correlation between its input and output distributions, i.e., between its source and destination totals. In other words, a congruent network is one in which units on the average tend to produce messages in proportion to the extent that they receive messages.

(49) Congruence: the congruence of any network is the degree of correlation between its source and destination probabilities when these are ordered according to corresponding units.

$$C_n = rp\,(.a)\ p\,(a.)$$

where r is the correlation coefficient, $p(a.) = \sum_b p\,(a,b)$, the probability of a being a source, and $p(.a) = \sum_b p\,(b,a)$, the probability of a being a destination.

When this congruence measure has its maximal positive value of 1.00, the most active units as producers of messages are also the most active units as receivers of messages. When congruence has its maximum negative value of -1.00, the most active producers of messages in the network are the least active receivers of messages. When congruence has its minimal value of zero, there is no relation between the rate at which units produce messages and the rate at which they receive messages. It is possible to obtain equivalent measures of congruence for the single units within a network, in which case we would speak of the specific source balance, the specific destination balance, and the specific congruence of a communicating unit within a network. These computations would be limited to the particular row and column defined

---

*It should be noted that although messages received by a given unit are part of its input, the sum of messages being received by all units is the output of the network, viewed as a single system.

by corresponding source and destination labels, i.e., we would be obtaining the relative entropy of the sources and destinations of messages handled by a single unit and their correlation.

Another measure obtainable from column and row totals of the matrix is what might rather picturesquely be called the filter/condenser ratio. It is simply the ratio of the relative entropy of the output (destinations) to the relative entropy of the input (sources) of a network.

(50) Filter/condenser ratio: the filter/condenser ratio of a network is the ratio of its destination balance to its source balance.

$$F/C = \frac{H_{rel}\,(B)}{H_{rel}\,(A)}$$

If a network operates in such a way that a few communicating units produce most of the messages (H_{rel} (A) is near zero) but all of the units share equally in the reception of messages (H_{rel} (B) is near 1.00), then F/C ratio will be large and this network may be characterized as a filter system, i.e., it operates to "fan out" information from a few sources to many destinations. The typical teacher-student group is such a filter system, as is the leased-wire news service. If a network operates in such a fashion that its units share equally in the production of messages (H_{rel} (A) is near 1.00) but only a few of these units are the destinations of messages (H_{rel} (B) is near zero), then the F/C ratio will be small and this network may be characterized as a condenser system, i.e., it operates to concentrate information from many sources onto a few destinations. Information-gathering staffs of organizations such as newspapers and the State Department would thus be condenser systems. Similar measurements may be obtained for single units within networks (by restricting calculations to the row and column defined by corresponding labels), in which case we would refer to specific filter/condenser ratios. Viewed within the combat army as a network, "observers" might appear as filters of information; within business organization networks, public relations men (or "mouthpieces") might appear as condensers of information. It should be carefully noted, however, that the specific f/c ratio is that of the relative entropy of the row (output of a unit) to the relative entropy of the corresponding column (input to the same unit). If a unit's output is relatively more balanced than its input, it is functioning as a filterer of information in the network; if its output is relatively less balanced than its input, it is functioning as a condenser of information in the network.

What about the predictability of the destination of a message in a network from knowledge of its source? This measure, which we shall call the organization of a network, is formally identical with that for redundancy (cf., 9,10) in a single system.

(51) Organization: a network is organized to the degree that the destinations of messages are predictable from knowledge of their sources.

$$ORG = H(B) - H_A(B)$$

and in analogous fashion, the relative organization of a network is given by

(52) Relative organization: the relative organization of a network is the ratio of the organization measure to its maximum value, assuming H (B) is a constant.

$$ORG_{rel} = \frac{ORG}{H(B)}$$

a measure ranging from zero (complete uncertainty of destinations given sources) to a maximum of 1.00 (complete dependency of destinations upon sources). In what we might call a relatively unorganized network, source a might tend to send messages at a relatively high rate to destination b, but this rate would be approximately that predictable from the high output of a and high input of b, i.e., that expected from the row and column totals bounding the ab cell. In a relatively organized network, on the other hand, source a might tend to send messages to destination b at a rate either significantly higher or significantly lower than that expected from the marginal totals bounding the ab cell, i.e., there is an unexpected closeness or absence of association between particular sources and destinations in the network. The relation of this type of measure to contingency, chi-square, and interaction measures will be apparent. In the study of social groups, this measure might be evidence for the existence of 'cliques' or specializing sub-networks within the more inclusive one. Finally, just as we are sometimes interested in specific redundancy, so here we may be interested in specific unit organization (i.e., lack of randomness in the messages sent and received by particular units in the network). The measures used here will be formally the same as those given for specific and relative specific redundancy (11, 12).

If the types of messages permit, one may obtain an index of the degree of self-feeding characteristic of the network, the degree to which units send messages to themselves.

(53) Network self-feeding: the self-feeding of a network is the degree to which the sources of messages are their own destinations.

$$SF = \sum_{a,b} p(a,b = a)$$

This will be recognized as equivalent to the fidelity estimate of a system. It is simply the sum of all ρ-values falling in the diagonal cells of the network matrix. The degree of self-feeding characteristic of any unit within a network can also be estimated limiting computations to the row and column representing that unit.

(54) Unit self-feeding: the self-feeding characteristic of a particular unit within a network is the degree to which it is the destination of its own messages.

$$SF_d = p\,(d,d) \;/\; p\,(.d)$$

where $p\,(d,d)$ is the probability of d receiving a message from itself and $p\,(.d)$ is the probability of d receiving a message from any source, this ratio being the proportion of all messages received which d sends to itself.

Particular units may also vary in the relative degree to which they function as feeders of information into the network or storers of information within the network; if they tend to send more messages than they receive they are feeders of information into the network; storers of information, on the other hand, receive more than send.

(55) Unit input/output ratio: the input/output ratio of a communicating unit within a network is the ratio of the probability of it receiving a message to the probability of its sending a message.

$$I/O_d = \frac{p\,(.d)}{p\,(d.)} = \frac{{}_a\,p(a,d)}{{}_b\,p(d,b)}$$

However, since individual units within a network may vary considerably in the total number of messages they handle, i.e., their traffic, it is useful to include this factor in estimating the degree to which a unit functions as a storer of information within a network.

(56) Network storing by a unit: the degree to which a unit in a network stores information circulated in the network is a multiplicative function of its traffic and its input/output ratio.

$$NS_d = T_d \times I/O_d$$

When applied to the complete set of units within the network, this index will order them according to the overall degree to which they are feeding information into the network (low values) or storing information (high values).

Communication Among Psychological Journals

In this application we treat each of 21 psychological journals as a communicating unit; a journal receives information from other journals to the extent that its authors cite other journals as references and it sends information to other journals to the extent that its articles are cited by the authors of other journals. All citations for the year 1950 were examined and serve as the raw data.* Thus for any pair of journals, a and b, we have the frequency of citation of a by b and of b by a, and since the journals being cited are sources and the journals doing the citing are destinations, we can apply the analytic techniques described in this section, treating each citation as a message. There were, of course, citations to and from journals outside the system as well as books, and the degree of closure (eq. 44) of the psychological journal network could have been estimated, but for convenience we have analyzed the data as if for a closed network.

The total traffic in the entire network (eq. 46) was

$$T_N = 4{,}046 \text{ citations}$$

The distribution of this traffic among the 21 individual journals (T_d, eq. 45) is given in Table 2. These values are about what one would expect. The Journal of Experimental Psychology published a large number of short research articles with extensive citation of other relevant research, the Psychological Bulletin is a review journal for the most part, summarizing research activities over considerable periods, and the Psychological Review reports theoretical papers usually replete with bibliographical support. The Psychoanalytic Quarterly, the Psychoanalytic Review, and the American Journal of Orthopsychiatry often publish detailed case studies and theoretical speculations, which may account in part for their low T_d values, but they are also out of the main stream of professional psychological work (and we were counting only intra-network citations).

Since we have 21 journals in the network, maximum uncertainty (eq. 3) is

$$H_{max} = \log_2 21 = 4.39.$$

The absolute uncertainties (eq. 4) for sources (A) and destinations (B) are

$$H(A) = 4.02 \text{ and } H(B) = 4.11$$

and from these values we can compute the source balance and destination balance

*These data were collected by Mr. John J. Ball.

TABLE 2

Measures of Specific Characteristics of Psychological Journals

Measure	T_d	C_d	SF_d	I/O_d
Equation No.	(45)	(49)	(54)	(55)
Amer. J. Orthopsych.	168	0.49	0.41	1.04
Amer. J. Psychol.	343	0.70	0.24	0.66
Educ. & Psych. Meas.	190	0.74	0.46	1.86
J. Abnorm. & Soc. Psychol.	513	0.80	0.26	0.95
J. App. Psychol.	325	0.87	0.42	0.84
J. Clin. Psychol.	218	0.90	0.34	1.79
J. Comp. & Phys. Psychol.	520	0.63	0.35	0.73
J. Cons. Psychol.	400	0.47	0.42	1.16
J. Educ. Psychol.	174	0.31	0.14	0.56
J. Exp. Psychol.	962	0.66	0.39	0.88
J. Gen. Psychol.	418	0.86	0.16	1.37
J. Genetic Psychol.	312	0.34	0.38	1.90
J. Personality	158	0.82	0.32	1.83
J. Psychol.	379	0.11	0.11	0.70
J. Soc. Psychol.	210	0.64	0.24	1.16
Psychiat. Quat.	124	0.69	0.51	1.51
Psychiatry	113	0.78	0.48	0.89
Psychoanal. Quat.	81	0.65	0.35	0.66
Psychoanal. Rev.	55	0.57	0.21	1.17
Psych. Bull.	612	0.26	0.20	1.51
Psych. Rev.	590	0.56	0.21	0.67

(eq's. 47 and 48) in this psychological journal network,

$$H_{rel}\ (A) = 0.92 \text{ and } H_{rel}\ (B) = 0.94$$

These values approach their maxima and indicate that this network is quite well balanced, i.e., that frequencies of both citing and being cited are rather evenly distributed over the 21 journals.

The congruence (eq. 49) of this network is an index of the degree to which source activity (being cited) is correlated with destination activity (citing). The correlation value of

$$C_N = 0.88$$

indicates that to a considerable degree the units which send the most messages also receive the most messages, i.e., there are 'active' journals and relatively 'inactive' journals with respect to their exchange of citations in the reference network. Measures of specific congruence (eq. 49 applied to the rows and columns for individual units) are given as the second column in Table 2. The fact that all these correlations are positive shows that in general each journal tends to cite other journals in proportion to the extent that they cite it. This is least true of the *Psychological Bulletin*, which could be expected from its essential review function.

The filter-condenser ratio (eq. 50) of this network, the ratio of source to destination balances, is

$$F/C = 1.02$$

i.e., the network is functioning neither as a filterer nor a condenser of information, but rather the "sharing" in being cited is equal to the "sharing" in citing. This would not be the case in a professional journal network where one or two journals concentrated on original research and the rest on theoretical and review articles utilizing this information.

To what extent can the journal doing the citing (destination) be predicted from knowing the journal cited (source)? The measure of organization (eq. 51) of this network,

$$ORG = 1.23$$

and the derived index of relative organization,

$$ORG_{rel} = 0.30$$

indicate a relatively low order of predictability of destination from source. However, the fact that the value does differ significantly from chance* exchange of messages is indicative of some tendency for the formation of "cliques", i.e., journals which cite each other above chance and other "outsider" journals less than chance. As might be expected, these "cliques" are organized around specialization of content -- general experimental journals (Journal of Experimental Psychology, Journal of Comparative and Physiological Psychology, the American Journal of Psychology, and the Journal of General Psychology) cite each other with above chance frequencies but psychoanalytic and clinical journals below chance and vice versa.

The network self-feeding characteristic (eq.53) is the proportion of total messages which are received by the same unit as sent them. For the psychological journal network this index has a value of

$$SF = 0.30$$

In other words, 30% of all the citations made in these journals were to other articles previously published in the same journal. This is, again, indicative of a tendency for psychological journal s to be quite selective contentwise; the author of a research paper or a theoretical paper in this field is likely to send his piece to that journal which, in the past, has carried papers on the same topic or of the same type (and certainly psychologists are well aware of this characteristic of their publication network). Inspection of the values for unit self-feeding, given as the third column of Table 2, support this impression. Psychiatry and the Psychiatric Quarterly

TABLE 2 ABOUT HERE

have relatively high self-feeding indices, as do the Journal of Educational and Psychological Measurement (specializing in statistical contributions), The Journal of Applied Psychology, the Journal of Consulting Psychology, the Journal of Experimental Psychology, and the Journal of Genetic Psychology. Most "ecclectic", as might be expected from their titles, are the Journal of General Psychology and the Journal of Psychology. Also low in the self-feeding characteristic were the American Journal of Psychology, the Journal of Educational Psychology and the Journal of Social Psychology.

The fact that both the Psychological Bulletin and the Psychological Review are "ecclectic" (low in self-feeding) is understandable from their functions in the network,

* For the sample size and number of degrees of freedom, a value of ORG of .083 would deviate significantly from 0 at the 1% level (9,pp. 99-100).

and this may apply to the Psychoanalytic Review as well.

The input/output ratios (eq. 55) for the individual journals are given as column 4 in Table 2. This is the ratio of the probability of receiving (i.e., citing) to that of sending a message (i.e., being cited). Presumably journals carrying original contributions of either an experimental or theoretical nature would have low values here, i.e., would be used as references more often than the reverse. All of the journals having less than unity I/O_d values seem to fit this description (e.g., American Journal of Psychology, Journal of Experimental Psychology, and the Psychological Review). These dominantly are feeders of information into the network. On the other hand, the one journal that we know to be a storer of psychological information, the Psychological Bulletin, has an appropriately high I/O_d value. But why organs like Educational and Psychological Measurement, the Journal of Clinical Psychology, the Journal of Genetic Psychology, the Journal of Personality, and the Psychiatric Quarterly should have such high I/O_c ratios is not readily apparent. It is a fact that they cite other journals much more than they are themselves cited, but why this should be so we do not know. The network storing characteristic (NS_d, eq. 56) of each unit can be derived simply by multiplying its I/O_d value by its T_d value, and hence they are not reproduced in Table 2. As would be expected, the Psychological Bulletin has the highest value here.

These measures seem to provide a fairly extensive and, we feel, reasonably valid description of the communication network of psychological journals -- at least to the extent that citations are an adequate message unit. But just as was the case with the inherent characteristics of single systems (e.g., English phonemes), here also we would like to compare the psychological journal network with other professional, or even non-professional, publication systems. Is the psychological network comparatively well balanced? Does it show more or less organization into "clique-like" sub-systems than other professional journal networks? One can see where evidence of this sort could be of value to a policy-making group such as the Board of Editors of the American Psychological Association. As far as the individual editor of a journal within the network is concerned, he too might derive some new insights from such a descriptive picture as this -- particularly were a time-sequence of the network's operation through a number of years prepared. Could it be shown that the role of the *Psychological Bulletin* is changing with respect to its network storing characteristic? Is the *Journal of Experimental Psychology* becoming progressively more of a self-feeder? Is the *Journal of General Psychology* becoming more and more dependent upon other experimental journals? In other words, it is feasible to make comparisons across time samples with these measures as well as across networks.

THE PROBLEM OF STRUCTURE

Whether dealing with networks, with single systems, with transfer of information between coupled or non-coupled systems, or with the transmission characteristics of single systems, we would like to have some quantitative method for describing the over-all structure or pattern of the set of units or states involved. In the psychological journal network, for example, it is intuitively apparent that there are certain *clusters* of journals which are functionally similar to one another, which are 'close' to each other in terms of reciprocating citations, similar balance and congruence values, etc., yet clearly distinguishable from other clusters. The *American Journal of Psychology*, the *Journal of Experimental Psychology*, and the *Journal of Comparative and Physiological Psychology* form such a cluster, as do a group of psychoanalytic and psychiatric journals. There are a number of methods for computing and representing structural relationships among variables in matrices, including the simple 'distance' measure of Osgood and Suci (7) as described in our analysis of communication via facial expressions and used elsewhere in semantic studies. However, we have made no attempt as yet to generalize such approaches to the types of communication problems.

SUMMARY

A possible lexicon for talking about a variety of human communication situations has been elaborated and a set of measurement operations has been associated with these terms, derived for the most part from information theory.

The generalized types of situations in which students of human communication are likely to be interested have been described with this lexicon and in relation to these measures. They include (1) the inherent characteristics of single systems, (2) the transmission characteristics of coupled systems, (3) the transmission characteristics of single systems, (4) the transmission characteristics of non-coupled but corresponding systems, and (5) the inherent characteristics of communication networks. A number of applications to concrete experimental problems have been made to illustrate the potential scope of this kind of analysis and to show that the information measures are more or less sensitive to the common aspects of a wide variety of communication situations. Of course, this generality is bought at the "price" of lack of sensitivity to many unique aspects of particular situations, and experimenters would, therefore, want to supplement these descriptions with other measures.

March, 1961

REFERENCES

1. Garner, W. G. and Hake, H. W. The amount of information in absolute judgments. Psychol. Rev., 1951, 58, 446-59.

2. Jakobson, R., Fant, C. G. and Halle, M. Preliminaries to Speech Analysis: The Distinctive Features and their Correlates. Cambridge, Mass.: Acoustic Laboratory, Massachusetts Institute of Technology, Technical Report #13, 1952.

3. Miller, G. A. What is information measurement? Amer. Psychologist, 1953, 8, 3-11.

4. Miller, G. A. and Frick, F. C. Statistical behavioristics and sequences of responses. Psychol. Rev., 1949, 56, 311-24.

5. Newman, E. B. The pattern of vowels and consonants in various languages. Amer. J. Psychol., 1951, 64, 369-79.

6. Osgood, C. E. and Sebeok, T. A. (ed.) Psycholinguistics: A Survey of Theory and Research Problems. Bloomington: Indiana Univ. Publications in Anthropology and Linguistics, 1954.

7. Osgood, C. E. and Suci, G. J. A measure of relation determined by both mean difference and profile information. Psychol. Bull., 1952, 49, 251-62.

8. Osgood, C. E. A behavioristic analysis of perception and meaning as cognitive phenomena. Colorado Symposium on Cognition. Harvard University Press, 1957.

9. Quastler, H. (ed.) Information Theory in Psychology. Glencoe, Ill.: Free Press, 1955.

10. Saporta, Sol. Frequency of consonant clusters. Language, 1955, 31, 25-30.