1987

A Spectral Analysis of Relations, Further Developments

Klaus Krippendorff
University of Pennsylvania, kkrippendorff@asc.upenn.edu

Philadelphia: The Annenberg School of Communications, University of Pennsylvania, 1978 (mimeo).

A Spectral Analysis of Relations, Further Developments

Comments
Philadelphia: The Annenberg School of Communications, University of Pennsylvania, 1978 (mimeo).

A spectral analysis of relations, further developments

KLAUS KRIPPENDORFF
The Annenberg School of Communications, University of Pennsylvania
Philadelphia, USA

1. INTRODUCTION

Part of my background is in the social sciences where the concept of "relation" is a central one. Almost anything social is said to be transacted and has relational qualities. Processes of communication are relational in the sense of connecting, however tenuously, two or more individuals. Authority, power and control can only be defined in relational terms. And the very notion of organization, social or otherwise, implies boundaries within which certain patterns of interaction are maintained and across which matter, energy and information are exchanged which in turn creates, maintains or destroys other patterns of interaction. While much of social science data contain evidence about relations, it is surprising that empirical techniques for analyzing complex relations are quite underdeveloped.

In a previous paper [9] I developed the idea of a spectral analysis of relations in part because I feel quite strongly that we might miss important insights if we restrict our vision through existing analytical techniques to very simple kinds of relations, correlations, associations, factor loadings, proximities, differences, causes, etc., all of which are essentially binary in nature. This restricted vision stands in marked contrast to ordinary conceptions of social life which recognize human communication as a multi-channel affair, social interaction to be rich and complex, and the environment of social organizations to be turbulent in the sense that feedback loops are multiple and intermeshed and rates of change are unequal and interdependent.

A spectral analysis of relations does not promise all the badly needed insights into what everyone agrees are deep and hidden complexities (see Simon [12], for example). For once, it is constrained by current computational limitations. But it does offer a modest improvement in analytical vision.

2. A SPECTRAL ANALYSIS OF RELATIONS

Generalizing from physics, one could say that a spectral analysis entails a calculus or accounting system for certain magnitudes such that the magnitude of a whole can be regarded as the algebraic sum of the magnitudes of its component parts. The decomposition of a source of light into additive spectra is a classical example but so is the Fourier analysis of oscillations, mechanical, musical, economic or social. I am applying the general idea to the study of relations.

The relations of interest in this paper are manifest in distributions of data in multivariate spaces. Illustrating this notion, we are all familiar with scatter diagrams of data points in a plane. One can easily extend this notion by visualizing a distribution of data points in a three or four dimensional space. But when more than four variables are involved other conceptual devices are needed. Nevertheless it should not be difficult to imagine that multi-variable distributions may assume very complex forms. Some distributions are inherently simple in the sense that either some variables are redundant or the whole distribution can be explained as a conjunction of distributions of lower ordinality. Distributions that represent linear relations are of this kind. They are usually depictable in a series of scatter diagrams each representing a binary relation. Figure 1 exemplifies such a case.

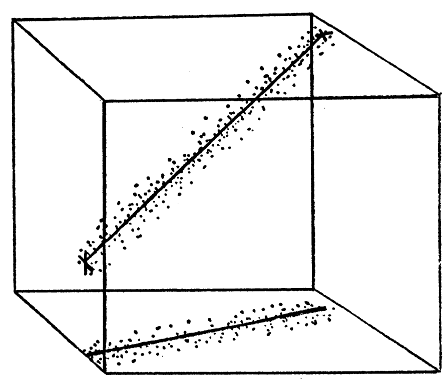


Figure 1. Example of a Linear Relation Typically Assumed in Social Science Methods

Some relations are inherently complex in the sense that all variables are needed, not in pairs, not in triples, all in conjunction, to account for that relation. Figure 2 depicts such a non-decomposable relation in three dimensions. Most non-linear relations are of this kind. Part of the aim of a spectral analysis of relations is to ascertain whether a multi-valued relation can be simplified without loss, where the non-decomposable complexities lie in the data, the ordinality of the explanation required, how much would be lost if one were to impose an unjustified simplification on an empirical fact, etc.

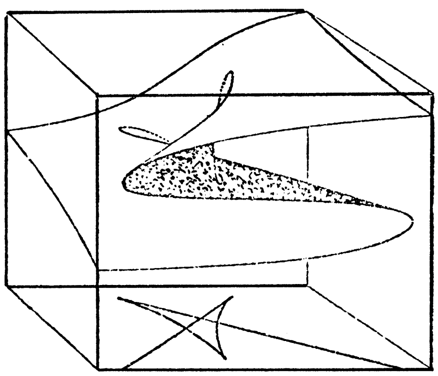


Figure 2. Example of a Non-Linear Relation of Concern to Catastrophy Theory Among Others

The qualitative components in the spectral analysis of relations consist of all possible relations of equal or lower ordinality than are inherent in a given data. Generally, for relational data within m variables there are:

1	zero-order relation
m	1st-order relations (properties)
$\frac{m(m-1)}{2}$	2nd-order relations (binary relations)
$\frac{m(m-1)(m-2)}{6}$	3rd-order relations (tertiary relations)
. . .	
${}_mC_r = \frac{m!}{(m-r)!r!}$	rth-order relations
. . .	
1	mth-order relation
2^m	total number of possible relations

This can be a large number, even with a moderate number m of variables. While the number of possible relations might set computational limits for a complete analysis of relational data, this fact can hardly serve as a justification for statistical techniques in the social sciences to assume as it were that social phenomena are essentially linear and, hence, decomposable into binary relations or at best constrained by a third variable.

The magnitudes in a spectral analysis of relations reflect deviations of observed frequencies from what would be expected by chance. Again, we are all too familiar with the two-variable case. Two variables, A and B, are regarded as independent when the joint probability distribution is fully explainable from its individual probability distributions, i.e., for all values $a \in A$ and $b \in B$" $p_{ab} = p_a p_b$ or $p_{ab}/p_a p_b = 1$ where $p_a = \sum_b p_{ab}$, etc. The extent of the deviation of p_{ab} from $p_a p_b$ is taken to be indicative of the strength of an association between the two variables. In the literature I have found no extension of this test to higher order dependencies. Bartlett [4] might be mentioned as an exception. However, his test concerns only 2 by 2 by 2 contingency tables. In the spectral analysis of relations the above comparison of observed and expected probabilities is generalized by an expansion of the probabilities in distributions of varying ordinality into a series of tests, each pertaining to a different subspace of the original distribution, and all are logically independent of each other (see (1) on next page). The probability p_{abcd} of a data point in a four dimensional space can be seen as the product of the tests for the presence of one zero-order relation (which is an artifact here and of no consequence in the following), four unary relations (properties or distributions in one variable), six binary relations (involving all pairs of variables), four tertiary relations, and one qua ternary relation.

The accounting equation for the spectral analysis of relations is obtained by first dividing both sides by the set of first order probabilities and then summing the average logarithm of these probabilistic expressions for all data points. While there are probably several tests that could be employed to establish the presence of relations of higher ordinality, the logarithmic function is justified by Shannon's [11] proof that it is the only function leading to additive quantities. Thus, the fundamental accounting equation for the spectral analysis of relations is:

$$T(AB) = Q(AB)$$

$$T(ABC) = Q(ABC) + Q(AB) + Q(AC) + Q(BC)$$

$$T(ABCD) = Q(ABCD) + Q(ABC) + Q(ABD) + Q(ACD) + Q(BCD) + Q(AB) + Q(AC) + Q(AD) + Q(BC) + Q(BD) + Q(CD)$$

$$T(m\text{ variables}) = \sum_{r=2}^{m} \sum_{i=1}^{{}_mC_r} Q(r\text{th-order})_i \tag{2}$$

In it, the T-measures assess the total amount of relation (relatedness, constraint, information transmission, multivariate association, etc.) in the data as a whole:

$$T(AB) = \sum_a \sum_b p_{ab} \log_2 \frac{p_{ab}}{p_a p_b}$$

(continued on p. 72)

$$
\begin{aligned}
1 &= 1 \\
p_a &= 1\, p_a \\
p_{ab} &= 1\, p_a\, p_b\, \frac{p_{ab}}{p_a p_b} \\
p_{abc} &= 1\, p_a\, p_b\, p_c\, \frac{p_{ab}}{p_a p_b}\, \frac{p_{ac}}{p_a p_c}\, \frac{p_{bc}}{p_b p_c}\, \frac{p_{abc}}{\dfrac{p_{ab} p_{ac} p_{bc}}{p_a p_b p_c}} \\
p_{abcd} &= 1\, p_a\, p_b\, p_c\, p_d\, \frac{p_{ab}}{p_a p_b}\, \frac{p_{ac}}{p_a p_c}\, \frac{p_{ad}}{p_a p_d}\, \frac{p_{bc}}{p_b p_c}\, \frac{p_{bd}}{p_b p_d}\, \frac{p_{cd}}{p_c p_d}\, \frac{p_{abc}}{\dfrac{p_{ab} p_{ac} p_{bc}}{p_a p_b p_c}}\, \frac{p_{abd}}{\dfrac{p_{ab} p_{ad} p_{bd}}{p_a p_b p_d}}\, \frac{p_{acd}}{\dfrac{p_{ac} p_{ad} p_{cd}}{p_a p_c p_d}}\, \frac{p_{bcd}}{\dfrac{p_{bc} p_{bd} p_{cd}}{p_b p_c p_d}}\, \frac{p_{abcd}}{\dfrac{p_{abc} p_{abd} p_{acd} p_{bcd}}{\dfrac{p_{ab} p_{ac} p_{ad} p_{bc} p_{bd} p_{cd}}{p_a p_b p_c p_d}}}
\end{aligned}
\tag{1}
$$

$$
p_{r\text{th-order}} = \prod_{i=0}^{r} \prod_{\substack{\text{all } i\text{th-} \\ \text{order in } r}} \frac{p_{i\text{th-order}}}{\dfrac{\prod_{\substack{\text{all } (i-1)\text{th-} \\ \text{order in } i}} p_{(i-1)\text{th-order}}}{\dfrac{\prod_{\substack{\text{all } (i-2)\text{th-} \\ \text{order in } i}} p_{(i-2)\text{th-order}}}{\dfrac{\vdots}{1}}}}
$$

$$T(ABC) = \sum_a \sum_b \sum_c p_{abc} \log_2 \frac{p_{abc}}{p_a p_b p_c}$$

$$T(ABCD) = \sum_a \sum_b \sum_c \sum_d p_{abcd} \log_2 \frac{p_{abcd}}{p_a p_b p_c p_d}$$

etc.

And the Q-measures assess the unique contribution each relation makes to the total:

$$Q(AB) = \sum_a \sum_b p_{ab} \log_2 \frac{p_{ab}}{p_a p_b}$$

$$Q(ABC) = \sum_a \sum_b \sum_c p_{abc} \log_2 \frac{\dfrac{p_{abc}}{p_{ab} p_{ac} p_{bc}}}{p_a p_b p_c}$$

$$Q(ABCD) = \sum_a \sum_b \sum_c \sum_d p_{abcd} \log_2 \frac{\dfrac{p_{abcd}}{p_{abc} p_{abd} p_{acd} p_{bcd}}}{\dfrac{p_{ab} p_{ac} p_{ad} p_{bc} p_{bd} p_{cd}}{p_a p_b p_c p_d}}$$

etc.

Mainly because it will not cause confusion in this paper, my notation condenses the conventional forms established by Ashby [1,2]. Accordingly, T(ABC) = TA:B:C), Q(ABC) = Q(A:B:C), and in what follows H(ABC) = H(A,B,C), T($\overline{AB}$C) = T(A,B:C), etc.

Some of the properties of these measures have been discussed and exemplified previously [9]. Here I will state them very briefly.

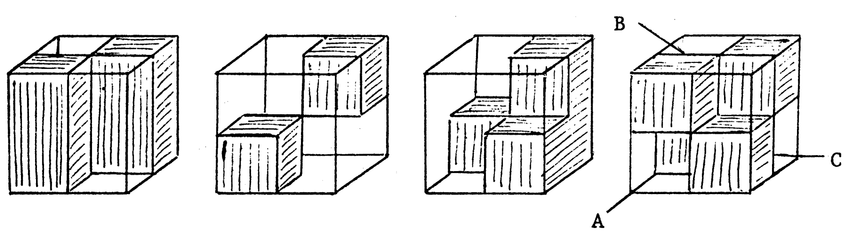
First, the Q-measures are indicative of the magnitude of a relation. If such a measure turns out to be zero or approximately zero, then the relation so assessed contributes nothing to the data and may be ignored. If it equals the total then it is the only relation accounting for the data.

Second, all Q-measures are independent. The finding that all binary relations are absent does not imply anything about the presence or absence of a higher order relation in data and, vice versa, the presence of a higher order relation suggests nothing about the magnitude of any of its relations of lower ordinality.

Third, Q-measures assume negative values when its immediately lower-order relations overdetermine the distribution (include redundant accounts of the total) and they assume positive values when its immediately lower-order relations underdetermine the relation (are insufficient as aggregate account of the total). In other words, each Q-measure of a relation compensates for the errors of commission or the errors of omission committed by the conjunction of its component relations.

For examples, consider the four relations in a three dimensional space (see Figure 3). The values of all expressions of the accounting equation are computed under the assumption that the probability p_{abc} of the shaded cells is 1/n and the others are zero.

The leftmost distribution shows variable B to be redundant. The whole distribution can be explained in terms of A and C without loss of generality. This is indicated in the corresponding Q-measure. The accounting equation here reduces to T(ABC) = Q(AC), suggesting that explanations based on the pairs of variables A and B or B and C yield nothing. The next distribution is seen as fully explainable by the binary relations within pairs of variables. Actually only two such binary relations are required, the third is implied and redundant. Taken together, the three binary relations therefore overdetermine the whole. It is the amount of overdetermination which now appears as the negative value in the test for the presence of the tertiary relation. The third relation from the left shows the binary account for the distribution to be important but not sufficient to provide a complete account of that distribution. An attempt to describe the whole in terms of the three binary component relations will miss about one-seventh of the total amount of constraint in the data. The rightmost distribution is the more complex of the four. All projections of the distribution on the three



Q(AB)= 0	Q(AB)= 1	Q(AB)=.122	Q(AB)= 0
Q(A C)= 1	Q(A C)= 1	Q(A C)=.122	Q(A C)= 0
Q(BC)= 0	Q(BC)= 1	Q(BC)=.122	Q(BC)= 0
Q(ABC)= 0	Q(ABC)=-1	Q(ABC)=.067	Q(ABC)= 1
T(ABC)= 1	T(ABC)=22	T(ABC)=.433	T(ABC)=11

Figure 3. Four Examples of Relations

two-dimensional planes yield uniform distributions. The Q-measure for all binary relations are all zero. A unique tertiary relation accounts for the total constraint in the data: T(ABC) = Q(ABC). This last example demonstrates the point made earlier that the presence or absence of a relation of one ordinality is independent of the presence or absence of a relation of another ordinality. The Q-measures are not ordered, do not imply each other.

The idea of representing a distribution of data in a many-variable space in terms of simple relations is based on Ashby's "Constraint Analysis of Many-Valued Relations" [1] which suggests a qualitative method for analyzing complex relations into simple ones. Ashby starts by identifying a relation with a proper subject of a product set. When an empirically obtained relation is projected onto several subspaces and reflected back into the original space, the conjunction of these reflections form a new relation that contains the empirically obtained one as a proper subset. The set-theoretical difference between the two subsets can be taken as the loss incurred by the projection. Constraint analysis simplifies a relation by identifying those subspaces for which the loss is minimal or absent. A spectral analysis of relations realizes many of Ashby's intentions in a probabilistic context. Transmission measures, T, clearly are measures of constraint:

$$T(ABC\ldots) = H(A) + H(B) + H(C) + \ldots - H(ABC\ldots)$$

H(A) + H(B) + H(C) + . . . is the maximum entropy in a multi-variable space that would be observed if all variables are independent and H(ABC...) is the entropy actually obtained. In terms of the four examples given in Figure 3, the second from the left exhibits the largest constraint and the third from the left the smallest. It is the use of an accounting equation for separating the magnitudes associated with all logically distinct and unique contributions of each subset of variables that makes the spectral analysis different from Ashby's approach.

The quantitative part of this spectral analysis also owes much to the groundwork laid by Ashby in various extensions of information theory [2,3] which were influenced by McGill's work on multi-variate information transmission [10].

3. BASIC BUILDING BLOCKS OF ORDER

One of the interesting consequences of the accounting equation is that it points to the Q-measures as possible candidates of what one might call basic building blocks of order. While these measures are far from simple and quite removed from the entropies of a distribution, they are at least as appealing as entropy measures are because all information theoretical measures can be expressed as the <u>algebraic sum</u> of several Q-measures, whereas entropies require additions <u>and</u> subtractions. So, if one is willing and capable of computing all Q-measures for an empirically obtained distribution of data points, one can gain considerable insights into its relational properties which ordinary H-measures would hide. For example, in the five dimensional case of Figure 4, all possible Q-measures are seen as forming a lattice. In it one can identify all entropies, H, all information transmission terms, T, and all interactions, Q. The figure shows three examples. The Q-measures adding up to H(BDE) are connected by solid lines, those adding up to $T_A(BCD)$ by broken lines, and those adding up to $Q_B(ACDE)$ by a chain.

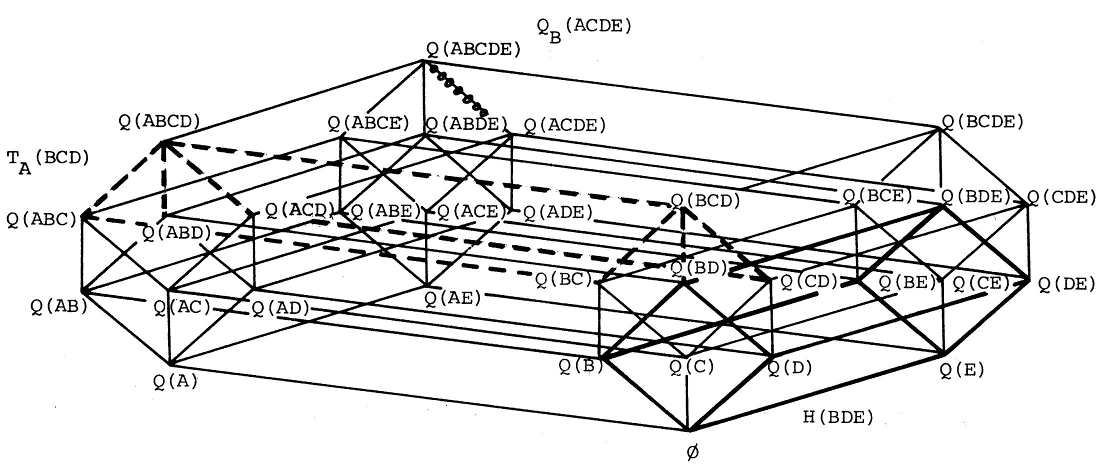


Figure 4. Lattice of Q-measures for Five Dimensions

The basic Q-measures are computable from entropies H:

$$Q(A) = -H(A) = \sum_a p_a \log_2 p_a$$

$$Q(AB) = H(A) + H(B) - H(AB) = H(A) + H(B) + \sum_a \sum_b p_{ab} \log_2 p_{ab}$$

$$Q(ABC) = -H(A) - H(B) - H(C) + H(AB) + H(AC) + H(BC) - H(ABC)$$

etc.

$$Q(m \text{ variables}) = \sum_{j=1}^{m} \sum_{k=1}^{{}_mC_j} H(j \text{ variables})_k \, \Delta_{m-j} \quad \text{where } \Delta_x = \begin{cases} +1 & \text{for uneven } x \\ -1 & \text{for even } x \end{cases} \tag{3}$$

whereupon all entropies, information transmissions and interactions are expressable as the algebraic sum of these basic Q-measures:

Transmissions

$$T(m \text{ variables}) = \sum_{j=2}^{m} \sum_{k=1}^{{}_mC_j} Q(j \text{ variables})_k \tag{4}$$

e.g.:

$$\begin{aligned} T(ABCD) = \; & Q(AB\;\;) \\ & +Q(A\;C\;) \\ & +Q(A\;\;D) \\ & +Q(\;BC\;) \\ & +Q(\;B\;D) \\ & +Q(\;\;CD) \\ & +Q(ABC\;) \\ & +Q(AB\;D) \\ & +Q(A\;CD) \\ & +Q(\;BCD) \\ & +Q(ABCD) \end{aligned}$$

Entropies

$$H(m \text{ variables}) = \sum_{j=1}^{m} \sum_{k=1}^{{}_mC_j} Q(j \text{ variables})_k \tag{5}$$

e.g.:

$$\begin{aligned} H(BDE) = \; & Q(B\;\;) \\ & +Q(\;D\;) \\ & +Q(\;\;E) \\ & +Q(BD\;) \\ & +Q(B\;E) \\ & +Q(\;DE) \\ & +Q(BDE) \end{aligned}$$

Conditional Interactions

$$Q_{r \text{ variables}}(s \text{ variables}) = \sum_{j=0}^{r} \sum_{k=1}^{{}_rC_j} Q(j+s \text{ variables})_k \tag{6}$$

e.g.:

$$Q_{AB}(CDE) = Q(\;\;CDE) + Q(A\;CDE) + Q(\;BCDE) + Q(ABCDE)$$

Conditional Transmissions

$$T_{r\ \text{variables}}(s\ \text{variables}) = \sum_{j=2}^{s} \sum_{k=1}^{{}_sC_j} Q_{r\ \text{variables}}(j\ \text{variables})_k \tag{7}$$

e.g.:

$$\begin{aligned}
T_{AB}(CDE) = Q_{AB}(CD\ \ ) &= Q(\ \ CD\ \ ) + Q(A\ CD\ \ ) + Q(\ BCD\ \ ) + Q(ABCD\ \ ) \\
+Q_{AB}(C\ E) &\quad +Q(\ \ C\ E) + Q(A\ C\ E) + Q(\ BC\ E) + Q(ABC\ E) \\
+Q_{AB}(\ DE) &\quad +Q(\ \ \ DE) + Q(A\ \ DE) + Q(\ B\ DE) + Q(AB\ DE) \\
+Q_{AB}(CDE) &\quad +Q(\ \ CDE) + Q(A\ CDE) + Q(\ BCDE) + Q(ABCDE)
\end{aligned}$$

Conditional Entropies

$$H_{r\ \text{variables}}(s\ \text{variables}) = \sum_{j=1}^{s} \sum_{k=1}^{{}_sC_j} Q_{r\ \text{variables}}(j\ \text{variables})_k \tag{8}$$

e.g.

$$\begin{aligned}
H_{AB}(CDE) = Q_{AB}(C\ \ ) &= Q(\ \ C\ \ ) + Q(A\ C\ \ ) + Q(\ BC\ \ ) + Q(ABC\ \ ) \\
+Q_{AB}(\ D\ ) &\quad +Q(\ \ \ D\ ) + Q(A\ \ D\ ) + Q(\ B\ D\ ) + Q(AB\ D\ ) \\
+Q_{AB}(\ \ E) &\quad +Q(\ \ \ \ E) + Q(A\ \ \ E) + Q(\ B\ \ E) + Q(AB\ \ E) \\
+Q_{AB}(CD\ \ ) &\quad +Q(\ \ CD\ \ ) + Q(A\ CD\ \ ) + Q(\ BCD\ \ ) + Q(ABCD\ \ ) \\
+Q_{AB}(C\ E) &\quad +Q(\ \ C\ E) + Q(A\ C\ E) + Q(\ BC\ E) + Q(ABC\ E) \\
+Q_{AB}(\ DE) &\quad +Q(\ \ \ DE) + Q(A\ \ DE) + Q(\ B\ DE) + Q(AB\ DE) \\
+Q_{AB}(CDE) &\quad +Q(\ \ CDE) + Q(A\ CDE) + Q(\ BCDE) + Q(ABCDE)
\end{aligned}$$

Grouping of Variables in Interactions

$$Q(\overline{r\ \text{variables}} + s\ \text{variables}) = \sum_{j=1}^{r} \sum_{k=1}^{{}_rC_j} Q(j{+}s\ \text{variables})_k \tag{9}$$

e.g.

$$\begin{aligned}
Q(\overline{ABC}DEF) = \ & Q(A\ \ \ DEF) \\
& +Q(\ B\ DEF) \\
& +Q(\ \ CDEF) \\
& +Q(AB\ DEF) \\
& +Q(A\ CDEF) \\
& +Q(\ BCDEF) \\
& +Q(ABCDEF)
\end{aligned}$$

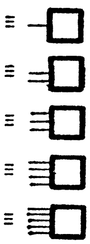
It should be noted that the number of possible Q-measures equals the number of possible H-measures, and since all information theoretical measures can be expressed in either of the two kinds of terms, there seems to be no immediately apparent advantage. In fact, one might favour the entropies on account of their simplicity.

However, I am suggesting that Q- and H-measures provide complementary accounts of empirical facts. Q-measures approach an empirical phenomenon from the point of view of what can be explained within the set of observational variables. They locate the relations of different ordinality that account for given data. H-measures approach the same phenomena from the point of view of what can not be explained within these variables. They locate the uncertainties within different subspaces of a multi-variable distribution. The spectral analysis elucidates structure, pattern, dependency and constraint, whereas an account in terms of entropies emphasizes the freedom of variation, uncertainty, lack of structure and error. Both are essential for gaining any understanding of systems.

4. BASIC BUILDING BLOCKS OF FUNCTIONAL COMPOSITION

Another interpretation of the results of a spectral analysis of relations is that it quantifies the extent to which functional components of varying complexity underly a process that is responsible for the given data. $Q(AB) > 0$ indicates that there exists a binary relationship between the variables A and B that cannot be explained in terms of the properties associated with A and/or B. $Q(ABC) > 0$ indicates that there exists a tertiary relationship among the three variables A, B and C over and above what the three binary relationships between A and B, A and C, B and C, and the three properties in

A, B and C can account. Q(ABCD) > 0 indicates the presence of a unique and non-decomposable quarternary dependency between A, B, C and D, etc. All variables among which a Q-measure exhibits a positive value could be said to stem from a single source that maintains the pattern which the Q-measure quantitatively assesses. Since the Q-measure reflects the unique co-occurence of values on all variables thereby assessed, it identifies the coordinative effort by a source, an effort that cannot be understood by fewer variables. To make this correspondence between Q-measures and functional components apparent, I am using Klir's [8] graphical symbols:



Q(A) assesses a property ≡ a single-valued component

Q(AB) assesses a binary relation ≡ a two-valued component

Q(ABC) assesses a tertiary relation ≡ a three-valued component

Q(ABCD) assesses a quarternary relation ≡ a four-valued component

Q(ABCDE) assesses a quintenary relation ≡ a five-valued component

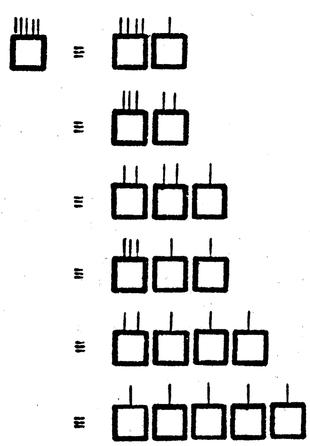
etc.

By evaluating all components that could possibly generate data of a given dimensionality and by associating a magnitude with each, the accounting equation in effect outlines the dependency structure (the wiring diagram) of the components of a complex data source.

An accounting equation for m-valued data has 2^m-m-1 distinct Q-measures. Since any one Q-measure may be positive and since each configuration of Q-measures describes a different dependency structure, there are in fact 2^{2^m-m-1} different dependency structures to be considered. This can be a tremendously large number. Happily, not all Q-measures need to be evaluated one-by-one. Some can be evaluated en bloc, which provides the basis of a more efficient algorithm.

The first such shortcut is suggested by Ashby's [2] theorem stating that if a transmission term between two sets of variables is zero, so are all Q-measures that make reference to variables in both sets. Thus, if a set of variables can be broken down into two, having, say, r and s variables respectively, then T(r variables + s variables) > 0 implies $(2^r - 1)(2^s - 1)$ of the 2^{r+s} Q-measures are known to be zero, leaving only $2^r + 2^s$ Q-measures to be evaluated. This means a considerable saving of computational efforts. To find independent subsystems in the set of variables should be the first and in a sense preliminary step of any spectral analysis of relations. A system of, say, five variables may be decomposed in the way shown below.

In what follows I am therefore ignoring all zero dependencies across independent subsystems and focusing instead on the interdependencies within a system of variables that cannot be partitioned as shown below. I suppose this is where the real power of a spectral analysis of relations lies.



≡ iff T(5 variables) = T(4 variables)

≡ iff T(5 variables) = T(3 variables) +T(2 variables)

≡ iff T(5 variables) = T(2 variables)$_1$ +T(2 variables)$_2$

≡ iff T(5 variables) = T(3 variables)

≡ iff T(5 variables) = T(2 variables)

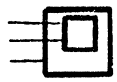
≡ iff T(5 variables) = 0

The second shortcut is suggested by the way functional components are conceptualized. Suppose the lattice of Q-measures reveals that one relation is embedded in another. For example:

Q(ABC)>0

Q(AB)>0

or



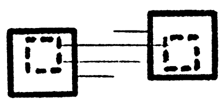
This finding would suggest that a tertiary relation between variables A, B and C contains a binary relation between variables A and B. To realize (program or build) a functional component that would represent (generate or simulate) the given data would require a function of no less than three values or a box with no less than three variables attached. All constraints among these three variables will have to be programmed into that box, including any binary dependency that might be manifest in the data. Only if these binary relations are shared through communication with other components might they have to be considered separately. The example:

Q(ABC)>0

Q(AB)>0

Q(AB D)>0

or

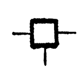
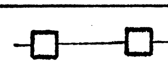
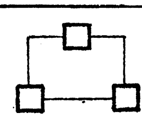


leaves open where the binary relation is to be realized, in the ABC-component or in the ABD-component. The dependency structure is not affected in either case.

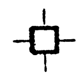
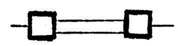
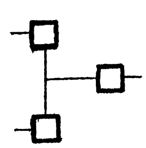
Summarizing, one can say that a relation is embedded in another when the set of variables in which that relation is manifest is a proper subset of the set of variables in which the other relation is manifest. If it is the task to identify dependency structures among functional components, then a spectral analysis of relations can stop with the location of postive Q-measures for relations none of which is embedded in another. If it is the task to quantify the information each functional component requires to generate given data, then the quantities associated with embedded relations need to be examined. Clearly, the task of identifying the dependency structure is prior to that of quantifying what is involved inside and across each component. The illustration to follow pertains to the structural identification task only.

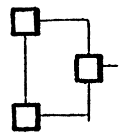
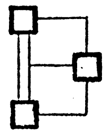
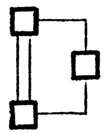
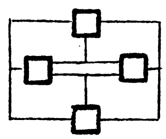
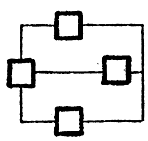
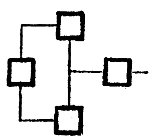
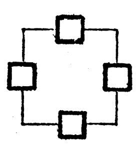
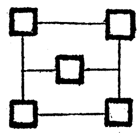
A final and in a sense still preliminary point is that structural ambiguities may arise from the fact that Q-measures assess the extent of a unique relation among variables. To determine structure types requires suitable decision criteria. In the above example, if Q(ABC) were positive but very small compared with Q(AB) then one has the choice of representing the data either by the most inclusive positive relation, here involving A, B and C, or by tolerating the error Q(ABC) and representing the data in terms of the variables that dominate the total constraint, here involving A and B only. The choice of a suitable decision criterion is not a simple matter. But I am assuming here that such a criterion does exist. It disambiguates the emergence of structure types.

With these preliminaries, I will now exemplify the dependency structures a spectral analysis of relations will reveal. The two-variable case is actually of no particular interest. Assuming a suitable decision criterion given, a binary relation is either present in the data or it is not. A spectral analysis of a three-variable system, on the other hand, could result in any one of three dependency structure types which are listed together with the Q-measures of relations that are ignored for lack of statistical significance or overdetermination, and Q-measure configuration defining the structure. I am ignoring the permutations of variables throughout.

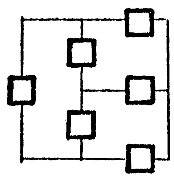
Statistically insignificant or negative Q-measures	Defining configuration, non-embedded	Q-measure, embedded	Dependency structure	Required components
	Q(ABC)	Q(AB) Q(A C) Q(BC)		1 tertiary
Q(ABC) Q_B(AC)	Q(AB) Q(BC)			2 binary
Q(ABC)	Q(AB) Q(A C) Q(BC)			3 binary

A spectral analysis of a four-variable system could result in any one of the following fourteen dependency structure types:

Statistically insignificant or negative Q-measures	Defining configuration, non-embedded	Q-measure, embedded	Dependency structure	Required components
	Q(ABCD)	Q(ABC) Q(AB D) Q(A CD) Q(BCD) Q(AB) Q(A C) Q(A D) Q(BC) Q(B D) Q(CD)		1 quarternary
Q(ABCD) Q_D(ABC) Q_C(ABD) Q_{CD}(AB)	Q(A CD) Q(BCD)	Q(A C) Q(A D) Q(BC) Q(B D) Q(CD)		2 tertiary
Q(ABCD) Q_D(ABC) Q_C(ABD) Q_B(ACD) Q_{CD}(AB) Q_{BC}(AD)	Q(BCD) Q(A C)	Q(BC) Q(B D) Q(CD)		1 tertiary 2 binary
Q(ABCD) Q_D(ABC) Q_C(ABD) Q_B(ACD) Q_A(BCD) Q_{CD}(AB) Q_{BC}(AD) Q_{AB}(CD)	Q(A C) Q(B D) Q(BC)			3 binary
Q(ABCD) Q_D(ABC) Q_C(ABD) Q_B(ACD) Q_A(BCD) Q_{CD}(AB) Q_{BC}(AD) Q_{AC}(BD)	Q(A C) Q(BC) Q(CD)			3 binary

Q(ABCD) Q_D(ABC) Q_C(ABD) Q_B(ACD) Q_{CD}(AB)	Q(BCD) Q(A C) Q(A D)	Q(BC) Q(B D) Q(CD)		1 tertiary 2 binary
Q(ABCD) Q_D(ABC)	Q(AB D) Q(A CD) Q(BCD)	Q(AB) Q(A C) Q(A D) Q(BC) Q(B D) Q(CD)		3 tertiary
Q(ABCD) Q_D(ABC) Q_C(ABD)	Q(A CD) Q(BCD) Q(AB)	Q(A C) Q(A D) Q(BC) Q(B D) Q(CD)		2 tertiary 1 binary
Q(ABCD)	Q(ABC) Q(AB D) Q(A CD) Q(BCD)	Q(AB) Q(A C) Q(A D) Q(BC) Q(B D) Q(CD)		4 tertiary
Q(ABCD) Q_D(ABC) Q_C(ABD) Q_B(ACD)	Q(BCD) Q(AB) Q(A C) Q(A D)	Q(BC) Q(B D) Q(CD)		1 tertiary 3 binary
Q(ABCD) Q_D(ABC) Q_C(ABD) Q_B(ACD) Q_A(BCD) Q_{CD}(AB) Q_{BC}(AD)	Q(A C) Q(BC) Q(B D) Q(CD)			4 binary
Q(ABCD) Q_D(ABC) Q_C(ABD) Q_B(ACD) Q_A(BCD) Q_{CD}(AB) Q_{AB}(CD)	Q(A C) Q(A D) Q(BC) Q(B D)			4 binary
Q(ABCD) Q_D(ABC) Q_C(ABD) Q_B(ACD) Q_A(BCD) Q_{CD}(AB)	Q(A C) Q(A D) Q(BC) Q(B D) Q(CD)			5 binary

Q(ABCD)	Q(AB)
Q_D(ABC)	Q(A C)
Q_C(ABD)	Q(A D)
Q_B(ACD)	Q(BC)
Q_A(BCD)	Q(B D)
	Q(CD)



6 binary

It should be noted that the algebraic sum of the defining Q-measures (of embedded and non-embedded relations incorporated in the dependency structure) equals the amount of transmission accounted for by the components of the structure. Thus, in the first dependency structure in this list (the undifferentiated case) this sum is T(ABCD) and in the last it is the sum of the six binary transmissions between the four variables.

The aim of the above exercise was not to enumerate dependency structure types. Although these are far fewer in number than the configurations of possible Q-measures, they grow exponentially with the number of variables involved. A spectral analysis does not test for distinct structure types. They simply emerge from or are implied by the configuration of statistically significant and non-embedded relations as represented by their Q-measures. The aim of the above was merely to show the correspondence between configurations of Q-measures and dependency structures, and to thereby demonstrate what a spectral analysis (of interconnected systems) of relations may reveal.

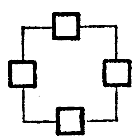
One way to organize dependency structure types is by the number of components required to represent each structure. The above list is an example of this. Another and far more promising way is to follow the path of an algorithm for structure identification from the most complex component of the highest possible ordinality to the smallest set of least complex (binary) components all of which could conceivably represent the interconnections among the variables of a non-decomposable system. The algorithm iteratively evaluates the consequences of removing non-embedded relations (in decreasing order of ordinality) and brings thereby into focus those previously embedded relations that are now considered candidates of a more efficient representation of dependencies.

This algorithm is applied here to the four-variable system of the above example (see Figure 6). When the removal of a relation is indicated by the transition this invokes from one dependency structure to the next, the following lattice emerges. It assigns each structure type a unique place. To enhance readability, the removal of the quarternary relation is indicated by a horizontal line, of a tertiary relation by a vertical line, and of a binary relation by a 45^o line. The transitions are signed by the Q-measures representing the removed relation.

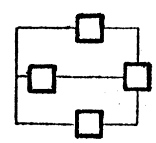
The first step of this algorithm involves testing for the statistical significance of the one quarternary relation encompassing the whole system. Its second and third steps involve deciding whether and which tertiary relations can be ignored as well. It might be noted that the removal of the second tertiary relation renders one of its embedded binary relations a component of the resulting structure. The fourth step offers a choice between removing the one binary relation or one of the two tertiary ones; etc. The process continues until a structure type emerges that is maximally simple and represents the total with a minimum of loss.

The functional component interpretation of the results of the spectral analysis of relations invites comparisons with Klir's work [7,8]. Its aim is very similar if not identical, but it deviates from the approach taken by the spectral analysis in at least two ways. First, Klir uses the sum of the absolute differences between expected and observed probabilities as measures of the degree to which a relation is approximated by the conjunction of its components. While this approach is well grounded in the tradition of statistical testing (e.g. X^2 methodology), it is biased towards a binary notion of constraint. In terms of the spectral analysis, the binary notion of constraint is implicit in the T-measures of information theory. But no longer in third or higher-order Q-measures. The fact that Klir's approach cannot lead to a calculus of additive quantities for many-valued relations need not be a disadvantage.

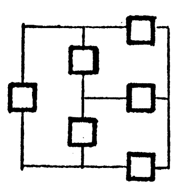
Second, Klir [8] adopts two rather stringent "axioms of structure candidates" that limit the structure types his procedure is able to differentiate. As a consequence, circular dependencies, e.g.:



one loop



three loops



seven loops

cannot be identified if they exist in the data. In the above example of an interconnected system of four variables, nine out of the fourteen structure types include circular or indirect mutual dependencies. Klir's approach would identify only the first five on the list of dependency structure types (not to be confused with the steps of the algorithm). The others are forced into these five types. The major advantage and presumably the reason for adopting this somewhat more restricted concept of

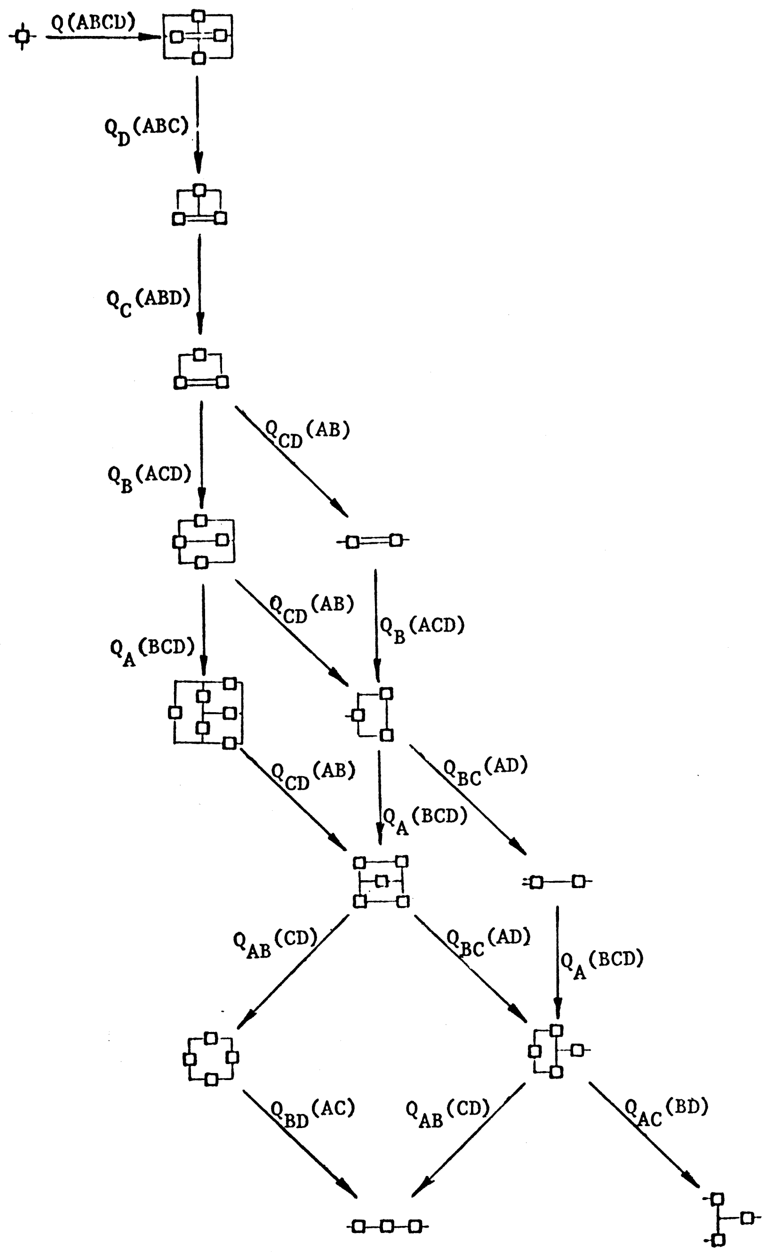


Figure 6. Lattice Resulting from Structural Identification Algorithm

structure is computational efficiency which is still a problem with spectral analysis. At this point, I cannot say how important it is and what it implies that the spectral analysis of relations identifies so many more alternative structures.

I should also acknowledge here recent work by Broekstra, who was so kind to supply me with copies of his work [5,6] after the presentation of this paper in Linz. Like myself, he approached the problem of structure identification from information theory. He too encountered the power of the Q-measures. But, unlike myself, he adopts a notion of structure that includes only binary relations and can be depicted as graphs between variables, although his second publication expands this notion to give a quantitative account of Klir's structure types. I believe the key to the difference between the spectral analysis approach and his lies in divergent generalizations of "statistical independence." The tests for the absence of a relation adopted by spectral analysis are spelled out in (1). His generalization of statistical independence to more than two variables is, in effect:

$$p_{ab} = p_a p_b$$

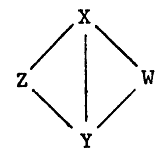
$$p_{abc} = \frac{p_{ac}p_{bc}}{p_c}$$

$$p_{abcd} = \frac{p_{acd}p_{bcd}}{p_{cd}}$$

etc.

In transmission terms, this condition is respectively: $T(AB) = 0$, $T_C(AB) = 0$, $T_{CD}(AB) = 0$, etc. He correctly argues that a direct dependency between two variables (represented by a graph) is born out by data only if the conditional (on all other variables) transmission between the two variables is positive. His approach does allow for circular dependencies to be detected. But this representation of structure is limited to binary relations only.

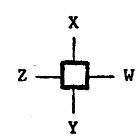
One of his examples may aid the comparison. From data he presented in [5, p. 76] as Case VI, he finds that the five binary transmission measures $T_{WX}(YZ)$, $T_{WY}(XZ)$, $T_{WZ}(XY)$, $T_{XZ}(WY)$, $T_{YZ}(XW)$ are positive while $T_{XY}(WZ)$ is zero, and concludes that of the six possible direct connections between the four variables only the W-Z connection is absent. In terms of his graph of binary dependencies, the result is:



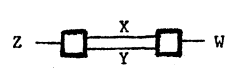
From Broekstra's data:

$Q(WXYZ) = -.0014$
$Q_Z(WXY) = .0192$
$Q_Y(WXZ) = -.0011$
$Q_X(WYZ) = -.0045$
$Q_W(XYZ) = .0192$
$Q_{YZ}(WX) = .0248$
$Q_{XZ}(WY) = .0510$
$Q_{XY}(WZ) = .0000$
$Q_{WZ}(XY) = .1065$
$Q_{WY}(XZ) = .0451$
$Q_{WX}(YZ) = .0510$

Following the algorithm for structure identification, the spectral analysis of relations would start with the following:

Statistically	Defining configuration, non-embedded	Q-measure, embedded	Dependency
	Q(WXYZ)=-.0014	Q(WXY)= .0206	
		Q(WX Z)= .0003	
		Q(W YZ)=-.0031	
		Q(XYZ)= .0206	
		Q(WX)= .0053	
		Q(W Y)= .0349	
		Q(W Z)= .0042	
		Q(XY)= .0667	
		Q(X Z)= .0053	
		Q(YZ)= .0349	
	total:	T(WXYZ)= .1883	

Removing the three relations that are overdetermined as indicated by negative (conditional) Q-measures and the one that is zero would zield the following dependency structure and quantitative account:

$Q(WXYZ) = -.0014$	Q(WXY)= .0206	Q(WX)= .0053	
$Q_Y(WXZ) = -.0011$	Q(XYZ)= .0206	Q(W Y)= .0349	
$Q_X(WYZ) = -.0045$		Q(XY)= .0667	
$Q_{XY}(WZ) = .0000$		Q(X Z)= .0053	
		Q(YZ)= .0349	
	equals:	T(WXY)= .1275	
		-T(XY)=-.0667	
		T(XYZ)= .1275	
	total:	.1883	

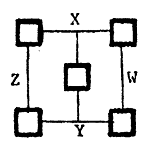
In this special case, the two tertiary components fully account for the total amount of transmission in the data. This need not, of course, always be achievable. There may be losses and redundancies. The example further illustrates the role of the embedded relations in the representation of data by functional components. The relation X-Y may be realized either in the WXY-component or in the XYZ-component. Since T(XY) is included in both transmission measures, it is redundant in the sum T(WXY) + T(XYZ) and must therefore be subtracted from the account.

If one were to further simplify the representation of the data, for example to the point of Broekstra's method, one would have to accept losses as evident in the account below. Evidentally, the tertiary relation that cannot be depicted in the form of a graph with variables at its nodes are removed in this result as a consequence of which about 22% of the accountable transmission in the system is lost.

In conclusion, let me say that my spectral analysis of relations is far from a state of completion. What I have learned is that while it may be difficult to conceptualize higher-order relations, when they are manifest in data, they cannot be analysed into pieces. Perhaps much of systems research into reality hangs on the ability to cope with relations of different (and usually higher) ordinalities. The spectral analysis of relations may be regarded as a stepping stone in this direction.

Removed	Non-embeded	Embedded
Q(WXYZ)=-.0014	Q(WX)= .0053	-
Q_Z(WXY)= .0192	Q(W Y)= .0349	
	Q(XY)= .0667	
Q_Y(WXZ)=-.0011	Q(X Z)= .0053	
Q_X(WYZ)=-.0045	Q(YZ)= .0349	
Q_W(XYZ)= .0192	T(WX)= .0053	-
	T(W Y)= .0349	
Q_{XY}(WZ)= .0000	T(XY)= .0667	
	T(X Z)= .0053	
	T(YZ)= .0349	
	.1471	

Dependency structure



REFERENCES

1. ASHBY, W. R., "Constraint Analysis of Many-Valued Relations," General Systems, 9, 99-105, 1964.
2. ASHBY, W. R., "Measuring the Internal Information Exchange in a System," Cybernetica, 8, 1, 5-22, 1965.
3. ASHBY, W. R., "Two Tables of Identities of Governing Information Flows Within Large Systems," American Society for Cybernetics Communications, 1, 2, 3-8, 1969.
4. BARTLETT, M. S., "Contingency Table Interactions," supplement to the Journal of the Royal Statistical Society, 2, 248-252, 1935.
5. BROEKSTRA, G., "Constraint Analysis and Structure Identification," Annals of Systems Research, 5, 67-80, 1976.
6. BROEKSTRA, G., "Constraing Analysis and Structure Identification II," Annals of Systems Research, 6, 1-20, 1977.
7. KLIR, G. J., "Identification of Generative Structures in Empirical Data," International Journal of General Systems, 3, 89-104, 1976.
8. KLIR, G. J., "Structural Modelling of Indigenous Systems," paper presented to the 22nd Annual Meeting of the SGSR, Washington, February, 1978.
9. KRIPPENDORFF, K., "A Spectral Analysis of Relations, Foundations," paper presented to the International Congress on Communication Sciences, Berlin, May 1977.
10. McGILL, W. J., "Multivariate Information Transmission," Psychometrica, 19, 1954.
11. SHANNON, C. E. and W. WEAVER, The Mathematical Theory of Communication, Urbana, University of Illinois Press, 1949.
12. SIMON, H. A., "The Architecture of Complexity," pp. 84-118 in his The Science of the Artificial, Cambridge, Mass., MIT Press, 1969.