4. PART AND WHOLE — THEORY OF STRUCTURE

If the birth of information theory should be marked by its independence from thermodynamics, the present author's paper of 1939 may be regarded as the first paper in the new discipline. Usually, information theory is considered to have been born in 1949 (Shanon, 1949). The paper of 1939 had two messages to convey: (A) the quantity expressed by the function (2) can be used as a measure of uncertainty in any area disconnected from thermodynamics; (B) the same function can be used as a measure of structure in an assembly of individual objects. The theory used in the paper of 1939 was formulated in the framework of the non-Boolean logic (quantum logic) and therefore was much more general than (hence included as a special case) the usual information theory, but we shall not discuss this last point in the present paper. Point (A) may have been made clear by the explanation given in the first three sections of the present paper. We shall now explain point (B) briefly in this and the next sections. The idea of redundancy discovered in the 1950's by the communication engineers is a simple case of what is meant by structure here.

As the 'assembly of individual objects', you can imagine, for instance, a human society or a piece of matter consisting of molecules. Nobody can predict the exact behaviour of individual persons in society or the exact positions and velocities of molecules. The state of each individual can therefore be considered as a chance phenomenon. The assembly, being a collection of these stochastic individuals, also behaves stochastically. When would we then say that an assembly is structured or organized? If the behaviours of individuals are correlated, and interdependent, we would say that there is some structure. The result of such correlation would be that in spite of individual disorder or uncertainty, the behaviour as a whole will show some regularity or predictability and less disorder or uncertainty. If this is the case, the degree of structure may be measured by the balance between the uncertainty of individuals and the uncertainty of the whole. For, the larger the individual disorder and the smaller the total disorder, the larger the structure.

The above consideration leads us to the following idea. The degree of structure K is expressed by the sum of the entropies (the S-function given in (2)) of individuals minus the entropy of the whole.

$$K = \Sigma\, S(\text{part}) - S(\text{whole}) \tag{3}$$

Let us test this idea with a very simple example. Let us consider two girls *a* and *b* in a dormitory. We observe the lounge and if *a* is there, we put 1 in the row corresponding to *a* and if *a* is not there we put 0. Similarly for girl *b*. Suppose we have obtained the following list as the result of eight observations:

	1	2	3	4	5	6	7	8
a	1	1	0	0	1	1	0	0
b	0	1	1	0	1	0	0	1

This shows that girl *a* has probability $\frac{1}{2}$ of being there and probability $\frac{1}{2}$ of not being there. Hence, her entropy is $S(a) = \log 2$. This can be obtained from (1), as there are two cases (presence and absence) with an equal probability. Similarly, the entropy of girl *b* is $S(b) = \log 2$. Now how about the state of the assembly of two girls together? The are four cases: both there (1,1), both not there (0,0), *a* is there and *b* is not there (1,0), and *a* is not there and *b* is there (0,1). Each of the four cases appears with equal probability in the above list, hence the entropy of the whole is $S(\text{whole}) = \log 4$. Hence, the strength of structure is $K = 2 \log 2 - \log 4 = 0$, i.e. there is no structure.

This result agrees well with our intuitive understanding of the situation. The above list will be obtained indeed if two girls come to the lounge randomly and independently with probability $\frac{1}{2}$.

Now consider the list:

	1	2	3	4	5	6	7	8
a	1	0	1	1	0	0	1	0
b	1	0	1	1	0	0	1	0

This list shows that each girl comes to the lounge with probability $\frac{1}{2}$ as before, but they are such good friends that they come always together. $S(a)$ and $S(b)$ remain the same, but $S(\text{whole})$ is different. In fact there are only two cases, both there (1,1) and both not there (0,0), with equal probability $\frac{1}{2}$. Hence $S(\text{whole}) = \log 2$. The result is that the strength of structure is $K = 2 \log 2 - \log 2 = \log 2$. The point is that, individually seen, the girls came to the lounge randomly with probability $\frac{1}{2}$, but seen collectively they show some predictable regularity, namely, they come always together.

5. EMERGENT STRUCTURE

The use of entropy functions as a measure of structure was started in the paper of 1939, but a more systematic study along these lines had to wait another twenty years (Watanabe, 1959; 1960; Garner, 1962). These later papers of the present author showed among other things that the entropic measure of structure had a power to reveal 'emergent' properties of an assembly. By an emergent property is meant here a property which does not exist when individuals are taken singly or in pairs but appears when individuals are taken as constituting a group of three or more.

Let us take an example. There are four girls living in a dormitory, and we observe as before their presence and absence in the lounge. The result of eight observations was as follows:

	1	2	3	4	5	6	7	8
a	1	1	1	1	0	0	0	0
b	1	1	0	0	1	1	0	0
c	0	0	1	1	1	1	0	0
d	0	1	0	1	0	1	0	1

The probability of each girl being in the lounge is $\frac{1}{2}$. Therefore their individual entropy is log 2. Take, next, a pair of girls, say, *a* and *b*. We find the situation exactly the same as the first list of the last section, i.e. there are four cases (0,0), (0,1), (1,0), and (1,1) with equal probability. Hence, their pairwise entropy $S(a,b) = \log 4$. If we consider the pair as the 'whole' and the individual girls as the 'parts', the structure K for these two is $K(a,b) = S(a) + S(b) - S(a,b) = 0$. There is no structure. This situation is the same for any of the six possible pairs you may take among them. The pairwise structure does not exist among these four girls.

Next, take three girls together, say, *a,b,c*. We discover from the above list, they show four different patterns (1,1,0), (1,0,1), (0,1,1), (0,0,0) with equal probability. Therefore, the entropy for the three as a group is $S(a,b,c) = \log 4 = 2 \log 2$. Applying our formula to this group of three we get $K(a,b,c) = S(a) + S(b) + S(c) - S(a,b,c) = 3 \log 2 - \log 4 = \log 2$. So, there is some structure in the group of (a,b,c).

Next, take another group of three, say, (a,b,d). We discover that they show eight different patterns. (1,1,0), (1,1,1), (1,0,0), (1,0,1), (0,1,0), (0,1,1), (0,0,0), and (0,0,1) with equal probability. Hence $S\ (a,b,d) = \log 8 = 3 \log 2$. The structure for these three is $K(a,b,d) = S(a) +$

$S(b) + S(d) - S(a,b,d) = 3 \log 2 - \log 8 = 0$. This means that the group of (a,b,d) shows no structure. There are two more ways of taking a group of three girls, and they do not show any structure either. $K(b,c,d) = K(a,c,d) = 0$.

Why does the group of (a,b,c) have structure while the other three do not? The reason is because the group of (a,b,c) lacks the patterns (1,0,0), (0,1,0), (0,0,1) and (1,1,1) which exist in the other groups of three. Such a situation can very well happen if the three girls a,b,c are emotionally linked by friendship and jealousy in such a way that each of them does not want to be alone in the lounge but does not want to see the other two together at the same time. Such a property disappears if we take girls pairwise but emerges if we take them in a group of three. The groups of three other than (a,b,c) do not have such an emotional tie, hence they come into and leave the lounge in a random fashion.

To simplify the explanation, we took simple and concrete examples, but the method of structure analysis based on the idea of information is applicable also to more abstract items such as grammatical entities, propositions, etc., to reveal stıuctures underlying them. We may discern two different streams of thought in the analysis of structure. One is the now popular 'grammatical' approach and the other is the 'statistical' approach. The first tries to formulate the generative rules from which all individual cases are supposedly derived by deductive instanciation, while the second tries to discover the underlying common generality or regularity inductively from a large number of individual cases. If the former is 'rationalistic' the latter is 'empiricistic'. These two, however, would not consider themselves as mutually exclusive but they should complement each other. Our method of structural analysis belongs of course to the latter stream.

$$S = - \sum_{i=1}^{n} p_i \log p_i \qquad (2)$$