SOME STATISTICAL PROBLEMS IN A RESEARCH PROJECT ON ORGANIZATION

By: Harrison C. White, University of Chicago

In order to assess the sampling variability of any measure of sociometric structure, one must define a priori probabilities in the sample space of structures. There is a curious divergence in the literature. In both approaches equi-probable elementary states are defined, but on the one hand only non-isomorphic structures are counted as equally probable,[1] while on the other hand all structures are counted as equiprobable elementary states.[2] It is ultimately an empirical problem to choose between these (and other) alternatives,[3] a problem which apparently has not been recognized in sociometric research. It is suggested that perhaps all structures must be counted if the population to which the measures refer is of groups (of fixed size) studied in concrete field research. On the other hand perhaps only non-isomorphic states should be counted when the target population is an abstract anhistorical one such as is implied by experimental small-group research or research in the evolution of institutional forms of small groups.

This problem can be regarded as one aspect of the general problem of recognizing what states are distinguishable from one another and thus legitimate candidates for a set of equiprobable states.[4] In the writer's research project, in addition to sociometric questions, managers in a firm were asked to estimate the allocation by a named manager of his attention at work among four categories. Here it seemed possible that the psychological process of allocation is such that the counting of only non-isomorphic states is appropriate rather than the more familiar counting of all states. Since this case is much simpler than the sociometric one it will be used as the detailed example to bring out the issues involved in the general problem.

We concentrate on the psychological process of estimation of a typical manager, and in particular on the statistical variations of his estimates around an assumed underlying long-term estimate. Let the null hypothesis be that in some underlying large population of estimates by a manager the mean allocation of attention perceived is the fractions p_1, p_2, p_3, and p_4 for the four categories. We examine one estimate e_1, e_2, e_3, and e_4. What is the probability of obtaining e_1, e_2, e_3, and e_4 if a sample of one is drawn at random from the null population?

We assert that the estimates are discrete rather than continuous variables, so that the problem is one in finite probabilities. The spacings of the possible discrete values of the variables is an empirical question. However, we assume in our crude psychological model that the manager in making an estimate is allocating a number N of "units" of attention among the four categories; it follows that $e_i = \frac{N_i}{N}$ and each e_i has the same equidistant spacing of possible values, $\frac{1}{N}$.

It may seem that the multinomial distribution is now the right form for the answer to our question: i.e.,

$$\binom{N}{N_1, N_2, N_3, N_4} p_1^{N_1} p_2^{N_2} p_3^{N_3} p_4^{N_4}$$

is the probability of obtaining the estimate e_1, e_2, e_3, e_4, where $N_i = Ne_i$. Here N is the number of values taken by each e_i; if the manager answer in integer percentages, N = 100. This form for the answer is equivalent to the following probability model of the process of allocation: the respondent sequentially assigns his N units of estimation to the four categories, each unit having a priori probabilities p_1, p_2, p_3, and p_4 of falling in the various categories, <u>independent</u> of where preceding units have fallen. The four dimensional sample of size one is thus reduced to a one-dimensional sample of size N which is sequential. In order to make clear our objection to this sequential probability model of the psychological process of allocation, and simultaneously to prepare the ground for the model believed more appropriate, we must redescribe the multinomial distribution, in terms of the basic process of counting equi-probable states. This new description is in an abstract form, and a number of different specific empirical interpretations of it can be given.

N balls are to be distributed among Z distinct boxes. The Z boxes are divided into r groups, the j^{th} group containing Z_j boxes. In the concrete problem r would be 4, and Z is the number of micro-categories of attention assumed to exist in the respondent's thinking about attention allocation. Let N_j be the number of balls in the j^{th} group. Then

$$\sum_{j=1}^{r} N_j = N, \qquad \sum_{j=1}^{r} Z_j = Z.$$

If the balls are distinct there are $(Z_j)^{N_j}$ ways of distributing N_j balls among Z_j boxes, with no restrictions on the number of balls allowed in any one box. Thus there are $\prod_{j=1}^{r} (Z_j)^{N_j}$ ways of distributing N_1 particular balls in the first group of Z_1 boxes, and N_2 particular balls in the second group, etc. But there are

$\binom{N}{N_1, N_2, \ldots N_r}$ ways of splitting N different

balls into r distinct groups such that there are N_1 balls in the first group, N_2 in the second, etc. Finally, there are

$$\binom{N}{N_1, N_2, \ldots N_r} \prod_{j=1}^{r} (Z_j)^{N_j}$$

ways of assinging N different balls to Z boxes with the restrictions that there be N_1 balls in the first group of Z_1 boxes, etc.

On the other hand there are Z^N ways of putting N different balls in Z boxes with no restrictions.

If we make the fundamental assumption that each distribution of N distinct balls among the Z boxes is equally likely, it follows that

$$\frac{1}{Z^N} \binom{N}{N_1, N_2, \ldots N_r} \prod_{j=1}^{r} (Z_j)^{N_j} \qquad \text{(A)}$$

is the probability that there will be N_1 balls in the first group of Z_1 boxes, etc. Two distributions of N distinct balls in Z boxes are different if even one of the balls is in a different box.

Formula (A) is seen to be equivalent to the multinomial formula if we recast it as

$$\binom{N}{N_1, N_2, \ldots N_r} \prod_{j=1}^{r} \left(\frac{Z_j}{Z}\right)^{N_j}$$

and equate $\left(\frac{Z_j}{Z}\right)$ to the a priori probability p_j.

This interpretation of the multinomial model asserts that there are Z elementary categories of attention perceived by the respondent, Z_j of them being included in the j^{th} category of the r categories we defined as observers. The assertion that there is an a priori probability p_j of the respondent assigning one of the N units of attention to the j^{th} category is now equivalent to the assertion that he is equally likely to assign one of the N units to any of the Z elementary categories, if $\frac{Z_j}{Z} = p_j$.

We object to this multinomial probability model of the allocation process because it assumes that the N segments of the attention being allocated are distinguishable, as we can see from the derivation of formula (A), where the N "balls" are assumed distinct. This is not plausible psychologically. It is the categories of activity that are logically distinct in themselves. And the N segments of attention are not distinguishable by sequential order, since in our view the allocation by the respondent of the attention of the named manager among the four categories is not a sequential choice process.

The appropriate probability model for the attention allocation estimate in our opinion is obtained by assuming that the N units of attention are identical. The number of ways of distributing N_j identical balls among Z_j distinct boxes is $\binom{N_j + Z_j - 1}{N_j}$. Our fundamental assumption is that each of the $\binom{N + Z - 1}{N}$ ways of allocating the N indistinguishable segments of attention among the Z elementary categories is equally likely. It follows that the probability of obtaining the percentage allocation estimate e_1, e_2, e_3, e_4 is

$$\prod_{j=1}^{4} \binom{N_j + Z_j - 1}{N_j} \Big/ \binom{N + Z - 1}{N}, \qquad \text{(B)}$$

where $N_j = Ne_j$. For the same N, Z, and Z_j, formulae B and A give entirely different probabilities for each set of N_j.

Formula (B) is often called the Bose-Einstein distribution, and the multinomial distribution in the form of (A) is often called the Maxwell-Boltzmann distribution (in both cases for zero energy differences among the Z microstates).[3] It should be noted that as $Z \to \infty$ for fixed N, formula (B) can easily be shown to approach formula (A) by the use of Sterling 's approximation.

The determination of Z can be treated as an empirical problem similar to the determination of N. Our guess is that Z and N for the concrete problem of managers making allocation estimates would be approximately the same size, about 20, for any of the respondents. With such a value for Z and N it is not possible to approximate formula B by formula A or by a continuous distribution in which Z_j and N_j can be separated.

The major objection to this Bose-Einstein model is that for a given respondent there may not exist elementary categories which are psychologically meaningful and yet equi-probable in the sense given above. The specification of

$\frac{Z_j}{Z}$ is equivalent to stating a null hypothesis

for the given respondent in our new language; i.e. the p_j are replaced by Z and the set of Z_j.

The $\frac{N_j}{N} = e_j$ are of course the observed sample of one estimate.

The nature and total number Z of elementary categories of attention perceived may vary from one manager respondent to another. Whether or not this is true, the number of elementary categories Z_j assigned to the category j imposed by the observer will be expected to vary from one respondent to another: it is this variation which reflects the difference in underlying judgments as to how the named manager allocates his attention (which presumably results from the difference in experiences with the named manager). It should be made clear that the Z elementary categories are of an abstract nature not correlated with the categories imposed by the observer: the simplest example of such abstract elementary categories would be concrete segments of time such as various hours in the day. Selective recall and selective observation are two possible psychological mechanisms which could be used in specific empirical interpretations of the abstract process of allocating N units among Z micro-categories.

This example is in many ways artificial but it is hoped that it clarifies the general problem discussed initially, and that it brings out the assumption of full distinguishability which can be regarded as implicit in the multinomial distribution.

Footnotes

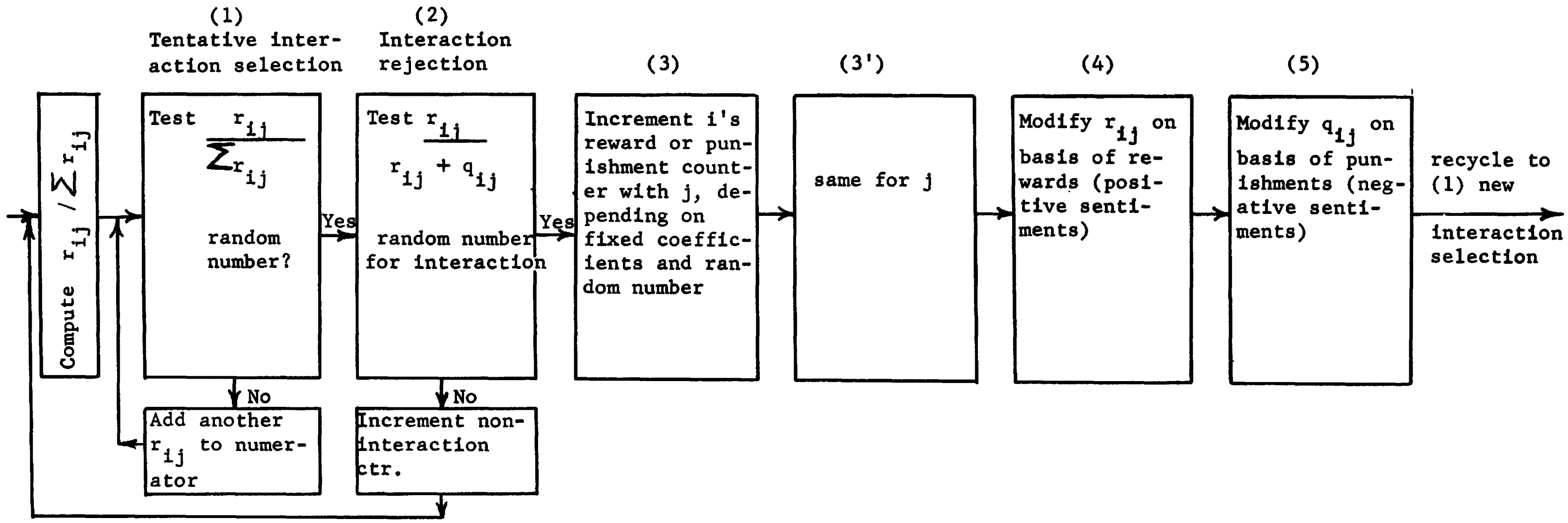
1. e.g., cf. F. Harary, "The Number of Linear, Directed, Rooted, and Connected Graphs," Trans. Amer. Math. Soc. 78 (1956), p. 445.

2. e.g., cf. L. Katz and J. H. Powell, "The Number of Locally Restricted Directed Graphs," Proc. Amer. Math. Soc. 5 (1954), p. 621.

3. W. Feller, An Introduction to Probability Theory and its Applications I (Wiley, New York) 1950, Chapter 3.

4. J. Riordan, An Introduction to Combinatorial Analysis (Wiley, New York) 1958, Chapter 6.

FIGURE 3: Flow chart for stochastic interaction within a triad, depending on rewards and punishments from past interactions.



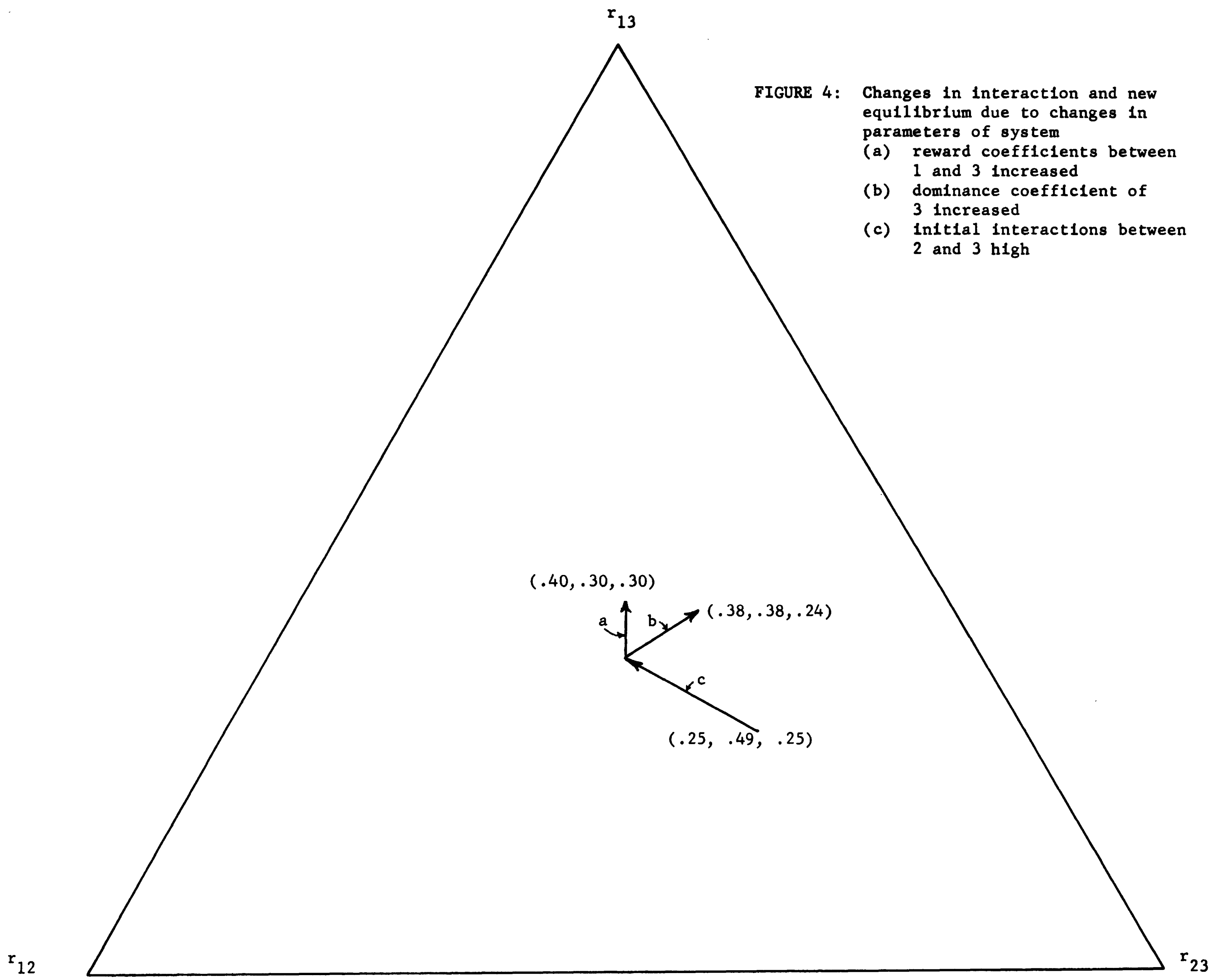


FIGURE 4: Changes in interaction and new equilibrium due to changes in parameters of system
(a) reward coefficients between 1 and 3 increased
(b) dominance coefficient of 3 increased
(c) initial interactions between 2 and 3 high

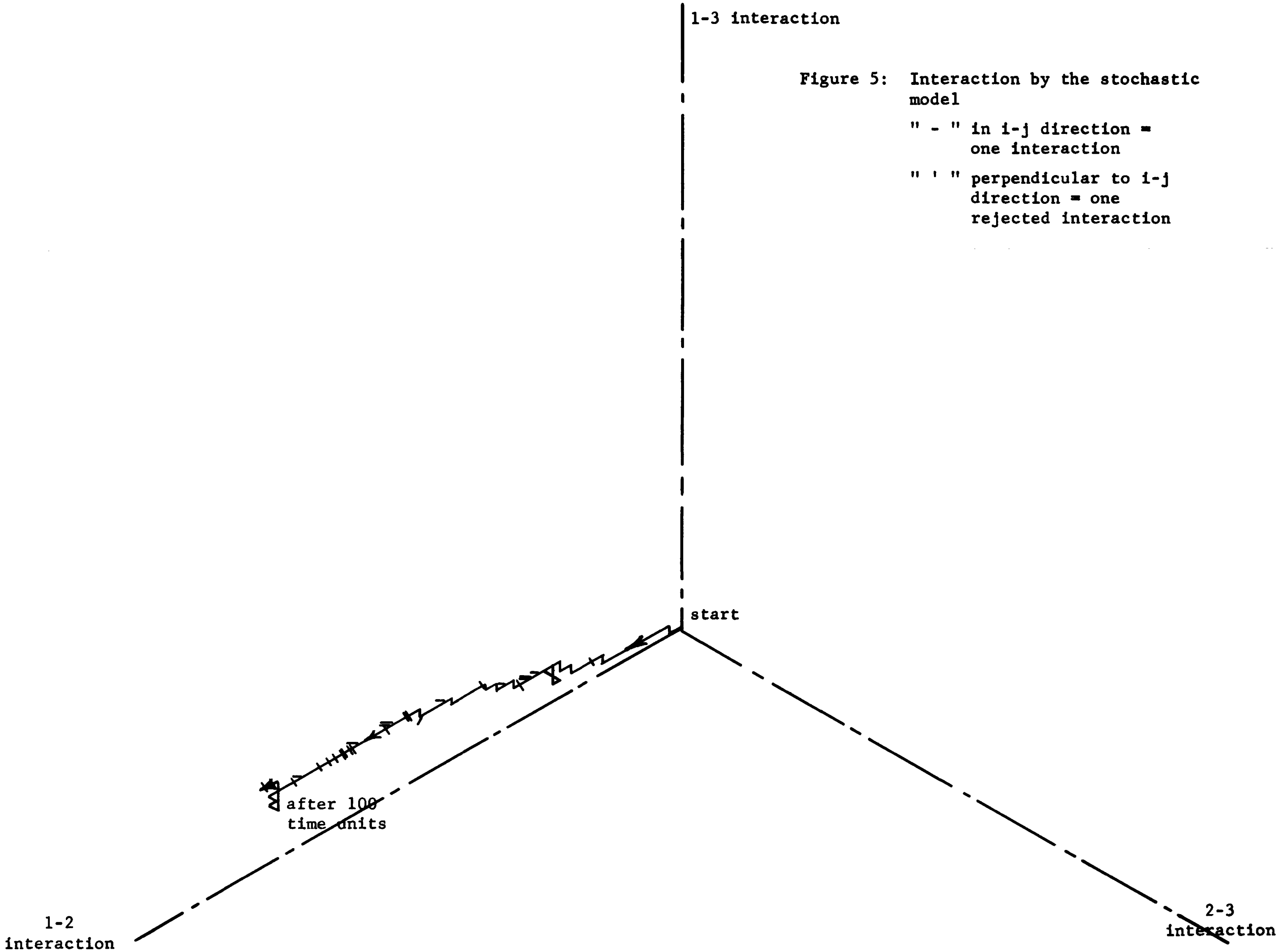


Figure 5: Interaction by the stochastic model

" - " in i-j direction = one interaction

" ' " perpendicular to i-j direction = one rejected interaction