also many checks between departments. Data are also checked against previous averages, which serves as a rough check, if not a precise one.

In the field itself, besides fieldmen, there are field supervisors, who in their turn are controlled by the field manager. The field supervisors' work consists of no actual work except constant supervision over the fieldmen's thoroughness of collecting data.

Over and above these checks in the field, and during the time when the data are passed through the Tabulator and processed for presentation in the form in which they are wanted by our clients, there is another department—" Inspection Department "—whose job it is to check the purchase, sales, stocks, price and other tables for internal correspondence and plausibility, having regard to known data about market conditions, etc. We often discover at the very last stages mistakes which should have been picked up weeks before. That reflects neither on the field investigators' " intellectual competence " nor on the " intellectual competence " of the checking and supervisory staff within the office itself. These mistakes, we find, are inevitable, and hence they are allowed for. Extra staff are employed to see that they do not go through. There are, of course, cases when deliberate dishonesty is noticeable. I recall an instance where the field investigator, who was supposed to have gone to certain shops and recorded their stocks of certain branded goods, never in fact went there. He filled in the field information sheets at his seaside hotel. In certain cases, where the fieldman detects inaccuracy in purchase data or his stock count, he is supposed to inform the office about it. We in the office know that a percentage of shop data will, over a period, be marked as inaccurate by the fieldmen, and if any one fieldman's audits are extraordinarily accurate, from that very fact alone, we sometimes suspect that the audits are not being conducted as thoroughly as they might have been. This particular fieldman went so far as to advise us from time to time that such and such shops should be thrown out of the sample for that period owing to inaccurate records, etc., and thus our confidence in his work was kept up. He was at last found out by certain discrepancies in his stock position which did not tally with our knowledge of the then existing zoning scheme for the distribution of certain products. There are thus instances of deliberate dishonesty which one can quote.

It is also very easy to question the " intellectual competence " of field investigators. In our work, for instance, present stocks are subtracted from past stocks by fieldmen themselves, so that if present stocks are higher than past stocks, the difference is shown up as a negative quantity. Obviously, this increase in stock, or this negative quantity, must be covered by an equivalent, or higher, purchase figure. It is physically impossible for stocks to increase without equivalent purchases, unless the previous stock-taking is erroneous. Now, one might say that all fieldmen should be intelligent enough to see that these minus quantities are covered by equivalent purchases. In fact, in a very large number of cases honest and very competent fieldmen fail from time to time to account for increases in stocks. Perhaps they would be called " physical fluctuations," but anyhow physical fluctuations are very difficult to explain in this context.

One word about the size of samples. The speaker has indicated that he often takes samples with a sampling fraction of 1 : 4,175, and his largest samples have been of the order of 1 : 168. The samples with which we deal in our particular field of Food and Drug trades, and which we know, from experience over a period of time, yield satisfactory results, are sometimes as large as 1 : 50 and even 1 : 25, but sometimes as small as 1 : 250. I should have thought the size of samples would be related to the variability of the data, among other factors, and consequently the sampling fraction is perhaps not in itself as significant as might be thought.

PROFESSOR MAHALANOBIS made a brief reply at the Meeting, and amplified it later, as follows :—

I am extremely gratified by the kind way in which my talk was received. Unfortunately, the paper could not be circulated nor presented before the meeting in a written form. This naturally left many gaps, to which references were made in the course of the discussion. Relevant material was in certain cases actually with me, but I had no time to discuss it at the meeting.

I greatly appreciate Dr. Yates's emphasis on the importance of close contact between mathematical statisticians and field workers and the need for intimate contact with numerical material. This is a point which I have always striven to emphasize in our Institute. My own view has always been that no one should be considered to have qualified as a statistician without having gone through an apprenticeship as a computer.

I am particularly interested in Dr. Yates's remarks about bias in estimating crop yields. I entirely agree with him in thinking that such troubles can probably be eliminated if a permanent organization of trained workers can be built up. Our problem has been to find out what can be done until such trained workers become available. I may also mention in passing that the size of the cut has been the subject of sharp controversy between the Indian Statistical Institute and the Imperial Council of Agricultural Research of the Government of India. I have been pressing for several years for a scheme to study and compare the differences in technique under the joint supervision of the I.S.I. and I.C.A.R., as this is the only way in which agreed conclusions can be reached. The proposal for joint investigations was strongly supported by Professor R. A. Fisher in his memorandum submitted to the I.C.A.R. in March 1945. Unfortunately, we have not yet succeeded in persuading the I.C.A.R. to agree to this proposal. I

spoke to a number of statisticians in the United States a few weeks ago and I hope to take the matter up there again shortly.

Dr. Oswald George as well as Dr. Bartlett have asked what I meant by physical fluctuations. I have already referred to this point in the text. Conceptually, sampling variations may be considered to have arisen from the fact that only a small portion or a fraction (and not the whole) of the universe is being surveyed. Recording mistakes arise from the unconscious bias, or sometimes conscious and deliberate negligence or dishonesty, of the human agency. Even when these two sources of error are eliminated there still remains a certain margin of fluctuations which I have called physical. For example, it would not serve any useful purpose to try to determine the weight of a living human being correct to, say, a milligram. When the administrator demands an estimate for a certain entity which he calls "the acreage under rice in Bengal in a particular year" or "the per capita consumption of cereals, etc.," there is a certain margin of fluctuation which is residual in the dual sense of being something independent of sampling variations and recording mistakes and also of being of no importance for the practical purposes in view. Operationally, the different types of error are mixed up in practice; but it is possible to some extent to separate them by appropriate analysis. I may add that what I have in mind is something like the permissible margin of fluctuations in the quality control sense, where production may be considered to be under control even when the measurements do not strictly conform to statistical randomness. Determination of the limits of physical fluctuations (in relation, of course, to the purpose in view) is difficult and requires a purely experimental approach. We realize its importance, but cannot claim to have made much progress in the actual determination.

Dr. George referred to my use of the phrase "unitary and zonal, configurational," etc. I had given a preliminary account in the paper "On large-scale sample surveys" in *Philosophical Transactions*, and I had included a more up-to-date description of the terminology and notation of sample surveys in my original draft, which I had with me at the meeting, but to which I could not refer for lack of time. This has been included in the present paper.

Dr. George has pointed out the need for small sample theory. I agree, of course, that it might sometimes become necessary to use it. We found, however, that too fine a stratification was often uneconomical in the sense that the additional cost was not commensurate with the gain in precision. In fact, in large-scale surveys we usually found large sample theory to be adequate for practical purposes. But I agree with Dr. George that in appropriate cases small sample theory would no doubt prove to be of great value. Incidentally, I may mention that one of the important lines of work in our Institute has been the development of the exact or small sample theory of distribution for multivariate analysis by my colleagues R. C. Bose and S. N. Roy.

I am glad to have Dr. George's support about the importance of errors other than sampling error. I also agree that errors arising through faulty working of machines cannot be neglected. I did not refer to this, simply because there was nothing peculiar in the Indian experience to make it worth mentioning.

Dr. George has referred to my remarks on the defects of complete enumeration. Certain observations which I had no time to make at the meeting will be found in the text of the paper. I may add one or two more remarks. Complete statistics are often obtained almost as a by-product of routine administration—for example, the number of criminal cases sent up for trial. Apart from mistakes in compilation or similar clerical errors I have no doubt that such statistics would be substantially accurate. Consider another example, which is nearer to the subject of the present paper. In temporarily settled provinces of India, crop records are compiled by a permanent, and often hereditary, staff of village recorders (called *patwaries*) primarily for revenue purposes, as the assessment is fixed on the basis of crop production. In such provinces, crop-acreage figures are claimed to be sufficiently accurate for all practical purposes. I have no personal experience of the quality of the work done by *patwaries*, but it may very well be true that crop records prepared by them are substantially accurate. This is quite different from what happens in a permanently settled province such as Bengal or Bihar. I had in mind the so-called complete enumeration made by an *ad hoc* staff whose number is large (of the order of fifteen or twenty thousand) and who are scattered widely over a province like Bengal. All available evidence points to the fact that work done under such conditions is unreliable. What I had wanted to emphasize was that the fact that complete paper records had been made ready was no guarantee that such records would be reliable. I may perhaps quote one concrete example. In 1939 the Government of Bengal decided to prepare a complete record, plot by plot, of the land sown with jute. After these records were prepared the Government arranged to have certain portions checked by permanent Government officers. The primary records, when checked, were found so unreliable that the Bengal Govenrment ordered all the records to be destroyed.

I agree with Dr. George, who thinks that suitable systems of sample checks and controls can be used with complete enumeration with great advantage. Some of the checks described in this paper may be quite suitably taken over.

Dr. George has referred to the results of the Bihar crop estimates. He is quite right in pointing out that the area under any given crop is not determinate unless the date is specified. I have

discussed this point elsewhere.* In a province like Bengal no machinery exists for obtaining the crop acreage on a specified date. The estimate obtained by a sample survey spread over a lengthy period must necessarily be somewhat indeterminate. In Bengal, by the alternative method of obtaining an estimate of complete enumeration, the result would be equally indeterminate, for the simple reason that the complete enumeration would have to be made by an *ad hoc* staff whose work would be spread over more or less the same period as that of a sample survey.

A little consideration would, however, show that although a specified date makes the crop acreage determinate it renders the estimate of total production necessarily somewhat vague. Whatever may be the crop acreage on a specified date it is likely that the harvested area would be somewhat different. As already mentioned, such differences are likely to be quite appreciable in India, especially in the case of rain-fed crops. In this situation we thought it advisable to develop a method for ascertaining the total production by a direct estimation of production per grid. This seems to be as far as one can go in existing conditions in Bengal.

I may mention here one particular procedure with which we have been experimenting for some time. At the time of harvesting, the yield per acre is determined by crop-cutting experiments in a number of grids. The total cropped area in each grid is estimated at the same time, and in this way the total yield for each such grid is ascertained. This enables a direct estimate to be made of the total production in each cell, zone or district, or in the province as a whole. All lands which go out of cultivation would not in this method contribute anything to the estimate.

In showing the progressive estimates for the *ad hoc* crop survey my object was to show the kind of data supplied to administrators. The sampling at each stage was not individually or independently representative of the whole "universe" as spread over space and time. Suitable modifications in the theory of estimation to cover such cases have not been worked out. Different progressive estimates necessarily referred to different regions of the province at different times, and were not strictly comparable. Differences were due partly to sampling variations and partly to actual physical differences in crop acreage arising from the different geographical coverages at different periods of the survey. My real point was that, in spite of such intrinsic differences, the figures supplied to administrators were probably quite useful for administrative purposes.

Dr. George has enquired about the proportional sampling errors. These, naturally, were large when the total acreage or the proportion of land under a particular crop was small. I have discussed this point in greater detail for the Bengal crop surveys, for which material was available for a number of years. The Bihar crop survey was, unfortunately, terminated after one year and sampling errors had not been fully investigated. I hope, however, to give an account of the effective margin of error in crop surveys in a subsequent paper.

I am glad Dr. Bartlett has referred to what I called space correlation functions, which I mentioned in the paper in *Phil. Trans.*, although I had no time to speak on this topic at the meeting. Since that paper was written certain other concepts have been developed; and, as already mentioned, some theoretical work has been done by R. C. Bose. The subject appears to be of considerable interest; but, unfortunately, owing to heavy pressure of other work, it has not yet been possible to tackle it on fundamental lines.

It may be useful, however, if I add a few brief remarks. Consider a field of, say, square or rectangular shape consisting of a very large number of rows and columns of elementary cells or quads. Along any single row (or column), the problem of space correlation would be one-dimensional, and identical with that of serial correlation in time series. The fact that space distributions are intrinsically two-dimensional, however, makes it necessary to have more generalized concepts. The approach at present adopted in the Statistical Laboratory is as follows.

Consider a single row or a single column consisting of a very large number of cells. (The boundary difficulty can be avoided by making the row or column endless by joining the two terminal cells.) Imagine any particular row or column divided into a number of segments, each segment itself still consisting of a very large number of individual cells or quads. The correlogram or serial correlation for different gaps can be then determined for each segment. The form of the correlogram or serial correlation for different gaps can then be determined for each segment. The form of the correlogram or the serial correlation function may then differ from segment to segment or may happen to be statistically the same for different segments. The serial correlation is usually called stationary if it happens to be statistically the same for different segments in whatever way the segments are formed (subject, of course, to the provision that each segment continues to have a sufficiently large number of quads to enable the serial correlation to be determined with the required precision).

Now consider the adjoining (or some other) row. The correlation function or serial correlation in this row may again be stationary, but may differ from the stationary serial correlation in the preceding row. Thus, although the serial correlation in each row is stationary, they may be all different. In a simplified case, the serial correlation along each row may be stationary and may be the same for all rows. Such a field may be called "uniform along rows."

* Presidential Address to the Section of Mathematics and Statistics, *Indian Science Congress*, Baroda 1942, p. 25.

Now consider the columns. The serial correlation along each or some of the rows may differ from segment to segment. If they happen to be the same over all segments in any particular column, then the serial correlation is stationary over that column. In the same way, the serial correlations may be stationary for each column and finally may be the same for all columns, in which case the field would be "uniform along columns."

It is possible, however, to consider the serial correlation not merely along rows or columns but in different directions, making different angles with the direction of, say, rows. If the field is divided up into two-dimensional segments (each segment still consisting of a sufficiently large number of elements or quads) it is possible to determine the serial correlation in different directions. In this way it is possible to construct what may be called the surface of the serial or space correlation in such a way that a section of the surface in any particular direction supplies the co..esponding curve of the serial correlation. If such surfaces of the correlation function happen to be the same for all segments, then the field may be called homogeneous. It might not, however, be isotropic. In fact, in each segment the field would be isotropic only if the correlation function is identical in all directions or, in other words, the surface of the correlation function is a surface of revolution.

The interesting fact then emerges that the surface may be isotropic for each segment, but the actual surface may be different in different segments, in which case the field would not be homogeneous, but would be isotropic. It is obvious that the field may be homogeneous, but not isotropic. Finally, when the field is both homogeneous and isotropic it may be called uniform.

Instead of working with the correlation function it is also possible to work with the variance function in the same way. As the two functions are mathematically connected the results based on them would be equivalent. For practical purposes the variance function would often be more convenient, as it usually enables the sampling errors to be calculated more directly.

I may add that the study of the surface of the correlation or variance function is of interest in connection with what I have called mapping problems, in which it is desired to ascertain the actual distribution of the variate under study over different elements or regions of the field under survey.

In the paper in *Phil. Trans.* I explained that fields may be of non-periodic, quasi-periodic or periodic types. I also explained the use of the patch number in characterizing the different types of fields. From the point of view of patch numbers, fields may be broadly considered as belonging to three types, namely, (*a*) fields in which the patch number is very small, corresponding to what I called "patterned" fields; (*b*) fields with patch numbers lying within an assigned range on either side of the model or expectation value, which I called fields of the random type; and finally, (*c*) fields in which the patch number was larger than the limit assigned for random fields, which I called the crystalline type and which would correspond to quasi-periodic and periodic types from the point of view of correlation function.

It is easy to see that previous knowledge of the type of the field can help in improving the efficiency of the sampling design, especially in the case of mapping surveys. For example, if the field is of a random type then a purely random sampling would be appropriate. If the field is however of quasi-periodic, or periodic, or what I have called of the crystalline type, then cluster or configurational sampling must prove quite useful. On the other hand, if the field is known to be of the patterned type then it should be possible to improve the efficiency by suitable zoning and stratification.

I am particularly grateful to Major Elphinstone for his kind remarks. I entirely agree with him in thinking that the apathy of administrators and the peculiar difficulties in which statistical work has to be carried out in India has to be experienced in order to be properly appreciated. The average administrator in India expects the scientific or statistical technician to supply evidence or proof in favour of what the administrator thinks to be right, rather than to give independent advice on objective grounds. Intellectual dishonesty, to which Major Elphinstone has referred, would in such circumstances be an actual advantage in securing promotion in official posts. This is why I have never favoured the idea of the Statistical Institute being run as a Government department or under predominating Government control. I think it is of the utmost importance that scientific organizations like the Statistical Institute should maintain an independent position, so as to be able to offer impartial technical service and advice on an objective basis. Mr. D. M. Sen has referred to the size of samples. I am afraid I had not made the position quite clear in the paper. It is not correct to say that we have used a sampling fraction of 1 : 4175. This was the fraction of the sample (in Table 1) on which we supplied a purely provisional first estimate. In the sample survey of crop acreage we are at present using a sampling fraction of the order of about 1 : 150 or 1 : 200. In estimating the yield of crops per acre the sampling fraction is necessarily far smaller and of the order of 1 : 6,000,000. On the other hand, in small-scale work one may go up to such large fractions as 1 : 20. Mr. Sen is entirely right in thinking that the size of samples would depend on the variability of the material. Naturally, it also depends on the precision it is desired to attain in the final results, and on permissible budget limits.