79

II. THE STATISTICAL METHOD IN PSYCHICAL RESEARCH.

By R. A. Fisher, Sc.D., F.R.S.

I have been invited to make a few comments on M. Sudre's interesting criticism of Miss Jephson's card-guessing experiment, but after reading Miss Jephson's own reply,* it is clear that she is so well able to justify the experimental technique she has adopted that there is very little for me to add on her behalf. The little that I should like to say is rather on behalf of, and in explanation of, the statistical method in general, in relation, not specially with psychical research, but with all phenomena the experimental demonstration of which is not so simple that the facts are universally accepted.

* Jephson, I. A reply to M. Sudre's article, 'An experiment in card guessing'. Proc. Soc. for Psychical Research, 39, 185-189, (1929).

M. Sudre goes to the heart of the matter with the paradoxical statement that the report of a single case in which five cards in succession were correctly guessed on five separate occasions "would have been vastly more convincing to the incredulous" than (presumably) the demonstration that a large number of persons working independently should, while making many mistakes, consistently make better guesses than could be ascribed to chance. How paradoxical this statement is must be apparent to anyone who has watched the reactions of an audience to a good conjurer. All, with their own eyes, have seen him produce a living rabbit from a hat which they have also seen to have been empty; most of them are confessedly unaware of the "normal" means by which the trick is performed. Yet they do not feel that their knowledge of the nature of things has been enlarged, or that new potentialities in their environment have been revealed to them. The unexplained phenomenon is discounted because it is only performed by a special person, the conjurer, presumably after special but unknown preparations. If Mr. Saltmarsh had really performed the feat, which M. Sudre ascribes to him, before our eyes, his performance would still have carried little conviction to the incredulous, who would, however, be rationally impressed if they were to discover a much slighter power of clairvoyance in repeated trials with ordinary people like themselves.

It is important that the incredulous should react in this way, but it is more important that they are right so to react. If any one of the conjurer's audience, after witnessing phenomena beyond his immediate comprehension, were to face the problem whether they were produced by "normal" or by "supernormal" means, he would see at once that he had not obtained the data necessary for a decision. What is needed is for him to frequent the company of amateur conjurers, and of learners in all stages of acquiring the art. The exhibition phenomena will then no longer be isolated and inexplicable; they will be brought into relation with special apparatus and special adroitness, which he can understand and accept, without necessarily being able to imitate it. In the opposite possibility that they were actually produced by the possession of powers hitherto unknown to science or common experience,

he would receive, what the isolated performance could never give him, adequate and convincing evidence of their reality.

In the investigation of living beings by biological methods statistical tests of significance are essential. Their function is to prevent us being deceived by accidental occurrences, due not to the causes we wish to study, or are trying to detect, but to a combination of the many other circumstances which we cannot control. An observation is judged significant, if it would rarely have been produced, in the absence of a real cause of the kind we are seeking. It is a common practice to judge a result significant, if it is of such a magnitude that it would have been produced by chance not more frequently than once in twenty trials. This is an arbitrary, but convenient, level of significance for the practical investigator, but it does not mean that he allows himself to be deceived once in every twenty experiments. The test of significance only tells him what to ignore, namely all experiments in which significant results are not obtained. He should only claim that a phenomenon is experimentally demonstrable when he knows how to design an experiment so that it will rarely fail to give a significant result. Consequently, isolated significant results which he does not know how to reproduce are left in suspense pending further investigation.

If I have rightly expressed the principles upon which scientific truth is established, it will be obvious why Miss Jephson did not confine her attention to one subject, but rightly extended her enquiry to many. She was trying out an experimental method which might (and may) prove capable of raising clairvoyance to the level of a demonstrable scientific fact. I believe that few will doubt the importance of such a fact, if true, or the need, in a sceptical world, of giving it a critically sound demonstration.

The value of such an experimental method depends on its chance of success if the phenomena is there to be revealed. It depends, in fact, on being sensitive to whatever powers of clairvoyance, however slight, might happen to be possessed by her subjects. It is this that M. Sudre has overlooked in suggesting that only complete success, an exact reproduction of the Suit and Value of the card drawn, should be taken into account. A subject who could always perceive the colour

of the card correctly, but could see no further, would possess very remarkable clairvoyant powers. Such a subject would score complete success only once in 26 trials, whereas one with no clairvoyant powers whatever would score complete success once in 52 trials, or just half as often. With M. Sudre's system of scoring, even so remarkable a clairvoyant could only count on scoring a significant result after several hundred trials; and since data could not be collected on this scale, all such cases would be missed. On Miss Jephson's system the same subject would score an average of 4·01 above chance expectation, and after 25 trials, with a standard error of 2, this is significant; the score, in fact, would serve to call attention to the astonishing special powers possessed by the subject.

This example, though conclusive in showing that Miss Jephson's procedure, in taking account of all the chief points to be perceived by a clairvoyant, is much more sensitive than the crude method of ignoring all but complete success, suggested by M. Sudre, does not really do full justice to her system of combining all powers of perception in one composite test; for the subject might well perceive, sometimes colour, and sometimes form the more clearly, and all tendencies to be right rather than wrong are allowed to reinforce one another in proportion to their power of excelling the results of random guessing. In this matter, as M. Sudre severely says, I have done nothing more than translate into that language of figures the instructions which I have been given; I take, however, full responsibility for judging this to have been worth while, as a contribution to experimental science.

In testing significance, a lower standard should not be taken than twice the standard deviation, and in M. Sudre's sentence: "Hence the score of a series of 25 tests may be attributed to chance if it falls between 9·18 and 13·18," one should read 7·18 and 15·18 as the more reasonable limits. M. Sudre's attempt to demonstrate by example the "univerisimilitude" of the scoring system therefore fails; as indeed any attempt to show that significant results could frequently be obtained without the action of a real cause, is necessarily foredoomed to failure.