EXPERIMENTAL DESIGN IN PARAPSYCHOLOGY

ROBERT H. THOULESS (*Great Britain*)

A well designed experiment is one that is: (1) *Valid;* it proves what it is claimed to prove; (2) *Economical;* maximum results are obtained with the minimum expenditure of time spent in experimenting and in subsequent working out; (3) *Fruitful;* the method of experimenting is such that positive results are to be expected if they are validly obtainable. Much published work in experimental parapsychology fails to some extent to satisfy these criteria, as does much work in experimental psychology and other sciences.

These defects are often due to experimenters, embarking on their work without sufficient consultation with others experienced in methods of experimenting. A set of experiments indicating a negative conclusion may, for example, be unfruitful because they were not done under conditions favorable to positive results. Fruitful experimentation depends on such factors as the mental attitudes of subject and experimenter, the length of experimental sessions, the time of day chosen for experiment, the health of the experimental subject, and his previous experience. Absence of a hostile atmosphere appears to be desirable; the ideal is that of adequate experimental precautions without fuss.

While the design of many modern parapsychological experiments seems to be satisfactory, certain defects are still too common. For example:

1. *Neglect of systematic randomization of conditions.*

If the purpose of an experiment is to discover whether a different condition of experiment produces a difference in scoring rate (e.g., whether it makes any difference whether the experimenter does or does not look at the target card),

this cannot validly be done by any arrangement of the experiment so that one of the two conditions may be differently influenced by chronological decline or any other cause operating differently at different times. It is necessary, therefore, to avoid such an arrangement as that of testing one condition in one set of experiments and the other in a later set of experiments. The comparison will be most economically made if the two conditions are suitably alternated or randomized within every experimental session.

2. *The use of an inappropriate method for assessing the significance of a difference between two groups of subjects.*

To solve a problem of the type: "Are colored or white people more successful in card-guessing experiments?" a method often adopted is to give each subject in both groups a number of runs through packs of ESP cards and to infer the superiority of one group if there is a significant difference between the total (or mean) scores of the two groups. This method is erroneous and its use explains much apparent contradiction between results obtained by different experimenters.

The error lies in the fact that each separate *guess* has been treated as the unit for the purpose of answering a question for which each separate *person* is the relevant unit. That the method is erroneous becomes apparent if we consider that by its use apparent evidence for a difference between the groups would be found if a large enough number of runs were made by each subject even if only one member of either group were scoring beyond mean chance expectation.

Two valid methods of dealing with such a problem are available. First, the experimenter can make a fourfold contingency table of the number of persons in each group scoring above and below the general mean of both groups. This table can then be tested for significance by the chi^2 method. Alternatively, he can calculate the mean score for each experimental subject and the standard error for each group from the observed deviation of each individual's score from the mean of his group. The significance of the difference between

the groups can then be determined by the use of these observed standard errors.

3. *The use of simple types of experimental design when dealing with problems for which more complicated designs would be superior in economy.*

Whether to prefer to carry out experiments by the use of simple designs or by more complex designs involving such statistical procedures as the analysis of variance is a matter of controversy amongst experimental psychologists as well as amongst parapsychologists. I suggest that the right answer is that we should use that method, whether simple or complex, which is most appropriate to the problem in hand. While the advantages of complex designs are often exaggerated, they may be really preferable in the fact that they may enable an experimenter to study simultaneously a number of variants instead of the two compared in more familiar experimental methods. They may thus effect a considerable economy in experimental time by enabling a single experiment to do the work of many. One may expect that the more complex experimental designs will be increasingly used in parapsychology.

It has not been my object in this paper to suggest that anything is radically wrong with most current parapsychological experimentation. But sometimes its methods could be improved. Experimental workers in our subject are so few that we cannot afford wasted effort.