I am sure that all of us appreciate the importance of the activity under discussion, that is, social experimentation. I hope that the complexities of the total process of defining an experiment, performing it, collection of appropriate data, data interpretation, and, finally, the drawing of conclusions, are ade quately appreciated. I mention this specifically because I am led by my reading both in the technical literature and in the semi-popular and popular press to the opinion that the complexities are <u>not</u> appreciated.

The past two decades of the United States and of the world have been remarkable for the wide concern that has been felt for human problems. It is interesting in this connection to note that concern about poverty "of the masses" and all the difficulties of people all over the world was felt before this century and even before these two decades only by very few. Indeed, it is more than interesting, it is quite remarkable in the history of mankind. One hundred years ago, the few people who were concerned were considered by most of "educated" humanity to be crackpots. There were thousands of so-called educated people, e.g. graduates of universities, who simply ignored the problems, and of those who did, the majority approached the problems with a value system which from our present point of view we can only term appalling. One could write a fascinating book on this. An example which struck me a few years ago was Galton who was a near-genius to be sure, but whose writings reflect strongly the prejudices and value system of academia of Great Britain of the 19th century.

Several of the crackpots performed what were and are called experiments, though, in fact, they were not experiments in the modern sense of comparative experiments, but consisted almost entirely of the implementation in a community of the ideas of a leader, with no defined "treatment" protocol, no replication and no control in the sense that the idea is used in the modern comparative experiment. The books on these "experiments" make fascinating reading to be sure, but there is almost a complete absence of any sort of experimental inference.

The area of social experimentation is of vast importance because every Dick, Tom and Harry is concerned with social programs. Legislators promulgate new social programs by the score. Our governmental apparatuses are involved deeply and politics is all-pervasive. The intrusion of politicians and the bureaucracy is inevitable. There is considerable risk that political influence will be exerted to induce conclusions from social experiments that are acceptable from a particular political viewpoint. The piper may try to call the tune, the piper in this case being a political high appointment in the supporting federal agency.

My contact with social experimentation at a professional level has been almost zero,

Oscar Kempthorne, Iowa State University

and the only basis for my contributing to the present session is that I have had intimate contact with experimentation in biology and agriculture and moderate contact with physical science and engineering. I am able, therefore, only to talk about general principles of experimentation with the hope that some of the remarks I make will have some relevance.

I was involved many years ago in what may be thought of as a very simple social experiment. The question was what would be achieved by supplementing the diet of children in grade school by giving them a glass of orange juice each morning at school. This is ludicrously simple in comparison with the experiments that are presented in this session. But I found great difficulty in developing a rationale for the design, for the choice of data to be taken, and for the analysis of the data. I mention this because the problem is extremely simple relative to the type of problem considered as social experimentation. The difficulties are compounded by the eagerness of workers, the public in general, politicians and public executives for definitive answers. I was struck in this connection by a statement in the Jensen paper. This paper has aroused a huge amount of controversy, and, I believe, justifiably so, for a multitude of reasons. In connection with the present discussion, we find the strongly assertive statement:

"Compensatory education has been tried and it apparently has failed."

This is followed by a long exposition of a hypothesis for the failure. I have to record my antagonism to this kind of statement in science generally and to this particular statement in its specific context. I ask:

Is compensatory education a well-defined treatment, like for instance a treatment protocol in an acceptable medical experiment? Is it not the case merely that some very ill-defined procedures which have some of the appearances of what we all think of as compensatory education were tried in an uncontrolled experiment? What were the possible biases of the studies? What was the sensitivity of the studies?

I am of the opinion that the making of such a statement is more in the nature of demagoguery than reasoned scientific evaluation. I question also whether it is consonant with the social responsibilities of scientists. Quite apart from any other aspect, it is clear to the near-idiot that compensatory education does succeed to a considerable extent. We can see this at all levels. Reading recently C. P. Snow writing on G. H. Hardy, one of the top mathematicians of the world of this century, I saw that the acquisition of the best possible tutor was an important act in the seeking of high status in the Cambridge University Mathematical Tripos. Why do we seek teachers to give compensatory (and improving) education to our children - we of the "upward bound" or "upward oriented" sections of society? Because we know that compensatory education of some sort does in fact work. So what is the status of the assertion of Jensen? If there is a germ of truth in it, as there may well be, let that germ be stated clearly.

It is essential that science be responsible and that it state knowledge and lack of knowledge very precisely, particularly in the context of deeply human affairs. The same holds with regard to heredity and environment, but this is outside the present arena.

In thinking about experimentation, it is useful, I believe, to run over the spectrum of experiments from Galileo's trials with balls rolling down inclined planes, to rates of reaction in chemical kinetics with varying concentrations of reactants, to determining the effects of nutrients on plants, to determining the effects of nutrients on animals, to determining the effect of physical nutrients on man, to determining the effects of physical, mental and economic nutrients on humans, which is what the present discussion is aimed at. This whole spectrum may be partitioned in various ways, but one partition is, I think, of critical importance. On one pole we have the experiments on physical objects, or on objects which do not have mental apparatus. On the other pole, we have objects or entities which cannot be observed, let alone experimented on, without the observation process itself producing an effect. The phenomenon was surely known back to antiquity in the case of humans and animals. It is interesting that it came to light so recently in physical science with the Heisenberg Uncertainty Principle.

If experimentation is done on humans, they know that they are being studied and subjected to chosen stimuli. This, alone, regardless of the nature of the stimuli, will affect their behavior and their reactions. In the case of humans and drugs, it is often possible to use placebos to measure some of the effects irrelevant to the drug being tested. But in the areas of social experimentation that are now under discussion, it seems very difficult to obtain any indication of the effects of experimentation, regardless of the stimuli. It will be not at all unusual for an experiment to show effects that are not found in a general societal program, because of such effects, which have been given, I surmise, special names. Social experiments are in a real sense psychological experiments with all the difficulties that these encounter.

Another factor is the occurrence of very long term effects. The simple physical (e.g. temperature varying) experiment on a piece of material, or the simple experiment on the effect of plant nutrients on the yield of, say, wheat or potatoes gives a nearly complete answer in a short interval of time, e.g. a season. But in the case of social experimentation, the experimental material may show effects for decades. The immediate effects may be encouraging, but these may 'wash out' over time, or there may be effects over a long period of time that are evinced <u>only</u> over time and cannot be predicted from the short term experiment. I feel that general remarks of this sort are necessary, because it is so easy to say "Let's do an experiment and we will see what happens."

It is also relevant to note that even in areas which do not present the difficulties alluded to, the progress of knowledge through experimentation has been very slow. I surmise that no one has made a census of the number of experiments that have been conducted on a "simple" problem, such as what nutrients to feed to wheat or to pigs. One would think that a simple experiment involving say, 50 pens of pigs would give all the necessary answers. But the fact is that no one such experiment or not even 100 such similar experiments have given definitive answers. So to hope that 2 or 3 social experiments will give definitive answers is very naive. This needs to be said because our citizenry and our politicians (and perhaps some of our scientists) will think so.

Related to this is the additional fact that the structure and technology of presentday experiments has been developed only after very laborious pains-taking steps. Technique of experimentation has taken a long time to develop and is still developing. So it may be surmised strongly that we do not know how to do social experiments. We learn how to do experiments only by doing them and learning from mistakes. So there is a very long road ahead for the social experimenters, and it is critical that our citizenry and politicians have some appreciation of the vast difficulties.

There were in the 20's and 30's several books on what was called field plot technique, i.e. how to perform agricultural and biological experiments. I would like to see a book on "social plot technique" if I may use a crude analogy, by which I mean how to choose experimental units, how to take care of border effects, how to measure the experimental entities, and so on. I say this because the technology and the body of now-accepted principles came about only with many workers with different ideas and working in different directions. Many of the experiments were aimed at the building of knowledge, and the formulation of action programs came later. In the social area there is now the natural desire to do one or two simple experiments and to hope to obtain from them action programs which will solve the ills of society. If furthermore, one or two simple experiments do not give a social prescription, then it is inevitable that many will say that experimentation has failed.

In view of my background of knowledge and experience, I have to confine myself to the philosophy, logic and technology of exper-

iments. The simple notions that arose in agricultural experiments have, I think some force in potential social experiments. The initial step is surely the making of a choice of what is termed experimental units, and the related choice of a pattern of imposition of treatments on the experimental units. In the case of an agronomic experiment, the experimental units are simply field plots. In the case of an animal experiment, the unit may be a single animal or a pen of animals. In social experiments, the unit could be a single person, a family, the families on a city block, the families of a certain type, of a town of a certain type, and so on. If it is possible to give one person one treatment and another person another treatment, the person could be the experimental unit. In contrast, it may be that one can apply a treatment only to the family, or the household, or the apartment block, or the city. The next step is to realize that it is impossible to obtain units which are identical. There will be variability between units. The prescription that has been accepted universally (apart from some of the neo-Bayesians) is that the only valid way to obtain some statistical control of this variability in comparisons of treatment effects is to adopt the two ideas of blocking or stratification of the units. So we have designs like the randomized block design, the split-plot design, the Latin square design, the incomplete block design. The logic of the process was given in highly heuristic terms in the classic by R.A. Fisher, "The Design of Experiments". The simple point is that the experiment is regarded as a single trial of a population of trials, in which the contribution of variability between units to comparisons between treatments is guaranteed by the conduct of the experiment. This is to be contrasted with the control of variability between experimental units and its effects by means of assumed linear models and processes of linear or non-linear least squares or some other method of fitting a model to data. The thrust here then, is to obtain validity of the experiment, in the sense that those forces which contribute to treatment differences must contribute equally to the estimation of error by which one assesses the reality of observed treatment differences.

It is important in this connection to note that if an experimental unit consists of several, say 5 animals, then the variability of response within the experimental units may be strongly misleading as a measure of the variability to which treatment differences are subject. If then a school were to be the experimental unit, with different treatments on different schools, the variability between observed treatment differences would depend on the variability between schools, over and above the variability between pupils within schools. This type of thinking, if it be accepted, has a devastating effect on one's ideas of size of experiment. Two social programs that are compared on 4 cities, with 2 cities receiving one program and 2 cities

receiving the other, has only 2 replicates with regard to the variability between cities. And this is the case whether the cities have one thousand or one million inhabitants. If, however, one can assume that there are no city differences but only variability within cities, then the replication with regard to that source of error will be the relevant one and may well be large.

It is also important to make a distinction between exploratory and confirmatory experiments. In the exploratory experiment one will impose several different treatments, and one will subject the resulting data to as wide a variety of analyses as one considers to be worth exploring. In the confirmatory experiment on the other hand, the whole protocol of the experiment both in design and in analysis must be prespecified. In the noisy sciences almost every experiment has two as pects, in that insofar as it is used for confirmation the analysis must be prespecified, but insofar as it is exploratory, the field of possible analyses is wide open.

If an experiment-experimenter interaction produces a new idea, then that new idea becomes an input for a new confirmatorv experiment. The philosophy of knowledge and statistical theory have not achieved, it appears, a mode by which hypotheses suggested by a set of data may be confirmed on the basis of that set of data. On the other hand, significance tests may be applied which enable one to make a judgment of the extent to which an apparently aberrant result is actually aberrant on the basis of an assumed model.

So I see a wide variety of social experiments, many of which will be inconclusive, most of which I hope, will be suggestive. I see a great need for repetition of experiments which have been suggestive. I see also a need for the formulation of the ethics of social experimentation, just as there has been a formulation of ethics for medical experimentation. I surmise that there are very considerable difficulties in this respect which cannot be ignored. The notion of an era of social experimenters using our citizenry as guinea pigs is offensive, but the idea that experimentation should not be done will delay the development of the sort of society that we seek.