INCENTIVE AND DISINCENTIVE EXPERIMENTATION FOR INCOME MAINTENANCE POLICY PURPOSES

By GUY H. ORCUTT AND ALICE G. ORCUTT*

Man's progress on the technological front has been enormous, but in spite of this—or in part even because of this—his progress on the social front seems dangerously inadequate. Nor does it appear that present research in the physical and biological sciences, valuable as it is, is likely to contribute more than marginally to the solution of those social problems which man must solve if he is to have a continued and attractive footing on this planet.

To create the world of his dreams man must learn to control himself, and he must learn to do so in a way which is compatible with his values. A deeper knowledge of individual and social behavior seems essential, for this. We need a development of the social sciences directed towards finding how to utilize and control the forces put into our hands by the physical and biological sciences. We need to know how incentives and disincentives, for instance, can be used to keep our social behavior under reasonable control without the sacrifice of values which are central to our well-being.

Policy makers in our social institutions are able to exert control over certain variables; they are expected to achieve some control over still other variables. It follows that policy makers need to know how these other variables are related to the variables which they do know how to regulate.

For instance policy makers know how to set minimum wage rates, and they need to know how the employability and earned income of individuals are affected by changes in these rates. They know how to legislate a personal income tax law, and they need knowledge of how

* Mr. Orcutt, a professor of economics at the University of Wisconsin during part of the time this article was being written, is a Senior Adviser at the International Bank for Reconstruction and Development. Miss Orcutt was an undergraduate economics major at the University of Wisconsin. The authors wish to express their gratitude to the National Science Foundation, the Ford Foundation and the Social Systems Research Institute of the University of Wisconsin for facilitating this study. A December 1966 draft of an exciting proposal by the United Planning Organization of Washington, D. C. provided the original stimulus for this paper. Entitled, "A Proposal for a Demonstration of New Techniques in Income Maintenance," the draft was written by Miss Heather Ross, a Brookings Research Fellow in 1966–67. William Baumol, Martin David, James Duesenberry, Robert Ferber, Robert Lampman, Robert Solow, Henri Theil, James Tobin, Harold Watts, and William H. White were of great help in developing the substance of our current paper. We hope it has and will continue to contribute to the undertaking and execution of planned experimentation in this important area. An appendix to this paper, dealing with "Sample Size and Cost Calculations," is available from the authors upon request.

labor force participation, occupation and job choice, effort, and productivity depend upon such things as the level and progressivity of taxation. Policy makers know ways of legislating a guaranteed minimum income for individuals or families, and they need to know how participation in the labor force and earned income of individuals would be affected by this guaranteed income. Also, they may want to know things such as how fertility, family composition, nurture, and education of children would respond to such a change.

By virtue of taxing, spending, loaning, and regulatory powers, governmental policy makers are in a position to modify the incentives and disincentives brought to bear on individuals, families, firms, and other groups. Presumably government policy makers know, or can ascertain, suitable objectives. But do they know how to reach these objectives? The physical means appear at hand, but how can the means be mobilized? What is needed from the social sciences are guide lines for how policy tools such as incentives and disincentives could be used to bring about our dreams, in place of nightmares.

In Section I we present the basis for hoping that well-designed experiments could assist in reducing ignorance about effects that would follow from the introduction of new economic policies or programs. In Section II we seek to encourage the idea that an experimental approach is feasible by discussing three of the primary reasons why many economists have ruled out any serious consideration of experimentation as a research approach which they might use. Sections III, IV, and V deal briefly with the choice of experimental variables, experimental units, and methods of analysis. Section VI suggests ways in which relevance of experimentation to policy issues may be enhanced. This section thus represents a natural extension of the discussion in Section II on feasibility of an experimental approach. Section VII presents some economizing possibilities and thus also represents an extension of Section II. Recommendations are presented in the final section.

I. *Whys and Whats of an Experimental Approach*[1]

Given that policy makers need more knowledge of the results of using available incentives and disincentives, how can the social sciences

[1] This paper focuses on issues that should be understood and faced up to by economists or other social scientists interested in carrying out social experimentation. However, the problem of designing an efficient experimental plan is a rather specialized one and certainly calls for the services of a statistician especially skilled in this area. The serious student will find that the following books provide an excellent introduction to the statistical literature on the design of analysis of experiments: W. G. Cochran and G. M. Cox, *Experimental Designs*, 2nd Ed., New York 1957; D. R. Cox, *Planning of Experiments*, New York 1958; W. T. Federer, *Experimental Design*, New York 1955; R. A. Fisher, *The Design of Experiments*, 7th ed., New York 1960; O. Kempthorne, *The Design and Analysis of Experiments*, New York 1952; H. B. Mann, *Analysis and Design of Experiments*, New York 1949; K. C. Peng, *The Design and Analysis of Scientific Experiments*, Reading, Mass., 1967.

meet this need? The research problem is a general one but, in the interest of clarity, we shall discuss it in a specific context.

Among the various income maintenance proposals which have been made, negative income tax proposals are receiving a great deal of attention and support. These proposals appear to hold great promise, but important questions remain unanswered. Uncertainty may lead to an indefinite delay in implementing a program which could become a milestone in our struggle with poverty.

Experimentation for policy purposes is needed to attack questions of interest to policy makers. Hence, with regard to negative tax experimentation researchers might begin by asking: "What are the concerns of policy makers?" Hundreds of questions come to mind, of course. Some have to do with administrative methods. Others revolve around the possibility of changes in labor conditions. From our standpoint, however, the most interesting group of questions are those concerning the responses of the people who would actually receive negative income tax transfers:

1. Does money have different incentive effects on people from different socio-economic classes?
2. Will the substitution and income effects associated with this tax both exert their pull in favor of more leisure rather than work?
3. Will individual work effort be affected by the level of assistance payments, the welfare tax rate or by both?
4. Will transfers associated with a negative income tax, affect the future earning powers of the poor?
5. On what will these transfer payments be spent?
6. What percentage of those helped by such a program could be expected to become self-supporting?
7. What long-range effects will these transfers have on the children of the poor?
8. What effect does the work status of parents have on the development of their children?
9. Should a negative tax program include some system of bonus payments to be awarded for finding and keeping a job?
10. What can be said for group bonuses and maintenance payments?
11. Will the stigma associated with welfare be avoided with a negative tax program? What are the implications of this stigma?
12. What effect will this proposed tax have on family stability?
13. What are the relationships between income level, the extent to which negative tax payments are geared to family size, and the birth rate?
14. And finally, how might such a tax effect the mobility of the poor?

There is general agreement that the present welfare system provides,

at best, no work incentive. President Johnson recognized this problem in his 1967 Economic Report to Congress: "With minor exceptions, payments under public assistance are reduced dollar for dollar of earnings by the recipient, removing any incentive to accept part-time work. We should encourage self-help, not penalize it. It is time to put an end to this 100 per cent tax on the earnings of those on public assistance." The need to improve upon our present welfare system seems clear. Answers to the above research questions would help in designing a better welfare system, but how are such answers to be obtained?

One highly important approach to the study of the impact of negative income tax and other income maintenance programs on earned income is the use of nonexperimental data with multiple regression techniques, in an effort to sort out the determinants of earned income of individuals and of family units. This approach has been, and should continue to be, pursued. Nevertheless, it has not yet served to eliminate our uncertainties. In part the difficulty stems from a failure to obtain, and effectively use, appropriate time series data for panels of individual and family units. Failure also may stem from difficulty in identifying historical policies which can be considered comparable with negative income tax laws. But even avoiding this, problems would remain. From the standpoint of application, not only is it important to know that two or more things are typically related; it is important also to know the direction of causation. Does individual A receive welfare payments because he has a low earned income, or does he have a low earned income because welfare payments are available to him? Or does causation run in both directions? An experimental approach could resolve this matter, and greatly reduce the area and range of uncertainty.

The basic inductive problem which faces us is that we wish to know the impact of a negative income tax law on earned income. We can observe earned income in the presence of a tax, but how can we know what earned income would have been in the absence of the tax. We can observe earned income before introduction of a negative income tax, and earned income after introduction of the tax. But can we attribute changes in earned income to the tax, or are they due to other factors which have changed in the meantime? We might observe earned income of one group before and after introduction of a tax, and earned income of a second control group not exposed to the tax over the same time period. Would it be reasonable then to ascribe the differential change of earned income of the first group relative to the second group to introduction of the tax? This would be a sensible procedure. This could lead us astray though if the characteristics of those exposed to the tax were systematically different from those not exposed either as a result of self-selection or as a result of selection by others.

A well-developed experimental approach has some problems of its own but can overcome uncertainties of the above types by virtue of our ability to draw random samples of experimental units from populations, use randomization in assignment of treatments to experimental units, and obtain unbiased estimates of population characteristics from sample characteristics.

Direct measurement of change in the earned income of any group, due solely to the introduction of a tax law, is out of the question. It is impossible simultaneously to apply, and not apply, a tax law to the same group. However, change in earned income of a group could be estimated by randomly drawing a sample from the group, and then using randomization in the assignment of tax treatments to members of this sample. In effect two non-overlapping samples would be drawn from the same group, without regard for subsequent treatment. The first of these two samples would be assigned a negative income tax; the second would not be given this treatment. Differences between the earned incomes of the two sample groups then would be due solely to sampling variability and to differences in the tax treatment. From the first sample we could obtain an unbiased estimate of what the earned income of the entire group would have been if it had been exposed to a negative income tax. From the second sample we could get an unbiased estimate of what the earned income of the same group for the same time period would have been if the entire group had not been exposed to the negative income tax. The standard error of the estimates would depend on the sample sizes, and will go to zero as the sample sizes go to infinity.

II. *Is an Experimental Approach Feasible?*

The need for incentive and disincentive experimentation is clear, but is it feasible? Will the public accept it? Can a useful level of experimentation be financed? Can the Hawthorne effect be neutralized? Other and more specific inquiries also must be made with respect to the technical feasibility of any particular experimental plan, but uncertainties in this area do not cast doubt on the general feasibility of incentive experimentation.

Public Acceptance

The experimental approach to acquisition of knowledge is eminently respectable due to the enormous success of experimentation in the physical and biological sciences. As a matter of fact, even experimentation potentially dangerous to the subjects is tolerated, and often widely lauded, if the research objectives are regarded as socially important.

Incentive experimentation, by its very nature, recommends itself for public acceptance. This kind of experimentation permits a satisfactory

disclosure of intended treatments to prospective subjects, and can be conducted in such a way as to ensure that each experimental subject makes a net gain. Furthermore, a strong case can be made for the social benefits that would follow from incentive experimentation. Disincentive experimentation poses some special problems, but even in this case it frequently would be possible to conduct experiments in such a way that each experimental subject at least makes an overall net gain.

Financial Feasibility

Realistic incentive and disincentive experimentation is likely to be costly, of course, if each subject must make a net gain, and if incentives are to be varied sufficiently to enable their effects to be distinguished from the effects of environmental changes. In fact, judging from the several proposals currently being put forward for investigating income maintenance plans, it would appear that possibly several million dollars per year are needed for a moderately comprehensive and highly attractive program of systematic incentive experimentation in the United States.

Relative to the billions being spent on research in the physical and biological sciences, a few million per year seems paltry. Yet is this much feasible? To us it is obvious that incentive research is of greater social importance than an accelerator costing an initial two hundred and fifty million and requiring an annual research budget of perhaps an additional sixty million. What do policy makers controlling the availability of research funds think though? No doubt we are now rich enough to afford both kinds of research, but will we? Time will tell if our society is mature enough to use that most basic and powerful of all scientific tools, experimentation, in learning to master critical social problems.

The Hawthorne Effect

As a result of a now classic set of experiments, social scientists are aware that even inference on the basis of experimental data is not always safe. In the experiments being referred to above, the experimenters thought they were investigating the effect of various kinds of environmental conditions—such as room color—on productivity. They quickly found that by altering the room color they could improve productivity. Being good scientists they investigated still farther, and found that the improved productivity could be maintained or even increased by other environmental alternations as well. They continued to alter environmental conditions, until they thought they had restored the original environmental conditions present at the start of the experiment. But, lo and behold, productivity remained far above what it had originally been. Apparently the workers involved were responding to something

other than the intended experimental treatment. It finally was evident that they thrived on attention.

In view of the above, some have concluded that the experimental method cannot be used successfully in studying effects of incentives. Yet this sort of danger also is present in the physical and biological sciences, and the methods used by physical and biological scientists to minimize this danger can be adopted by social scientists.

The difficulty underlying what social scientists refer to as the Hawthorne Effect is that treatments applied by experimenters may have a richness which is not fully captured by the specification of experimental variables. In the case of the Hawthorne experiments, the experimenters systematically and consciously altered room color—but inadvertently altered the attention paid to the experimental units receiving the intended treatment. As long as changes in room color occurred along with changes in attention, the incorrect association of room color with response was a likely error. In fact, if this one-to-one correspondence is maintained, the resulting error in association probably will not be discovered. Nor will there be any consequences. However, serious problems occur when the measured treatment variables and the unmeasured treatment variables are tightly linked during experimentation—but are not linked, or are linked in quite a different way, during application.

In the context of negative income tax experimentation, various negative income taxes would be applied to groups of experimental subjects. The fear is that the response of the experimental subjects might be due to attention or some other aspect of the experimental treatment that would not be duplicated if a real negative income tax law were enacted.

The first and most obvious step in minimizing the danger behind the Hawthorne Effect is the use of treatments which are as similar as possible to actions that might actually be taken by policy makers. In this way the same associations between measured and unmeasured treatment variables will occur during both experimentation and application. The experimenter may decide why something works for the wrong reasons, but at least it will probably work for the same wrong reasons during application by policy makers. In the case of negative income tax experimentation, this means that a serious effort should be made to avoid giving excessive attention to experimental tax units.

A second major line of defense is the thorough use of control groups. Three types of control groups come immediately to mind. The first is obtained by randomly allocating experimental subjects between treatment groups and a control group. An effort is then made to treat both the control group and the treatment groups in exactly the same way, except for the treatments of interest. For example, if attention is not considered part of the treatment, an effort must be made to give mem-

bers of the control group as much attention as members of the treatment groups. This will only be feasible, of course, if the experimenter at least recognizes the possibility that attention might play a significant role in influencing behavior.

Two additional types of control groups can be obtained by observation of the treatment groups both prior to treatment and after treatment has been stopped. In effect it was really by the use of such self-control groups that the Hawthorne Effect was discovered. Intended treatment stopped but the observed effect continued. This made it possible to avoid incorrectly attributing changes in productivity to changes in room color. Thus instead of regarding discovery of the Hawthorne Effect as pointing up a weakness of the experimental approach it might even be possible to regard it as showing how effective a well-designed experimental approach could be.

The third major line of defense is the use of experimentation in depth. Treatments are systematically regarded as involving many dimensions. By controlling more aspects of administered treatments the experimenter reduces the possibility that some unconsidered aspect of a treatment is what is really important. Depth of knowledge is sought and, with achievement of depth, less room is left for unpleasant surprises.

III. *Policy Variables and Choice of Primary Experimental Variables*

Choice of experimental variables lies at the heart of any experimental strategy. Two different possibilities suggest themselves. The most obvious choice to the social scientist grounded in the lore and spirit of the dominant physical and biological sciences, is to select, develop, and use experimental variables which appear in well-developed hypotheses. No one knows for certain how new hypotheses suggest themselves, but the search is for knowledge and the scientist follows wherever the trail leads.

A second possibility, however, which may suggest itself more forcibly to economists is to use as experimental variables those variables which policy makers can and are willing to control, or might reasonably learn to control. The objective, in this case, may not be knowledge so much as the more utilitarian objective of learning something that would be useful in developing and applying social policies.

Given unlimited resources and time, no conflict would exist between the above two strategies. The difficulty, however, is that in our present state of ignorance the number of interesting hypotheses which might be investigated in the social science area comes close to being infinite, while the resources available for investigating them are not. The world could wait for the discoveries of the physical and biological sciences, but once having acquired our present powers in these areas it is not certain that

mankind can afford to wait indefinitely for the knowledge needed to bring our social activities under some reasonable control.

No doubt the tools available to governmental policy makers can, and should, be enlarged. Yet really new policy tools of major significance suggest themselves but rarely, and are difficult to incorporate and master successfully. In addition, a wide and varied collection of policy tools are available already. We should attach great importance to finding which of these tools are suitable for use, and how they may be effectively used singly and in combination to achieve what needs to be achieved.

Types of variables which suggest themselves as experimental variables, by virtue of the fact that they might to some extent be controlled by governmental policy makers, include the following: payments of money and/or economic goods and services; availability and terms of employment; tax rates and schedules; availability and terms of loans; prices at which goods and services are available; rate of return offered on investments; and availability of information. For negative income tax experimentation, the experimental variables to be used would be payments of money combined with application of simulated tax rates and schedules.

It is obvious, of course, that the range of possibilities open to the experimenter is subject to some restrictions that don't apply to the government policy maker. Government can require compliance with a tax law, whereas the experimenter must obtain the voluntary cooperation of experimental units. Hence, determination of the effect of rewards seems more feasible than similar experimentation with penalties and/or prohibitions. However, experimentation of the latter sort might be feasible in cases where a penalty or prohibition is already in use by government. The experimentalist might investigate the effect of the penalty or prohibition by experimenting with its removal or reduction.

Another approach to investigating the effect of a basically unpleasant variable, such as a tax rate, might be to use two or more experimental variables in combination. For instance, various tax rates could be combined with payment of a specified sum. In this way the combination might be made attractive to prospective experimental units. This approach would be ideal for negative income tax experimentation since both types of treatment are aspects of a single law tax. Thus, in negative income tax experimentation, the experimenter is seeking to determine the separate effects of both the base support level and the marginal tax rate. However, if the objective were to learn the effects of marginal tax rates at higher levels of income, it might still be necessary to combine a variety of marginal tax rates with each of a variety of lump sum transfers in order to obtain any experimental units.

IV. *Choosing Experimental Units*

A prime consideration with respect to many of the policy variables that can be controlled by government is the resulting ratio of benefits to costs. This means that studies considering the effects of policy variables should seek not only to measure benefits in relation to costs, but also to uncover ways in which the ratio of benefits to costs could be increased. In this connection one objective of incentive experimentation should be the determination of how response depends upon the type of unit receiving the incentive, as well as on the manner in which the incentive is applied.

Evidence suggests that individuals respond most strongly to incentives and disincentives which, by their nature, can only be given by other individuals. Thus the possibility arises that the most effective way for governmental policy makers to apply economic incentives may be to provide group rewards for individual behavior. By so doing, groups may be motivated to provide their members with attentions individuals desire but cannot buy—i.e. respect, appreciation, attention, concern, love and affection.

Behavioral units which should be considered as possible candidates for experimental study include the following: individuals, families, household units and other spending units; work oriented groups, neighborhood groups, villages, towns and cities and social, religious, and other such associational groups. In negative income tax experimentation, the experimental units presumably would be either individual adults or married couples that might be tax units if a negative income tax law were to be enacted.

V. *Multiple Regression, Analysis of Variance, and Use of Parameters*

The following considerations are of a general nature but have been presented in a specific context for expositional convenience. Earned income during time period t of tax unit i, Y_{ti}, may be thought of as some unknown function of t and i. This may be expressed as follows:

$$Y_{ti} = F(t, i).$$

Given an interest in determining the influence on earned income of a negative income tax, as characterized by two tax law parameters, α and β, experimentation might be used in estimating and evaluating the following specification of the above equation.

$$Y_{ti} = F_1(\alpha_{ti}, \beta_{ti}, Y_{t-1,i}) + F_2(t) + U_{ti}.$$

Let us suppose the experimental use of α and β will start with $t=1$ and that prior to this both α and β are equal to zero. Let us also suppose

that a sample of experimental units is selected for observation during $t=0$ and that selected combinations of α and β are assigned to these experimental units by a random process at the end of $t=0$. Particular interest should center on the determination of the influence of α and β on earned income, but failure to utilize the predictive value of $Y_{t-1,i}$ would result in a sizeable and unnecessary increase in the error variance. For any given sample size the precision with which the influence of α_{ti} and β_{ti} on Y_{ti} could be determined would be reduced.

Unless we wish to merge experimental evidence from different time periods, $F_2(t)$ will have the same value for all experimental units and can be treated as a constant. Even if we do wish to merge evidence from different time periods, $F_2(t)$ does not pose much of a problem since it can be represented perfectly using a single parameter per time period.

The real question arises when a decision is made about the form of $F_1(\alpha_{ti}, \beta_{ti}, Y_{t-1,i})$. Given an unlimited number of observations, assumptions or approximations about the form of F_1 could be avoided by replicating each relevant combination of α, β, and Y until achievement of the desired accuracy in the determination of F_1. Having used a factorial experimental design, the resulting data could be subject to a three-way variance analysis with Y_{ti} as the dependent variable. Each combination of α, β, and Y would lead to the estimation of one response parameter. If ten values of α, β, and Y were of interest, then 1000 parameters would be required. The number of experimental units required would be some multiple of 1000.

An alternative approach to the estimation of $F_1(\alpha_{ti}, \beta_{ti}, Y_{t-1,i})$—and one which is far more attractive given high experimental unit costs and several continuous variables—is to approximate F_1 by means of either a linear or a quadratic function of α, β, and Y. The linear approximation would take only three parameters while the quadratic would take a maximum of nine. How particular negative income tax laws could be represented by the values of two or three parameters is described in a subsection of Section VI of this paper.

A parsimonious use of parameters is important because it becomes impossible to estimate parameters with precision as their number approaches the number of experimental units. In fact, several hundred or even thousand experimental units per parameter estimated may be required in some cases to achieve acceptable levels of precision. This does not mean that the usual analyses of variance approach should be discarded. However, a microanalytic multiple regression approach is almost certainly going to be attractive, and should be provided for. In such an approach, primary interest will center on the accuracy with which the influence of α and β can be determined; not on the accuracy with which individual Y_{ti} can be predicted.

The policy maker needs to know the extent, if any, to which tax units will reduce their earned incomes in response to a tax law. He does not need to know how other individual and environmental factors influence earned income. Such other variables are only to be included in the analysis in so far as they significantly reduce error variance, and thereby improve the precision with which the influence of the tax parameters α and β can be estimated. It is on this basis that the use of earned income, prior to introduction of an experimental tax treatment, is justified. The use of income of control individuals or groups also might be justified in this manner.

VI. *Relevance*

The steps from field experimentation to policy applications appear more certain than similar steps from nonexperimental research. Still experience indicates that there are pitfalls. The following subsections indicate ways in which the relevance of experimental findings may be increased.

Selection of Experimental and Observational Variables

The obvious point to be made here is that selection of experimental and observational variables should take into account what it is that policy makers could control, and what it is that they want to achieve control over. The experimental variables then should include variables which policy makers can control, and the observational variables should include those which policy makers wish to control. The experimenter may, of course, hold constant or vary additional variables in order to eliminate unwanted sources of variation, and to assist in clarifying the role of the primary experimental variables. The experimenter also may wish to observe other variables besides those which interest policy makers. One purpose of doing this is to guard against the possibility of unsuspected and deleterious side effects.

Selection of Experimental Units

The concern here is not so much with the type of experimental unit as it is with the choice of a sample of experimental units to be selected from a population of individuals, or group of individuals, of the same type. The issues are similar to those involved in the choice of respondents in sample surveys.

If all members of the population of units sampled respond in identically the same way to treatemnts, except for purely random variation from trial to trial, then many methods of selecting experimental units will be satisfactory. However, if members of the population respond in systematically different ways, then the results obtained may depend

critically on the choice of experimental units. In this situation, relevance of experimental findings can be enhanced by selecting a set of experimental units which are as representative as possible of the population with which policy makers must deal.

If experimentation is to shed light on what would happen if a specific negative income tax law were enacted, there is some point to using as experimental units a probability sample of tax units drawn from the entire U.S. population. Since there is no way of sampling future populations, it is never possible to sample exactly the populations that will be affected by future policy decisions. Nevertheless, it is clear that the relevance of experimental findings can be greatly increased by careful selection of experimental units. Randomization in the assignment of treatments to experimental subjects can guard against biases in making inferences about the responses of experimental units. It cannot guard, however, against errors that might follow from applying results obtained with one type of individual, or group, to individuals or groups of a drastically different type.

Part of the problem of conducting relevant experiments is to achieve an adequate correspondence between experimental treatments and potential negative income tax laws. One point of difficulty is that under actual tax laws individual tax payers are able to remain fairly anonymous and to keep their tax payments a matter of personal rather than public knowledge. Can these features of real tax laws be duplicated in an experimental study? Use of a geographically dispersed sample, such as could be achieved if experimental subjects were themselves a probability sample of the relevant segment of the entire U.S. population, would help a great deal on this score as well as yielding a representative group of experimental units. Maintaining the anonymous position of experimental units also would be facilitated if no information were released even about the city and state of residence of experimental subjects.

Selection of Environment for Experiments

Laboratory experimentation is attractive because of the possibilities it seems to offer of achieving control over the experimental variables. Nevertheless, in the case of policy oriented incentive experimentation, it seems clear that primary reliance should be placed on field experimentation. The primary drawback of placing much reliance on laboratory experimentation is that the conditions under which research results need to be applied cannot be satisfactorily duplicated in the laboratory. Many variables besides potential policy variables influence behavior. In fact, the influence of those potential policy variables of concern may be minor relative to the effect of other variables which vary but little

in a distributional sense from year to year for the population as a whole. Nevertheless the levels of these other variables in a laboratory setting might be very different from what they are in real life. If the effects of these other variables and the experimental variables of central interest are additive, this would not be disastrous. However, the assumption of additivity, while frequently satisfactory over relatively small ranges of variation, is often unsatisfactory over large ranges. The mere fact that the values of these other variables might be held constant in the laboratory would be of little help, since the basic difficulty is that they would be held constant at the wrong levels.

In the case of variables which are singled out for explicit consideration, it might be possible to guard against falsely assuming additivity. However, in the present state of our knowledge, many variables of importance are bound to be unknown. The role of these variables can not be taken into account in any adequate way. Hence, the advisability of applying laboratory results directly in making policy decisions is questionable. To guard against such surprises it is important for experimentation to be carried out in as realistic settings as possible, and with samples of experimental units which are representative of the populations with which policy makers must deal.

Parameterization and Generality of Results

In the original negative income tax experimentation proposed by the United Planning Organization, each tax law considered was treated as an essentially different treatment, rather than as a composit treatment involving a different combination of levels of parameters used to specify the essential features of a tax law. Unfortunately results obtained in this manner could not be easily generalized. If it were true that a specific negative income tax law could be selected as better than all alternatives, then the proposal might be justified in the same way that building a pilot plant might be. A last chance would be provided to see if the relevant theory holds. And at the same time, unforeseen operating problems could be identified and overcome. In this case, however, the scale and coverage of the suggested experiments would provide for only a tiny pilot experience. Moreover the theory which is available is far from adequate to guide us to the best one, or even to the best three or four, negative income tax possibilities. Instead, what we need is a more fundamental understanding about how such variables as the income of tax units are related to the significant characteristics or parameters of any negative income tax law.

Parameterization of negative income tax laws could be carried out in a variety of ways, but the following approach strikes us as the best. Its main virtue, aside from the fact that it is operational, is that the param-

eters singled out are expected to have independent, or almost independent, roles in respect to behavior. This orthogonality of role may not be an overriding requirement, but it will facilitate specifying a maximum amount of relevant information about a tax law with a given number of parameters.

For a given type of tax unit with a given initial income, possibly the most significant aspect, as far as income earning effects are concerned, of any linear or nonlinear negative income tax law would be the dollar transfer which that tax unit would receive in the absence of any adjustment of pretax income. Let this amount be the first parameter characterizing any negative income tax law as it applies to this tax unit. Possibly the second most significant aspect of a negative income tax law for any specific tax unit would be the marginal tax rate that would apply to changes in income earned by the unit. Let this marginal tax rate be the second parameter. In practice the above two parameters would appear adequate to describe the relevant features of all linear, and many nonlinear, tax laws for any specific tax unit. In the case of nonlinear tax laws they would provide a linear approximation of any law in the region which is most relevant to the specific tax unit. If, however, a third parameter seemed desirable, it might be the base support the tax unit would receive if it had no other income minus the base support it would receive under a linear tax law specified by the first two parameters. The effect of negative income tax on family composition and child bearing probably would depend on how the tax law provided for dependents. An additional one or two parameters might well be needed to capture this aspect of a tax law.

In the next section we discuss how a microanalytic regression approach would facilitate generalizing results obtained by study of one population to those to be expected for other populations. Suffice it to say that the microanalytic regression approach would likewise facilitate generalization of results to any tax law which could be locally approximated for individual tax units by use of the two or three experimental parameters

Microanalytic Regression and Generality of Results

A second weakness of the proposed experimentation, or at least of the method of analysis anticipated in the original proposal of the United Planning Organization, is that primary emphasis is placed on the determination of group means and on the accuracy with which such group means could be determined. This emphasis might be in order if each experimental group was in fact a probability sample from the same population to which any future negative income tax law would be applied. Since, however, future populations subject to any negative income tax

legislation may differ in vital respects from the population to be sampled in the proposed experimentation, it does not appear that primary significance should be attached to the means of the experimental groups. Nor is it evident how generalization from the responses of these groups to the responses of other groups would be achieved.

Use of a microanalytic regression approach would facilitate generalization from the responses of experimental units to the responses of groups having a composition which differs in some respects from the group from which experimental units are drawn. With this type of analysis the dependent variables in the regression equations should be individual, or individual tax unit, response variables. The predictor variables in such regression equations should include the values of the tax parameters which are being treated as experimental variables, plus individual tax unit, and environmental characteristics which may play important roles in determining the values of the dependent variables.

Use of a regression approach based on data from micro units would permit estimating relations which were intended to apply to micro units, and, by doing so, facilitate generalizing from results based on particular samples to the results which might be expected for other populations and other specific negative income tax laws. Such multiple regression equations, once estimated, could be used to predict the response of each member of a new population to a specific tax law. Appropriate values for the tax law parameters and appropriate values for those variables needed to describe the members' individual and environmental characteristics would simply be plugged into the regression equations. Having predicted the response of each member of a new population, or at least of a random sample of the members of a new population, predictions about the aggregative implications of the new tax could be obtained by summation of the micro predictions.

Relevance and Duration of Experiment

A real negative income tax law might go on forever. An experiment must terminate. Does this introduce an inescapable difference between any experimental situation and the situation that would be associated with a real negative income tax law?

One way of minimizing this problem is to conduct experiments which do extend over as long as from three to five years. This would seem to provide as much certainty about the future as is provided by congressional action.

A second way to deal with this problem would be to let the length of the experimental treatment also be an experimental variable. If essentially the same responses or lack of responses result from experimen-

tal treatments extending over three, four, and five years, then the fact that a real negative income tax might go on forever would not seem to matter. If responses were found to depend upon the length of experimental treatment, then some basis for making the necessary extrapolations also would have been obtained.

VII. *Economizing Possibilities*

Costs are the single most important barrier to a program of planned incentive experimentation. Incentives applied experimentally must be large enough to be noticed, and to be in the range of potential application. Moreover, since all experimental units must be willing subjects, it follows that each experimental unit must come out ahead. This means that the possibility of minimizing costs by use of penalties as well as rewards is limited. This section mentions several ways, however, in which experimental costs might be held down.

Sequential Experimentation

The basic idea behind sequential experimentation is to start with a small number of experimental units, and gradually increase the number of replications of the experimental treatments until the desired precision of results is achieved. This sort of experimentation permits postponing a final decision about sample size until initial experimentation has generated estimates of error variances and magnitudes of effects. Otherwise the experimenter must pick a sample size which will be large enough under the worst possible assumptions. This leads frequently to the use of samples which are larger than needed. Or else, the sample size may be inadequate, and the whole experiment may fail.

Another advantage of starting each experiment with a small number of experimental units is that gross errors in technique or experimental design may be discovered before the success of an entire costly experiment is jeopardized.

Joint Use of Control Groups

A group of experimental units which is not exposed to experimental treatments could, of course, serve equally well as a control group for judging the effects of experimental treatments performed on any number of other experimental units. Thus, if several experiments were performed simultaneously, some economy might be achieved by using a single control group for the whole set of experiments.

Joint Use of Treatment Groups

In many cases it seems reasonable, at least as a first approximation, to assume that the effects of two or more quite distinct experimental variables are additive. Use of an experimental design which achieves

orthogonality between the different experimental variables permits the use of the same experimental units as subjects for each of several experiments.

Joint Use of Experimental Treatments

Sharing the costs of a given set of experiments among several customers is another way of lessening the costs problem. In the case of negative income tax experimentation, for instance, there are some who are interested in work related responses, others who want expenditure information, and still others who are interested in the effects of negative income taxes on family formation and fertility. By collecting information on several observational variables, all of the above potential customers could obtain knowledge which they desire from the same basic set of experiments.

Joint Use of Specialists, Field Staff, and Facilities

An individual experiment may require the services of experts in sample design, sample selection, data collection, statistical analysis, computer use, etc. But any one experiment is likely to use only a fraction of the time of several such specialists. Clearly, economies of scale are possible. In particular, since sample survey organizations already require an expert staff, economies might be achieved by combining a program of incentive experimentation with a sample survey program. This might even permit the use of sample survey respondents as a control group for a wide variety of experimental studies.

Problem Selection and Specification

In addition to the above possibilities for economizing, there may be many additional ones having to do with problem selection and specification. For instance, if policy makers are interested in knowing whether or not some policy variable has an appreciable effect, the experimenter can focus on the detection of linear effects. In general this will be less expensive than determination of a nonlinear response surface.

VIII. *Recommendations*

In building knowledge and securing experimental evidence of vital concern to policy makers the full force of the scientific, experimental method should be brought to bear. The supply of pressing, unanswered social questions is enormous. How can any given experiment be expected to more than nibble at the mountain of our ignorance? Yet the nibbles can be enlarged, given some coordination, and stepped up in frequency. This surely is what has resulted in the marvelous flowering we are enjoying of the physical and biological sciences.

Incentive, disincentive, and motivational experimentation, of course,

are not the only research areas of crying importance to policy makers. But they are perhaps the most important. In any case, it is clear that policy oriented incentive experimentation should be undertaken in a systematic way, on an operationally viable scale, and *on a long-run basis.*

A long-run approach to social experimentation, however, needs—and in fact, demands—long-run financing for its assured success. This is necessary for wise planning. And it is essential to building up, and keeping together, an effective research team. A commitment of at least one or two million dollars per year under an arrangement providing for a three- to five-year moving horizon, with the horizon to be extended after each annual review, is needed for each major center of field experimentation established in the United States.

Some economies and increased effectiveness might be achieved by developing experimental, field research programs in close association with major survey research organizations. At no extra cost, sample survey respondents might in this way be made available for use as members of control groups. Also specialists and special facilities needed by sample survey organizations could be used to meet the needs of field experimentation. Regardless of the specific arrangements though, ways should be found to ensure that those major centers of experimentation and surveying inaugurated will be of use to social science researchers across our nation. It is time we set up some Mt. Palomars for the social sciences!