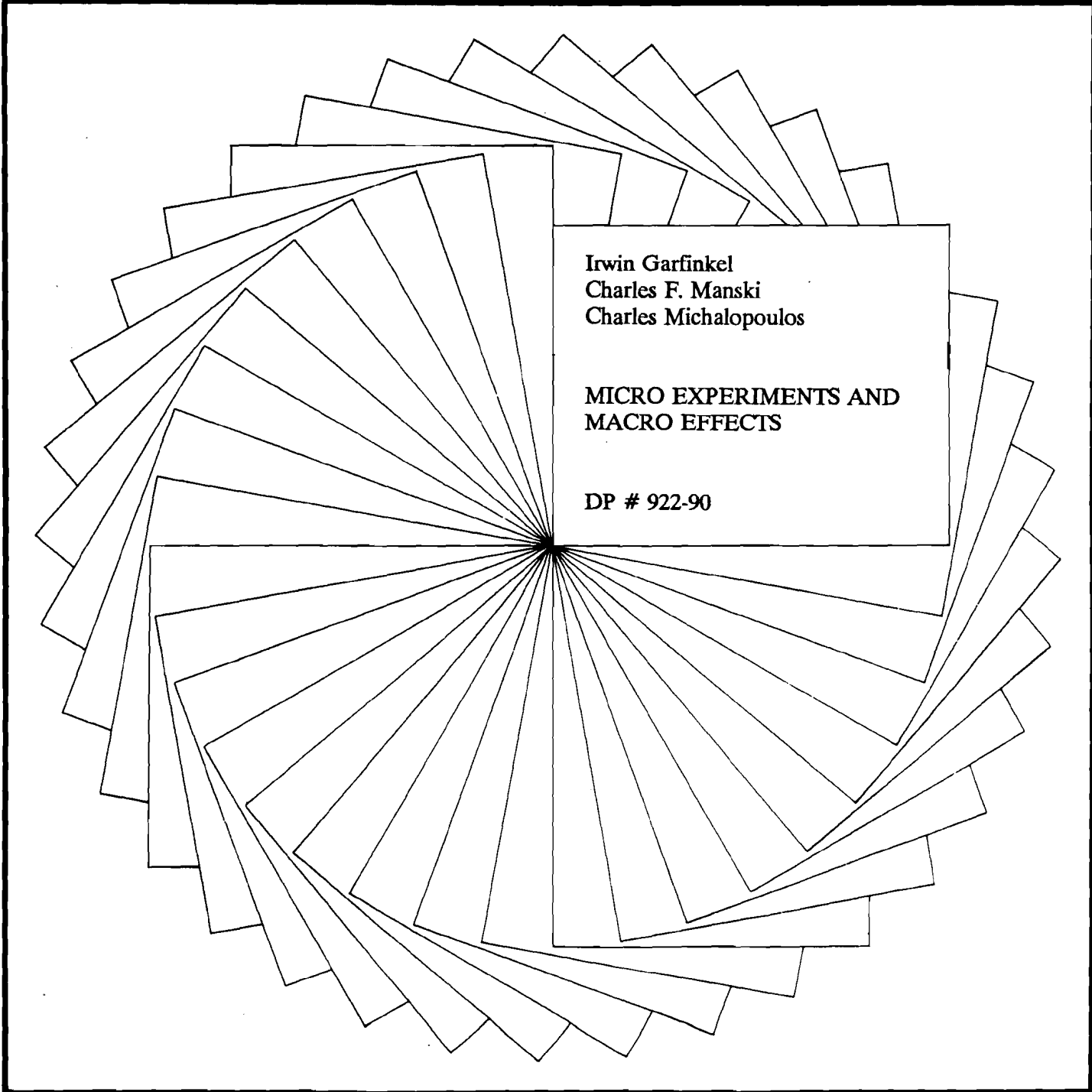# Institute for Research on Poverty

Discussion Papers

Irwin Garfinkel
Charles F. Manski
Charles Michalopoulos

MICRO EXPERIMENTS AND
MACRO EFFECTS

DP # 922-90

MICRO EXPERIMENTS AND MACRO EFFECTS

Irwin Garfinkel
Charles F. Manski
Charles Michalopoulos

Institute for Research on Poverty
University of Wisconsin-Madison

July 1990

Abstract

Micro experimentation, which is widely believed to be the best way to evaluate changes in social policy, fails to capture the macro effects that are likely to occur when social policy changes are implemented in the real world. Some phenomena unlikely to be measured are changes in market equilibrium, the effects of the diffusion of information, and the effects of norm formation.

A micro experiment to measure the effectiveness of a child support assurance system, for example, would have a number of shortcomings which would prevent it from adequately accounting for macro effects. Similarly, experiments to measure the effects of income maintenance programs and work and training programs fail to estimate the likely macro effects.

Because macro effects may be important consequences of many changes in social policy, those who carry out evaluations of social policy must design their experiments to measure the magnitude of these effects. Macro experiments, phased-in experiments, micro experiments with cluster sampling, and analysis of natural variation, while not trouble free, may provide some estimate of the magnitude of macro effects.

I. INTRODUCTION

In recent years, the belief that micro experimentation is the best way to evaluate social policy changes has become widespread in both the academic and policy worlds. Academic proponents of randomized micro experiments cite the ability to compare experimental and control groups without having to resort to statistical or behavioral models as the overriding advantage of experimentation. (See Ashenfelter, 1987; Lalonde, 1984; and Lalonde and Maynard, 1987.) This argument has persuaded policymakers to require that new social programs be evaluated by micro experiments. Fishman and Weinberg (forthcoming) describe the efforts of the federal Interagency Low Income Opportunity Advisory Board during 1987 and 1988 to persuade states to evaluate experimentally their new work and training programs targeted at AFDC recipients. The Family Support Act of 1988 goes beyond urging by requiring that its JOBS component be evaluated by micro experiment.

Consensus on the preferred design of evaluations nevertheless remains elusive. Ethicists challenge the moral appropriateness of experimentation under certain conditions. Some economists argue that less expensive nonexperimental micro-evaluation designs might provide comparable estimates of program impacts. (See, for example, Heckman, Hotz, and Dabos, 1987.) Economists also point out that micro experiments misrepresent or ignore the effects of policy changes on program entry. (See Heckman, forthcoming; and Moffitt, forthcoming.) Even proponents of experiments recognize that biases can arise out of systematic, unplanned differences between experimental and control groups. Burtless and Orr (1986), for example, mention nonresponse, limited duration, the voluntary nature of participation, and Hawthorne effects as sources of potential bias in classical experiments.[1]

The present paper challenges micro experimentation from a different perspective: micro experiments inherently cannot capture the macro or community feedback effects of a real social policy change.

Section 2 summarizes the all-too-sparse literature on macro effects. Section 3 illustrates the potential importance of macro effects and the corresponding deficiency of micro experimentation through a case study of a new child support assurance system. Section 4 extends the discussion to income maintenance programs and work and training programs.

The discussion in Sections 3 and 4 is more speculative than we would like. We suspect that macro effects are often substantial and that, correspondingly, the policy impacts measured by micro-experimental evaluations are often seriously biased. But the only way to determine the magnitude of macro effects is to measure them, something that has not been done. Section 5 considers alternative evaluation approaches that might provide information on macro effects.

II. MACRO EFFECTS: A REVIEW

The term "macro effects" embraces a wide spectrum of phenomena. We call attention to four here. These are market equilibrium effects, information-diffusion effects, social-interaction effects, and norm-formation effects. Of these effects, the first two are uncontroversial and therefore are discussed only briefly. We devote the most space to norm formation, which is the most contentious idea.

A. Market Equilibrium Effects

Perhaps the most widely recognized macro deficiency of a micro experiment is its inability to detect the changes in labor market equilibrium that might follow the introduction of a full-scale work or training program. For example, a job-counseling experiment conducted in a labor market with inelastic labor demand may detect positive employment effects on experimental subjects but would miss the associated "job displacement" effect, that is, the reduction in the employment opportunities available to other individuals in the community.

B. Information-Diffusion Effects

A second macro effect misrepresented by a micro experiment is the information-diffusion process by which potential participants learn that a new program exists and form impressions of its characteristics. Individuals presumably obtain much of their information by word of mouth, learning from the experiences of others. (See, for example, Lerman and Manski, 1982; and Manski, 1990.) In an experimental setting, where a new program is available to only a small subset of the community, diffusion of information about the program may be slow. The diffusion process is likely to be more rapid when the policy is implemented in a comprehensive fashion.

C. Social-Interaction Effects

Various social scientists have proposed models of nonmarket social interaction which suggest that a small exogenous impetus can yield a large social effect. Two notable examples are Gunnar Myrdal's principle of cumulation and Thomas Schelling's tipping model.

Although Myrdal's exposition of the principle of cumulation was written in 1944, it remains of interest today. Myrdal suggests that social phenomena may not be well characterized by the stable equilibrium ideas that dominate economic thinking. He speculates that the interaction among individuals may be so intense that the economy, and society as a whole, are better described by unstable equilibria with complex dynamics. The idea behind the principle of cumulation is that a push on any of the dimensions of a society--such as the economy--will have effects on other dimensions of the same society. Myrdal's application of the principle to race discrimination in the South provides an excellent example of the process:

> If ... the Negro plane of living should be lowered, this will--other things being equal--in
> its turn increase white prejudice. Such an increase in white prejudice has the effect of
> pressing down still further the Negro plane of living, which again will increase prejudice,
> and so on, by mutual interaction between the two variables, *ad infinitum*. A cumulative

process is thus set in motion which can have final effects quite out of proportion to the magnitude of the original push (p. 1066).

In Myrdal's model, a change in one facet of a society has a cumulative effect because the various dimensions of society affect one another. In contrast, Schelling (1971, 1972, 1973) presents a model in which decisions of individuals along one societal dimension are affected by previous decisions of other individuals along the same dimension. Schelling's tipping model (1971, 1972) attempts to explain de facto housing segregation. Here individuals have preferences which make them unwilling to live in neighborhoods in which the percentage of residents of other races is above some threshold level. Using simulations, Schelling demonstrates that extreme segregation can result even when residents are only slightly averse to the presence of neighbors of other races. Suppose housing patterns are in equilibrium, but that a nonwhite person moves into a predominantly white neighborhood. If all residents of the neighborhood are tolerant, then the equilibrium is stable. However, if some residents are averse to other races, they may move. Their movement results in an increase in the proportion of nonwhites in the neighborhood. This secondary effect on the racial composition of the neighborhood may induce more residents to leave. If this process continues, extreme segregation may result.[2]

In general, a micro experiment will misrepresent the Myrdal and Schelling type social-interaction effects that would prevail when a program is implemented in a comprehensive fashion. The prevailing practice in micro experimentation has, in fact, been to ignore such effects entirely. Social-interaction effects imply that the introduction of a program not only directly affects the behavior of the target population but also indirectly affects the behavior of the general population. But the practice has been to monitor only the behavior of the target population.

D. Norm-Formation Effects

Information-diffusion effects and the social-interaction effects described by Myrdal and Schelling pose specific nonmarket channels, or externalities, by which each individual's behavior depends in part on the behavior of other individuals. The idea that the behavior of individuals is mutually dependent is perhaps carried furthest in the social psychological literature on norms.

Stripped to its basics, the idea is that an individual's valuation of a given behavior increases with the fraction of the population who engage in that behavior. Suppose that implementation of a new program generates an exogenous change in the behavior of a target population. If behavior is norm-dependent, this begins a feedback process wherein the exogenous change in behavior induces other members of the community to change their behavior and this in turn reinforces the change in behavior of the program's original target population. In general, a micro experiment will underestimate the norm-formation effect that would be found if the program were implemented universally.

The notion that individual preference may depend in part on norms, which are themselves endogenous, has generally been dismissed by economists but does seem to have some empirical foundation. Given the contentiousness of this view, we review aspects of the relevant literature here.

Lamm and Myers (1978) summarize empirical findings on the phenomenon of group-induced polarization of ideas and actions. They cite research indicating that group discussion of ideas can solidify an individual's beliefs in the prevailing attitude. Related research indicates that individuals are often persuaded to change or suppress opinions--even strongly held opinions--if enough peers express the opposing point of view.

Studies by psychologists also suggest that imitation is an important mechanism by which groups influence individual behavior. West (1981) cites several relevant experimental findings.

For example, Asch (1952) conducted experiments in which subjects were shown a line and three comparison lines and were asked to indicate which comparison line was the same length as the original line. The subject was then placed in a group with seven "stooges," each of whom was instructed to respond incorrectly. In the group setting, 33 percent of the subjects confirmed the incorrect answer of the stooges, while a control group, answering in isolation, responded incorrectly only 7 percent of the time. Crutchfield (1968) describes similar experiments in which experimental subjects, who were exposed to uniform group sentiment, expressed different opinions from those of the control subjects, who responded in isolation. See also Kiesler (1969).

If we accept the premise that individual behavior is affected by norms, then strong potential consequences follow. Granovetter (1978) applied a model of norm formation to explain the discontinuous nature of riot behavior. Crane (1988) adapted this model to general delinquent behavior. In these models, individual decisions to participate in riots or other delinquent acts depend in part on the proportion of the relevant group who are already delinquent. Individuals vary in their threshold levels: some individuals will commit delinquent acts even if few others do so, whereas others will not become delinquent even if delinquency is prevalent in the population. Much in the manner of Schelling's tipping model, Granovetter and Crane demonstrate that a small change in the distribution of thresholds can turn a peaceful crowd into a rioting one or a peaceful neighborhood into a crime-infested one.

To close this section, it should be said that concern with norm formation is not entirely foreign to the literature on social experimentation. Harris (1985), arguing for randomized macro experimentation, recognizes that micro experiments underestimate the effects of social interactions among individuals. He asserts that "changes in life-style are likely to involve social learning, the diffusion of information, the changing of norms, and other phenomena that render individuals' responses interdependent" (p. 154). Earlier, in a discussion of the income

maintenance experiments, Kurz and Spiegelman (1973) expressed concern that the responses of

isolated individuals to the negative income tax would provide an inappropriate forecast of society's

response if the program were implemented broadly. In a research memo, they asserted, "An

argument in favor of saturation experimentation is that individual conduct is conditioned by social

norms which either discourage or reinforce him" (p. 18).

III. POTENTIAL MACRO EFFECTS OF CHILD SUPPORT

This section begins by describing the child support system in effect through the 1980s and a

proposed child support assurance system (CSAS). Next we take total child support payments as

the key outcome variable and argue that the effects of CSAS on child support payments would,

for several reasons, be misestimated by a micro-experimental evaluation of CSAS. The direction

of bias would probably be to underestimate the change in payments. We also speculate that a

CSAS micro experiment would underestimate the resulting decreases in poverty and welfare

dependence.

A. The Child Support System in the Past and a Proposed New System

By child support, we mean the transfer of income to a resident parent of a child who has a

living nonresident parent. Transfers paid for by the nonresident parent are referred to as private

child support and those paid for by the government are referred to as public child support.

Assessments in the 1970s and 1980s of the U.S. child support system--composed of 50

different state systems and innumerable different county systems--indicated that this system

condoned parental irresponsibility and contributed to the poverty and welfare dependence of

single mothers and their children. In 1978, only 60 percent of women with an eligible child had a

child support award (U.S. Bureau of the Census, 1987). Among unmarried mothers, only about

one in ten had a child support award. Among the 60 percent of mothers with awards, only half received the full amount due and over a quarter received nothing. One estimate of the ability of nonresident parents to pay child support concluded that, according to guidelines adopted by states, nonresident fathers should have been paying about four times the amount they were paying (Garfinkel and Oellerich, 1989).

The failure of the system to ensure that nonresident parents paid child support contributed to the impoverishment of children and shifted the burden of their support to the public sector. Nearly half of all children living in female-headed households were poor and on welfare.[3] If these families had received all the private child support to which they were entitled under the prevailing child support standards, both the poverty gap and the costs of AFDC to the U.S. Treasury would have been reduced by about 25 percent (Oellerich, Garfinkel, and Robins, 1989). Finally, because they had little education and experience, and would have had child care expenses if they did work, a large proportion of mothers receiving AFDC could not earn enough to lift their family from poverty even if they worked full time (Sawhill, 1976).

To rectify these shortcomings of the child support system, a group of researchers at the Institute for Research on Poverty at the University of Wisconsin-Madison, in conjunction with civil servants in the Wisconsin Office of Child Support Enforcement, developed a proposal for a new child support assurance system (CSAS) (Garfinkel, Melli, et al., 1982). The philosophical premise underlying CSAS is that parents are responsible for sharing income with their children, and government is responsible for assuring that children who live apart from their parents receive the share to which they are entitled.

The three major components of CSAS are a child support standard, routine income withholding, and an assured child support benefit. The share of income, or child support obligation, is determined by a simple legislated standard. Child support payments are routinely

withheld from wages and other sources of income. The child's custodian receives either what the nonresident parent pays or an assured child support benefit, whichever is higher.

B. The Effects of CSAS on Child Support Payments

Total child support payments are the product of the proportion of eligible children with awards, the level of awards, and the proportion of awards paid. That is,

(1)   CS = % with awards X award level X % paid.

To simplify the discussion we shall assume that mothers get residential custody of children and that fathers are the nonresident parents; in fact, introduction of CSAS might alter custody arrangements, an impact which would further strengthen the arguments to be made here. Also for simplicity, we assume that mothers know the award level and that awards are paid either in full or not at all.

Economic reasoning predicts that the decision of a mother to seek a child support award should depend on the value of having an award, which in turn depends upon the level of the award, the probability that an award will be paid, and the level of the publicly assured child support benefit, which is available only to those who have awards. (The decision to seek an award should also depend on the costs of seeking an award, which depend in turn on the father's resistance to the award. We ignore this resistance to simplify the exposition here. It is discussed in a later section.) Sociological theorizing about norms suggests that the mother's propensity to seek a child support award will also depend upon the norms in the community. Putting economic and sociological thinking together suggests that

(2) % Awards = f(award level, % paid, assured benefit, norms).

Exactly what norms are and how they influence behavior is the subject of debate.[4] We abstract from the details of this debate and simply assume that the norm in period t with regard to seeking a child support award increases with the proportion of eligible mothers in previous periods who actually had awards. That is,

(3) Norm$_t$ = f(% awards$_{t-1}$, % awards$_{t-2}$ ...).

Thus, entering norms makes the model dynamic.

To complete the specification, we note that the level of child support awards will depend on the child support standards that are adopted. That is,

(4) Award level = f(child support standards).

Similarly, the proportion of awards paid will depend upon how effectively the government collects child support payments. That is,

(5) % paid = f(govt. efficacy).

Equations (1) through (5) provide a simple framework for considering the changes in child support payments that would result from implementation of a child support assurance system. Previous research suggests that the child support standard in CSAS will increase award levels (Oellerich, Garfinkel, and Robins, 1989). Experience with wage withholding, as well as some

research, suggests that routine withholding of child support obligations will increase the proportion of awards paid (Garfinkel and Klawitter, 1990). These impacts of CSAS, in combination with the increased security provided by the assured benefit, suggest that implementation of CSAS would increase the proportion of mothers who seek child support awards.

The direct impact of CSAS could, in principle, be learned from a micro-experimental evaluation of CSAS. However, the direct impact might also change norms with regard to securing child support awards, thereby yielding an additional indirect impact. This macro effect cannot be learned from a micro experiment. In what follows, we consider this and other problems of micro experimentation in some detail.

C. The Problems of a Micro Experiment

What would a micro-experimental evaluation of CSAS look like? If we follow the practice in recent micro experiments, a sample of mothers potentially eligible for child support would be chosen in a few cities. Half of them would be provided CSAS and half would be provided the previously existing child support system. Within a period of one to two years the samples of experimentals and controls would have been enrolled. The experiment would continue for one to five years. If there was an interest in long-run effects, a small subsample of the experimental group--randomly chosen, of course--would be eligible for the experimental program for eighteen years.[5] The discussion that follows describes potential shortcomings of such an experimental design.

1. The duration problem.

Nearly all micro experiments last only a few years. Metcalf (1977) analyzes the problem of inferring long-run micro effects from short-run, micro experiments. Our concern about short

duration is different. The usual short duration of micro experiments truncates the entry effects and the dynamic macro feedback effects described below.

2. The entry problem.

Our description of a micro experiment did not specify how the sample of mothers would be selected. One possibility is to randomly assign mothers who come to court to obtain a divorce or separation or to establish paternity. The problem with this selection procedure, however, is that a large proportion of mothers with children born out of wedlock never enter the courts. Similarly, a substantial proportion of separated mothers do not enter the court system. Increases in award levels, payment rates, and the assured benefit all work to increase incentives to obtain a child support award. If the sample is drawn only from those already in the system, the micro experiment will miss the effects of CSAS on entry into the system.

Another possible point of random assignment is among women receiving Aid to Families with Dependent Children (AFDC). A fairly large proportion of mothers with children born out of wedlock receive AFDC benefits. Randomization of AFDC cases presents a better method of detecting the effects of CSAS on the entry of AFDC mothers into the court system. Of course, effects among those who are neither in the court system nor receiving AFDC will still be missed.

Even if we succeed in randomizing among eligible mothers, the entry problem disappears only if CSAS has no effects on divorce, out-of-wedlock births, and child custody decisions. To assess these effects, a micro experiment would have to randomize over the entire population of potential parents. It is difficult to conceive of a practical method of randomizing over the entire population of potential parents other than to randomize by site--that is, to conduct a macro experiment.

3. The information-diffusion problem.

Even if an appropriate point of randomization can be found, the stimulus in the micro experiment is likely to differ from the stimulus that would be received in a real program. In the micro experiment, the experimenter explains the advantages of the new system to the mother in the experimental group. At that point she can choose to enter the program. This stimulus may be repeated periodically. For the small group that is eligible for eighteen years, the stimulus could be repeated every year.

In the real world, mothers are likely to find out about the new system from friends and relatives who experience the improvements of the new system. A micro experiment, by its nature, precludes learning from the experiences of others. We speculate that whereas information diffusion would make the real-world stimulus grow stronger over time, the salience of the micro-experimental stimulus would be likely to diminish over time. Unfortunately, because we know so little about how the stimuli in micro experiments compare to the stimuli in real programs, we can do no more than speculate about this.[6]

4. The problem of detecting norm-formation effects.

a. Norms and the experimental subjects. A micro experiment, in which only the experimentals and controls are observed, precludes estimation of the effects of individual changes in behavior on social norms. If, as hypothesized, an individual's valuation of a given behavior does increase with the fraction of the population who engage in that behavior, then each additional mother who secures a child support award increases the normative value of securing awards. In the next period, the increase in the normative value of securing awards leads to a further increase in the number of mothers who choose to secure awards. Where this process stops depends on the specific relationship between micro behavior and norms. The social change

induced by a micro experiment might be close to or much smaller than the change induced by a real shift in policy.

b. <u>Norms and other actors in the community</u>. So far we have only discussed the influence of norms on the behavior of mothers. Norms may also influence the resistance of fathers. As noted above, the mother's decision to seek a child support award will depend on the costs as well as the benefits of seeking an award. The costs will depend upon the resistance of fathers. The effects of the provisions of CSAS are asymmetrical on resident and nonresident parents. The asymmetry arises from the fact that, under CSAS, government weighs in heavily on the side of the resident parent and the children. It follows that CSAS may induce some resistance on the part of fathers. To the extent that fathers' resistance is an individualistic behavior, a micro experiment can capture its influence. But the evidence of collective behavior on the part of nonresident fathers seems too strong to ignore. Fathers' rights groups have sprung up all over the country. It is conceivable that these groups will succeed in rolling back some of the initial increases in award levels achieved by CSAS.

On the other hand, fathers' resistance to child support awards may decrease in the long run due to changes in norms. The father's resistance to establishing paternity and to a child support award may depend upon how common paternity establishment is among his friends and relatives. Thus an initial increase in the proportion of fathers with child support obligations could set off a dynamic process which results in a larger ultimate increase. Once again, a micro experiment cannot capture this effect.

c. <u>Normative effects on earnings and welfare use</u>. Microeconomic theory predicts that CSAS will decrease the labor supply of mothers who would otherwise not have been on AFDC and increase the labor supply of those who would have been on AFDC in the absence of CSAS.[7] We focus on mothers who, in the absence of CSAS, would have been AFDC recipients. CSAS

promotes work among this group because it has both a lower guarantee and a lower tax rate.[8] A micro experiment can capture this initial effect.

But the initial decline in welfare use decreases the proportion of single mothers who are AFDC recipients. If mothers' decisions concerning welfare depend upon how common welfare use is, the initial decline in welfare use is reinforced by a macro feedback effect. That is, as the proportion of mothers dependent on AFDC declines, the acceptability of being a welfare recipient may decrease and the stigma of AFDC increase.[9] Once again, it is impossible to say how big the initial micro effect will be relative to the long-run total effect. All that is certain is that the micro experiment cannot capture this macro effect.

5. The civil servant morale problem.

There is another way in which a micro experiment might misestimate the effects of CSAS on earnings and welfare use. This reason is more speculative, but should not be dismissed out of hand. As AFDC caseloads begin to shrink, the morale of the civil servants who run our public assistance systems may increase. There is probably no more dispirited group of public servants in the country. A decline in caseloads brought about by an improvement in conditions outside of welfare may improve the morale of caseworkers, middle-level bureaucrats, and even the top civil servants for two reasons. It will reduce their feeling of being overwhelmed by the numbers of cases they have to deal with. It will also make them feel successful in both increasing economic well-being and decreasing dependence on AFDC. Improvements in morale in turn could lead to greater efforts on the part of our civil servants to promote the economic well-being and independence of poor mothers. A micro experiment cannot tell us anything about this morale effect.

IV. MACRO EFFECTS IN INCOME MAINTENANCE PROGRAMS AND WORK AND TRAINING PROGRAMS

In Section III we examined in some detail the possible effects of a child support assurance system and argued that a micro experiment would be likely to yield biased estimates of these effects. In this section, we briefly consider the difficulties that arise in experimental evaluations of two other kinds of social programs: income maintenance programs and work and training programs.

A. Macro Effects in Income Maintenance Programs

During the 1960s and 1970s four different income maintenance experiments were conducted. A principal interest in these experiments was to learn the effect of negative income tax (NIT) programs on the labor supply of beneficiaries.

Suppose that individual labor supply behavior is a function not only of income and prices but also of the behavior of others. Then the results of the income maintenance experiments are biased because they cannot measure norm-formation effects. That is, the decrease in the labor supply of prime-aged men that results directly from the economic disincentives of an NIT weakens the norm of full-time full-year work for men by reducing the percentage of them who actually fulfill that norm. It follows that the ultimate labor supply effect of the program is stronger than that measured by the micro NIT experiment.

Is there any evidence to suggest that norms are an important determinant of individual labor supply behavior? We think so. The income and substitution elasticities of labor supply for prime-age able-bodied men are much lower than those of married women and of men of college age and retirement age.[10] It is plausible that the difference is due to differences in norms with respect to full-time work. Prime-aged, able-bodied men are expected to work full time full year, whereas the

other groups have more socially acceptable alternatives to market work--keeping house and

raising children, going to school, and being retired. It may be the case that the other groups show

greater response to economic incentives because norms give them more options.

B. Macro Effects of Work and Training Programs

At least two kinds of macro effects may result from work and training programs. One--the

so-called displacement effect--has received a great deal of attention and has led the government

to sponsor a matched-site evaluation.[11]

The displacement effect is easiest to understand in the context of a work program. A new

job created by the government may be at the expense of another job. When an unemployed

person is paid to repaint buildings that have not been painted in years, a professional painter may

lose future work.

A micro-experimental test of a work program will overstate the effects of the program

because it cannot capture displacement effects. The micro-experimental evaluation measures the

effects on individual experimental subjects. It does not capture the effects on individuals outside

the purview of the experiment. On the other hand, a macro experiment, with communities rather

than individuals the focus of study, can capture displacement effects.

Training programs may also suffer from displacement effects. The simplest, but most

extreme, argument is that training does not lead to an improvement in productivity but merely

provides credentials that alter individuals' order in the job queue. If so, a micro experiment might

find differences between the earnings of the experimental and control groups even if there are no

social benefits to training. A less extreme argument is that the short-run difference in earnings

between the experimental and control groups overstates the long-run effects on productivity

because of ordinary supply and demand responses of markets.

The discussions of Sections II and III suggest a second macro effect of work and training programs, one implying that a micro experiment may underestimate the benefits of such programs. If work and training programs succeed in increasing the work and productivity of individuals in the community, then these programs may also lead to a change in norms with respect to work and training. This might set off the kind of reinforcing process we have described earlier. We can only conjecture about the magnitude of such a feedback effect.

V. MEASURING MACRO EFFECTS

We believe that the various macro effects discussed in this paper may well be important consequences of many social policy changes. The arguments we have presented seem to us forceful enough to shift the burden of proof to advocates of micro experiments. They must show that the macro effects of the program under evaluation are likely to be small.

Given the long history of concern with displacement, information diffusion, social interaction, and norm formation, we find it frustrating that those who carry out evaluations continue to act as if macro effects are negligible or, at most, to treat them as unquantifiable caveats to be placed on the findings from micro-experimental analyses. Progress will be made in understanding the magnitude and nature of macro effects only if we obtain appropriate data and apply suitable modes of analysis.

How, in fact, can we measure macro effects? To conclude this paper, we pose four alternatives. None of these is problem-free. Nevertheless, they all warrant consideration.

A. Macro Experimentation

In theory, macro experimentation would seem to provide the most effective means of data collection. Ideally, the unit of analysis in a macro experiment is a closed system, one in which

entry and exit is impossible. In practice, a macro experiment usually refers to an experiment in which the unit of analysis is a local site, perhaps a city or county. Some randomly selected sites are designated as experimentals and the policy is implemented comprehensively in these sites. Other randomly selected sites are designated as controls.[12]

Macro experiments can be difficult to implement. Harris (1985) observes that randomization of sites is problematic if communities must be voluntary participants in an experiment. Harris also notes that, given the high cost of administering experiments over entire communities, there are likely to be few sites in a macro experiment. As a result, it may be difficult to distinguish experimental effects from preexperimental idiosyncracies and from extraneous events that occur during the experiment (e.g., a plant closing).[13]

Some situations do lend themselves to macro-experimental evaluation. In particular, states frequently decide to implement new programs even when they have insufficient funds to implement the programs universally. The usual solution to this difficulty is to randomly select individuals who will be allowed to participate in the program. A valid alternative is to randomly select communities or counties in which the program will be implemented. For example, if the state determines that its funds allow only half of eligible individuals to participate in the program, then it can hold a lottery in which the program is implemented in the winning counties. If counties within a state are sufficiently similar, this strategy will allow evaluation of the effects of the program at the aggregate level.

B. Phased-in Experimentation

An alternative is an experimental design that combines aspects of both micro and macro experimentation. The experiment could start with two matched communities, one assigned as a control and one as an experimental community. The eventual goal would be saturation of the

experimental program in the experimental community. If saturation is gradual, then micro experimental evaluation could occur within the context of the macro experiment. At the beginning of the experiment, half of the individuals potentially eligible for the program might be . assigned randomly to a control group. The other half would constitute the initial experimental group. Several years later, the control group would be further subdivided, half of that half remaining in the control group and half being placed in a second experimental group. Several years after this subdivision, complete saturation could be effected.

The gradual saturation of the experimental community permits the collection of enough data to conduct a micro-experimental evaluation. A comparison of the responses of the original experimental group to those of the group that is phased-in later could provide information regarding macro effects. After saturation, the experimental and control communities could be compared again. If the differences between the first and second experimental groups and those between the experimental and control communities can be explained solely by responses predicted by the initial micro experiment, this would be evidence that macro effects were absent. On the other hand, if the differences could not be explained in this manner, then one would have a measure of the net magnitude of the macro effects.

C. Micro Experimentation with Cluster Sampling

Some kinds of macro effects, particularly social interaction and norm-formation effects, might be measured by conducting a micro experiment and observing the behavior of persons who live in the vicinity of persons participating in the experiment as experimentals and controls. In the absence of macro effects, the behavior of those sampled persons living near experimentals should not vary systematically from the behavior of those who live near controls. If macro effects do exist, one may in principle detect them in the differential behavior of the two sets of sampled

persons when the experiment is under way. It is not clear, however, whether the differences would be substantial enough for them to be detected in practice.

D. Analysis of Natural Variation

An obvious alternative to any kind of randomized experiment is to observe natural variation in policy across political jurisdictions. This method is already commonly used to study micro effects because it avoids much of the expense of experimental evaluation. It can readily be applied to study macro effects as well.

Analysis of natural variation sometimes seeks to mimic experimental conditions by matching sites that are observationally similar before the policy intervention and then attributing later differences to the change in policy. In the absence of matching, the evaluator must bring to bear prior knowledge that permits him to distinguish the impacts of differences between communities from the impacts induced by variations in policy.

Notes

[1]Hawthorne effects are behavioral changes that are induced by the process of observation rather than by the treatment under evaluation.

[2]Schelling (1973) presents a model of binary choice with externality. In this model, the choices of individuals affect the payoffs for other individuals. Although it may be optimal for all to choose one way when everyone else chooses that way, unanimity may not be Pareto efficient. If enough individuals make the opposite choice, then everyone may be better off than when all made the same choice.

[3]See U.S. Bureau of the Census (1982, Table 11). For the proportion dependent on welfare, see Garfinkel and McLanahan (1986, p. 138). Note that because of low earning capacity and high child care costs, most women on AFDC could not earn enough to lift their family out of poverty even if they worked full time.

[4]Some theoretical models of norm formation in sociology and economics include Opp (1979, 1982), Demsetz (1967), Akerlof (1980), and Axelrod (1986).

[5]Eighteen years would appear to be the maximum period for eligibility because CSAS is relevant only until the child reaches age 18.

[6]See Knudsen et al. (1977) for a discussion.

[7]CSAS increases the income of resident parents, which in turn leads to a decrease in labor supply. See Garfinkel et al. (1990).

[8]The lower guarantee applies to Wisconsin and the nation as a whole. In some states, however, a national assured benefit would exceed the current level of AFDC.

[9]See Moffitt (1983) for a model of the effects of stigma on welfare use.

[10]See Pencavel (1986) and Killingsworth and Heckman (1986). Both of these articles present elasticities from a number of studies. While estimates of uncompensated wage elasticities for men

range from -0.29 to 0.14, nearly half of the estimated elasticities for women are greater than 0.50. See Masters and Garfinkel (1977) for an estimation of elasticities for both men and women which uses common data.

[11]The Youth Incentive Entitlement Pilot Projects (YIEPP) were conducted in Baltimore, Cincinnati, Denver, and rural Mississippi with comparison sites in Louisville, Cleveland, Phoenix, and rural Mississippi. For an overview, see Gueron (1984).

[12]Local sites obviously are not closed systems, so migration into or away from experimental communities can be a problem in macro experimentation. The experimental intervention may promote inmigration to take advantage of the new benefit being offered. For example, the availability of CSAS in one Wisconsin county might provoke a single mother residing in a neighboring county without CSAS to relocate. Experimentally induced migration could provide biased estimates of the national costs.

Burtless and Orr (1986) and Harris (1985) observe that micro experiments are also subject to experimentally induced migration. It is likely, however, that the migration bias arising from a micro experiment will be less severe than the migration bias from a macro experiment. In a macro experiment, individuals can move from the control to the experimental group and vice versa, while in a micro experiment, attrition is a problem only if there are differences in attrition between the control and experimental groups.

[13]In a comment on Harris, Orr (1985) adds that, given a limited number of sites, the researcher must have a very specific experiment in mind. In contrast, in a micro experiment, many different treatments can be applied to different sections of the experimental group.

REFERENCES

Akerlof, George. 1980. "A Theory of Social Custom, of Which Unemployment May Be One Consequence." Quarterly Journal of Economics 94 (No. 4), 749-775.

Asch, Solomon E. 1952. Social Psychology. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Ashenfelter, Orley. 1987. "The Case for Evaluating Training Programs with Randomized Trials." Economics of Education Review 6 (No. 4), 333-338.

Axelrod, Robert. 1986. "An Evolutionary Approach to Norms." American Political Science Review 80 (No. 4), 1095-1111.

Burtless, Gary, and Larry L. Orr. 1986. "Are Classical Experiments Needed for Manpower Policy?" Journal of Human Resources 21 (No. 4), 606-639.

Crane, Jonathan. 1988. "An Epidemic Model of Social Problems in Ghettos." Unpublished manuscript, Center for Health and Human Resources Policy, John F. Kennedy School of Government, Cambridge, Mass.

Crutchfield, Richard S. 1968. "Conformity and Character." American Psychologist 10, 191-198.

Demsetz, H. 1967. Toward a Theory of Property Rights, reprinted in E. Furobotn and S. Pejovich, eds., The Economics of Property Rights, Cambridge, Mass.: Ballinger, 1974.

Fishman, Michael, and Daniel S. Weinberg. Forthcoming. "The Role of Evaluation in State Welfare Reform 'Waiver' Demonstrations." In Evaluating Welfare and Training Programs in the 1990s, ed. Irwin Garfinkel and Charles F. Manski.

Garfinkel, Irwin, and Marieka Klawitter. 1990. "The Effect of Routine Withholding of Child Support Collections." Journal of Policy Analysis and Management 9 (No. 2), 155-177.

Garfinkel, Irwin, and Sara McLanahan. 1986. Single Mothers and Their Children: A New American Dilemma. Washington, D.C.: Urban Institute Press.

Garfinkel, Irwin, Marygold Melli, et al. 1982. Child Support: Weaknesses of the Old and
Features of a Proposed New System. Institute for Research on Poverty Special Report #32,
University of Wisconsin-Madison. Report prepared for the Division of Economic Assistance,
Wisconsin Department of Health and Social Services, Madison, Wisconsin.

Garfinkel, Irwin, and Donald Oellerich. 1989. "Noncustodial Fathers' Ability to Pay Child
Support." Demography 26 (No. 1), 219-233.

Garfinkel, Irwin, Philip Robins, Patrick Wong, and Daniel Meyer. 1990. "The Wisconsin Child
Support Assistance System: Estimated Effects on Poverty, Labor Supply, and Costs,"
Journal of Human Resources 25, 1-31.

Granovetter, Mark. 1978. "Threshold Models of Collective Behavior." American Journal of
Sociology 83 (No. 6), 1420-1443.

Gueron, Judith. 1984. Lessons from a Job Guarantee: The Youth Incentive Entitlement Pilot
Projects. New York: Manpower Demonstration Research Corporation.

Harris, Jeffrey E. 1985. "Macroexperiments versus Microexperiments for Health Policy." In
Social Experimentation, ed. Jerry A. Hausman and David A. Wise. Chicago: University of
Chicago Press.

Heckman, James J., V. Joseph Hotz, and Marcelo Dabos. 1987. "Do We Need Experimental
Data to Evaluate the Impact of Manpower Training on Earnings?" Evaluation Review 11
(No. 4), 395-427.

Heckman, James J. Forthcoming. "Dynamic Selection Bias." In Evaluating Welfare and Training
Programs in the 1990s, ed. Irwin Garfinkel and Charles F. Manski.

Kiesler, Charles A. 1969. "Group Pressure and Conformity." In Experimental Social Psychology,
ed. Judson Mills. London: The MacMillan Company.

Killingsworth, Mark R., and James J. Heckman. 1986. "Female Labor Supply." In Handbook of

Labor Economics, Vol. 1, ed. Orley Ashenfelter and Richard Layard. New York: North-

Holland Publishing.

Knudsen, Jon Helge, John Mamer, Robert A. Scott, and Arnold R. Shore. 1977. "Information

Levels and Labor Response." In The New Jersey Income-Maintenance Experiment, Volume

II: Labor Supply Responses, ed. Harold W. Watts and Albert Rees. New York: Academic

Press.

Kurz, Mordecai, and R. G. Spiegelman. 1973. "Social Experimentation: A New Tool in

Economics and Policy Research." Research Memorandum 22, Stanford Research Institute,

Menlo Park, Calif.

Lalonde, Robert. 1984. "Evaluating the Econometric Evaluations of Training Programs with

Experimental Data." American Economic Review 76 (No. 4), 604-620.

Lalonde, Robert, and Rebecca Maynard. 1987. "How Precise Are Evaluations of Employment

and Training Programs: Evidence from a Field Experiment." Evaluation Review 11 (No. 4),

428-451.

Lamm, Helmut, and David G. Myers. 1978. "Group-Induced Polarization of Attitudes and

Behavior." Advances in Experimental Social Psychology 11, 145-195.

Lerman, Steven, and Charles F. Manski. 1982. "A Model of the Effect of Information Diffusion

on Travel Demand." Transportation Science 16 (No. 2), 171-191.

Manski, Charles F. 1990. "Dynamic Choice in a Social Setting." Social Systems Research

Institute Paper 9003, University of Wisconsin-Madison.

Masters, Stanley, and Irwin Garfinkel. 1977. Estimating the Labor Supply Effects of Income-

Maintenance Alternatives. New York: Academic Press.

Metcalf, Charles E. 1977. "Predicting the Effects of Permanent Programs from a Limited Duration Experiment." In The New Jersey Income-Maintenance Experiment, Volume III: Expenditures, Health, and Social Behavior, and the Quality of the Evidence, ed. Harold W. Watts and Albert Rees. New York: Academic Press.

Moffitt, Robert. 1983. "An Economic Model of Welfare Stigma." American Economic Review 73 (No. 5), 1023-1035.

Moffitt, Robert. Forthcoming. "Evaluation Methods for Program Entry Effects." In Evaluating Welfare and Training Programs in the 1990s, ed. Irwin Garfinkel and Charles F. Manski.

Myrdal, Gunnar. 1944. An American Dilemma. New York: Harper and Row.

Oellerich, Donald, Irwin Garfinkel, and Philip Robins. 1989. "Private Child Support: Current and Potential Impacts." Institute for Research on Poverty Discussion Paper No. 888-89, University of Wisconsin-Madison.

Opp, Karl-Dieter. 1979. "The Emergence and Effects of Social Norms: A Confrontation of Some Hypotheses of Sociology and Economics." Kyklos 32 (No. 4), 775-801.

Opp, Karl-Dieter. 1982. "The Evolutionary Emergence of Norms." British Journal of Social Psychology 21, 139-149.

Orr, Lawrence L. 1985. "Comment: Choosing between Macroexperiments and Microexperiments." In Social Experimentation, ed. Jerry A. Hausman and David A. Wise. Chicago: University of Chicago Press.

Pencavel, John. 1986. "Labor Supply of Men." In Handbook of Labor Economics, Vol. 1, ed. Orley Ashenfelter and Richard Layard. New York: North-Holland Publishing.

Sawhill, Isabel. 1976. "Discrimination and Poverty among Women Who Head Families." Signs 2, 201-211.

Schelling, Thomas. 1971. "Dynamic Models of Segregation." <u>Journal of Mathematical Sociology</u> 1, 143-186.

Schelling, Thomas. 1972. "A Process of Residential Segregation: Neighborhood Tipping." In <u>Racial Discrimination in Economic Life</u>, ed. A. Pascal. Lexington, Mass.: Heath.

Schelling, Thomas. 1973. "Hockey Helmets, Concealed Weapons, and Daylight Saving: A Study of Binary Choices with Externalities." <u>Journal of Conflict Resolution</u> 17, 381-428.

U.S. Bureau of the Census. 1982. <u>Characteristics of the Population below the Poverty Level: 1980,</u> Current Population Reports, Series P-60, No. 133. Washington, D.C.: U.S. Government Printing Office.

U.S. Bureau of the Census. 1987. <u>Child Support and Alimony: 1985.</u> Current Population Reports, Series P-23, No. 152. Washington, D.C.: U.S. Government Printing Office.

West, Charles K. 1981. <u>The Social and Psychological Distortion of Information.</u> Chicago: Nelson-Hall.