An Economist's View of the Income Maintenance Experiments

*Robert M. Solow**

I am cast on this panel as offering an economist's view of the policy lessons to be drawn from the income maintenance experiments. That will be true enough if you take the word "an" seriously. I am indubitably a card-carrying economist. But I have the feeling that — on these particular matters at least — my views are not always those of the typical economist. Part of the difference, but only part, arises because I am primarily a macroeconomist, so I look at the questions now at hand as a partial outsider.

Any time you are about to utter heresy it is a good idea to quote some highly respectable authority. So let me remind you of something John Stuart Mill wrote in the Preface to the *Principles of Political Economy.* "Except on matters of mere detail, there are perhaps no practical questions, even among those which approach nearest to the character of purely economical questions, which admit of being decided on economical premises alone." Actually, I intend to go even further than Mill: there are very few analytical questions about mass human behavior which admit of being decided on economical premises alone. And those few are not the ones we are dealing with at this conference.

The formal purpose of social experiments is to provide knowledge about mass human behavior that will be useful in the design of policies. So the lessons for policy depend on what we read the experiments as saying about behavior. In commenting on that issue, I am drawing also on my experience with other social experiments, in particular those dealing with supported work and with various work-welfare schemes, designed, conducted and analyzed by the Manpower Demonstration Research Corporation.

* Professor of Economics, Massachusetts Institute of Technology.

The surest generalization that emerges from all such experiments is that the implications for mass human behavior are weak. That does not mean they are not important; but they are uniformly weak. I use the non-technical word "weak" to cover a couple of distinct characteristics. In the first place, statistically significant response-coefficients are hard to come by. If there is any signal in the experimental results, it is rarely audible above the noise. This is hardly surprising; it is a very common outcome in cross-section studies with individuals as the unit of observation. No doubt it reflects both the inherent variability of each individual's behavior and the variation among individuals in their average response above and beyond what can possibly be related to observed and observable characteristics. Research workers usually get their kicks from large t- and F-statistics, so this general lack of statistical significance is usually a disappointment; but I think it has its policy uses, as you will see. It might sometimes be possible to design an experiment for adequate precision or power at some favored point in parameter space. But that is hardly ever what we really want.

Even when a statistically significant response-coefficient surfaces it is usually small. The elasticity of this with respect to that is rarely large for the thises and thats we are concerned with. To take two examples: (1) Gary Burtless shows convincingly that the labor-supply response to a realistic negative income tax is in the expected direction but fairly small; (2) in the supported-work experiments, even when a favorable post-program response-coefficient could be estimated, as with the sub-sample of mothers receiving aid to families with dependent children, it was far from dramatic. Gains in employment and earnings were small, but definite. I could document an analogous statement about the work-welfare experiments Manpower Demonstration Research Corporation is now doing. We are so accustomed to this sort of outcome that our first reaction to a large statistically significant elasticity is to say: Wait a minute; this must be spurious, the result of a misspecification. And almost always we can convince ourselves that this skeptical reaction is correct. I don't find this sort of outcome terribly discouraging either. We are talking about fairly commonplace aspects of behavior, not about responses to exotic stimuli or extreme situations. If sharp responses were to be expected we would already know about them; nobody spends millions of dollars to verify the obvious. (It goes without saying that sometimes common knowledge will be wrong and the obvious is not only not obvious, but false. That's life, but it does not undermine the reason I gave for expecting small responses.)

If I am right about this, that the typical outcome of a social experiment is some weak conclusions about response elasticities, it has an important implication. The prevalence of small effects opens the way to alternative interpretations of the research findings. The interpretation adopted will depend a lot on the interpreter's ideological and doctrinal preconceptions and only a little on the detailed experimental results themselves. It is the same principle that governs those personality tests in which you are shown a picture of an ambiguous scene and different people, interpreting the same picture, will tell you the story of their own lives. Thus I think Gary Burtless hit the nail right on the head when he concluded his survey of labor-supply responses by observing that policymakers "seem far more impressed by our certainty that the efficiency price of redistribution is positive than they are by the equally persuasive evidence that the price is small." Because the price is small and not well-defined, policymakers can find what they are looking for. We all know what they are looking for these days.

One obvious reason for the prevalence of weak results is Mill's dictum. These experiments do not take place in a test-tube and they do not involve identical individuals. There is just a lot more going on than can possibly be controlled. And many of those things are not even economical at all. I thought of this especially while reading Glen Cain's meticulous and carefully inconclusive evaluation of the well-known findings about marital instability. Of course current and prospective income under various circumstances is one of the forces pushing a family one way or another. But there are many others, probably more acute, impossible to measure and control for, and — this is the important point — very likely correlated with some of the things we do measure, but in complicated ways. Since those unmeasured forces necessarily get parked in the noise, it is no wonder that clear, comprehensible, and robust response-coefficients are hard to find in social experiments.

If the experiments usually offer little basis for discrimination among quite different interpretations of the outcome, do they therefore offer no lessons for policy? No, I don't think that is the right conclusion to draw. I think that the income maintenance experiments and others like them tell us a lot about social policy, though maybe not the things we are accustomed to look for. Once again it is Mill's dictum that points the way.

Why was there, back in the 1960s and 1970s, a brief flicker of interest in the negative income tax, intense enough to give rise to expensive experiments? It is hard to think back across the ethical desert that is American national government today, but I think I remember some of the reasons. There was a feeling that we were at last in a position to eliminate poverty, that it was the right thing to do, and that the direct way to do it was to transfer income directly to people who would otherwise be very poor. This was combined with a feeling that the existing hodgepodge of categorical transfer programs involved the bureaucracy deeply, meanly, and inefficiently in running the lives of the recipients. In particular, it was thought that the rules governing AFDC had the effect of inducing two-parent families to split up. This strand could be thought to fall under the second category of reasons, but it was prominent enough to be worth independent status. The negative income tax seemed to be a way of fixing all of these things at once, and doing it through a bureaucracy with which *everybody* was involved. It attracted people at both ends of the normal political spectrum. This may have been very important at the time. After all, we didn't get a negative income tax, just some experiments. The unusual sight of Milton Friedman and James Tobin agreeing (but not really) about something could help bring that about. I think Wildavsky's categorization of the political spectrum is too *ad hoc* to be fundamentally useful; but I also think he captures a lot of what was going on in this particular episode.

The one possible hitch was the fear that a decent guaranteed income level, combined with the high tax rates necessary to keep from transferring a lot of money to people above the poverty line, would induce many recipients to withdraw from work. That outcome would fit badly with a very important strand of the American ethic. The experiments were designed primarily to test that possibility. (At least so it seems to me. I may here be exhibiting the economist's occupational bias.)

If that is what the experiments were about, they did provide something of an answer. There is a labor supply effect, as every economist thought there would be; but it could hardly be described as large enough to jeopardize the work ethic. Besides, the inducement to withdraw from the labor force seemed to be stronger among women than among men; it might easily have weakened over time as women generally have become increasingly involved in the world of work. A culture that can pat itself on the back unceasingly after having gone seven years with an unemployment rate higher than 7 percent can hardly complain that its foundations are being eroded by so small a withdrawal from the labor force. There is no particular reason to suppose that a return of those lost souls to the labor force would have increased *employment* perceptibly under current conditions.

The possible effects on marital instability are some cause for alarm, though Glen Cain makes one wonder if they are real. No one would want the transfer mechanism to contribute to the breakup of marriages. By the way, is there any reason to connect any measured increase in marital splits with the negative income tax *mechanism* itself, as distinct from the change in current and expected disposable income? If the couples that did split had encountered the same income possibilities in the private market, would they have split up? There is no merit in insisting that the transfer mechanism keep couples together who want to break up for other reasons, and do so by inflicting unnecessary poverty on them. The negative income tax experiments may not have been causing anything, just telling us what was already there, but suppressed.

Maybe this view of the experiments provides a model for social experiments in general. Suppose society wants to do something because it is the right thing to do, not for purely economical reasons. There will sometimes be economical doubts, worries that doing the right thing might be unsuspectedly costly. A well-designed experiment can help find out, and the prevalence of weak results is not an obstacle. It only means that a lot of the time the experiment will tell us that, for all the reasons discussed earlier, the kinds of changes we contemplate will not turn the world upside down. Many of the most important things people do they do for reasons that have little to do with the price mechanism. I think that the great shift in educated opinion, away from belief in income maintenance-type solutions toward belief in multi-part tailored, work-related programs, had little to do with the outcome of the negative income tax experiments. The generic fact that results are weak merely permits the social consensus to work itself out and to convert the research community without undue strain on its conscience. It must also be remembered that the nature of the poverty problem seems to have changed during 20 years and that should change the preferred solutions.

I should add that experiments have the added advantage of providing some information about the system's capacity to administer and operate a new policy mechanism. This is important; it is something that the Manpower Demonstration Research Corporation has always paid a lot of attention to in its experiments and demonstrations.

The main disadvantage of the social experiment as a policy tool is that it may often leave us having to explain to ourselves why we do not do the right thing, when it is costly but not terribly costly. But that is not a serious problem either. One of the things Americans are best at is kidding themselves along.

Discussion

*Edward M. Gramlich**

I would have thought that this conference would commemorate the income maintenance experiments, but in fact the experiments have taken quite a beating. Yesterday Arnold Zellner made extensive criticisms of their statistical properties. Today Lee Rainwater criticized their initial conception, method, and policy interpretation. Elmore found the notion that policy experiments of this sort could be used to leverage policy support patently absurd. Nathan feels that neither the negative income tax nor the experiments with it were well-advised. Robert Solow's paper was also critical of the experiments. When he wrote it, he probably thought that his paper was being too critical, and he was quite apologetic. But compared to the others, Solow's paper is mild. To quote another famous social philosopher (Solow himself, in his 1974 Ely lecture), he must have felt "like a nice independent rat, trotting down to the sea, and suddenly discovering that he's a lemming."

Solow is only rarely a lemming, a feeling I have much more experience with. Today, however, I want to try to play the rat and stick up for the experiments. I begin with a standard Robert Hall proposed yesterday when he suggested that the experiments be compared to what the government was doing at the time. Hall compared the experiments to the Post Office, which seems like a cheap shot. But we could compare the experiments to other major social experiment efforts of the Office of Economic Opportunity (OEO) at the time. There were three altogether. One, on educational performace contracting, tested a policy that soon became viewed as a failure, using an experiment that was very flawed but ultimately arrived at useful "nail in the coffin" type data because of

*Acting Director, Congressional Budget Office, and Professor of Economics and Public Policy, University of Michigan.

the well-known inadvertent finding. That finding was that it became very hard for OEO to litigate the performance contracting claims with private companies, raising the ugly spectre of small school districts trying to fight it out with the legal departments of national teaching companies. Another, for educational vouchers, never even got off the ground because OEO could not find any school boards interested in conducting an experiment. The third, the negative income tax, was at the time viewed as the jewel in OEO's crown. The experiment was working, and the policy was interesting to many people, even some who were not economists. In this sense, while we can look back, rub our beards, and pronounce the negative income tax a failure, we should give the experiment its due — unlike the others, it at least lived on to get roasted 20 years later.

But the experiment did more than that. There are basically two reasons why many are now pronouncing the experiment a failure:

1) the research was inconclusive;
2) interest in the policy under investigation waned.

Yesterday we discussed the first reason extensively, and I take it by now the consensus is that not *all* the research was inconclusive, and that it is pretty hard even to imagine what conclusive research would be on matters such as marital splits. Today we are talking about the second issue, one on which I would like to spend most of my time.

The reason why interest in the negative income tax as a policy option has waned also brings up Solow's remark about the "ethical desert that is the American national government today." (Since I am now part of that national government, this remark does put me in an awkward position.) Back in 1970, and probably also today, Solow would have voted for a pure negative income tax without any work requirement because "it was the right thing to do." A decent negative income tax would cost about $20 billion, $100 per capita, about one-fifth of what the average American spends on voluntary charity these days. Most people give to charity without any effective monitoring for leaky buckets, and one would think we could do the same for income transfers. Solow appears to adopt this line of reasoning when he calls income maintenance the right thing to do. I certainly agreed with him back in 1970, and would still today but for another problem that I was not then very sensitive to: unfortunately, a pretty basic problem.

Theorists rationalize the existence of income transfers in the first place on the basis of taxpayer-donors' altruistic motive. But any altruist knows that one important thing to worry about is free riding: if there is some other altruist willing to make support payments, an altruist can free ride, see the donation given, and save his money.

It takes just a slight extension of this logic to realize that with income support payments, there is a potential free rider — the recipient himself.

It may be that free riding is minor, that it is not costly to the donor, and that leaks in the bucket are modest. But if free riding exists at all, it can be tremendously destructive. The donor-taxpayer is working a little bit harder to support a poor person, and here we have the poor person working a little bit less. Any donor-taxpayer, even a very ethical one, is likely to become so enraged by such an outcome that he may cut off all support. And even if he does not, his democratically elected representative might be inclined to. And even if that representative does not, he may have to run against an opponent who could make hay out of the issue. When all is said and done, income support under those terms will probably be very limited.

This, I think, is what blew the pure negative income tax out of the water politically. Unless donors could *assure* themselves that responsibility for supporting recipients would be shared by recipients themselves, in the form of work requirements, child support enforcement, and other things that sounded punitive back in 1970, they are simply not interested. And it is not even unethical for them to be uninterested. This is why, I think, Nathan tells us that politicians did not like the pure negative income tax even in 1970, and Burtless tells us that politicians are more impressed by the positive price than by the small price. And this is why I also think we should stop talking about ethical deserts. They may not be ethical deserts after all, and whatever they are, they have been around for a while.

I should point out that by responsibility-sharing, I do mean sharing. Unlike Charles Murray, I do not think that poor people need to provide *all* of their own support. There can be altruism, and it can be highly satisfying to donors as long as donors perceive that poor people are doing their share. And such a feeling is probably best for the long-run self-esteem of the poor too, another important value that should not be ignored.

Unlike all pessimists who are down on both the negative income tax and experiments with it, this notion of responsibility-sharing shows me a silver lining in all these gray clouds surrounding the world of income support these days. It is just possible that national policy, nudged along by state governments (the true social laboratories) and even unwittingly by the Reagan administration (mainly in the form of enabling legislation passed in 1981), is evolving a successful income support strategy that contains heavy doses of responsibility-sharing. States are more and more requiring welfare recipients to work or search for jobs as a condition of getting benefits. They are also enforcing child support obligations on absent fathers. Both make sense in terms of responsibility-sharing; the first may even pass the Manpower Demonstration Research Corporation's benefit-cost test, and evidence indicates that it is viewed as fair by recipients.

Once the perception that welfare involves responsibility-sharing takes hold, we may even see governors begin to bring welfare out of the closet and brag about their "humane and responsible" approaches to the problem. One such governor is in the state just south of us, Dukakis of Massachusetts. And with governors and other politicians talking up this new view of public assistance, these programs may begin to grow in popularity, and real benefit levels may even stop losing ground.

There is a silver lining for social experiments too. The Manpower Demonstration Research Corporation's interesting and successful evaluations of employment for welfare recipients have somehow managed to take root in this desert. They have benefited technically and administratively from earlier experiments with the negative income tax, and they now have the huge political advantage that their policies under investigation seem to be popular and have stayed so, seemingly because they have worked out the sharing of responsibility more satisfactorily. To respond to those who proposed retrospectively a resident anthropologist, we now even have at least one new large-scale project, by Bill Wilson at Chicago, that combines survey and ethnographic work. Hence for all the mistakes made by the earlier generation of social experimenters, and all the bad political luck that was suffered, we might even be getting another chance. Let's hope things work out better this second time around.