The Policy Laboratory: The Income Maintenance Experiments and Trial-and-Error in American Public Policy

Connor Stangler
2018-2019 Ostrom Workshop Fellow

Abstract: The Income Maintenance Experiments (IMEs) of the late 1960s and 1970s were the first large-scale social experiments in American policy history. Using federal funds, two research institutes tested the effect of a guaranteed annual income on several neighborhoods — urban and rural — across the United States, measuring divorce rates, happiness, rates, birth rate, and workforce participations. Following the IMEs, policy experimentation expanded significantly among American policy circles. This paper explores the logistics and results of those experiments and poses two questions: (1) what effect did those experiments have on actual policy programs, and (2) how did these first experiments both limit and initiate a new movement in policy experimentation during the final twenty-five years of the twentieth century? While the IMEs informed the content of Richard Nixon’s Family Assistance Plan, their shortcomings and design flaws undermined policy scientists’ interests in testing radically new policy alternatives.
The purpose of this project is to investigate the origins and effects of social experimentalism in the American social sciences and public policy from the 1950s to the 1970s. Leaders and supporters of New Frontier and Great Society liberalism fostered policy experimentalism, but increasingly scientific and quantitative methods called into question some of the original assumptions of their liberal policy goals. I am interested in three central questions. First, why did economists and policymakers come to embrace the experiment as a method for making claims to social truth? Second, how did the experimental form change and reify certain perspectives of anti-poverty policy and welfare? Finally, did experimentalism ultimately compromise and narrow community- and grassroots-oriented liberalism as social scientists acquired more intellectual and political capital? This paper is part of a larger project that will become my dissertation.

Experimentation as a social scientific method began in the behavioral sciences but eventually took root in public policy. Large-scale social experimentation started with the Income Maintenance Experiments (IMEs) of the 1960s and became a standard tool for policy scientists in the 1970s. Indeed, according to later social scientists, the 1970s were the “Decade of the Social Experiments.” Why did policymakers come to embrace experimentation? It should not have become as popular as it did. Economists initially doubted its efficacy, preferring to believe in thoroughly rational models unmarred by empirical data on actual human behavior. Moreover, policy experiments were not demonstration programs, which were supposed to popularize a new and workable idea. Experiments, on the other hand, left the results an open question. Political uncertainty did not make for convincing rhetoric.

Beginning in the late 1960s, psychologist Donald T. Campbell started making calls for social scientists to become “methodological servants” in the “experimenting society.” They needed to serve a political purpose, not as partisan hawker or even bureaucrat, but as directors of the democratic laboratory and emissaries of the democratically elected lawmakers. “The job of the methodologist for the experimenting society,” Campbell wrote, “is not to say what is to be done, but rather to say what has been done.” Campbell’s references to experiments were not metaphorical, a la the rhetorical tradition of the “American experiment” or states as the “laboratories of democracy.” Rather, Campbell called for actual
experiments wherein American citizens were the subjects, their economic conditions the variables.

Lawmakers, according to Campbell, would not promise political panaceas. Rather, these experimental programs would be explicitly amendable and unfinished. In short, “we try out new programs designed to cure specific social problems, in which we learn whether or not these programs are effective, and in which we retain, imitate, modify, or discard them on the basis of apparent effectiveness on the multiple imperfect criteria available.” The goal would be to form a more scientifically perfect union through the evaluation of empirical data culled from live, randomized experiments.1

Campbell’s vision was not preordained. Postwar economics did not rely on experimentation for verification. The field was torn between two camps: empirical (or as some scholars refer to it, “applied”) and theoretical. Though at the time economics was not as mathematically advanced as it would become later in the twentieth century, some of its practitioners, such as Wesley Clair Mitchell and his colleagues at the National Bureau of Economic Research, preached the necessity of data. These economists used empirical measurements to gauge the relationship among wages, employment, and, on the macroeconomic side, public investment. Theorists, such as Kenneth Arrow and Milton Friedman, believed rational, positivist inferences provided the most accurate insights into economic behavior, not data. Their “profit-maximizing theories of the firm” relied on assumptions of universality, on the primacy of the thoroughly rational homo economicus. Though Friedman was not averse to data, he prioritized theory, as one of his more famous metaphors in defense of rational models makes clear.2

2 Beatrice Cherrier, and Roger Backhouse. “The Age of the Applied Economist: The Transformation of Economics since the 1970s,” History of Political Economy 49 (2017), 10; Friedman’s pool player analogy compares an expert billiards play to an optimizing firm. Over time and over the course of lots of individual events, the billiards player and the firm optimize not because they are capable of perfectly mathematical calculation but because they act as if they are. In Friedman’s words, “Our confidence in this hypothesis is not based on the belief that billiard players, even expert ones, can or do go through the process described; it derives rather from the belief that, unless in some way or other they were capable of reaching essentially the same result, they would not in fact be expert billiard players.” Friedman, “The Methodology of Positive Economics,” in Essays in Positive Economics (Chicago: University of Chicago Press, 1953), 15. Much later, the behavioral economist and “nudge” proponent Richard Thaler used the pool player analogy to argue for economic theory’s irrelevance: “My main point was that economics is supposed to be a theory of everyone, not only experts. An expert billiard player might play as if he knows all the relevant geometry and physics, but the typical bar player usually aims at the ball closest to a pocket and shoots,
Even at the height of theoretical economics’ reign in the 1960s, empirical methods were growing in popularity. One iteration of this expanding empiricism was experimental economics. Experimental economics involved both randomized control and naturally occurring experiments. Like in the physical sciences, experiments provided an empirical litmus test. They proved “that one can learn a lot about the world without imposing optimizing models.” Experimental economics did not settle the applied vs. theoretical debate. While it became an applied field, economists disagreed over whether it confirmed or muddied rational actor theory. Proponents of game theory, for example, clung fiercely to rationalist frameworks. Game theorists could not bother with the prospect of divergent and idiosyncratic outcomes.  

Experimental economics would not have become a fixture of the “policy toolkit” without government funding, however. The federal government has been a patron and producer of the social sciences since the late nineteenth century. Of all the social sciences, economics has been the greatest beneficiary of federal investment and trust. Economists proved vital not only to questions of industrialism and labor but also foreign affairs and immigration. Early in the twentieth century, the economist Jeremiah Jenks served on the Dillingham Commission, the investigatory body that helped make immigration statistically legible. Both world wars elevated the status of economists. The social scientists’ theories and research complemented and justified a growing federal role in the economy. Wartime planning relied on the statistical perspective of social scientists, as well as business leaders. Even prior to the massive mobilization of World War II, former engineer Herbert Hoover demanded statistics such as unemployment rates in order to project how federal programs might alter the economy. New Deal planners such as Raymond Moley used tables, graphs, and surveys to render complex economic problems managerial ones.

---

The statistical resources available at the time pale in comparison to the data and studies generated by every federal agency today, but as the government expanded its reach and ambition, its officials sought the increasing quantification of all aspects of social, economic, and political life. Social scientific experts provided that quantification, but their alliance with politicians was not free of suspicion. In 1922, even as social scientists provided more data than ever before to lawmakers, critic Walter Lippmann wrote that the new division of policy labor rankled some lawmakers. The latter were “slow to call in the social scientists.” Unwittingly presaging the experimental society, Lippmann argued that social scientists did not have the ability to foresee the consequences of their suggestions. As a result, the lawmaker, “observing that the social scientists knows only from the outside what he knows…from the inside, recognizing that the social scientist’s hypothesis is not…susceptible of laboratory proof, and that verification is possible only in the ‘real’ world, has developed a rather low opinion of social scientists.” It is not clear whether Don Campbell and others had Lippmann in mind when they called for policy experimentation, but they promised verification in the “real” world through randomized selection.  

Despite the success of social scientists in government, their role was neither safe nor explicitly defined. Don Campbell was not the only social scientist during the 1960s wondering about the role of his profession in politics and policymaking. In his 1964 presidential address to the American Economic Association, George Stigler addressed the role of economists in policymaking up until that point. Confronting the debate between theory and empiricism head on, Stigler mourned the paucity of data on effective public policy. “From 1776 to 1964 the chief instrument of empirical demonstration on the economic competence of the state has been the telling anecdote.” Economic theories could not and did not by themselves guide the direction of policy; rather, they described “general relationships, and which part of a theory is decisive in a particular context is a matter of empirical evidence.”

---

6 Though Stigler’s modesty about economists’ historical sway was admittedly overstated. As he said, “Lacking real expertise, and lacking also evangelical ardor, the economist has had little influence upon the evolution of economic policy.” George Stigler, “The Economist and the State,” 11-12, 13.
The ‘uneasy partnership’ between the social sciences and the federal government grew more intertwined over the course of the century as administrative agencies multiplied and policymakers demanded more data. A triangular exchange developed among agencies, universities, and foundations. These partnerships, however, did not mean that all politicians and bureaucrats endorsed all means of scientific investigation. Political actors wanted, first and foremost, concrete and usable knowledge. Experimentation did not grant them that. Systematic analysis posed a threat to the intuition of the policymaker. At the same time, however, experimentation did promise eventual predictive capabilities. Policy scientists and economists interested in experimentation did not seek empirical anomalies. As explained above, some social scientists believed experimental design confirmed rational action while others argued it belied it. Either way, all sought repeatable and demonstrable rules applicable to policy designs. As Richard Thaler details in his account of his discovery of behavioral psychology, experimentation was a peek into recurring human action. The “big idea” of Daniel Kahneman and Amoos Tversky, the “one that made my hands shake,” was that “using…heuristics causes people to make predictable errors.” If errors were predictable, policymakers could anticipate them and plan accordingly. If economic actors were predictably irrational, then policies could help correct those mistakes.7

Policy experimentation as outlined by behavioral economists differed from past state attempts to be experimental or innovative. In 1915, the Bureau of Labor Statistics used a natural experiment — wages in Portland, Oregon that differed from the rest of the state — to assess the effect of minimum-wage laws on women’s labor force participation. Studies like these were early versions of “policy evaluation,” a practice that emerged concurrently with policy experimentation and the growth of the policy science profession starting in the 1950s. Policy evaluation entailed a before-and-after assessment of the introduction of a policy, sometimes using a control group and the treatment group. Professionally, policy evaluation emerged from the systems analysis revolution in the defense industry in the 1950s. The task of the evaluators was research and planning, which was distinct and independent of policy operations.

7 Lyons, 5; Thaler, 22.
Lyndon Johnson institutionalized systems analysis in non-defense offices in 1965, when he announced that the planning-programming-budgeting system (PPBS) would become the standard policy planning procedure throughout the executive branch. Modern evaluation formally leapt from defense to social policy that same year, when the Department of Health, Education and Welfare created the Office of the Assistant Secretary for Planning and Evaluation.

Evaluation did not mature quickly. In cases like the Portland study, data was often of dubious quality. There were too few women to evaluate, “and the time for adjustments…was too short to allow the results of the study to do more than show tendencies.” This problem would plague policy evaluation well into the 1970s. Major legislative programs, such as the New Deal, included some of the language of experimentation and evaluation, but not as systematically as Johnson-era administrators would use it. Progressive Era settlement houses and New Deal rural demonstration programs were attempts to reimagine the social order, whether to improve Americans’ moral conditions or their economic. Observers and critics called these programs “experimental” mainly because they were new. Even the tax cut that Walter Heller, chairman of the Council of Economic Advisors for Kennedy and Johnson, desired was called “one of the grandest experiments in social theory of this century.” The coming of the IMEs and other large social experiments represented a new attempt to connect policy intervention to social causation. Historians have argued that demands for better policy evaluation rather than any desire to “emulate medical science” drove social experimentation.⁸

The change in the nature and scope of experimentalism during the 1960s and 1970s represented a shift in social scientists’ expectations for a dramatic reshuffling of Americas’ system of social relations. The empowering of local residents through the Community Action Programs (CAPs) gave way to the more conventional power structures of Model Cities and the scientific incrementalism of the Income Maintenance Experiments (IMEs). The transition from spirit to method, from theory to practice, from

---

experimentalism to experiments, encapsulated American social policy’s increasing timidity and exactness in the second half of the twentieth century. Policymakers’ demand for more empirical data, in place of a checkered record of policy heavy on rhetoric but light on rigor, not only produced a policy paradigm that invested more in tweaks and nudges than fundamental shifts but also the consummation of the quantitative social sciences’ takeover of social policy in the U.S. Similar to the developments in systems analysis that took place in the Department of Defense in the 1950s and 1960s, as well as rise of cost-benefit accounting in the Department of the Interior and the U.S. Army Corps of Engineers in the 1930s, think tanks and policy experts installed objective evaluation as the standard arbiter of social policy.9

The IMEs were not like community action. They represented a significant maturation of commitment to planned, scientific experimentation and data-gathering. In many ways, the late Great Society was a reaction to the early one. The latter symbolized what Alice Rivlin, Assistant Secretary for Planning and Evaluation at Lyndon Johnson’s Department of Health, Education and Welfare, called “random innovation.” Since the decade turned, Johnson and, mostly, Kennedy, had thrown themselves into anything that might work. “Experiment” had been used as a flattering synonym for “new” or “innovative.” The Great Society had been an admirable if recklessly optimistic attempt at new strategies to improve lives. According to Rivlin, there was a point “at which the random innovation model breaks down.” The problem was that there were no postmortems. Policy analysts had to try something, measure its results, recalibrate, then try again. As Rivlin wrote in the early 1970s, “This is not an argument for less random innovation. Indeed, we need more of it. If we are to understand the healing process or the learning process, there is no substitute for support of undirected basic research by creative people who are permitted to follow where their instincts lead them.” Systematic experimentation and evaluation had to replace the “spirit” that had pervaded the 1960s.10

---

10 Alice Rivlin, Systematic Thinking for Social Action, (Washington: Brookings Institution, 1971), p. 73, 76. Throughout the era of social experimentation, economists and policy analysts have used a strict definition of “social experiment.” The central qualifier is “random assignment” of “treatment.” The 1983 final report of the Seattle-Denver Income Maintenance Experiments (SIME-DIME) defines “social experiment” as “a field test in the sense
The research question remains: why did the federal government endorse the IMEs? Some historians point to the increasing demand for policy evaluation as one reason for the growth of experimental studies, in addition to the “transparency” of rigorous empirical tests. Still, experiments did not promise perfect evaluation. Stigler, in his presidential address, remarked that economists and policymakers had been understandably reluctant to test theory in practice “because its implementation would pose large social or administrative costs.” As it turned out, the IMEs confirmed Stigler’s worries. Bureaucrats, economists, and employees tasked with collecting data over the course of the experiments all debated the process and worth of the IMEs. At the same time, economists engaged in a broader debate over what was “credible” evidence and whether that credibility would convince policymakers. Experiments risked failure and regression in quality of life. Communities and social scientists, not to mention lawmakers who answered to constituents, did not always embrace methods that appeared alternative, for “the map of experiments within the United States is partly the map of tolerance to policy innovation and social research.” No economic or national security crisis catalyzed the experimental society, but a willingness to embrace its methods spread out from the social sciences into American communities.11

Statisticians and policymakers designed the IMEs to have targeted subjects and variables. The New Jersey and Pennsylvania experiment was the first IME, and it was the most limited in scope and, thus, had the most statistically helpful outcomes. It restricted eligible families to those who had two or more members and who had a male head of household between the ages of 18 and 58. The Rural Income

that families or individuals are actually enrolled in a pilot social program offering some type of special benefit or service. It is ‘experimental’ in the sense that families or individuals are enrolled in each of the tested programs on the basis of a random assignment process, for example, the flip of a coin. To draw conclusions about the effects of the treatment, it is necessary to collect information about the people who are enrolled in each experimental program and about those who receive no special treatment (called the control group), and then to compared them on the basis of the collected information.” “Overview of the Final Report of the Seattle-Denver Income Maintenance Experiment,” Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, May 1983. https://aspe.hhs.gov/report/overview-final-report-seattle-denver-income-maintenance-experiment.

Maintenance Experiments were designed to adjust some of the errors and gaps in the original New Jersey experiment. The Rural IME included sites in Iowa and North Carolina. All communities had fewer than 2,500 residents. For this experiment, restrictions on participants were intentionally looser than those of the New Jersey experiment. It included households with a single resident, as well as those headed by women and residents over the age of 58. Variable choices such as these reflected the realities of rural America. Residents in rural areas were generally older than those in urban areas, and they suffered a different type of poverty. There were fewer job centers from which to choose. Agricultural ruin came in waves. Over the course of the century, rural areas had experienced severe outmigration. Policymakers wanted to know if something like the basic income could stem the tide of the “brain drain.”

Advocates of the negative income tax hoped that the experiments would prove that such a policy would increase the economic stability of families. Critics of the current welfare system pointed out that Aid to Families with Dependent Children (AFDC) perversely incentivized divorce since single mothers warranted greater transfers than if the father remained in the home. To critics of welfare, this trend captured the perfidy of complex social programs — and, for some, the poor themselves. A basic income payment that was uncoupled from marital status would not incentivize dissolution. Those same critics of programs like AFDC also favored two-parent households. They tended to favor behavioral over social and structural explanations of poverty. According to those critics, the state mistakenly intervened in family life and subsequently erected foolish incentive structures that undermined individuals’ abilities to work their way out of poverty. Supporters disagreed, saying that programs such as AFDC were necessary to address more fundamental problems with economic power structures.

Unfortunately, the results of the experiments disappointed welfare critics. Even though the designers of the experiment did not list family solubility as a central variable in their program, the administrators of the experiment collected data on several aspects of family life under the conditions of the basic income. In fact, according to Glen G. Cain, “the most influential research finding of the

---

experiments turned out to be not about labor supply but about marital stability, a family issue.”

During the Seattle-Denver IME, divorces increased among participating families compared to control groups. If these results were to be trusted, not only would policymakers expect divorces to increase if the negative income tax was applied to a wider area, such as a nation, but welfare rolls would also increase as a result since there would be more single mothers. The reasons for the increased divorce rates remain a mystery. Glen Cain hypothesized that because women who participated in the experiment, and then divorced, had the choice between the higher of either AFDC or basic income payments. Still, if partners stayed together over the course of the experiment, they would both receive basic income payments, which would seem to encourage marital stability. Other economists suggested that because two-parent families are more stable than one-parent families, the negative income tax was a greater subsidy to the single parent. Parents who remained married did not feel the positive effects of a basic income payment as much as a single parent did.

The social scientists who hoped the experiments would provide firm data on economic behavior encountered several obstacles. American communities, after all, were not natural laboratories. They had to be made into laboratories. This turned out to be challenging, from the perspective of administration, human subjects, and conclusions drawn from the data. Misreporting, experimental subjects suddenly dropping out, and political graft formed a layer of gritty reality beneath the veneer of scientific method. Scandal plagued the New Jersey IME. Project Manager David N. Kershaw was investigated by a grand jury for illegal welfare payments. An anonymous IME employee said that a Jersey City official stopped by one of the project’s offices and reportedly offered to “take care” of staff parking tickets if the project had more funds it could provide a couple of families. After the completion of the experiments, the data offered no clear or persuasive conclusions. Public policy professor David T. Ellwood admitted that, after

14 David T. Ellwood, 94.
looking at the data, “I come away mostly frustrated, unable to say anything but the most equivocal statements.”

Such stories represent the difficulties the experiments’ staff and architects had in turning the crooked timber of American society into sterile and constant lab variables. They also enrich our understanding of why those advocating for the experimental society wanted to build it in the first place. Social scientists had tired of politicians praising policy results that only benefitted them, of governing bodies that acted decisively on the merits of small and deceptive data samples. They wanted their methods and their results to be the lingua franca of political decision-making. In some respects, they were successful. Randomized experiments became a standard policy research tool. The IMEs became the object of many retrospective studies, and not just those dedicated to the subject of the universal basic income. The IMEs opened up the possibility of tweaking human behavior and readings the results.

The use of experiments for the purposes of public policy raises a question of political ethics: is the use of experiments to test life-changing policies reconcilable with democracy? On the one hand, all public policies risk detriment to some groups. Client politics benefit certain groups, and even well-researched and empirically grounded policies risk ruin. On the other hand, experiments use outcomes as means to an end, even if those outcomes are deleterious for the human subjects. The purpose of experimenting with policy alternatives is to find workable solutions, but if the design is scientific, then the results are undetermined prior to implementation. What if there is a bad outcome? This was the case with the IMEs. As described above, divorces increased. Concerns over the potential effects of divorce were not limited to Republicans, traditionally the party most concerned with family stability. After the experiments, social scientists wondered about the effect of the increased divorces on the children of those families participating in the experiment. For them, the experiment provided valuable insight into whether the negative income tax could ultimately harm children if it increased divorces over time. The experimenters demurred, however, “since the difficult question may require many years to elapse before it

---

can be answered.” Nevertheless, since marital dissolutions increased among experimental participants, and because “attrition rates were higher for control wives than experimental wives in the Seattle-Denver experiment,” it would seem that, for the duration of the experiment, the structure of the program incentivized divorce and whatever spillover effects it had on children.¹⁷

The policy proposal that emerged from the experiments and the broader policy discussions about a basic income was Richard Nixon’s Family Assistance Plan. That such a policy came from a Republican presidential administration or was even a serious piece of legislation in the first place, surprises most contemporary scholars. The plan was a significant departure from existing welfare and anti-poverty programs. There were no extant analogs for the program, hence the need to experiment similar programs, but there was substantial support among economists. The economist Milton Friedman was the most prominent advocate of a negative income tax and basic income. He supported a simplified cash transfer program over one that distinguished among different groups. Milton, along with 1200 other economists, signed a pro-negative income tax petition in 1968, just when policymakers were coalescing around the feasibility of such a plan. Economists in support of the plan suspected government bureaucrats more than they disapproved of “handing out” welfare to the “undeserving” poor. “Guaranteed income was seen as a technically superior mechanism; a simple, efficient visible transfer would replace a myriad of complex programs with their hidden costs.” The simplicity of the program pleased economists and egalitarian-minded pols who detested the perverse incentives of AFDC.¹⁸

Nixon’s Family Assistance Plan was not a pure universal basic income, however. Nixon announced it to the public in a televised address on August 8, 1969. The primary theme of the address was not vicious poverty but rather the moral injustice of the current system that rewarded not working. During the 1968 presidential election, Nixon proved willing to ally with voters who appreciated “hard work” and suspected minority communities, as the latter were most closely identified with welfare. In his

¹⁷ Glen G. Cain, 66.
¹⁸
address to the nation, Nixon valorized the hard-working American rather than highlighting the
pervasiveness of poverty, as liberals such as Lyndon Johnson and Michael Harrington did.

The present system often makes it possible to receive more money on welfare than on a low-paying
job. This creates an incentive not to work, and it also is unfair to the working poor. It is morally
wrong for a family that is working to try to make ends meet to receive less than a family across the
street on welfare. This has been bitterly resented by the man who works, and rightly so – the rewards
are just the opposite of what they should be. Its effect is to draw people off payrolls and onto welfare
rolls – just the opposite of what government should be doing. To put it bluntly and simply – any
system which makes it more profitable for a man not to work than to work, or which encourages a
man to desert his family rather than to stay with his family, is wrong and indefensible.19

The Family Assistance Plan’s primary purpose would be to solve the complex welfare system, as long as
it had the intended effect of re-wiring the economic incentives of transfers. The details of FAP reflected
Nixon’s concerns with work. The Plan would have included a guaranteed annual income of $1,600 that
decreased by 50% of earned income until a break-even point. Significantly, the plan also included work
requirements, which were anathema to welfare rights organizations. The latter rejected the FAP on the
basis that it was not enough and that it tied transfers to work. Left opposition to work guarantees and right
opposition to cash handouts combined to sink the FAP after it passed the House of Representatives.

The FAP tested the relationship between policy and the experiment as policy tool. The New
Jersey and Pennsylvania IMEs were not yet complete, and their data had not yet been analyzed, before
Nixon used it for his own purposes. “Well before it was completed, a President had embraced its
principles and hoped-for conclusions. … Inevitably, there arose a conflict between the methodological
demands of social science and the political needs of Congress and the Administration, and perhaps just as

19 Richard Nixon Foundation, “President Nixon’s Family Assistance Plan.”
https://www.nixonfoundation.org/2014/06/family-assistance-plan-families-can-succeed/
inevitably, the latter won out.” The President’s staff members raided (or in Moynihan’s terms, “broke into”) the early returns for anything that would bolster their case.20

Nixon’s Office of Economic Opportunity issued its report on the experiment, citing as evidence that the FAP would be a success on a national scale, despite the doubts of some economists. “There is no evidence that work effort decline among those receiving income support payments. On the contrary, there is an indication that the work effort of participants receiving payments increased relative to the work effort of those not receiving payments.” 21 Even members of Congress in support of the plan jumped on the early experimental results. Ways and Means Committee Chairman Wilbur Mills, a supporter of FAP, said on the floor of the House that “[the New Jersey experimenters’] final report will indicate the success of that experiment.”22 Unless Mills had access to the final report, statements like these undermined the intention of policy experiments. The experiment began with a hypothesis, not a conclusion. Don Campbell had advocated an experimental society in order to clarify political issues and evaluate the effectiveness of programs, but his recommended methods were used to project politically favorable outcomes. Later, the Office of Economic Opportunity would revise its initial claims about how much the experiment had proven that work effort had not gone down among participants. Instead, the office reported, “We have not picked up any precipitous decline in work effort. That is the major crux of our findings.” OEO Assistant Director John Wilson “emphasized the positive by claiming that ‘low-income people are strongly work motivated,’ basing that assertion not on behavioral evidence, but on an opinion survey of recipients.”23

The Family Assistance Plan failed because of staunch political opposition, but it was the first piece of legislation partly informed by a large, randomized policy experiment. From the late 1960s on, experimentation would become a standard policy tool. Soon after the IMEs, the federal government

22 Quoted in Coyle and Wildavsky, in Munnell, “Lessons from the Income Maintenance Experiments,” (179)
would fund an education experiment that was billed as the successor to Head Start, the early education program that started under President Lyndon Johnson. In a 1982 article, Edward Zigler, one of the architects of Head Start, directly invoked Donald Campbell on the “experimenting society,” calling Head Start “an exemplar of the best of what Campbell was proposing.” The “viability” of Head Start “has allowed it to serve as a stable based from which to experiment with a number of services and methods.”

While Head Start more aligned with the experimenting spirit of the early 1960s, Project Follow Through (PFT) represented a more social scientific attempt to objectify students as subjects. Originally a service program intended to expand Head Start, limited funding forced the OEO to make PFT one of the largest educational social experiments in U.S. history. Classrooms as “laboratories” dated back to John Dewey’s school at the University of Chicago, the first lab school in the nation. This attempt at making a laboratory, however, differed from earlier education innovation. The designers of the program envisioned PFT as part of community action but opted for research experimentation instead, though the initiative differed from the IMEs in that school districts chose from a menu of innovations. Thus, the experiments were not “random innovation” but “planned.” The experiment included 84,000 children as its subjects at its peak. Head Start and Project Follow Through represented a midway point between the aimless random experimentalism of the early 1960s and the empirical interests of the IME. As an OEO program, it retained an element of advocacy. The experiment also affords scholars another opportunity to re-periodize 1960s social reform, as PFT remained a large experiment until the mid-1970s, even continuing as a much smaller, and underfunded, project well into the 1990s.24

At the 1986 conference on the Income Maintenance Experiment hosted by the Federal Reserve Bank of Boston, Richard P. Nathan, a public policy scholar, wondered why the IMEs occurred in the first place. “Were they a delaying tactic supported by political officials who resisted this approach to welfare reform? Or were they an effort on the part of proponents of a negative income tax to put their idea on the

agenda and prove that it would work? I think it was the latter.”25 This definitely is a possibility. The negative income tax was a popular policy idea among those economists and policy scientists who worked in the Johnson and Nixon administrations and worked at the think tanks that supplied the intellectual labor. However, we should also consider whether the experiment functioned as a legitimating tool for economists. Considering the vociferous debates surrounding the significance of economic theory versus data, the experiment represented a new and uncertain future for policy evaluation. Social experimentation emerged at a time when bureaucrats and policy experts demanded more, and more precise, data.

Additionally, experimentation did not end with the IMEs. It became more popular as a policy too and persists to this day. Stockton, California started issuing $500 debit cards to select residents in February 2019. The program was designed as an experiment to test the workability of a broader basic income. Ottawa recently ended a guaranteed income experiment that had support across the political spectrum. Experimentation promises innovation, but its method in the public sphere invites political manipulation and ethical uncertainty. An understanding of the origins of policy experimentation can help policymakers better refine the process of evaluation.26

---

Bibliography
Ashenfelter, Orley and David Card, eds. *Handbook of Labor Economics* (Amsterdam: Elsevier, 2010), 111


———. “The Professionalization of Reform.” *The Public Interest,* no. 1 (Fall 1965): 6-16.


