Experiments in Democracy

Jeremy Rozansky

There is a timeworn joke about a man who, walking home after midnight, comes across a drunkard on his knees under a streetlamp, patting the pavement with his hands. "What are you looking for?" asks the passerby. "I can't find my keys," says the drunk. The passerby, being a kindly man, gets on his own knees and joins in the search—to no avail. Finally, after a quarter of an hour, he asks, "Are you sure this is the spot where you lost your keys?" "Oh, not at all," the drunk answers, pointing: "I dropped them on my front stoop across the street." "Then why are we looking on this side of the street!" "The light's better over here."

This dusty old story may be good for a laugh or a groan, but it also serves as a parable about social science and policymaking. The light shone by the methods of social science is limited, sometimes too dim, and may not illuminate what is important or useful for policymakers—but a social scientist, like the drunkard in the joke, is going to scour the ground he can see. For most public policy questions, the streetlamp under which social scientists search for useful and important knowledge is what is known as the "econometric method." The econometric method takes a large sample of observed human interactions and creates a model to predict what a particular policy or treatment will do, controlling for a large number of variables like race, sex, income, habits, location, and attitudes.

In his debut book *Uncontrolled*, entrepreneur and policy analyst Jim Manzi argues that social scientists and policymakers should instead adopt the "experimental method." The essential tool of this method is the randomized field trial (RFT), a technique that already informs many of our successful private enterprises. Perhaps the best known example of RFTs—one that Manzi uses to illustrate the concept—is the kind of clinical trial performed to test new medicines, wherein researchers "undertake a painstaking series of replicated controlled experiments to measure the effects of various interventions under various conditions," as he puts it.

> *Uncontrolled:*
> *The Surprising Payoff*
> *of Trial-and-Error for*
> *Business, Politics, and Society*
> By Jim Manzi
> Basic ~ 2012 ~ 300 pp.
> $28.99 (cloth)

The central argument of *Uncontrolled* is that RFTs should be adopted more widely by businesses as well as government. The book is helpful and holds much wisdom—although the approach he recommends is ultimately just another streetlamp in the night, casting a pale light that tapers off after a few yards. Much still lies beyond its glow.

The relative merits of different social science methodologies may seem to be a subject of merely academic interest. But Manzi persuasively shows why it is important for policymakers to think about science, beginning with a recent example illustrating the reality that policymakers often don't know whether or not their plans will work until they actually put them into place.

In early 2009, President Obama and congressional Democrats proposed to jumpstart the U.S. economy out of recession by passing a fiscal-stimulus bill. Whether this infusion of about $800 billion in deficit-financed government spending would improve the economy was not at all certain, but the White House could point to several economists, including a handful of Nobel laureates, who predicted that every dollar spent in stimulus would increase the national income to the tune of a dollar-and-a-half by "stimulating" demand. However, other Nobel laureates disagreed, arguing that the bill was not worth its price tag: The stimulative effects of the deficit spending would be diminished by (among other things) expectations of future taxation, and so the return on every dollar of deficit spending would be closer to zero. Paul Krugman, a pro-stimulus economist, lambasted anti-stimulus economists as priests from the "Dark Age of macroeconomics." Other academics hit back, alleging that it was the Keynesian Krugman who was himself stuck in the Dark Ages.

A policymaker or citizen looking at both arguments could see that each side made models of fiscal spending and GDP growth, and each had a smattering of econometric analyses of past fiscal spending that they said validated their models. But as Manzi wryly notes, "the only thing an observer could say with high confidence before the stimulus program launched was that at least several Nobel laureates in economics would be directionally incorrect about its effects." In such a policy stalemate—professor against professor, Nobelist against Nobelist, Princeton against Chicago, all appealing ultimately to their own authority—what can we lay citizens do but throw up our hands and concede ignorance? And what does this mean for democratic self-governance?

A similar policy stalemate arose during the 2012 presidential contest. Mitt Romney had promised both to cut tax rates by 20 percent and to maintain existing levels of revenue.

He claimed he would do so by eliminating unspecified carve-outs and loopholes. The Obama campaign cited a Tax Policy Center study claiming that the only way Governor Romney could achieve both goals would be by eliminating a variety of tax deductions popular with the middle class. Unless Romney did so, the study argued, he would either have to renege on his 20 percent rate-cut promise or fail to meet his revenue goals. When challenged on this subject in the first debate against President Obama, Romney fought study with study:

> Now, you cite a study. There are six other studies that looked at the study you described and say it's completely wrong. I saw a study that came out today that said you're going to raise taxes by $3,000 to $4,000 on middle-income families. There are all these studies out there.

Being a man for whom data analysis is perhaps a more intuitive craft than politics, Romney probably could have explained the differing assumptions behind the various studies and evaluated the validity of each. Perhaps he could have compared the studies in detail and shown the method of one to be obviously superior. Instead, he made two moves that well illustrate an all-too-common pathology of contemporary politics. First, he did not challenge the assumptions or methods used by Obama's preferred

study. Second, he merely cited *other* studies, presenting them as equal and opposite authorities to the Tax Policy Center's. Romney knew that even though the electorate is disposed to believe what "studies" show, most voters lack the expertise, the time, and the inclination to compare the validity of contradictory studies. Thus, so long as social science is not unanimous, its authority will be inconclusive for the voter.

But a self-contradictory authority is, in effect, no authority at all. Voters tend to ignore much of the daily business of government as hopelessly complicated—but while understandable, this state of affairs is a recipe for both shallow political debates and rule by technocrats. Manzi's book, which addresses the underlying assumptions of different social science methods, offers a solution to the problem of dueling Nobelists or think-tank studies—a solution that promises not only better policy but, more importantly, better democratic politics.

Uncontrolled is in many ways a book about the scientific method, and Manzi begins by staking out a position on an important question about science that most Americans rarely think to ask: What is science—in varieties both "hard" and "soft"—good for? Manzi's answer, which may startle some readers, and may even offend some scientists, is not "finding truth." Rather, the answer is

utility. Science, Manzi argues, chiefly aims to discover effective means for reaching the ends that human beings choose through other forms of reflection, such as philosophy, theology, or the arts. While common sense can be useful for identifying the means to accomplish desired ends, "the key value of science is that it provides causal rules that are nonobvious, that is, that extend beyond common sense."

Still, as Manzi describes, the causal rules that science gives us are not definitive. In fact, it is fundamentally impossible to know any causal rules with certainty: Even when we see one kind of event precede another kind of event under many conditions, we can never be completely certain that the first kind causes the second. Every time we see someone let go of a rock, we see it fall, and so we infer that dropping a rock causes it to fall. But although this apparent causal relationship has always served as a reliable rule, we cannot know with perfect certainty that the next time someone drops a rock it will fall, and even if it does fall, we cannot conclude that dropping it is what causes it to fall.

This limitation on both science and common sense, which philosophers call the problem of induction, can of course seem preposterous when it calls into question our ability to draw any conclusions about causation. But there are always hidden factors that can account for apparent cause-effect relationships. Any ancient thinkers who might have seen dropping and falling as inviolably linked would have been brought up short by the later discovery that a rock will float rather than fall when it is released in outer space. When we drop a rock here on the ground, the proximity to a large gravitational body is a hidden conditional. So too would the ancients have been unlikely to anticipate that some rocks let go here on earth sometimes will *not* fall, for example, when a large magnet is present.

So we cannot demonstrate causal rules to be true, and should not try to; we can only demonstrate causal rules to be false or in need of amendment, when we find a hidden conditional. The practice of science involves making useful assumptions and gradually and meticulously adding nuance to our assumptions in order to make them more useful. Manzi both narrows the focus of science and demystifies its methods, bringing it down to its rightful place among the useful arts.

As we will see, in lowering the sights of science, particularly social science, Manzi points toward why he prefers the experimental method over the econometric method. But the crucial reason for Manzi's preference for RFTs over econometrics is more technical—an idea Manzi calls "causal density." In seeking the cause of a given effect, the general approach of science is to isolate each

potential variable that might play a causal role and manipulate it while holding everything else equal. By performing experiments and measuring the apparent effects of each isolated cause, scientists can make useful assumptions about cause and effect. While we can never control for everything that could possibly be causally significant—in part because we never know where hidden conditionals might be lurking—we can be satisfied with an assumption if the cause seems well isolated and the effect is reliably observed when we replicate the causal conditions we think are relevant. The model we create from these observed rules can never completely capture the actual system, and we can never know we have included every hidden conditional, but it can be a useful predictor.

Still, some systems are more easily modeled than others; it all depends on how easily the relevant conditionals are isolated. Creating models is appropriate, and often relatively straightforward, in a field like astrophysics, where objects are far away from each other and replicating an observed rule is easy given the vast expanse of data. Astrophysics is a science of low causal density.

Social science, by contrast, has very high causal density. The subjects—human beings and their institutions—are complexly intertwined. It is very difficult to isolate a conditional, and it is impossible (or at least terribly intrusive and often unethical in practice) to hold *all* other things equal. Think of our debates over education, and all the variables that can affect whether a child becomes well educated: the resources available to his school, whether his parents value education and how much they do, the skill of his teacher, the influence of his classmates, his I.Q., self-motivation, nutrition, access to school supplies, the particular textbooks he reads (or doesn't read), and so forth. Now consider how all these variables interact with one another. Would anyone really expect two groups of students—one in, say, Moline, Illinois and the other across the river in Davenport, Iowa—to react exactly the same to a new fourth-grade math curriculum, even if virtually every feature of their lives that social scientists can measure were the same? Would we be reasonable to chalk up the observed differences to the differences between Illinois and Iowa? The answer to both these questions is most likely "no"—because human beings and institutions are just too complicated to justify such claims. Manzi vividly compares social science to medical research in which every test tube is poorly cleaned and contains a foreign residue—a hidden conditional.

The econometric method now dominates the social sciences because it helps to cope with the problem of high causal density. It begins with a large data set: economic records, election results, surveys, and other

similar big pools of data. Then the social scientist uses statistical techniques to model the interactions of sundry independent variables (causes) and a dependent variable (the effect). But for this method to work properly, social scientists must know all the causally important variables beforehand, because a hidden conditional could easily yield a false positive.

The experimental method, which Manzi prefers, offers a different way of coping with high causal density: sidestepping the problem of isolating exact causes. To sort out whether a given treatment or policy works, a scientist or social scientist can try it out on a random section of a population, and compare the results to a different section of the population where the treatment or policy was not implemented. So while econometric models aim to identify which particular variables are responsible for different results, RFTs have more modest aims, as they do not seek to identify every hidden conditional. By using the RFT approach, we may not know precisely *why* we achieved a desired effect, since we do not model all possible variables. But we can gain some ability to know that we *will* achieve a desired effect, at least under certain conditions.

Strictly speaking, even a randomized field trial only tells us with certainty that some exact technique worked with some specific population on some specific date in the past when conducted by some specific experimenters. We cannot know whether a given treatment or policy will work again under the same conditions at a later date, much less on a different population, much less still on the population as a whole. But scientists must always be cautious about moving from particular results to general conclusions; this is why experiments need to be replicated. And the more we do replicate them, the more information we can gain from those particular results, and the more reliably they can build toward teaching us which treatments or policies might work or (more often) which probably won't. The result is that the RFT approach is very well suited to the business of government, since policymakers usually only need to know whether a given policy will work—whether it will produce a desired outcome.

Manzi offers plenty of evidence of the efficacy of RFTs. In the business world, he himself has built a career by consulting with companies to help them run RFTs and improve their profits. One of his former consulting colleagues founded the credit card company Capital One; its rise in an industry with high barriers to entry is the result of following the RFT approach, conducting thousands of experiments. Outside of business, RFTs also contributed to some of the great broken-windows innovations in criminology that helped bring down the crime rate two decades ago. Some RFTs help by disproving theories.

Milton Friedman's idea that a negative income tax—essentially a guaranteed minimum income—could replace the welfare-state bureaucracy and eliminate welfare's perverse incentives was tested in several massive programs. Policymakers discovered that the negative income tax actually exacerbated some of the perverse incentives Friedman was hoping to fix. In the end, the experiment was most useful in that the failure of Friedman's hypothesis pointed to a different way to fix welfare: the work requirements in the 1996 reform bill. More recently, RFTs have made their way into electoral politics, first with Rick Perry's successful gubernatorial campaign and later with Barack Obama's reelection.

Manzi looks at the RFTs conducted for public policy questions—which are dwarfed by the number conducted in the business world—and draws a few general conclusions. First, most policy experiments don't work, so policymakers should not be too enthralled with their own designs. Second, programs that focus on "raising skills or consciousness" tend to fail because people's character is hard to change; the ones that do work tend to be the ones that focus on changing behavior by changing incentives. Third, grand counterintuitive or surprising causal effects that make up much of the pop-social-science literature are generally not true or are only half-

true; although science is supposed to find non-obvious rules, social science mostly confirms (or refutes) ideas already held by common sense. And so Manzi does not anticipate that the greater use of RFTs will revolutionize policymaking. His expectations are more modest.

With these rules in mind, Manzi urges policymakers to embrace the use of randomized field trials. He recommends not just that we use RFTs to test specific policies but also, more broadly, that we adopt an experimental disposition. Such a disposition entails a general deference to decentralized systems—like federalism and the market—that encourage trial-and-error improvements. This preference for decentralization should not be taken to extremes: some interventions, like changes in interest rates, must be done at a large scale. Rather, a disposition toward trial and error will encourage experimentation by the most local competent political authority, and by firms of all sizes.

Manzi's prescription is in many ways deeply conservative. History is often seen, incorrectly, as encompassing a few revolutionary moments when new truths are discovered. A more accurate view, Manzi posits, holds that history is simply a long record of trial and error that has built up and retained a reserve of "implicit knowledge." Under this view, the arrangements that have withstood experiments over the generations

must seem workable and wise. Thus a preference for the status quo is rational; Manzi places the burden of proof on those who advocate radical change.

There is, however, a tension between, on the one hand, the restless experimentation that Manzi recommends, and on the other, the conservative bias toward the present order. Manzi does not presume that he can cleanly resolve this tension, but he does offer some ideas for how it can—and has—been managed. For instance, the use of decentralized systems can offer an advantage because within them growth is more incremental and therefore less disruptive. Some government policies can also reduce the tension between innovation and social cohesion by improving the adaptability of vulnerable individuals to the jarring effects of economic growth. Public education is one way of improving adaptability, and redistributive policies can be another. Manzi warns, however, that the welfare state can stifle trial-and-error processes, especially given its tendency to be highly centralized. But since it is necessary to smooth over some of the harsh edges of innovation, Manzi writes that, as much as possible, the welfare state should be structured so as not to choke the very innovation that it is meant to make palatable.

There are of course some caveats that those disposed to trial and error should keep in mind. For one,

policymakers cannot always conduct a satisfactory experiment—one randomized and replicable—so they will sometimes have to settle for implementing an idea without a real record of success. In these cases Manzi simply encourages policymakers to try out new ideas on a smaller scale so as to reduce the risk.

But naturally, there are situations where even such small-scale attempts are impossible. To revisit the example of the 2009 stimulus, the theory behind the policy was that deficit spending would improve GDP growth in the midst of a recession. But the states, our laboratories of democracy, could not conduct their own stimulus experiments because nearly all of them are legally required to have balanced budgets, preventing them from running the deficits that the stimulus required. Nor could the federal government experiment with different ideas on a smaller scale. If, say, the federal government selected one hundred counties for workers to have payroll-tax breaks, workers would flood into those counties. Policies like the stimulus bill are not made by social scientists testing theories, but by politicians facing a national crisis. Many policy decisions are made under circumstances that force large-scale and unpredictable—and therefore risky—ventures. And crises do not give policymakers enough time to learn from their mistakes.

In the final chapter of the book, Manzi offers a handful of specific policy recommendations for how the government can "embed a trial-and-error process within humane constraints." First, and most obviously, we should decentralize and start conducting more and better experiments, especially at the state level. For example, we could have a better sense of the costs and benefits of universal preschool if several counties, states, or foundations were to undertake more rigorous experiments. Much of the hype about the benefits of universal preschool has been fueled by the intense Perry and Abecedarian preschool projects of the 1960s, but since so much of the success of those unusual projects was attributable to the singular talents of the individuals involved, it seems inappropriate to cite those projects as evidence for the proposition that preschool is generally a wise investment.

The federal government's involvement in policy experimentation has received little public attention. The Centers for Medicare and Medicaid Services (CMS) runs a number of demonstration projects, although they are generally less useful and replicable than true RFTs. Nonetheless, the CMS experience with experimentation is an apt illustration of Manzi's broader point: There seems to be a wall between the knowledge gained and broader policymaking. For instance, many of the savings in

Medicare that Obamacare hopes to achieve are gained through the introduction of what is known as "bundled payments": payments based on standardized rates for the treatment of specific clinical conditions, regardless of the services employed in treating that condition. Yet CMS has already run a demonstration project on this exact topic and learned that we should be skeptical of bundled payments saving even one percent of Medicare costs. As long as the experiments are conducted without fanfare and their results ignored, policymakers can go on hyping bad ideas as if they might work.

Manzi also wants to see a proliferation of different and creative policy experiments that allow state governments to better achieve the goals of federal mandates and programs. He therefore proposes that the states and the federal government agree to a simple trade: The federal government will broadly waive regulations governing program design and other federal mandates on a trial basis if the states accept certain standards of experimental rigor. Manzi also calls for the creation of an organization within the federal government to create and enforce standards for the design and interpretation of randomized public policy experiments, much like what the Food and Drug Administration does for clinical trials in the field of medicine.

Manzi also offers a set of several proposals to build human capital.

He calls for a universal voucher program for public education, a bias in visa-granting toward highly skilled immigrants, and an increase in the share of the federal budget dedicated to research and development. He also suggests that policymakers involved in education and R&D be more "ruthless" in killing off ideas that do not lead to helpful results, while "pointing a fire hose of money at those that succeed." Unfortunately, he has no ideas for how to craft such a bureaucracy.

He also urges a rethinking of the welfare state, in which policymakers list all its discrete tasks, creating separate programs for each that employ market mechanisms where feasible. For instance, Social Security is both a means of forcing workers to save their income and a safety net for the vulnerable elderly. These are separate tasks, and the first could be accomplished by a managed market of private accounts, while the second could be achieved through direct federal transfers.

What is striking about Manzi's policy ideas is how commonsensical and unoriginal they are in the world of conservative policy analysis. Of course we would rather have a Ph.D. in chemistry than a high-school dropout join our workforce! Of course we should introduce vouchers and transparent transfer payments where possible. Yet *Uncontrolled* is billed as revealing the "surprising payoff of trial-and-error." Nothing is too

surprising in this last chapter. In fact, Manzi's concrete policy proposals do not directly follow from knowledge gained through specific trial-and-error processes or experiments. With the exception of school vouchers, which Manzi discusses earlier in the book, these policy prescriptions have little experimental evidence to support them.

The lack of cited experiments should lead us to wonder if new data will find Manzi is wrong about the positive effects of high-skilled immigration or the unbundling of Social Security. Manzi's method of choice does not lead directly to his policies of choice. Instead, his method follows from principles which also more or less guide him, in parallel, to his policies. This could be seen as the great error of *Uncontrolled*: The method he uses many pages to argue for is really not very proximate to the policies the reader is encouraged to prefer. Instead, this should be seen as a point in the book's favor. Social science has always had a hubristic ambition to draw a straight line from its general findings to particular policy proposals. Even at its best, however, social science can offer us only partial knowledge of human affairs. But because we live in the world, we must still muddle along—and Manzi's book offers a few sensible ways to muddle better, without claiming to do much more than that.

To Manzi's great credit, he never describes the randomized field trial

as a silver bullet, only a sensible alternative to the overrated econometric method. He summarizes the academic critiques of RFTs—most notably from economist James Heckman—and basically agrees with them. In practice, it is terribly difficult to truly randomize a field test for a public policy. Small-scale experiments will also not be useful in every policy context. As social scientist Robert A. Moffit wrote about experiments with welfare reform, "the RFT methodology is poorly suited to measuring the effects of structural, system-wide reforms." And, again, the great strength of RFTs—that they can show whether a policy works—does not mean that they offer insight into *why* it does or does not work. The econometric method still has the advantage of breaking results down into the sort of piecemeal conclusions that can inform a theory of the actual reasons for the effectiveness of a treatment or policy. If researchers have a sense of the why, they can perhaps design programs that are even more effective, and more narrowly targeted.

Manzi concedes all of this. For him, however, the RFT is a second streetlamp to the econometric method. Together, the two barely illuminate a city block, but the RFT reveals some parts of the pavement that econometrics leave in shadow.

By arguing for a more common-sense approach to social science,

Manzi denies the privileged place that experts with arcane econometric models occupy in contemporary politics. His appeal is rightly understood not merely as an endorsement of randomized field trials but rather as an argument for the empiricist disposition characterized by incrementalism, libertarianism, caution, and epistemic modesty.

This disposition is intuitively American. Alexis de Tocqueville observed nearly two centuries ago that although "there is no country in the civilized world where they are less occupied with philosophy than the United States," Americans do have certain inclinations: the use of "tradition only as information, and current facts only as a useful study for doing otherwise and better," and striving "for a result without letting themselves be chained to the means." Tocqueville identifies this cast of mind with the skeptical and rationalistic French philosopher René Descartes, calling Americans natural Cartesians.

The greatest work of American political philosophy, *The Federalist*, joins practical wisdom gathered from the study of history—empirical evidence, of a sort—with philosophic reflections on human nature. Neither of these is especially helpful without the other, and Alexander Hamilton assails "the reveries of those political doctors whose sagacity disdains the admonitions of experimental instruction." Hamilton even quotes the great

Scottish empiricist David Hume in the final paragraph of the final paper of the *Federalist*. Hume writes that the craft of creating a constitution is such a difficult work "that no human genius, however comprehensive, is able by the mere dint of reason and reflection, to effect it. The judgments of many must unite in the work: EXPERIENCE must guide their labour: TIME must bring it to perfection: and the FEELING OF inconveniences must correct the mistakes which they *inevitably* fall into, in their first trials and experiments."

All these refractions of American thinking seem to argue that, given the limitations of human reason, trial and error is very often the best means toward the practical improvement of any human undertaking. The experimental method lends itself to incremental progress, with experiment after experiment adding to the useful knowledge of mankind. The econometric method gathers the data of the past to create a predictive model of human behavior. But this predictive model too often takes the place of the accumulated experience of many generations and becomes a single, supremely-but-unjustifiably confident reference for policymaking. It uses history, but it uses history in order to wipe history away.

In our increasing deference to econometric studies and technocratic experts, we Americans have come a long way from being the "natural Cartesians" who so impressed Tocqueville. Experimental

knowledge—the residue of trial and error—can also be haggled over and made esoteric, but, being more empirical and concrete, it is less liable to become mired in abstractions or to lead us down the dangerous alleyway of a false positive.

If Manzi gets his way and such experimental social science becomes more common, it will become the norm in political debates to say "show me." Rather than be stuck in a stalemate of whether or not the "studies say" a proposed policy will work, experiments will be conducted, giving the public a sense of the costs, benefits, and tradeoffs associated with that policy. Then we can argue about whether the policy is worth implementing on a wider scale instead of engaging in rhetorical theatrics, with politicians bludgeoning one another with studies. We should return to making moral arguments—the sort of arguments that self-governing citizens ought to have basic competence to judge.

Jim Manzi says he offers a modest solution for some of the problems of poor policymaking. Whether or not his proposal for changing the way social science is done in America improves policy outcomes, by advancing the experimental method for social science, he may just help to revive the intellectually independent disposition of the American citizen.

Jeremy Rozansky *is assistant editor of* National Affairs.