

RICHARD T. ELY LECTURE

The Economist as Plumber[†]

By ESTHER DUFO*

Economists are increasingly getting the opportunity to help governments around the world design new policies and regulations. This gives them a responsibility to get the big picture, or the broad design, right. But in addition, as these designs actually get implemented in the world, this gives them the responsibility to focus on many details about which their models and theories do not give much guidance.

There are two reasons for this need to attend to details. First, it turns out that policymakers rarely have the time or inclination to focus on them, and will tend to decide on how to address them based on hunches, without much regard for evidence. Figuring all of this out is therefore not something that economists can just leave to policymakers after delivering their report: if they are taking on the challenge to influence the

real world, not only do they need to give general prescriptions, they must engage with the details.

Second, details that we as economists might consider relatively uninteresting are in fact extraordinarily important in determining the final impact of a policy or a regulation, while some of the theoretical issues we worry about most may not be that relevant. This sentiment is well summarized by Klemperer (2002, p. 170), who presents his views on what matters for practical auction design, based on his own experience designing them and advising bidders: “in short,” he writes, “good auction design is mostly good elementary economics, [whereas] most of the extensive auction literature is of second-order importance for practical auction design.”

It seems appropriate to open an essay on plumbing with an actual plumbing example that illustrates the two points I made above (Devoto et al. 2012). Many cities in the developing world seek to improve citizens’ access to home water connections. Even when there are public taps, urban households without a connection at home spend several hours per week collecting water, and this burden causes them considerable stress and tension. The typical policy for improving water access is to build the necessary infrastructure, and then to encourage “end of pipe” connections through subsidized tariffs and/or subsidized loans. In 2007 in Tangier, a firm called Amendis (the local subsidiary of Veolia Environnement), which was in charge of the water and sanitation for the city, had spent considerable resources building large pipes and installing toilets in each house, and, in collaboration with the city, was offering interest-free loans to poor households to make it possible to cover the marginal cost of new water connections. But take-up of the subsidized loan program was very low (less than 10 percent); applying for the program required a trip to the municipal office with supporting documents,

* Department of Economics, MIT, 77 Massachusetts Avenue, Building E52-544, Cambridge, MA 02139 (e-mail: eduflo@mit.edu). This essay is based on the Ely lecture, which I delivered at the AEA meeting in January 2017. Many thanks to Alvin Roth, both for his invitation to give this lecture, and for his license to get our hands dirty with policy work. I thank Abhijit Banerjee and Cass Sunstein, master plumbers, for their encouragements, for several conversations that shaped this lecture, and for detailed comments on a first draft. David Atkin, Robert Gibbons, Parag Pathak, and Richard Thaler, provided useful comments. Vestal McIntyre gave wonderful and detailed comments on style; all awkwardness remains mine. Laura Stilwell provided excellent research assistance. Plumbing is not a solitary activity, and I would like to acknowledge the many people who have plumbed and discussed with me over the years, notably: Abhijit Banerjee, Rukmini Banerji, James Berry, Raghabendra Chattopadhyay, Iqbal Dhaliwal, Rachel Glennerster, Michael Greenstone, Rema Hanna, Clement Imbert, Michael Kremer, Shobhini Mukherjee, Karthik Muralidharan, Nicholas Ryan, Rohini Pande, Benjamin Olken, Anna Schrimpf, and Michael Walton.

[†] Go to <https://doi.org/10.1257/aerp20171153> to visit the article page for additional materials and author disclosure statement.

and this proved a real barrier. When the research team randomly visited households and offered procedural assistance by photocopying the required documents at home and delivering them to the municipal office, take-up of the loan and the water connections increased to 69 percent. As a result of this small additional expenditure, poor residents of Tangier gained access to water, and therefore recovered a considerable amount of time to do other things. They were ultimately much happier and less stressed, despite a large increase in their water bill.

Improving access to private water connections was a sensible policy idea, and the entire effort was broadly well designed. But the lack of attention to the very last step (the administrative steps to sign up) had been preventing this large investment in both physical and financial infrastructure from paying off.

These kinds of practical design questions are ubiquitous in policy design. When new frontline workers are hired to implement a program, will emphasizing wage and career prospects discourage publicly minded individuals or encourage the most talented to join? When thinking about an immunization policy, should policymakers assume that parents understand the full costs and benefits and immunization and rationally internalize them, or assume they may be ill informed and/or present biased? When designing the health exchanges for the Affordable Care Act, should the health-plan options be labeled with precious metal (platinum, gold, or bronze) or will that inadvertently bias the participant choice toward one type of plan versus the other? At the outset, it will be difficult to know the answer to these questions, especially when the policy problem is fairly new.

Paying attention to the details of policy requires a mindset that is slightly different from that which graduate school instills in economists. Banerjee (2007, p. 161–162) summarizes the reluctance of economists to engage with those details very well in his essay “Inside the Machine.” Economists, he writes, tend to think in “machine mode”: they want to find out the button that will get the machine started, the root cause of what makes the world go round. He writes:

The reason we like these buttons so much, it seems to me, is that they save us the trouble of stepping into the machine. By assuming that the machine either runs on

its own or does not run at all, we avoid having to go looking for where the wheels are getting caught and figuring out what small adjustments it would take to get the machine to run properly. To say that we need to move to a voucher system does not oblige us to figure out how to make it work—how to make sure that parents do not trade in the vouchers for cash (because they do not attach enough value to their children’s education) and that schools do not take parents for a ride (because parents may not know what a good education looks like). And how to get the private schools to be more effective? After all, at least in India, even children who go to private schools are nowhere near grade level. And many other messy details that every real program has to contend with.

When we are concerned with such details, *ex ante*, we will have some priors on what features will be important, and this guides our first-pass attempts at design. But it is not clear that either the policymakers or the scientists will correctly identify the most important choices. Those may not have been the focus of either practical or theoretical attention, and thus may have been completely ignored. So an economist who cares about the details of policy implementation will need to pay attention to many details and complications, some of which may appear to be far below their pay grade (e.g., the font size on posters) or far beyond their competence level (e.g., the intricacy of government budgeting in a federal system). It will sometimes appear that the extensive training they received is underused if, as Klemperer notes, the theoretical complexities turn out to be second order. On the other hand, they will have a chance to apply their economist’s mind, since many of the details have implications for issues that are an economist’s bread and butter: incentives, information, imperfect rationality, etc. They will also need to be very observant, and keep a close eye on the impact of any change they recommend.

Inspired by this exhortation to go inside the machine, Alvin Roth’s image of the economist as engineer, and Banerjee’s (2002) image of the economist as an experienced craftsman,¹ I label

¹ Banerjee in fact mentioned plumbers, in defending the reputation that economists will provide good advice because, like plumbers, they care about their reputation.

this detail-focused approach as the “plumbing” mindset. The economist-plumber stands on the shoulder of scientists and engineers, but does not have the safety net of a bounded set of assumptions. She is more concerned about “how” to do things than about “what” to do. In the pursuit of good implementation of public policy, she is willing to tinker. Field experimentation is her tool of choice.

What I will try to argue in this lecture is not that all of us should be plumbers, or even that any of us should be plumbers *all the time*; instead, I will try to show that there is value for economists to take on some plumbing projects, in the interest of both society and our discipline.

I. Scientists, Engineers, Plumbers

A. Definitions

Alvin Roth’s seminal 1999 Fisher-Schultz lecture (Roth 2002) invited economists to adopt an “engineering” approach to their craft. Economists, he pointed out, are increasingly called upon, not just to analyze real world institutions, but also to design them. Roth’s focus is the design of markets, but economists are also called upon to help design incentives schemes for firms and regulation and social policies for governments. In this paper, I will consider the role of economists in the design of policies and regulation.

For Roth (2002, p. 1341), intervening in the real world should fundamentally alter the attitude of the economist and her way of working. He sets the tone in the abstract of the paper (emphasis added):

Market design involves a responsibility for detail, a need to deal with all of a market’s complications, not just its principle features. Designers therefore cannot work only with the simple conceptual models used for theoretical insights into the general working of markets. Instead, market design calls for an engineering approach.

The scientist provides the general framework that guides the design. In Roth’s Fisher-Schultz lecture, which again is focused on market design, game theory provides the general principles. For all the other domains where economists will be called to provide inputs, there exists a relevant body of theory (or at least general insight) that they can use to guide design.

The engineer takes these general principles into account, but applies them to a specific situation. This requires careful attention to the details of the environment being studied, but also new tools: the economist-engineer cannot shrug off the fact that a particular situation is not covered by the assumptions of the theorem, and cannot ask agents to change their preferences so the assumptions hold. If the specific real-world problem at hand cannot be solved analytically, then she will reach for other tools—in particular computation and laboratory experiments—and will simulate the behavior of a market.

For example, in the matching for doctors who are applying for their first residency to a hospital, which is the focus of Roth (2002), the simple matching theory does not accommodate the fact that some of these new doctors come as married couples, and need to be assigned to the same town. Roth refined the theory to accommodate married couples, but at that point, he could not solve the problem analytically: in particular, theory suggests that, with couples, there could be a situation without stable matching, and that the sequencing of the decision could in principle affect results. So Roth and his colleagues used computation to design an algorithm, and examine the impact of different rules (using data from previous years), including the potential impact of sequencing. The computations suggested that the algorithm never failed to converge to a stable match, and that sequencing effects were small and unsystematic. This allowed them to suggest a matching algorithm that would work (and has worked in practice), even though the theory for why couples were probably not a big problem after all had not been fully developed.

The plumber goes one step further than the engineer: she installs the machine in the real world, carefully watches what happens, and then tinkers as needed. At the time she inherits the machine, the broad goals are clear, but many details still need to be worked out. The fundamental difference between an engineer and a plumber is that the engineer knows (or assume she knows) what the important features of the environment are, and can design the machine to address these features—in the abstract, at least. There may not be a theory fully worked out to accommodate these features, but she can use computation and lab experiments to simulate how they will play out. When the plumber fits the machine, there are many gears and joints,

and many parameters of the world that are difficult to anticipate and will only become known once the machine grinds into motion. The plumber will use a number of things—the engineering design, his understanding of the context, prior experience, and the science to date—to tune every feature of the policy as well as possible, keeping an eye on all the relevant details as best he can. But with respect to some details, there will remain genuine uncertainty about the best way to proceed, because the solution depends on a host of factors he cannot easily quantify, or sometimes even identify, in the abstract. (These are the “unknown unknowns”: all the issues we can’t predict but will arise anyway). Thus, another difference between plumbers and engineers is that engineers will start from the outcome they are seeking to attain and engineer the machine to reach it. Plumbers, on the other hand, will have to adopt a more tentative approach, starting from the machine’s characteristics and identifying their effect (compared to another possible set of choices).

For example, in the early 2000s the city of Boston decided to change how it assigned students to schools. There was an important engineering part to choosing the mechanism that would be used, initiated by Abdulkadiroglu and Sönmez (2003) and followed by a considerable literature (see Pathak 2011 and Abdulkadiroglu and Sönmez 2013 for reviews). But once city leaders settled on a mechanism, they still had to make many decisions. How to communicate the change to parents? How to persuade them that they could reveal their true preferences when they ranked schools? Should there be a limit on how many schools parents can (or are required to) rank? Should parents living near a school receive preferential treatment, and how should that be set up? And so on. According to Pathak (forthcoming, p. 3), who reviewed his experience working in several cities and who modeled his conclusion on Klempner’s:

What really matters for school choice market design are basic insights about straightforward incentives, transparency, avoiding inefficiency through coordination of offers and well-functioning aftermarkets, and influencing inputs to the design, including applicant decision-making and the quality of schools. Some of the issues examined in the extensive theoretical literature on school choice matching market design are less important for practical design.

What is more, we may not even know *ex ante* which of these decisions will in fact matter, and our models offer fairly limited guidance on what to pay attention to. Pathak (forthcoming, p. 3) writes: “I will discuss a handful of issues examined in the theoretical literature on matching mechanisms that have proven to be first-order. It’s worth emphasizing that it is only with the benefit of several design case studies that we’re beginning to understand which issues are quantitatively important.”

This uncertainty—concerning what the true model is—has consequences for policy engineering very similar to the problems discussed in the macroeconomics literature on “robust” policy (Cogley et al. 2008). Since we don’t know what the true model is, we need to design policies that are as robust as possible to this model uncertainty. For example, Chetty (2015) argues that public policy that includes “nudges” (such as default) is more robust, in the Hansen and Sargent sense, to a specific model uncertainty (are people rational or not?): rational people will chose what they want, and “behavioral” agents will be in default that should largely work for them. A smart nudging policy is thus a fairly robust policy, in general. It will have an effect only if people suffer from behavioral biases, but have no effect otherwise.

There is no general theory of how to design policy under this kind of model uncertainty, however, and in many cases, even the best educated guess will still be just that, a guess. The economist-plumber will use all they know (including model uncertainty), to come up with the best guess possible, and then pay careful attention to what happens in reality. The uncertainty in the environment creates a highly stochastic world: the natural way to “pay attention” to what happens, as I will argue below, is thus to analyze natural experiments or set up field experiments to try out different plumbing possibilities.

B. Designing the Taps, Laying Down the Pipes: Two Examples

Although I will often talk about *details* as a general way to describe the handiwork of plumbers, there are two different kinds of plumbing in policy design. First is the “design of the tap” work: taking care of apparently irrelevant details, such as the way the policy is communicated or the default options offered to customers.

Second is “layout of the pipes” work: important logistical decisions that are fundamental to the policy’s functioning but often treated as purely mechanical, such as the way money flows from point A to point B, or which government worker has sign-off on what decisions.

Banerjee et al. (2015) worked with the Indonesian government on a “design of the tap” problem, in the context of the Raskin program, a subsidized rice distribution scheme that, with an annual budget of US\$1.5 billion and a targeted population of 17.5 million households, is Indonesia’s largest targeted transfer program. The Raskin program is plagued with corruption: the data in Banerjee et al. show that beneficiary households pay more than they should and receive less than their full quota of rice, to the extent that they receive only about one-third of the subsidy they are entitled to, overall. Working with the government of Indonesia, the researchers designed a set of field experiments to provide information directly to eligible households. In 378 villages (randomly selected from among 572 villages spread over 3 provinces), the central government mailed “Raskin identification cards” to eligible households to inform them of their eligibility and the quantity of rice that they were entitled to. In addition, the team varied several features of the design of the cards and the way they were introduced: whether the copay price was also listed on the card, whether information about all the beneficiaries was posted in the village, whether cards were sent to all eligible households or only to a subset, and whether the cards came with coupons to acknowledge the delivery of the rice.

A first plumbing issue is whether or not households got a physical card. Traditionally, households had not gotten one—the program was designed to be implemented by local leaders, who were given considerable discretion, and any effort at information dissemination had focused primarily on them, not on citizens. The optimistic policymaker who initially designed this machine clearly envisioned a beneficent local administration dedicated to the good of the people. The very specific design features of the cards and their distribution are further plumbing issues, and would normally have been ignored, even after the government had been convinced that sending cards was a good idea. Someone (a graphic designer maybe?) would have made a decision on whether or not to list on the card the price to be paid for rice. Distribution of the cards

would have been up to the local leaders, and so on. But there are good reasons to think that each of these features matters. The price to be paid for rice is one piece of information that recipients themselves do not have, so local leaders can easily skim off the program by inflating the copay. Making the list widely available will create communal knowledge, potentially changing the nature of the bargaining game between the local leader and the citizen. Distributing physical cards may not even be necessary, if lists of those eligible are posted in the village.²

In practice, distributing cards made a large difference: on net, across all of the variations of the program, the cards led to a large increase in subsidy received by eligible households. Eligible households in treatment villages received a 26 percent increase in subsidy, stemming from both an increase in quantity and a decrease in the copay price. This occurred despite imperfect implementation: eligible households in treatment villages were only 30 percentage points more likely to have received a card relative to the control. Moreover, several of the variations did in fact matter: gains were substantially larger when the information was public and prominent, when the prices were displayed on the card, and when people actually got a physical card. The coupons feature did not seem to be particularly relevant. This piece of plumbing research was influential: since the success of the cards was blatant, the government decided to scale them up and immediately started distributing them to everyone, for Raskin and other government transfer programs. In total more than 60 million beneficiaries got cards. The upside of the fact that plumbing is not a prime focus of governments is that when you hand them a good plumbing solution, it stands a better chance of being widely adopted than when you present a brand new program, which is prone to come up against someone’s ideology.

Banerjee, Duflo, Imbert, Mathew, and Pande (2016) focus on the hidden pipes. In many governments worldwide, some programs (such as local welfare programs or infrastructure building) are financed centrally but implemented by local communities. This causes a basic plumbing

²In a different setting, the government of India made the conscious decision not to necessarily issue an identity card when it established biomarker-linked unique identifiers for all Indians.

problem: how to transfer money from the center to the outposts, which are normally strapped for cash and unable to front the expenses for the programs. Traditionally, due to slow transmission both of financial orders and information, the only practical way to transfer funds is to give the local community an advance and, after they provided proof of how they used it, give them another advance. But this system is fraught with difficulties. First, there is a basic financial management issue: some money lies idle in one locale, whereas another might lack money. The advances need to be large enough to avoid paralyzing the program, but will add up to the government's expenditure. Second, and perhaps more importantly, the ex post justification of expenditures is often delayed and patchy, making auditing and accounting difficult. This creates opportunities for leakage and embezzlement, and the monitoring structure that is put in place to control it may become part of the problem, creating its own red tape, inefficiencies, and leakage opportunities (Banerjee 1997).

We had the chance to experiment with a system that was largely the brainchild of Santhosh Mathew, one of the coauthors of the study. Mathew is a rare creature—a bureaucrat-plumber. A career civil servant with a Masters in economics and a PhD in development studies, Mathew is deeply concerned with the pragmatic issues of any program he is involved with. In the context of the MGNREGS, a public workfare program financed by the Federal government but implemented at the village level, a unified financial platform was put in place in Mathew's adopted state of Bihar, whereby the money transfer could be made directly to the local government from the bank in the state capitol. This solved half of the problem, that of the transmission of money. The other half (transmission of information on who should get money) was originally not solved. Figure 1 shows the flow of authorizations under the traditional system: the local government (panchayat) transmits a request for funds to the next administrative level, the block, which ratifies it and transmits it to the next level up, the district, which ratifies it and enters it in the financial management platform, which generates an invoice to the Central Bank of India, leading to a transfer right back down onto the panchayat account. Mathew designed a reform to this system. The Panchayat now could log in directly into the financial management platform, enter an invoice

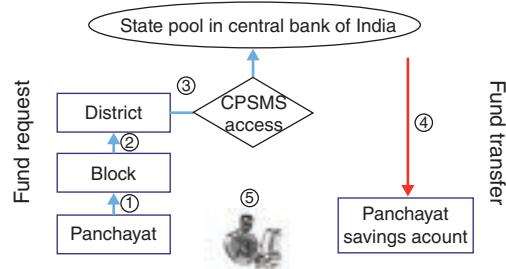


FIGURE 1. FLOW OF AUTHORIZATIONS UNDER THE TRADITIONAL SYSTEM

for a specific activity ("we have hired Mrs. Kumar for 14 days"), and receive a corresponding payment on a "zero balance account," which they would immediately use to pay Mrs. Kumar. We had the opportunity to design and implement a large-scale experiment to evaluate the impact of this reform. We worked in 120 districts, in 1,000 treatment and 2,000 control local governments, covering a population of 30 million people.

The main benefit that Mathew foresaw in this system was that it would remove the uncertainty on whether and when money would come, and he believed that it would empower local governments to be more proactive in implementing the program. As it turned out, unforeseen events meant that this could not happen in this context. A few days after the program was launched, the central government cut Bihar's MGNREGS funds, due to a dispute on data entry, and massive uncertainty ensued on when the money would be available again. This took five months to be resolved, and since it happened just after the introduction of the system, many local officials (wrongly) believed the new uncertainty they experienced on whether money would come was the new system's fault.

The plumbing reform, however, changed something else: the reformed system was considerably more transparent than the status quo. The local official was putting his signature on a document where he said that he had recruited specific people. This made it easy for auditors to go look for them, and punish the officials if they did not exist or had not worked for the program (many of the workers in MGNREGS are "ghost workers"). Although this was not the primary motivation of the reform, basic economics suggested that this could change the behavior of the local officials. We looked for evidence of a reduction in corruption, and we found plenty.

First, we saw a reduction of about 22 percent in the expenditures on the program, without apparent change in the number of workdays in an independent survey we conducted. Since the survey data could be noisy and miss a real decline in expenditure, we looked for smoking guns, and found two: first, we matched all the names reported in the public database for the program to a census that was conducted at the same time. We estimated that a match was less likely to be found for someone in the public database in the control group than in the treatment group, indicating that they were more likely to be a ghost worker. Second, we examined the wealth declarations of the Panchayat officials, and found a significant reduction in the median wealth (though we find no impact at the top and the average effect is insignificant).

Most anticorruption drives focus on explicit anticorruption tools (such as audits) or on incentives, but not on the way that the program is structured. This example showed that structure is not only relevant, but quantitatively important: during the course of the experiment, the reduction in expenditure alone saved \$6 million to the government. And as in the case of the Raskin program in Indonesia cited above, the experiment was actually instrumental in getting this program scaled up across all states, and then for the financial transfers in other types of program. Initially, before the corruption result came out, the program was canceled in Bihar: faced with a reduction in expenditure and the objections of the local officials, who for obvious reasons were not fans of the program, Mathew's successor decided to cancel it. But after it had become apparent that the reduction in expenditure stemmed mostly from a reduction in corruption, the story changed: When the Prime Minister was keen to show that his government was doing something to improve the implementation of public programs, this particular reform seemed uncontroversial enough for top bureaucrats—a “technical reform” they could get behind relatively easily. The program was scaled up to MGNREGS. It is now poised to be scaled up to other kinds of centrally sponsored schemes and has the potential to affect billions of dollars in central government spending.

II. Why Policymakers Need the Plumbers

It seems quite natural for policymakers to rely on economist-engineers to design complicated

policies and institutions. There is nothing obvious about designing a school choice mechanism, an auction, or an exchange market for kidney. Economic theory has developed the tools to think about those questions, and it stands to reason that policymakers would ask for economists' guidance, very much like economists provide advice on macroeconomic policy. It may however be less clear that policymakers need plumbers, or that economists would be particularly well suited to play this role. This section argues that policymakers do need plumbers. The next section will argue that our discipline prepares us well to think about many of the plumbing problem that they face.

Abhijit Banerjee, Gita Gopinath, and I recently spent a few hours with bureaucrats and consultants from the health department in the state of Kerala, in the south of India. Kerala is one of India's most successful states, at least as far as the social sector. It is starting to face first-world problems: an aging population, high prevalence of hypertension, diabetes, and obesity. The bureaucracy and the elected leaders are committed to try to do something about both rising health care costs and noncommunicable disease prevention. Their best idea is to reform the public primary health care system in order to make it more attractive to customers (who for the most part seek their health care in the private sector, like elsewhere in India), and to instate better practices of prevention, management, and treatment. They would try out a new organization of the health care sector. Nurses, volunteers from the local governments, and doctors would work seamlessly in a health team that would be in charge of keeping the population healthy, with a heavy focus on encouraging lifestyle changes and preventative activities. They are currently planning to try the system in 152 health centers.

Though the Additional Chief Secretary (top bureaucrat) in charge of health had invited us to the meeting, he was called away to deal with a doctors' strike around the time the conversation turned to the specifics of the reform. He handed us over to a retired professor and a retired doctor, who have been charged with designing the specifics of the policy. This in itself is symptomatic: top policymakers usually have absolutely no time for figuring out the details of a policy plan, and delegate it to “experts.” In our conversation, we started to push on some specific questions on the model that they had in mind:

why would patients pay attention to a nurse, given that until now they have only taken doctors seriously? Were they really sure that if the nurse started to take blood pressure and fill prescriptions, this would give her the authority she would need to dispense advice? Or that doctors would be willing and able to signal that nurses were to be respected, in a system that has always been heavily hierarchical? For that matter, did the planners really think it was going to be possible for health care professionals to spend a lot more time on public health and prevention when there were only 2 doctors for every 30,000 people?

What was striking was, not only did they not have any answer to these questions, but they showed no real interest in even entertaining them. Whenever we asked them to spell out what they thought their policy lever was (as opposed to their aspiration), the stock answer was that they did not really have one, that the local governments and medical officers could not be forced to do anything. Maybe a village committee would need to decide to organize yoga classes, but another one would not, so there was really no way to find out what really worked. This was entirely beside the point, since the presence (or absence) of yoga class would have been an outcome of what they could do at the central government (train local governments in the fundamental of public health, for example). But they seemed to have no understanding of a causal chain going from policy design to implementation and final outcome. Their position oscillated between presenting the illusion of the perfect system, and presenting the illusion of complete powerlessness in the face of local power and initiative.

We tried, and failed, to engage them on the details of policy. Not only did they have no understanding of plumbing issues, there was not even a realization that plumbing was an issue at all. When the top bureaucrat popped in and was apprised of the conversation, it was decided that we would be shown some details. We went away to another meeting and came back after three hours. They had set up the projector, a sign that things were about to get serious. They displayed a power point with each of the new UN Sustainable Development Goals, and a list of proposals to achieve them in Kerala. These amounted to a long, meritorious, and likely totally vacuous wish list (30 minutes of exercise per day mandatory in all schools, awareness

of obesity to be built in communities, etc.). Interestingly enough, it had nothing to do with the health care reform that we had discussed that morning. It appeared that the details would have to wait for another day.

I have since learned to avoid these kinds of meetings in general, but this encounter reminded me of my early days as a plumber. It turns out that most policymakers, and most bureaucrats, are not very good plumbers.

Part of this is because, contrary to most economists' generous beliefs, agents in general are not necessarily very skilled at what they do. This is the central argument of Banerjee (2002) in favor of normative economics. He points out that agents have not necessarily had the opportunity to experiment, even in their main line of trade. For example, farmers may not have stumbled on the best mix of fertilizer, because once they have something that works well enough, it may be optimal to stop experimenting. Even if they are inclined to experiment, the results may not be informative if the outcome data is very noisy, and learning from neighbors is difficult. This leaves a lot of scope for specialized expertise, including that of economists. Banerjee focuses on the example of microcredit, and the fact that, before the microfinance movement, bankers did not figure out how they could use some basic principles of mechanism design to make it possible to lend to the poor.

Hanna, Mullainathan, and Schwartzstein (2014) emphasize another barrier to learning: the tendency, even for experienced agents, not to notice some of the important features of the environment. They test a model of selective attention with seaweed farmers in Indonesia, which could just as well apply to the plight of policy practitioners. Seaweed is farmed by attaching strands (pods) to lines in the ocean. A large number of dimensions affect yield (e.g., length of the line, optimal distance between pods, optimal length of a cycle, size of the pod). The farmers have limited attention, and can only attend to a limited number of dimensions. They allocate their attention optimally, but a feedback loop may occur: if they start with the prior that something is not important, then they don't pay any attention to it, and they may never notice that it is in fact important. The authors present experimental evidence that farmers not only ignore some dimensions of farming, they continue to ignore them even when they are presented data

suggesting that they actually matter, unless they are told very specifically to pay attention to them. In this instance, they tend to ignore the size of the pod, and in interpreting the trial data, they do not think to compare trial by pod size. Very experienced farmers can get it terribly wrong, even when experimenting is not particularly costly, simply because they do not know what to look at. Very experienced policymakers often get it just as wrong.

This inability to notice details is maybe even stronger for the economist-scientists and the policy practitioners than for farmers. The economist-scientist is effectively trained to ignore them: the art of modeling requires simplifying reality to work out the logic of some essential assumptions. The spotlight is on general principles, not on details that may be very specific to an environment—indeed, such details are distractions. In order to focus on the general principles, the economist-scientist will tend to dismiss those nasty details (which in fact will make the policy work or not). This applies to theorists, but to us field experimenters as well, when we wear our scientists' hats. At those times, we run experiments that are as tightly controlled as possible to try to isolate a specific mechanism or test a theoretical principle. This is not a shortcoming, just a feature of the scientist's work, yet it means that scientists will not typically be the ones able to identify the key issues that hinder policy success. The policy practitioner is often not even willing to try to acknowledge details.

Many are, to one degree or another, afflicted by the specific blindness that affects the “high modernist” state described by James Scott in *Seeing Like a State* (Scott 1998). Policymakers and bureaucrats have a tendency to simplify the problems that face them until they fit some pre-conceived notion of what a human being should want, or need, or be. Scott’s book is mainly concerned with episodes where this high modernist approach went particularly awry (including the Soviet period and the Villagization policy in Tanzania), but the temptation to regulate people into what they want them to be affects policymakers more generally.

A striking recent example is the other unexpected and potentially calamitous event that took place on November 8, 2016. On that day, the prime minister of India, Narendra Modi, announced that, effective immediately, all

existing Rs500 and Rs1,000 notes would be replaced by new notes, and would become illegal tender by the end of December. The initial stated objective was to cut corruption by getting rid of black money, though other goals were mentioned later: making India cashless, fighting terrorists by removing their means of payment, creating a windfall for the government in the case that less money came back than was removed from circulation. What was absolutely remarkable is that this major move was made with zero preparation beforehand. The central bank had not printed enough money, creating a cash shortage and huge bank queues. The new money was a different size and most ATMs had to be retrofitted. Cash deposits in poor people’s “no frills” accounts skyrocketed, presumably because someone had given them the money for safekeeping. Between the time the policy was announced and the end of December, the government issued and changed more than 100 different rules concerning when individuals could deposit their cash, how much they could deposit, what kind of justification was needed, when they could withdraw, how much they would have to pay in penalty if they could not prove they had paid taxes and so on and so forth. The result was a combination of a flabbergasted public, a massive disruption in the economy and, in short, complete confusion.

The idea that you could remove 86 percent of the cash in an economy without first thinking through the logistics, including what would replace it (considering that mobile money, credit cards, and account-to-account transfers were not particularly well developed in India) is of course extreme. But it expresses the kind of magical thinking that many governments indulge in—and not just in developing countries—that if you have a generally good idea that makes sense in some abstract way, the details will work themselves out.³

The failure to experiment, the failure to notice, and the grip of ideology combine in what Banerjee and Duflo (2012) described as the “Three I’s” (ideology, ignorance, inertia). Policymakers tend to design schemes based on the ideology of the time, in complete ignorance

³An interesting side question is why bureaucrats cannot be given incentives to get the detail of the policy right, since unlike policymakers they are not elected and should not be wedded to a particular viewpoint and ideology.

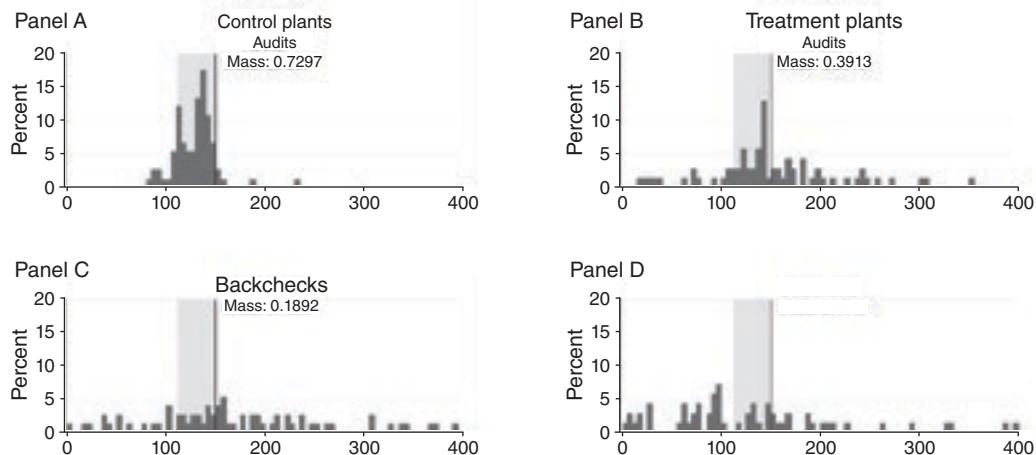


FIGURE 2. AUDIT READINGS FOR SUSPENDED PARTICULATE MATTER (SPM)

of the reality of the field, and once these policies are in place, they just stay in place.

There are two implications of this attitude of policymakers, a negative one and a more positive one. On the negative side, it means that many well-intentioned policies fail in practice (or don't succeed as well as they could have) even if their underlying principles were sound, because the details are not gotten right. Examples are numerous. Despite the considerable amount of political capital that was expended to launch the Affordable Care Act in the United States, and the importance of what was at stake, the roll-out of the health exchanges was plagued by entirely avoidable software glitches. Another example is the third-party audit system that the Supreme Court of Gujarat mandated in an attempt to improve the monitoring of industrial pollution (Duflo et al. 2013). Under the scheme, each firm with high pollution potential had to hire a private firm to complete a yearly auditing report. Each firm selected and paid its own auditor. Auditors completed three visits and submitted the resulting report to the government regulating body, the Gujarat Pollution Control Board. A number of rules were in place to ensure quality and limit corruption: auditing teams could not sign up more than a certain number of clients, there were qualification requirements for auditing team members, and dishonest auditors risked decertification. Yet, the system was a complete failure. As part of a randomized evaluation,

we performed backchecks on the auditors by sending an independent team to take pollution measurements right after them. The results for one particular pollutant (Suspended Particulate Matter) are reproduced in Figure 2. The auditors' reports show considerable bunching right below the threshold for acceptable pollution, which is completely absent in the backcheck. Even low pollution measurements, which are apparent in the backcheck, are absent in the auditors' readings. It seems that, in many cases, they were not even visiting the firms. This was reflected in the going rate for auditors' report, which was below the minimum cost of running an actual pollution test. The likely explanation is that the system was plagued by conflicts of interest: since the firms were selecting and paying the auditors, the auditors had every reason to give them the report they wanted, if possible for very little money. Although the system was designed with worthy general principles in mind, basic design flaws made it essentially useless.

On the positive side, policymakers' inattention to detail creates a space for policy improvements even in environments where the policy discussion is paralyzed, or where institutions are not good (Banerjee and Duflo 2012). When there are existing laws on the books, finding ways to make them work better meets with little opposition. In the United States, even as Congress and the executive branch were fighting over every single line item in the budget, millions

more children were certified between 2008 and 2014 as part of the school lunch program by successive improvements in direct certification for children of parents who benefited from other programs: SNAP, TANF, or Medicaid (Moore, Conway, and Kyler 2012; Moore et al. 2013). The Child Nutrition and WIC Reauthorization Act of 2004 mandated direct certification, but rate of enrollment accelerated greatly after the Healthy, Hunger-free Kids Act of 2010. The act contained a set of “plumbing” reforms that led to significant increases in direct certification: mandating name matching of different databases three times a year; removing the option of “letter certification,” where a letter was sent to the parents which had to be forwarded to the school; establishing awards for the best-performing states. In 2011–2012, 1.7 million children were directly certified (a 17 percent increase from the previous year), and 740,000 were added the following year (Moore, Conway, and Kyler 2012; Moore et al. 2013). This did not raise eyebrows in Congress, and probably did more to help kids in poverty than a more publicized measure would have.

III. Why Economists Make Good Plumbers

All of these arguments suggest that we need someone smart to take care of the plumbing. But why does it need to be an economist?

The reason is that many plumbing issues are actually fundamentally economists’ problems. Not all economists need to be plumbers; we need scientists and engineers too. And not all plumbers need to be economists. Sometimes, what a policymaker needs is a good software engineer, lawyer, or subject expert. But economics, as a discipline, provides domain expertise for some of the fundamental plumbing issues in public policy.

There are three broad classes of reasons why plumbing matters for public policy, and all three involve economics.

First, the citizens (or the supposed beneficiaries of the policies) are humans, with conflicting objectives, limited information sets, limited attention, and limited willpower. This means that the specific way that policies are presented and implemented will potentially have tremendous influence on whether they will work or not. Therefore, when a new policy is put in place, it is essential to think about those details. For

example, if households receive a cash transfer, it may matter whether the woman or the man of the household receives the money (Field et al. 2016; Benhassine et al. 2015; Almås et al. 2015). It may matter if the transfer is presented as something that can help children go to school (Benhassine et al. 2015), or if households are reminded that they can save it, and so on.

The success of the nudge agenda, inspired by the work of behavioral economists, and popularized by Thaler and Sunstein (2008) has given prominence to “design of the tap” issues among economists. Indeed, the last two Ely lectures centered on this issue. Campbell (2016) argued that the regulation of household finance must take into account the fact that households do not understand finance very well, and also that they are subject to behavioral biases. Chetty (2015) argued that the “pragmatist” public finance will pay attention to those issues regardless of their scientific foundations, because they do lead to greater success in what the policymakers are trying to do. It is not just economists advocating subtle behavioral intervention. Policymakers are also aware of them, at least in the United States and United Kingdom, where “Nudge Units” have been created in the White House and at 10 Downing Street. The main idea behind “choice architecture” is precisely that what may appear to be irrelevant details (default options, ease of making a decision, the way some numbers are presented, etc.) actually influence final consumer choices. Policymakers therefore need to be thoughtful about all those choices. (It is worth noting that many tap-design issues don’t have anything to do with behavioral economics: the issues explored by Banerjee et al. in Indonesia’s Raskin program have much more to do with political economy and corruption, for example).

The economist’s job does not end once policymakers are made aware that these tap-design issues matter, however, particularly because policymakers love a free lunch, and some of the behavioral ideas may stroke this disposition. They are thus wont to treat new behavioral ideas as general “solutions to every problem,” rather than as tools in a plumber’s kit, perhaps necessary to tune any kind of policy properly. A plumber’s mindset is required to carefully watch how these ideas play out in practice, and sometimes an economist may be need to analyze how it works. We will see several examples of cases where “obvious” improvements did not turn out

to be improvements at all. Ben Handel's and Jonathan Kolstad's work on health insurance illustrates the subtle economic analysis necessary to analyze the plumbing of a comprehensive health reform such as the Affordable Care Act (a.k.a. Obamacare). They are concerned with issues of practical relevance: in 2015, they wrote a policy document with two specific proposals for the health exchanges (Handel and Kolstad 2015), a consumer search tool that personalizes choice framing and recommendations based on an individual's specific characteristics, and a "smart default" policy, targeted to particular consumers, to which the government could default consumers during open enrollment when considerable evidence indicates that they are in the wrong plan. Those proposals are quite detailed, and they draw from a substantial academic literature (to which Handel and Kolstad have contributed before). Because the proposals are realistic, they need to tackle issues that are beyond the simple observations that consumers are ill informed and tend to be characterized by inertia (although this clearly is part of the problem). For example, Handel (2013) observes that, due to inertia, many consumers remain in dominated plans, where they lose a substantial amount of money (consumers leave approximately \$2,000 on the table, on average). However, Handel also finds that, if consumer inertia is reduced, adverse selection will likely increase in a marketplace with no insurer risk-adjustment transfers: this means that any proposal that helps consumer solve the inertia issue needs to address the adverse selection issue at the same time.

Second, those who implement policies, often government workers, are humans too. They may suffer from the same limitations as the final subjects of the policy, and their incentives are not necessarily to work very hard or in the best interests of the citizens they serve. In much the same way as plumbing details create a choice architecture for the final consumers, they also create an "incentive architecture" for the government workers who implement the policies. Despite the current vogue of catchphrases such as "state capacity," these problems have received fairly little attention in economics until relatively recently. There are, however, several recent interesting experiments on the "personnel economics" of the state (see a review in Finan, Olken, and Pande 2015). Indeed, the decisions

on how to hire, fire, and reward government employees creates the basic environment in which government work is happening.

Consider what is involved in the decision to hire new front-line workers (community liaison workers, health workers) to carry out a new program. While the bureaucracy will generally be quite good at determining how many new workers are needed and may be considering how to set their wages, it will almost surely not be thinking about what the job-announcement posters should advertise. And yet, Ashraf, Bandiera, and Lee (2015) show that a small change in the framing of a front-line health worker job on the advertisement poster—namely, emphasizing the career path that the job can lead to rather than the opportunity to do good for the community—makes a large difference on the observable and unobservable characteristics of those who apply. The "go-getters" who applied for the job in the career-path framing treatment conduct more visits, accomplish more during visits, and ultimately lead to better health outcomes than the "do-gooders." The overall goal of this new program was to create a new layer of health workers close to people in the community; however, the seemingly minor advertisement decision turns out to have very significant effects.

We are far from understanding how this plays out. In contrast to the results just discussed, Deserranno (2015) finds that financial incentives can lead to a less socially motivated applicant pool. She randomized the advertisements for a new position of health promoter for BRAC, a large international NGO. In one treatment arm, the job advertisement mentioned the minimum amount that a health promoter was expected to earn ("low-pay" treatment) and, in another treatment arm, the maximum ("high-pay" treatment). A third treatment advertised the mean of the expected earnings distribution ("medium-pay" treatment). Note that the difference was only in the framing: the actual pay was the same in all treatments. The study finds that while the high-pay treatment attracted 30 percent more applicants relative to the low-pay treatment, the applicants had much less experience as health volunteers and were much more likely to state "earning money" as the most important feature of the job. Applicants under the medium- and high-pay treatments were also 24 percentage points less likely to make a donation to a public health NGO in the context of a dictator game.

Some aspects of personnel economics of the state such as salary and incentives are very salient and politically charged, and I would not characterize them as plumbing.⁴ Some are much less salient, such as the framing of the positions (as described above) or transfer policy. Transfers and posting to more or less desirable locations are routinely used by the governments of developing countries to reward performance or punish people who are not liked, but this is not done in a systematic way. Banerjee et al. (2012a) find some positive effects of reducing the frequency of transfers of police officers on their performance. Banerjee et al. (2012b) find that the promise of a transfer out of a punishment posting induced police officers to work harder on conducting random sobriety checkpoints. Khan, Khwaja, and Olken (2016a) propose, implement, and test one particular system—performance-ranked serial dictatorship—in which tax inspectors select their most favored location in an order that is determined by performance. The authors show that the mechanism does in fact work, increasing tax collection by 40 to 80 percent. Moreover, the people who, *ex ante*, their model predicts will be the most sensitive to this incentive are in fact those who increase their collection the most. But note that none of these studies investigate an essential aspect of this policy in equilibrium: if transfers are used as rewards, or if postings are more secure, the most corrupt people may work very hard to get transferred to where they want to be. This could entirely change the welfare implication of the policy, so this is an area where more plumbing is needed.

But importantly, the incentive infrastructure issues apply to the design of any policy and regulation, not just incentives policy specifically or even personnel economics more generally. Regardless of whether the policymaker is drafting an education intervention, a banking regulation, a new tax, or anything else, there will be decisions to be made on who will carry it out, under what supervision, with what level of autonomy, and how implementation will interact with their other duties. Any particular set of

rules will, advertently or inadvertently, affect the ability and willingness of the front-line workers to implement the policy. Mindful incentive infrastructure requires understanding the government as an organization: any policy will necessarily take place in an organization that has power structures, and a culture that has large impacts on how the policy will play out. The specific way it is drafted, even the parts of it that do not formally address incentives, will determine its ability to be implemented (Gibbons 2010). This gives particular salience to the piping issues, which, having not benefited from the behavioral economics revival, are even more neglected than the tap-design issues.

Between 2001 and 2014, a group of us (Banerjee et al. 2016) experienced firsthand the power of incentive architecture as we attempted to mainstream a remedial education program through the government system. The basic idea of the intervention was that children must be taught at the level at which they currently are, rather than at the level prescribed by the curriculum for their grade. Several RCTs showed that it could be implemented by volunteers or paid NGO workers to very good effect, and it had been shown to raise learning outcomes when run by Pratham, a large NGO, outside of school. To mainstream it within schools, Pratham trained teachers in the pedagogy. We evaluated this attempt, as part of a large scale-up of the policy in the Indian states of Bihar and Uttarakhand over the two school years of 2008–2009 and 2009–2010. The results were disappointing: despite official endorsement, teachers did not seem to apply the methodology when they were back to their classrooms, and the intervention did not result in any increase in test scores. Our process monitoring and a set of qualitative interviews suggested that there was an incentive-architecture problem in the way Pratham and its partner government had implemented the program. While there was endorsement of the approach at the top, nothing else in the teachers' lives had really changed. In particular, their stated responsibility remained to complete the curriculum, and from their point of view, any time devoted to the remedial activities was essentially wasted.

To have a chance of success, something needed to be change, not in the basic philosophy of the program or in the content of the training, but in the plumbing. Working with the state of Haryana, Pratham implemented two key changes

⁴ Although they are extremely important, and the subject of important field experiments: see, for example, Dal Bó, Finan, and Rossi (2013) on front-line worker recruitment and Khan, Khwaja, and Olken (2016b) on financial incentives for tax collectors.

to the way the intervention was run. First, the program would be implemented during an extra hour in the day (added in all schools, not just those where the program was implemented, but earmarked to this in treatment schools). Since this hour was new, it was easy to make salient that the responsibility of the teachers was to carry out the remedial activity during that time. Second, a cadre of supervisors (from within the government system) were recruited and trained to act as both supporting people and monitors for the teachers as they were implementing the program. Nothing changed in a teacher's pay or her formal role, but these two adjustments made the difference between signaling that the program was an optional add-on or part of the teacher's core duty. We evaluated this version of the program in an RCT in 400 schools (200 treatment, 200 control), and found significant increases in test scores.

Finally, any regulation that seeks to affect the way that firms behave, or any policy that relies on firms for implementation at any point must also take firm behavior into account. This behavior is one reason why "obvious" solutions do not always work the way they would be expected to. Willis (2013) makes this point in the context of default enrollment in an overdraft protection plan (which Campbell 2016 describes as one of the signature attempts to incorporate behavioral economics into the regulation of household finance products). Defaults are settings or rules about how product, policies, or legal relationships function that are in place unless users actively change them. Based on their success in retirement savings in particular (Madrian and Shea 2001; Chetty et al. 2014), default-in programs have been come popular with policymakers. Since people tend to stick with the option they are presented with, policymakers can get the outcome they want by setting the default carefully, while leaving people the freedom of choice (this is what Sunstein and Thaler have called "libertarian paternalism"). When drafting policy, picking a default is often unavoidable, and hence, regardless of whether one is convinced of their stickiness, choosing the right default is always a plumber's problem. But in some cases, policymakers can attempt to influence the relationship between consumers and firms by mandating that the firm offers a particular default option. In the consumer finance area, legally required default terms have

been put in place for checking account overdraft coverage, credit card over-the-limit practices, and some home-mortgage escrows. In particular, a 2010 regulation imposed on banks a policy default regarding overdraft coverage. Under this regulation, banks were not allowed to charge an overdraft fee for when an account went into a negative balance due to an ATM transaction or a nonrecurrent debit card transaction unless the consumer had opted out of the default and chosen the bank overdraft protection plan. The idea was that the overdraft coverage is basically a high-interest, low-risk loan charged by bank and is not in the interest of most customers. Banks get away with high overdraft coverage fees because customers do not expect to go into overdraft, and thus pay no attention to this feature when they sign up for an account. Moreover, they don't necessarily realize when they are going into overdraft (Stango and Zinman 2014).

The 2010 regulation required banks to offer no overdraft protection by default, and prohibited them from charging overdraft fees in the default if the account ended up being overdrawn anyway.⁵ In doing so, the regulators cited the research on the stickiness of the default, and expressed the hope that most consumers would stick to the new default of no overdraft protection. To enhance the stickiness of the default, the regulator set some safeguards: to get out of the default, customer needed to take an "affirmative choice" (talk to someone or click a box), and before allowing them to do so, the bank needed to show them some documents.

Thus, the regulator had taken care of the obvious plumbing concerns they could think of. And yet, the default did not work that well. Willis (2013) cites data from the Consumer Finance Protection Bureau suggesting that around one-half of the bank's clients had opted out of the default, and that those who switched were those more likely to then pay high overdraft fees. Banks achieved that by using any number of strategies that Willis describes in detail, such as minimizing the costs of opting out as much as the law allowed them to, stationing employees at ATMs to convince people to opt out, using forms for new clients that made sticking with

⁵This applied to ATM or one-time transactions only. Banks could still authorize and charge for overdraft occasioned by checks or recurring transactions.

the default just as complicated as opting out, and using language that made the default scary and the alternative attractive.

The bottom line is that, if the objective of the government was to protect customers from the bank's abusive use of overdraft policy, then the policy should have been drafted with an understanding of the economics of the bank's likely response. This would have probably led to a different set of rules. The 2009 Credit Card Accountability Responsibility and Disclosure (CARD) Act was less respectful of the banks' freedom. It contained some "nudges," but also hard ceilings on the late fees that banks were allowed to charge. Agarwal et al. (2015) find important effects of the CARD Act on interest and fees paid by customers (as the ceilings were binding on some products and were not offset by increases on other products), and relatively little change of behavior related to the nudge.

Economists have some domain expertise to understand those dynamics, and therefore can help draft policies that take those incentives into account. Duflo et al. (2013) not only studied the (faulty) third-party audit system for environment that the court had ordered in Gujarat, but set out to improve it. Given what we know from economics, it was not rocket science to identify the key problem: since the system was plagued by conflicts of interest, the performance of the auditors could be expected to improve if the financial link between them and the firm was severed, and if they were instead made liable to the regulator, who would then reward them for accuracy. Similar proposals have been made for reforming financial audits in the United States and Europe, though they have not been implemented. The Gujarat Pollution Control Board (the regulator in charge) implemented a reformed system, where auditors were randomly assigned to firms, paid from a central pool, and independently monitored in a randomized experiment in 473 firms (233 of which were assigned to the new system). Compared to audits conducted under the status quo (even with the same auditor operating in both systems), we found those "reformed" audits to be significantly more accurate. This additional accuracy led to a reduction of pollution among the worst offending plants. This work inspired a reform of the entire third party audit system in Gujarat, once again demonstrating the convincing power of plumbing.

These two examples may give a somewhat too simplistic view of the problem: firms, like governments, are organizations themselves and may thus not always respond optimally to a change in the environment. Taking into account how a regulation will affect firms as organizations further underscores the role of details rules and how they interact with the environment.

To summarize, economists have the disciplinary training to make good plumbers: economics trains us in behavioral science, incentives issues, and firm behavior; it also gives us an understanding of both governments and firms as organizations, though more work probably remains to be done there. We economists are even equipped to think about market equilibrium consequences of apparently small changes. This comparative advantage, along with the importance of getting these issues right, makes it a responsibility for our profession to engage with the world on those terms.

IV. Plumbing and Experimentation

Roth (2002, p. 1342) made the point that "in the service of design, [lab] experimental and computational economics are natural complements to game theory." Lab experiments and computational work are essential tools for the economist-engineer because sometimes real life will not oblige with the nice assumptions that are necessary for elegant models with closed-form solutions. If that is the case, simulations are needed to understand the behavior of the proposed systems under the more complicated set of assumptions.

Similarly, in the service of fitting policies for the real world, field experiments are a natural complement to theory and economic intuition.

Evaluation in the field is necessary for the plumbers because intuition, however sophisticated and however well grounded in existing theory and prior related evidence, is often a very poor guide of what will happen in reality. Therefore, the economist-plumber is eminently fallible. We have already seen several interventions that did not initially succeed, despite being designed as well as possible, and with economists' inputs (e.g., the 2010 default regulation to prevent the banks from using overdraft policy, and the first attempt to scale up Pratham's "teaching at the right level" intervention). My career is peppered with such examples, and

occasionally, the best-intentioned efforts have backfired.

In Banerjee, Duflo, and Glennerster (2008), we helped the local government in Udaipur district, Rajasthan, design an incentive program to get nurses to come to work more often (a baseline survey had shown that they were absent more than one-half of the days). Each nurse was given a time-and-date stamp and asked to stamp a register affixed to the wall of the center several times in the course of a day, one day a week (on Monday) to prove her presence. The government announced that those who didn't show up at least 50 percent of the time on Mondays would get their wages docked. This policy was evaluated in an RCT. Initially, everything went according to plan: nurse attendance, which was around 30 percent before the launch of the program, jumped to 60 percent after the program was introduced in centers where it was in place, while it remained unchanged elsewhere. However, a few months later, the tide turned. Nurse attendance in the monitored centers started to drop, and kept dropping. Eight months into the program, the monitored and unmonitored centers were performing at an exactly equal—and very poor—level. When we looked into what happened, we were struck by the fact that recorded absence remained low even after the program fell apart. What went up sharply were “exempt days”—days when the nurses claimed that they were not required to come in for some reason (training and meetings were the most common reasons). There seems to have been a strong collusion between the nurses and the doctors in charge of supervising them. At some point, attendance in the monitored centers actually fell below the unmonitored ones and remained lower until the end of the study: by the end, the nurses in monitored centers came to work only 25 percent of the time. We had completely failed to foresee how this collusion would undermine the program, and would in fact make the situation worse than it had been before, as it dawned on nurses that they could get away with anything.

For some people, this very fallibility is a reason why we should not intervene at all. But everybody makes mistakes, so this is hardly an excuse for inaction (Banerjee 2002). However, because the economist-plumber intervenes in the real world, she has a *responsibility* to assess the effects of whatever manipulation she was

involved with, as rigorously as possible, and help correct the course: the economist-plumber needs to persistently experiment and repeat the cycle of trying something out, observing, tinkering, trying again.

This underscores the need for evaluation, but of course, randomized controlled trials are not the only way to evaluate policies. For example, Handel and Kolstad's work, which I discussed, largely exploit nonrandomized situations, although they have some experimental papers as well. Conversely, most RCTs are not plumbing experiments—they are designed and implemented carefully with as little disturbance as possible, to learn a specific parameter or measure the effect of one particular intervention. They are not concerned with the specifics of how such an intervention would be implemented at scale.

Still, randomized field experiments are an economist-plumber's natural tool, for several reasons.

First, there are the traditional identification arguments: it is generally difficult to form a proper counterfactual when a policy was launched in the world, and hence we may not know the real effect of many policy attempts.

Second, precisely because policymakers are not particularly interested in plumbing, they will in general make a large number of decisions at the same time. Even when it is possible to evaluate one policy or one regulation with a credible identification design, it may be difficult to “unpack.” For example Agarwal et al. (2015) estimate the impact of the 2009 CARD Act, which regulated the fees that banks could charge on consumer business cards, but not business credit cards, and also had some nudging provisions. They compare fees and interest paid before and after the acts on consumer cards versus business cards. This gives us a convincing estimate of the effect of the CARD Act taken as a whole, but it is a little difficult to separate out the impact of its different provisions. In contrast, with a carefully designed experiment, it is possible to manipulate one lever at a time. Thus, randomized experiments are a very natural complement to policy experimentation. This is nicely illustrated by the Banerjee et al. paper on the Raskin information card, to name just one example.

Third, an explicit experiment, with a start date and an end date by which a report must

be provided and a decision made, has important benefits for the process of government. Greenstone (2009) argues for a “culture of persistent regulatory experimentation and evaluation,” where every regulation should be subject to experimentation, and an automatic sunset clause (by which it must be reviewed to determine its costs and benefits during the experimental period) and an expansion clause (by which it will be expanded if it turns out to have large benefits and low costs). Greenstone argues that, in the United States, there are opportunities to reform the system for evaluating, adopting, and getting rid of policies. Most evaluations are done *ex ante*, with very little relevant data to determine the cost and benefits, and very little serious review *ex post*. Almost no federal regulation comes with a sunset clause, although they are used more frequently by states (Baugus and Bose 2015). Partly inspired by this article and by Greenstone himself (who hung up his professor’s gown for a time to engage in full-time plumbing), the Obama administration called for a “look back,” a comprehensive review of all the costs and benefits of regulations in place by all agencies (Executive Order 13563, 2011). This was a bold move, but without proper data available and without any counterfactual, the exercise had a narrower scope than was originally envisioned. Moreover, inertia is deeply ingrained in governments everywhere. Once an apparatus is installed to support a policy or a regulation, it is unlikely to go away. The case is different if the apparatus is the explicit subject of an experiment, with an anticipation of a review to come. Of course, what matters for this political economy process is that the process is viewed as experimental, not necessarily that the experiment be randomized. But once a regulator agrees to try something out (perhaps different variations of a program) on an experimental basis, the marginal cost of randomizing is often low, and the results of an RCT are more transparent and easier to explain.⁶

Fourth, the process of designing a randomized experiment in conjunction with a government is an important learning process, because

it forces the policymakers to think about the levers that they actually have control over. In our conversation with partners from the Kerala health department, which I recounted above, it was apparent that they were very confused as to what they could and could not do as a policy. On the one hand, they had an ideal vision of the nurse and the village officials working together to be key dispensers of health advice. On the other hand, they resisted any notion that they could actually tell local governments what to do. If they had to design an experiment, they would have needed to spell out what was in fact in their control (training the panchayats? Training doctors and nurses in a set of new guidelines and roles and responsibilities?). This would actually have forced them to write down a policy, rather than keep it suspended at the purely aspirational level.

Learning happens in the course of the experiment itself as well. With Banerjee et al. (2012a), we conducted a first set of randomized experiments, designed to improve the behavior of police officers with ordinary citizens in the field. The Rajasthan police top brass were very committed to the goal of improving the force, and ready to immediately implement reforms that we would have suggested. We convinced them to start with an experiment to test out some of the main proposals made by the various police commissions in India: training police officers, introducing community observers, rationalizing the workweek with a day off and a duty roster, and limiting the frequency of transfers. Even though this was a personnel experiment carried out on a large scale (more than 100 police stations, covering a population of over eight million people) and with a government, this was not a plumbing experiment, in that we did not pay much attention to the details of how each intervention was going to be carried out. The reason is that our partners in the police did not think that plumbing would ever be an issue at all. The police in India is structured like a military organization, and the top brass assumed that if they ordered something, it would happen. So they simply sent the order to carry out these different activities down the hierarchy. Unfortunately, this was not exactly what happened. Some of the reforms (the transfer freeze and the training) could be more or less directly implemented from the top, and were indeed carried out. But others (the rationalization of the workweek, and the

⁶Greenstone’s agenda found some resonance in the Obama administration. A 2013 memo from the Office of Management and Budget (Burwell et al. 2013) encouraged all heads of agencies to use existing evidence to draft new rules, but also to conduct randomized experiments.

community observer) relied on the station masters for implementation. If they had been carried out, they would have caused a direct limitation on the autonomy of those very station masters. So, with the benefit of hindsight, it is not entirely surprising that those measures were in fact never actually implemented in the field: neat reports made their way up, but when we sent monitors to the field, they found that nothing had changed in practice. Collectively, we entirely neglected the plumbing when we were designing the reforms and their implementation (and accordingly, we probably tested out reforms that were unworkable, although they had been recommended by various reform commissions), because we had a vision of the organization that turned out to be entirely different to what it was in practice. But this project was carried out as an RCT, rather than just rolled out (which had been our police partner's initial idea), so we were at least able to realize what we had gotten wrong: not only because we saw no change in the final outcome, but also because we had put in place a comprehensive data collection apparatus to see what was happening on the ground.

While we never got to learn whether the rationalization of the workweek or community observers could have worked if they had been implemented, we learned a fair amount about what kind of interventions were possible to carry out with the police. We took this lesson to the next project we carried out together, when Nina Singh got the responsibility to design an anti-drunk-driving project (Banerjee et al. 2012b). Some of the issues that we were concerned with are central to policing, and related to quite subtle questions of economics: should enforcement be always in a fixed place or vary locations? Should it follow a crackdown model or be at a lower level for a long period of time? We were interested in setting up an experiment to learn that. Based on the previous experiment, however, neither the researchers nor the police officials were ready to assume that traffic checkpoints to catch offenders would regularly happen at the specified time and place. So we thought hard about the police as an organization, with the constraints that it faces (in particular, the prohibition of merit pay) in order to find a space where it was possible to provide incentives. We formed drunk-driving enforcement squads out of officers assigned in a certain type of station perceived as a punishment posting. We promised that good behavior on this

squad would get them a ticket out of the doghouse. We fitted the police cars of this brigade with GPS, which allowed us to monitor whether they were in fact doing their job. Compared to regular police stations, officers in those squads were in fact more likely to be at their checkpoint and arrested more drunks. Thus, the first experiment gave the police enough understanding of what was under their control in their organization and what was not, information that allowed them to better design a new policy.

V. The Plumber and the Scientist

The knowledge that is generated in the course of a plumbing experiment is often quite context-specific (to a particular organization and institutions). One concern is that it may not generalize beyond this setting. In this case, even if a randomized experiment is used to pick the version of the policy that actually works, is it still science, or is it just glorified consulting? After all, firms also use randomized controlled trials to choose the price of their goods, the color of their yogurt packets, or what advertisement to feature on a landing page. The number and the scale of these experiments have exploded with the advent of the Internet and A/B testing. Many of these marketing experiments are of use to the companies that run them, but are not treated as research, do not receive IRB approval, and are never published. When they are (see Simester forthcoming for a review) it is because they manage to make a more general point on consumer behavior. Are plumbing experiments able to help us draw similarly able to help us draw general conclusions, valid in other contexts as well?

There are several answers to this criticism, which sets apart policy plumbing from usual A/B testing.

First, plumbing experimentation can create the counterfactual that allows for the evaluation of the policy itself. Many policies are implemented nationwide, which makes it difficult to evaluate them. Plumbing experiments can help by creating variation in how well they are implemented. This can generate the variation sufficient to evaluate the causal impact of the policy itself, even when it has already been rolled out. Consider Devoto et al. (2012), the experiment on water connection. By fixing some “design of taps” issues, and removing the barriers to access

to the loan (and hence the water connections), they generated an exogenous increase in water access in a randomly selected set of households (69 percent take-up for treatment households, versus about 0 in control households), which gave them the opportunity to evaluate the impact of access to water at home on health and household welfare.

In addition, a particular feature of plumbing experiments is that, because they can be carried out with governments, and can therefore be done on a large scale, they give us a window to evaluate the equilibrium effect of policies, which is otherwise very difficult to assess and model.⁷ For example, Muralidharan, Niehaus, and Sukhtankar (2016a) evaluate the impact of biometrically authenticated payment in the MGNREGS program in Andhra Pradesh. The new system replaced the traditional method of paying beneficiaries through their postal account, which required them to travel to the post office and often pay a bribe to an agent if they wanted to get paid. In the new system, they could withdraw money from an ATM in the village staffed by a local woman, using their fingerprint for identification. The rest of the implementation of the MGNREGS program (including how the money got to the Panchayat account) was not changed. The authors found that treatment improved the implementation of the program in many dimensions. In particular it reduced leakages and delays, leading to higher participation in the program. Importantly, the randomization was carried out at the subdistrict level (a group of 20–25 villages, covering over 60,000 people). This allowed the authors, in follow-up work (Muralidharan, Niehaus, and Sukhtankar 2016b), to study the impact of the program on wages. They found that their intervention led to an increase in private sector wages without reduction in employment. This general equilibrium impact is quantitatively important: 90 percent of the increase in individual's income due to the intervention could be attributed to the private wage increase. The question of the equilibrium impact of workfare programs like the MGNREGS is of central interest for the design of social safety nets. Of course, to generalize the finding beyond Andhra Pradesh, or beyond this

particular improvement, we would only need to be willing to assume that other reasons for intensive margin increase in program participation would have a similar effect on wages in the private sector. That assumption is more palatable than the assumption that smartcards themselves would work well in any context; this is a statement about the functioning of the private labor market rather than about the government as an organization.

Second, the plumbing experiments have also generated insights that are useful to pure science.

A first relationship between plumbing and pure science is that plumbing experiments can be designed to test mechanisms, precisely because the researcher can operate very specific levers. Thus, to employ the terminology of Ludwig, Kling, and Mullainathan (2011) and Congdon et al. (2017), plumbing experiments can be both mechanism and policy experiments. To be sure, this is not a specific advantage of plumbers. The most common use of field experiments these days is a careful test of a theoretical proposition, and this use of experiment as "science" usually takes place in a better controlled environment (see Banerjee and Duflo 2009 for a discussion of this use of experiments, and Glennerster 2017 for practical trade-offs between control and relevance that arise when running experiments). But plumbing does not preclude science, and presents some advantages. First, the policy context is usually not replicable outside of this policy environment. Second, the relative indifference of governments to plumbing questions often translates into a willingness to give the plumbers a relatively free reign on the design. Third, the scale also allows many subtreatments without losing power. The plumbers can then turn to their scientist friends (or to the scientist within) to design quite subtle "mechanism experiments," moving one parameter at a time to test a very specific scientific theory, in the policy-relevant context, at scale, and on a representative population. In the Banerjee et al. (2015) experiment on the Raskin card, for example, there were several manipulations that permit quite precise evaluation of the mechanism at play. One of the manipulations involved varying the visibility of the card distribution program to affect high-order belief (not only do I know the benefits I'm entitled to, but I know that the official knows, and I know that others know that we both know). This provides quite unique empirical evidence

⁷ Muralidharan and Niehaus (2016) make the same argument in their discussion of experimentation at scale.

on the impact of common knowledge, a concept in many models of game theory and political economy.

Another nice example is the Kling et al. (2012) study of comparison friction. We have seen that inertia is an important feature of consumer behavior. One reason may be that it is difficult for them to compare options, even when the information is available. There are a number of studies of “comparison friction” contrasting the Internet to other markets, suggesting that this is a general problem. Kling et al. (2012) set up an experiment that isolates comparison friction from other factors. In the treatment group, senior citizens in Medicare Part D received personalized information on the performance of different plans—information that they could, in principle, have acquired very easily by going to a website and entering the data. The intervention had large impact on plan switching (which increased from 27 percent to 38 percent). The design, which is focused on the very narrow channel of providing the results of the comparison, isolates the role of comparison friction from other mechanisms, such as difficulty of acquiring information, procrastination, etc.

More generally, well-designed plumbing experiments can sometimes introduce variation that does not exist in natural conditions, and thus generate a counterfactual to illuminate theoretical mechanisms that are not easily observable in nature. For example, it is a question of general interest whether people hired with prosocial motivations would generally make better employees, and specifically better public service employees. But since prosocial tendencies are not randomly assigned, this is not something where evaluating causality is easy. The Ashraf, Bandiera, and Lee (2015) study and the Deserranno (2015) studies that I mentioned earlier provide examples of how a plumbing experiment can help us make progress on difficult-to-settle empirical issues: in both studies, the authors used the different recruitment strategies to create variation in the type of health promoters that were recruited. Once employed, they were all tasked with the same responsibilities and faced the same incentives. Thus, the experiments clearly identify selection, controlling for incentive. Based on this design, Ashraf, Bandiera, and Lee (2015) find health workers attracted by career incentives are much more effective at delivering health services, as

measured by home visits and the organization of community meetings. In contrast, Deserranno (2015) provides evidence in support of the relationship between prosocialness and job performance. Compared to the high-pay treatment, the health promoters recruited under the low-pay treatment visited a larger number of households, organized more public presentations in the village, and provided more antenatal checks. She also finds that they were more likely to target the most vulnerable households. Clearly, these two studies do not settle the issue, as they have opposite conclusions, but they help us think about it better.

To be fair, it is not always possible to have very complex design in plumbing experiments: in some instances, it places too much strain on the implementing agency. For example, the Muralidharan, Niehaus, and Sukhtankar (2016a) experiment with smartcards in MGNREGS varies many things at the same time (they give the smartcards, they move the collection point to the village, they move it from the post office to the ATM, and they introduce local female “banking correspondents” who act as local agents for the bank). This makes it difficult in this setting to isolate one particular mechanism. But when possible, it makes the experiment particularly valuable, and my sense is that such trials will become increasingly feasible, especially as the first experiments provide examples to follow.⁸

A first relationship between plumbing and pure science is related to the “learning by noticing” concept that I referenced earlier. Many scientists-turned-engineers and later, turned-plumbers (Banerjee, Klemperer, Pathak) have said that experience taught them to pay attention to an entirely different set of notions than the ones they started with. The scientist’s focus on specific models and assumptions requires *not* noticing things that are unimportant, or that cannot be made sense of in their particular paradigm. The plumber must notice anything if it will matter for implementation. The plumber will share these observations with engineers and scientists (or with the engineer and scientist version of themselves) and thus hopefully generate new interesting and important questions to work

⁸For other examples of complex plumbing experiments with many treatments that look specifically at mechanisms, see Olken (2007); Alatas et al. (2012, 2016); Khan, Khwaja, and Olken (2016a); and Banerjee et al. (2012b).

on. Plumbing experiments shine the spotlight on new problems, and new theories can be developed to address those problems.

For example, Pathak (forthcoming, p. 32) concludes by reflecting on his many years of helping cities design and implement school assignment mechanisms, with which I opened this essay by saying: “Experiences from the field highlight new issues that have not been the focus of this earlier literature such as nonconsequentialist rationales for straightforward incentives, the importance of transparency and simplicity in influencing designs, the value of single-offer systems over multiple-offer alternatives, the need to streamline aftermarkets, and the importance of participation on the school and student side. It is my hope that further interaction between theory and practice will sharpen focus on these and potentially other significant issues that have received relatively little attention from the theoretical literature.” At the end of the paper, Pathak goes on to open yet another agenda on feedback between market-clearing algorithms, school attributes, and student preferences. Those are, I am guessing, difficult and interesting theoretical problems, but they would not have arisen without careful study of the property and behavior of actual school choice mechanisms

Indeed, the school choice literature has several examples of where questions that first arose as purely practical motivated further theoretical work, advancing market design as a field. The questions of how to break a tie in school assignment mechanisms originated as a practical question that parents and school administrators were very worried about, and led to an active literature (see Pathak forthcoming for a review). Similarly, researchers had not originally paid much attention to a central feature of school assignment in Boston, the “walk zone.” But when Mayor Menino promised that 50 percent of the students should go to a school they could walk to, researchers started simulating what would happen if the size of the “walk zone” was increased given the assignment mechanism that exists in Boston, and they were surprised to see that it did not lead to many more students assigned to their walk zone. This motivated theoretical work on what happens in an assignment system where a student can qualify for a school on two separate lists. The results are actually not obvious and are interesting theoretically. The basic mechanisms at play were not understood by

policymakers, market participants, or researchers before the research. In a working paper version, Dur et al. (2013) report how their research was influenced by the policy debate and how the policy debate was influenced by their research, which ultimately led to the abandonment of the “walk zone” in Boston on the grounds that the MIT-Boston College team had showed it made no difference, but it was perceived to unfairly advantage some students. For a last example, in many cases, participants in school choice mechanism have a different view than the scientists on whether a particular school assignment system can be “gamed” by being strategic about it. This is an example of a much more general problem, where participants’ perceptions of the optimal strategy in a system may be quite different than what the optimal strategy is. As scientists, we may have dismissed this, since agents are just wrong. But as plumbers, it obviously matters, since it will affect final outcomes. Motivated by these sorts of issues, Li (2016) actually tried to address this issue formally by defining the concept of “obviously strategy proof.”

This ability of experiments to put the spotlight on new problems is important. Many new fields in economics came about because someone decided that we could not continue to ignore the failure of assumptions that were staring us in the face: informational economics, behavioral economics, and network economics all come to mind. For this kind of innovation to happen, we need to leave a space for surprising information to come in. Of course, this requires being open to processing reality. Any field study, not just plumbing experiments (and not just experiments), can generate such insight. The plumbing studies may even have a particular handicap in this regard because, since they focus on details, they may often be difficult to theorize, even when they are important for policy (Banerjee 2002), beyond broad and rather simple conclusions (“information matters,” “default matters,” etc.). Plumbing experiments, since they are primarily motivated by pragmatism, must focus on what is important for the world, not necessarily on the very subtle issues that theorists would find worth discussing.

On the other hand, surprising information is perhaps more interesting to theorists, and thus likely to prompt more thinking when it is generated in very high-stakes environments and involving many people. The unexpected

behavior of a feature of a policy implemented at scale may motivate more interest than the result from a small experiment, however well controlled. Behavioral public finance is a good example of such back-and-forth between theoretical insights and findings from the plumbing experiments. Madrian and Shea (2001) were motivated by basic idea of procrastination to look at the role of default options in retirement savings. They found that employees enrolled in the 401k by default were much more likely to stay enrolled in the long term than those who were not defaulted in, and this prompted a much more systematic investigation of the prevalence of the default effect in retirement savings. The first wave of studies were empirical (see, e.g., Choi 2002, 2003; Choi et al. 2004, 2012; Chetty et al. 2014) and these, in turn, motivated theoretical work both on whether setting defaults is desirable (e.g., Carroll et al. 2009) and more generally on how to set them, since they had been shown to matter. The core theoretical question that defaults raise is motivated by welfare analysis: since economist-plumbers have demonstrated that defaults matter, the economist-engineer wants to set the default at a “reasonable” place. What is the right reasonable rate? This is not an easy theoretical question to answer, since it is likely to depend both on the underlying reasons why defaults matter in the first place, and because welfare analysis cannot rely on the standard tools when agents exhibit behavioral biases. Amador, Werning, and Angeletos (2006) start tackling the question of the optimal savings policy for an agent with hyperbolic discounting preferences. Bernheim, Fradkin, and Popov (2015) explicitly address the question of the optimal default policy, and show both how this depends on the underlying behavioral model, and how some relatively general results can nonetheless be obtained.

Last and perhaps most important, plumbing experiments, by their nature, are typically run on real policies and on a large scale. They are not just data points: they affect many real people and (potentially) have real benefits as they are carried out. When economists work on understanding how to design a policy that is going to affect millions of people and cost millions of dollars, the stakes are high enough that it is perhaps less important to know whether the findings will generalize beyond the particular work. Working with governments to evaluate versions of these

programs as they are being deployed represents a tremendous opportunity to generate evidence and improve the effectiveness of money that is already being spent. Moreover, as we have seen, the lessons that are generated from these partnerships are often actually acted upon, which means that gains from evaluation are not just potential, they are actual.

To take just one extreme example, the Affordable Care Act is a unique institution, and perhaps a less-than-ideal result of a long and complicated political process. It is very unlikely that anybody designing a health insurance system from scratch would adopt anything resembling it, so in a sense there is very little external validity to any work about the Affordable Care Act. Yet, when Handel, Kolstad, and others write papers that are concerned with detailed issues of implementation of a system like the ACA, it does not really occur to us to wonder whether this is perhaps a little too specific.

Similarly, most plumbing studies have first-order impacts. Getting the details right on school choice directly affects the welfare of millions of school children. The work of Banerjee et al. (2015) on the Raskin program led to the scale-up of cards to more than 60 million beneficiaries of public transfer programs. Obviously the scaled-up policy was adapted and slightly different than what was tested in the experiment. If the effects of this scaled-up version were the same as what they had estimated, they would translate in benefits anywhere between \$80 million and \$170 million annually. The work of Duflo et al. (2013) on environmental audits in Gujarat, which I discussed above, led to a reform of the regulation in the state, which has a population of 66 million. The evaluations that are described in Banerjee, Duflo, Imbert, Mathew, and Pande (2016) and Muralidharan, Niehaus, and Sukhtankar (2016a) directly affected hundreds of thousands of participants in the Indian welfare program, and eliminated millions of dollars of leakage, just while the research was underway, in the areas they covered. These studies also played a part in the adoption of a combination of those reforms as direct payment to beneficiary from the center in August 2015.⁹

⁹Unusually, a briefing presented by the Ministry of Rural Development to the Cabinet meeting actually cited a research paper (Ministry of Rural Development 2015). “As of today, 92 percent of Panchayats are brought onto e-FMS

This will affect the lives of 182 million beneficiaries, for a budget of \$5.6 billion.

Many of us chose economics because, ultimately, we thought science could be leveraged to make a positive change in the world. There are many different paths to get there. Scientists design general frames, engineers turn them into relevant machinery, and plumbers finally make them work in a complicated, messy policy environment. As a discipline, we are sometimes a little overwhelmed by “physics envy,” searching for the ultimate scientific answer to all questions—and this will lead us to question the legitimacy of plumbing. This essay is an attempt to argue that plumbing should be an inherent part of our profession: we are well prepared for it, reasonably good at it, and it is how we make ourselves useful. A feature unique to economics is that scientists, engineers, and plumbers all talk to each other (and in fact are often talking to themselves—the same economist wearing different hats). This conversation should continue: it is what will keep us relevant and, possibly, honest.

REFERENCES

- Abdulkadiroglu, Atila, and Tayfun Sönmez.** 2003. “School Choice: A Mechanism Design Approach.” *American Economic Review* 93 (3): 729–47.
- Abdulkadiroglu, Atila, and Tayfun Sönmez.** 2013. “Matching Markets: Theory and Practice.” In *Advances in Economics and Econometrics*, Vol. 1, edited by D. Acemoglu, M. Arellano, and E. Dekel, 3–47. Cambridge, MA: Cambridge University Press.
- Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel.** 2015. “Regulating Consumer Financial Products: Evidence from Credit Cards.” *Quarterly Journal of Economics* 130 (1): 111–64.
- Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, Ririn Purnamasari, and Matthew Wai-Poi.** 2016. “Self-Targeting: Evidence from a Field Experiment in Indonesia.” *Journal of Political Economy* 124 (2): 371–427.
-
- platform. The e-FMS has reduced the problem of parking of funds and increased transparency. Studies on the effectiveness of electronic fund flow in Bihar done by eminent economists (Banerjee et al. 2014) showed that e-governance reduced fund flow leakage by 25 percent.”
- Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias.** 2012. “Targeting the Poor: Evidence from a Field Experiment in Indonesia.” *American Economic Review* 102 (4): 1206–40.
- Almås, Ingvild, Alex Armand, Orazio Attanasio, and Pedro Carneiro.** 2015. “Measuring and Changing Control: Women’s Empowerment and Targeted Transfers.” National Bureau of Economic Research Working Paper 21717.
- Amador, Manuel, Iván Werning, and George-Marios Angeletos.** 2006. “Commitment vs. Flexibility.” *Econometrica* 74 (2): 365–96.
- Ashraf, Nava, Oriana Bandiera, and Scott S. Lee.** 2015. “Do-Gooders and Go-Getters: Career Incentives, Selection, and Performance in Public Service Delivery.” Unpublished.
- Banerjee, Abhijit V.** 1997. “A Theory of Misgovernance.” *Quarterly Journal of Economics* 112 (4): 1289–332.
- Banerjee, Abhijit V.** 2002. “The Uses of Economic Theory: Against a Purely Positive Interpretation of Theoretical Results.” Unpublished.
- Banerjee, Abhijit V.** 2007. *Making Aid Work*. Cambridge, MA: MIT Press.
- Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukherji, Marc Shotland, and Michael Walton.** 2016. “From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application.” National Bureau of Economic Research Working Paper 22931.
- Banerjee, Abhijit, Raghabendra Chattopadhyay, Esther Duflo, Daniel Keniston, and Nina Singh.** 2012a. “Can Institutions be Reformed from Within? Evidence from a Randomized Experiment with the Rajasthan Police.” Unpublished.
- Banerjee, Abhijit, Raghabendra Chattopadhyay, Esther Duflo, Daniel Keniston, and Nina Singh.** 2012b. “Improving Police Performance in Rajasthan, India: Experimental Evidence on Incentives, Managerial Autonomy and Training.” National Bureau of Economic Research Working Paper 17912.
- Banerjee, Abhijit V., and Esther Duflo.** 2009. “The Experimental Approach to Development Economics.” *Annual Review of Economics* 1 (1): 151–78.
- Banerjee, Abhijit V., and Esther Duflo.** 2012. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: PublicAffairs.

- Banerjee, Abhijit V., Esther Duflo, and Rachel Glennerster.** 2008. "Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System." *Journal of the European Economic Association* 6 (2–3): 487–500.
- Banerjee, Abhijit, Esther Duflo, Clement Imbert, Santhosh Mathew, and Rohini Pande.** 2016. "E-governance, Accountability, and Leakage in Public Programs: Experimental Evidence from a Financial Management Reform in India." National Bureau of Economic Research Working Paper 22803.
- Banerjee, Abhijit, Rema Hanna, Jordan C. Kyle, Benjamin A. Olken, and Sudarno Sumarto.** 2015. "The Power of Transparency: Information, Identification Cards and Food Subsidy Programs in Indonesia." National Bureau of Economic Research Working Paper 20923.
- Baugus, Brian, and Feler Bose.** 2015. "Sunset Legislation in the States: Balancing the Legislature and the Executive." <https://www.mercatus.org/system/files/Baugus-Sunset-Legislation.pdf>.
- Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen.** 2015. "Turning a Shove into a Nudge? A 'Labeled Cash Transfer' for Education." *American Economic Journal: Economic Policy* 7 (3): 86–125.
- Bernheim, B. Douglas, Andrey Fradkin, and Igor Popov.** 2015. "The Welfare Economics of Default Options in 401(k) Plans." *American Economic Review* 105 (9): 2798–837.
- Burwell, Sylvia, John Holdren, Alan Krueger, and Cecilia Munoz.** 2013. *M-13-17 Memorandum to the Heads of Departments and Agencies: Next Steps in the Evidence and Innovation Agenda*. Washington, DC: Executive Office of the President, Office of Management and Budget.
- Campbell, John Y.** 2016. "Richard T. Ely Lecture: Restoring Rational Choice: The Challenge of Consumer Financial Regulation." *American Economic Review* 106 (5): 1–30.
- Carroll, Gabriel D., James J. Choi, David Laibson, Brigitte C. Madrian, and Andrew Metrick.** 2009. "Optimal Defaults and Active Decisions." *Quarterly Journal of Economics* 124 (4): 1639–74.
- Chetty, Raj.** 2015. "Richard T. Ely Lecture: Behavioral Economics and Public Policy: A Pragmatic Perspective." *American Economic Review* 105 (5): 1–33.
- Chetty, Raj, John N. Friedman, Søren Leth-Petersen, Torben Helen Nielsen, and Tore Olsen.** 2014. "Active versus Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark." *Quarterly Journal of Economics* 129 (3): 1141–219.
- Choi, James J.** 2002. "Defined Contribution Pensions: Plan Rules, Participant Choices, and the Path of Least Resistance." In *Tax Policy and the Economy*, Vol. 16, edited by James M. Poterba, 67–113. Cambridge, MA: MIT Press.
- Choi, James J.** 2003. "Optimal Defaults." *American Economic Review* 93 (2): 180–85.
- Choi, James J., John Beshears, David Laibson, and Brigitte C. Madrian.** 2012. "Default Stickiness among Low-Income Individuals." RAND Working Paper WR-926-SSA.
- Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick.** 2004. "For Better or Worse: Default Effects and 401(k) Savings Behavior." In *Perspectives on the Economics of Aging*, edited by David A. Wise, 81–121. Chicago: University of Chicago Press.
- Cogley, Timothy, Riccardo Colacito, Lars Peter Hansen, and Thomas J. Sargent.** 2008. "Robustness and U.S. Monetary Policy Experimentation." *Journal of Money, Credit, and Banking* 40 (8): 1599–623.
- Congdon, William J., Jeffrey R. Kling, Jens Ludwig, and Sendhil Mullainathan.** 2017. "Social Policy: Mechanism Experiments and Policy Evaluations." In *Handbook of Field Experiments*, Vol. 1, edited by Esther Duflo and Abhijit Banerjee. Amsterdam: North Holland.
- Dal Bó, Ernesto, Frederico Finan, and Martin A. Rossi.** 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *Quarterly Journal of Economics* 128 (3): 1169–218.
- Deserranno, Erika.** 2015. "Financial Incentives as Signals: Experimental Evidence from the Recruitment of Health Workers." Unpublished.
- Devoto, Florencia, Esther Duflo, Pascaline Dupas, William Pariente, and Vincent Pons.** 2012. "Happiness on Tap: Piped Water Adoption in Urban Morocco." *American Economic Journal: Economic Policy* 4 (4): 68–99.
- Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan.** 2013. "Truth-Telling by Third-Party Auditors and the Response of Polluting Firms: Experimental Evidence from India." *Quarterly Journal of Economics* 128 (4): 1499–545.

- Dur, Umut M., Scott Duke Kominers, Parag A. Pathak, and Tayfun Sönmez.** 2013. "The Demise of Walk Zones in Boston: Priorities vs. Precedence in School Choice." National Bureau of Economic Research Working Paper 18981.
- Executive Order 13563.** 2011. "Improving Regulation and Regulatory Review." *Federal Register* 76: 3821–23.
- Field, Erica, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer Moore.** 2016. "On Her Account: Can Strengthening Women's Financial Control Boost Female Labor Supply?" Unpublished.
- Finan, Frederico, Benjamin A. Olken, and Rohini Pande.** 2015. "The Personnel Economics of the State." National Bureau of Economic Research Working Paper 21825.
- Gibbons, Robert.** 2010. "Inside Organizations: Pricing, Politics, and Path Dependence." *Annual Review of Economics* 2 (1): 337–65.
- Glennerster, Rachel.** 2017. "The Practicalities of Running Randomized Evaluations: Partnerships, Measurement, Ethics, and Transparency." In *Handbook of Field Experiments*, Vol. 1, edited by Esther Duflo and Abhijit Banerjee. Amsterdam: North Holland.
- Greenstone, Michael.** 2009. "Toward a Culture of Persistent Regulatory Experimentation and Evaluation." In *New Perspectives on Regulation*, edited by David Moss and John Cisernino, 111–25. Cambridge, MA: The Tobin Project.
- Handel, Benjamin R.** 2013. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." *American Economic Review* 103 (7): 2643–82.
- Handel, Benjamin R., and Jonathan Kolstad.** 2015. "Getting the Most from Marketplaces: Smart Policies on Health Insurance Choice." Brookings Hamilton Project Discussion Paper 2015–08.
- Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein.** 2014. "Learning through Noticing: Theory and Evidence from a Field Experiment." *Quarterly Journal of Economics* 129 (3): 1311–53.
- Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken.** 2016a. "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings." Unpublished.
- Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken.** 2016b. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors." *Quarterly Journal of Economics* 131 (1): 219–71.
- Klemperer, Paul.** 2002. "What Really Matters in Auction Design." *Journal of Economic Perspectives* 16 (1): 169–89.
- Kling, Jeffrey R., Sendhil Mullainathan, Eldar Shafir, Lee C. Vermeulen, and Marian V. Wrobel.** 2012. "Comparison Friction: Experimental Evidence from Medicare Drug Plans." *Quarterly Journal of Economics* 127 (1): 199–235.
- Li, Shengwu.** 2016. "Obviously Strategy-Proof Mechanisms." <http://siepr.stanford.edu/sites/default/files/publications/16-015.pdf>.
- Ludwig, Jens, Jeffrey R. Kling, and Sendhil Mullainathan.** 2011. "Mechanism Experiments and Policy Evaluations." *Journal of Economic Perspectives* 25 (3): 17–38.
- Madrian, Brigitte C., and Dennis F. Shea.** 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *Quarterly Journal of Economics* 116 (4): 1149–87.
- Ministry of Rural Development (MoRD).** 2015. "Note to Cabinet." Government of India.
- Moore, Quinn, Kevin Conway, and Brandon Kyler.** 2012. "Direct Certification in the National Lunch Program: State Implementation Progress School Year 2011–2012. Report to Congress: Summary." Alexandria, VA: US Department of Agriculture, Food and Nutrition Service, Office of Research and Analysis.
- Moore, Quinn, Kevin Conway, Brandon Kyler, and Andrew Gothro.** 2013. "Direct Certification in the National Lunch Program: State Implementation Progress, School Year 2012–2013." Report CN-13-DC. Alexandria, VA: US Department of Agriculture, Food and Nutrition Service, Office of Research and Analysis.
- Muralidharan, Karthik, and Paul Niehaus.** 2016. "Experimentation at Scale." Unpublished.
- Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar.** 2016a. "Building State Capacity: Evidence from Biometric Smartcards in India." *American Economic Review* 106 (10): 2895–929.
- Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar.** 2016b. "General Equilibrium Effects of (Improving) Public Employment Programs: Experimental Evidence from India." Unpublished.
- Olken, Benjamin A.** 2007. "Monitoring

- Corruption: Evidence from a Field Experiment in Indonesia.” *Journal of Political Economy* 115 (2): 200–49.
- Pathak, Parag A.** 2011. “The Mechanism Design Approach to Student Assignment.” *Annual Review of Economics* 3 (1): 513–36.
- Pathak, Parag.** Forthcoming. “What Really Matters in Designing School Choice Mechanisms.” In *Advances in Economics and Econometrics, 11th World Congress of the Econometric Society*.
- Roth, Alvin E.** 2002. “The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics.” *Econometrica* 70 (4): 1341–78.
- Scott, James C.** 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press.
- Simester, Duncan.** Forthcoming. “Field Experiments in Marketing.” In *Handbook on Field Experiments*, Vol. 1, edited by Esther Duflo and Abhijit Banerjee. Amsterdam: North Holland.
- Stango, Victor, and Jonathan Zinman.** 2014. “Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Bank Overdraft Fees.” *Review of Financial Studies* 27 (4): 990–1030.
- Thaler, Richard H., and Cass R. Sunstein.** 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.
- Willis, Lauren E.** 2013. “When Nudges Fail: Slippery Defaults.” *University of Chicago Law Review* 80 (3): 1155–229.