

Adaptive design, adaptive interventions

Behind every policy lie assumptions about why people behave the way they do. A policy that subsidizes fertilizer, for example, assumes that farmers find the price too high; that they can easily learn about price reductions once a subsidy is enacted; that they would benefit from using fertilizer and are aware of those benefits; that they are willing to invest some of their own money today and accept the associated risk to get payoffs at the end of the farming cycle; and that they have time to go purchase the product. But assumptions may often be incorrect, and solutions based on the wrong assumptions can lead to ineffective policies.

Concentrating more on the definition and diagnosis of problems, and expending more cognitive and financial resources at that stage, can lead to better-designed interventions.

For instance, as chapter 7 showed, farmers might find it difficult to translate their intentions to invest in fertilizer into concrete action at the time they need to purchase the fertilizer. The divide between intentions and actions may arise from the fact that farmers have cash in hand after harvest but do not need fertilizer until a few months later during the planting season. In Kenya, allowing farmers to prepay for fertilizer during the harvest and get it delivered during the next planting season proved as effective as offering a 50 percent

subsidy at the time fertilizer was applied (Duflo, Kremer, and Robinson 2011). In Malawi, allowing farmers to direct some of their harvest profits into commitment savings accounts, which held the money until the following planting season, increased investment back into crops and significantly increased the value of the subsequent harvest (Brune and others 2013).

Recognizing that individuals think automatically, think socially, and think with mental models expands the set of assumptions policy makers can use to analyze a given policy problem and suggests three main ways for improving the intervention cycle and development effectiveness. First, concentrating more on the definition and diagnosis of problems, and expending more cognitive and financial investments at that stage, can lead to better-designed interventions. For example, taking the time to figure out that application forms for financial aid for college might be the obstacle that depresses college attendance rates for low-income populations could lead to strategies that help students and their families fill out those applications—and could spare investments in an expensive and possibly ineffective information campaign (Bettinger and others 2012).

Second, an experimental approach that incorporates testing during the implementation phase and tolerates failure can help identify cost-effective interventions (Glennerster and Takavarasha 2013; Duflo and Kremer 2005). As many of the studies cited throughout this Report indicate, the process of delivering products matters as much as the product that is being delivered, and it can be difficult to predict what will matter in which context and for which population. For example, who could have predicted that weekly text-message reminders would improve adherence to a critical drug regimen for treating HIV/AIDS in Kenya better than

daily reminders (Pop-Eleches and others 2011) (see chapter 8). An experiment was required to learn that financial incentives were not effective in motivating the distributors of female condoms in Zambia (Ashraf, Bandiera, and Jack, forthcoming) (see chapter 7).

Third, since development practitioners themselves face cognitive constraints, abide by social norms, and use mental models in their work, development organizations may need to change their incentive structures, budget processes, and institutional culture to promote better diagnosis and experimentation so that evidence can feed back into midcourse adaptations and future intervention designs. Development practitioners must often act quickly and may thus feel compelled to skip a careful diagnosis and immediately apply “best practice.” Indeed, the intervention cycle typically allows neither the time nor the space to collect the data and perform the analysis needed to identify the problem properly, diagnose its determinants, and assess the fit between program and context or to make needed midcourse changes. As spotlight 3 and chapter 10 demonstrate, the mindsets of development practitioners can also differ substantially from those that prevail among low-income populations for whom they may be designing programs. Because development practitioners often have preconceived notions about a problem and its potential solutions, they may believe that they

know what should be done without having made their assumptions explicit and without having diagnosed the actual problem and its causes. While many development practitioners would agree that they often do not know what will work in a given context, their organizational environment may not allow them to admit as much (Pritchett, Samji, and Hammer 2013).

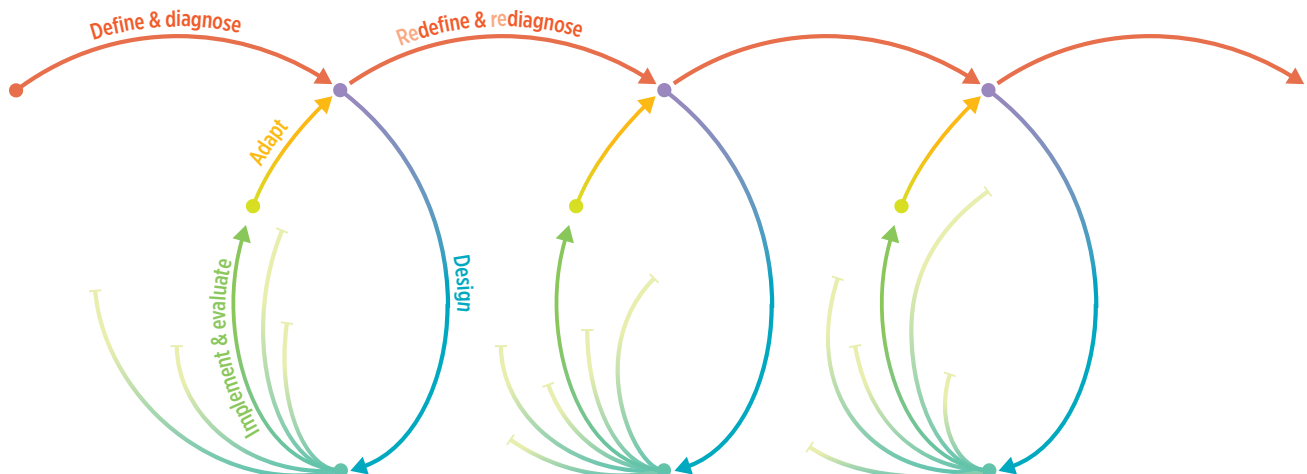
Delving deeper into the subject may lead to a better understanding of the underlying causes of an observed behavior and to identifying ways to intervene effectively. In a complex and iterative process (figure 11.1), problems may need to be redefined and rediagnosed, and multiple interventions may need to be piloted simultaneously—some of which will fail—before an effective intervention can be designed.

This chapter builds on the work by Datta and Mulainathan (2014), who discuss how to design development programs and policies in ways that are cognizant of and informed by the insights from the behavioral sciences, an approach that has been applied to design interventions for low-income populations across the United States (CFED and ideas42 2013).

To see how diagnoses and program design can evolve in the process of finding a solution to a challenge, consider the problem of ensuring access to clean water in rural Kenya and a series of field experiments that tested the effectiveness of different methods of

Figure 11.1 Understanding behavior and identifying effective interventions are complex and iterative processes

In an approach that incorporates the psychological and social aspects of decision making, the intervention cycle looks different. The resources devoted to definition and diagnosis, as well as to design, are greater. The implementation period tests several interventions, each based on different assumptions about choice and behavior. One of the interventions is adapted and fed into a new round of definition, diagnosis, design, implementation, and testing. The process of refinement continues after the intervention is scaled up.



averting the incidence of diarrhea among children (Ahuja, Kremer, and Zwane 2010). Lack of access to clean water was diagnosed as a problem, and thus an early intervention aimed to improve infrastructure at households' water sources, naturally occurring springs, which were susceptible to contamination from the surrounding environment. In particular, the springs were covered with concrete so that water flowed from a pipe rather than seeping from the ground. While this considerably improved water quality at the source, it had only moderate effects on water quality in households because the water could easily be recontaminated during transport or storage (Kremer and others 2011).

Thus the problem was not simply access to clean water; instead, it could be redefined as a problem of inadequate water treatment within the home. Another iteration of experiments demonstrated that providing free home delivery of chlorine or discount coupons that could be redeemed in local shops elicited very high take-up of the water treatment product at first but ultimately failed to generate sustained results. People needed to remember to chlorinate their water when they returned home from the water source, and they needed to continue to go to the store to purchase the product.

These results in turn suggested yet another diagnosis of the problem: households found it difficult to sustain the use of water treatment over time. This insight led to the design of free chlorine dispensers next to the water source, which made water treatment salient (the dispenser served as a reminder just when people were thinking about water) and convenient (there was no need to make a trip to the store, and the necessary agitation and wait time for the chlorine to work automatically occurred during the walk home). It also made water treatment a public act. This proved to be the most cost-effective method for increasing water treatment and averting the incidence of diarrhea (Abdul Latif Jameel Poverty Action Lab 2012).

As this example and other chapters demonstrate, context matters in particular ways. Seemingly small details of design and implementation of policies and programs can have disproportionate effects on individual choices and actions. Similar challenges can have different underlying causes. An approach that works in one country may not necessarily work in another. Indeed, evidence on the policy implications of a psychological and social perspective on development challenges is just now coming into view.

This Report does not advocate specific interventions. Instead, it argues for the need to change the process of arriving at solutions, regardless of the nature of the problem (acute, chronic, last mile, and so forth) or

the type of environmental or institutional setting (low- or high-income, low- or high-capacity).

This chapter discusses the components of the more complex and more iterative intervention cycle proposed in figure 11.1: (1) diagnosing and re-diagnosing psychological and social obstacles; (2) designing an intervention; (3) experimenting during implementation; and (4) learning from these previous steps and adapting future interventions accordingly.

Diagnosing psychological and social obstacles

While it goes without saying that identifying problems or obstacles must precede the design of solutions, there is less clarity on just how one should go about this process of diagnosis. Measuring an individual's lack of material resources or information, for example, is relatively straightforward, and countless household surveys provide data on these sorts of obstacles. In contrast, identifying the presence of psychological biases, cognitive burdens, social norms, and mental models may require more in-depth investigations.

Thick description, for example, and other forms of ethnography (spotlight 4) can be used to understand decision-making contexts. In traditional anthropology, ethnographic fieldwork consists of extensive participant observations, interviews, and surveys. More problem-driven forms of the ethnographic approach can be used to help development practitioners refine their hypotheses about what drives specific behaviors, as well as to monitor newly emerging behaviors. In Denmark, for example, a ban on indoor smoking shifted smokers to the areas just outside the doors of buildings. This posed a problem for Copenhagen Airport, since the secondhand smoke could easily find its way back into the building through doors and air vents. Simply creating a no-smoking zone around entrances did not help. Careful "fieldwork," however, in which the habits of those smoking at the airport were closely observed and mapped, was instrumental in finding solutions that cut smoking near entrances by more than 50 percent. Since smokers tended to come from inside the building and reach for their cigarettes as they were exiting the building, stickers with an icon of a lit cigarette and the distance to the smoking zone were placed on the floors right before the doors. Benches and trash cans, which tended to attract smokers, were placed farther from airport entrances in zones especially designated for smoking (iNudgeyou 2014). Along with "thick description" like this, another useful way to characterize decision-making contexts is the "Reality Check" (box 11.1).

More quantitative methods, such as surveys, can also be informative at this stage of the intervention

cycle. A number of measurement techniques can help reduce *courtesy bias* (where respondents provide answers they think the questioner wants to hear) and measure psychological patterns that respondents themselves may not be aware of (box 11.2).

Finally, there may be nothing as illuminating as the technique of *dogfooding*, discussed in chapter 10. In this practice, company employees themselves use a product they have designed to work out its kinks before releasing it to the marketplace. Policy designers could try to sign up for their own programs or access existing services to diagnose problems firsthand.

Designing an intervention

Once key obstacles have been identified, the task becomes designing an intervention that incorporates these insights. Sometimes the diagnosis phase of an intervention may reveal multiple obstacles but not their relative importance, and each of these would imply different designs for tackling the larger problem at hand.

Consider again the example of home water chlorination. Table 11.1 lists a number of different obstacles that could interfere with home water treatment and the corresponding interventions that could overcome them. An intervention designed for someone who knows the benefits of chlorine and can afford to purchase it but simply forgets to would look somewhat different from an intervention designed for someone

Box 11.1 Taking the perspective of program beneficiaries through the Reality Check approach

The everyday experiences, awareness, and aspirations of people living in poverty are often unmeasured and may in fact be dynamic. This challenges development professionals to keep in touch and up to date. An immersion program called the Reality Check approach has been used by donors, governments, and nongovernmental organizations (NGOs) to understand how poor people make decisions. Social science researchers live for several days and nights with a poor family, not as an important visitor but as an ordinary person, aiming to observe and build relationships, trust, and respect. This qualitative approach has uncovered important findings that might have been missed with more quantitative surveys. For example, in Bangladesh and Nepal, government health providers felt pressured every day to provide free medicine to people who they knew were not ill but who were selling it to others or who wanted it for their livestock. In northern Ghana, researchers learned that at certain times of year, the heat made it unreasonable to expect people to get inside a mosquito net.

Source: www.reality-check-approach.com.

who would use the product only if she saw someone in her peer group using it.

Table 11.2 presents a list of designs and related interventions that have been experimentally evaluated to identify effective interventions across a wide class of problems (Richburg-Hayes and others 2014). In light of this growing body of work, it has been argued that a science of design is emerging in which the

Box 11.2 Measurement techniques that can help uncover psychological and social obstacles

Techniques for eliciting sensitive information

- *Introduce personal distance.* Sometimes, answers are best elicited through questions that are asked indirectly. For instance, rather than asking an official whether he has ever accepted a bribe, the researcher can ask whether a person in his position typically accepts bribes. Eliciting information through vignettes or hypothetical situations about fictional people allows respondents to think about a situation in a way that is more emotionally removed from their personal concerns but that tends to reveal social expectations.
- *Allow a cover of randomness.* For instance, when asked a sensitive question that should have a yes/no answer, a respondent can be asked to privately flip a coin and say yes if it comes up heads or answer truthfully if it comes up tails. This can allow the person to answer truthfully and still allow the researcher to learn about the share of the population that engages in a potentially shameful behavior, even if she would not know about the behavior of any given individual. List experiments (Blair and Imai 2012; Droitcour and others 2011; Holbrook and Krosnick 2009; Karlan and Zinman 2012) are another method for measuring the share of a population that engages in a taboo behavior or holds an opinion that may not be freely admitted. Respondents are randomly

assigned one of two questions and asked to report the number of items that they agree with or that apply to them. The lists differ solely in the presence of the sensitive item or topic.

Measuring attitudes and social norms

- *Implicit association tests.* These tests measure automatic associations between concepts (such as the home or a career) and attributes (male and female) (Greenwald, McGhee, and Schwartz 1998; Banaji 2001; Beaman and others 2009; Banaji and Greenwald 2013). They are easy to administer and can be adapted for nonliterate populations. Demonstration tests can be found at www.implicit.harvard.edu.
- *Identifying social norms.* Survey questions in household surveys or ethnographic work can uncover perceptions about expected and prescribed behaviors. For example, questions like “Out of 10 of your neighbors, how many exclusively breastfeed their children?” can help reveal what people expect others to be doing. Questions like “If you decided to exclusively breastfeed your child, would you worry about anyone disapproving?” can help reveal the relevant network to which the social norm applies.

psychological and social sciences can play a key role (Datta and Mullainathan 2014).

Many of the quantitative and qualitative methods useful for diagnosing obstacles can also assist in the design phase—particularly in narrowing down options that could be tested at a larger scale. Two experiences from Zambia demonstrate this approach. A “mama kit” is a package provided to an expectant mother that contains all the materials she would need to ensure the clean and safe delivery of her child. The kits are typically used to encourage delivery in a health facility. Semi-structured interviews with women and a survey of local wholesale prices helped determine the kit contents that mothers would find desirable and that would be feasible to provide. This up-front qualitative work to optimize the content of the mama kit paid off. Ultimately, a randomized controlled trial found that the mama kits increased facility delivery rates by 44 percent (IDinsight 2014a).

Similarly, the Zambian government quickly experimented with different frequencies for household visits by community health workers to ensure that subsidized antimalarial bed nets were actually being used (IDinsight 2014b). Households that received bed nets through fixed-point distribution sites were ran-

domly divided into five groups that received a visit from a community health worker at different intervals after the distribution of the nets (1–3 days, 5–7 days, 10–12 days, 15–17 days, and six weeks later). Self-installation and retention rates were then compared across the five groups. These household visits revealed that bed nets were hung by recipients within the first 10 days; that nets that were not hung after 10 days were unlikely to be hung at all; and that retention was stable for the two months or so following distribution. These results provided the government a clear path to designing an optimal visit frequency, and it crafted guidelines specifying the optimal time to visit households and hang up the remaining bed nets as 10 days after distribution.

Mechanism experiments are another useful technique for narrowing down candidate policies for experimentation. Consider how such an experiment could be used to design a strategy for tackling the problem of obesity in low-income neighborhoods (Ludwig, Kling, and Mullainathan 2011). Suppose that policy makers were concerned about “food deserts”: that is, neighborhoods where there is plenty of food but none of it is healthy. One possible policy option would be to experiment with offering incentives for green

Table 11.1. Different obstacles may require different intervention design (Case study: increasing home water chlorination)

Design of intervention	Free bottles delivered at home	Discount coupon redeemable at local shop	Detailed instructions	Improved storage	Persuasion messages	Using promoters from social network	Chlorine dispenser at point of collection
Potential obstacles							
People do not understand how to use chlorine.							
Procrastination may cause individuals to postpone visits to the store where the chlorine is sold.							
People are not motivated to use chlorine because the effect on health is delayed.							
People forget to chlorinate the water.							
People are affected by what others in the community do.							
Product may be too expensive.							
Some people are not convinced about the importance of clean water.							

Source: WDR 2015 team.

grocers to locate in these areas. However, this would be an extremely costly intervention. An alternative experimental design would involve sampling low-income families and randomly selecting some of them to receive free weekly home delivery of fresh fruits and vegetables. This would be a much cheaper experiment and would provide valuable information before trying a more expensive variant. If, for example, people do not

consume any of the free produce, it is worth asking if they would take a more active step and purchase fruits and vegetables at a local store, even if the prices were heavily subsidized. If people did consume the produce but did not lose weight, then policy makers should perhaps not focus on the problem of food deserts as a priority in solving the problem of obesity in low-income neighborhoods.

Table 11.2 Experimental evidence is accumulating on the effectiveness of many psychologically and socially informed designs

Type	Strength of the evidence	Examples
Reminders	73 papers, appearing in 6 domains	A regular text-message reminder to save money increased savings balances by 6 percent (Karlán and others 2010).
Social influence	69 papers, appearing in all 8 domains	Homeowners received mailers that compared their electricity consumption with that of neighbors and rated their household as great, good, or below average. This led to a reduction in power consumption equivalent to what would have happened if energy prices had been raised 11–20 percent (Allcott 2011).
Feedback	60 papers, appearing in 5 domains	A field experiment provided individualized feedback about participation in a curbside recycling program. Households that were receiving feedback increased their participation by 7 percentage points, while participation among the control group members did not increase at all (Schultz 1999).
Channel and hassle factors	43 papers, appearing in 8 domains	Providing personalized assistance in completing the Free Application for Federal Student Aid (FAFSA) led to a 29 percent increase in two consecutive years of college enrollment among high school seniors in the program group of a randomized controlled trial, relative to the control group (Bettinger and others 2012).
Micro-incentives	41 papers, appearing in 5 domains	Small incentives to read books can have a stronger effect on grades than incentives to get high grades (Fryer Jr. 2010).
Identity cues and identity priming	31 papers, appearing in 5 domains	When a picture of a woman appeared on a math test, female students were reminded to recall their gender and performed worse on the test (Shih, Pittinsky, and Ambady 1999).
Social proof	26 papers, appearing in 5 domains	Phone calls to voters with a “high turnout” message—emphasizing how many people were voting and that the number was likely to increase—were more effective at increasing voter turnout than a “low turnout” message, which emphasized that election turnout was low last time and likely to be lower this time (Gerber and Rogers 2009).
Physical environment cues	25 papers, appearing in 5 domains	Individuals poured and consumed more juice when using short, wide glasses than when using tall, slender glasses. Cafeterias can increase fruit consumption by increasing the visibility of the fruit with more prominent displays or by making fruit easier to reach than unhealthy alternatives (Wansink and van Ittersum 2003).
Anchoring	24 papers, appearing in 3 domains	In New York City, credit card systems in taxis automatically suggested a 30, 25, or 20 percent tip. This caused passengers to think of 20 percent as the low tip—even though it was double the previous average. Since the installation of the credit card systems, average tips have risen to 22 percent (Grynbaum 2009).
Default rules and automation	18 papers, appearing in 7 domains	Automatically enrolling people in savings plans dramatically increased participation and retention (Thaler and Benartzi 2004).
Loss aversion	12 papers, appearing in 7 domains	In a randomized controlled experiment, half the sample received a free mug and half did not. The groups were then given the option of selling the mug or buying a mug, respectively, if a determined price was acceptable to them. Those who had received a free mug were willing to sell only at a price that was twice the amount the potential buyers were willing to pay (Kahneman, Knetsch, and Thaler 1990).
Public/private commitments	11 papers, appearing in 4 domains	When people promised to perform a task, they often completed it. People imagine themselves to be consistent and will go to lengths to keep up this appearance in public and private (Bryan, Karlán, and Nelson 2010).

Source: Richburg-Hayes and others 2014.

Note: The eight domains covered were charitable giving, consumer finance, energy and the environment, health, marketing, nutrition, voting, and workplace productivity.

Experimenting during implementation

Sometimes practitioners might not have the luxury of time for all the possible qualitative and quantitative work before implementation. Immediate action may be required. In such cases, it will still be important to embed experimentation during the implementation phase. Experimentation during the implementation process can still test psychological and social predictions and optimize impact within a particular intervention cycle. Moreover, while using evidence from elsewhere may be very useful at the preparation stage, it will not replace generating and using evidence from within the very policy intervention as it is being carried out.

One way to test the importance of implementation details, for example, would be to experiment with different modes of implementation. In 2009, the Kenyan government announced a nationwide contract teacher program that would eventually employ 18,000 teachers. In the pilot area, some schools were randomly chosen to receive contract teachers as part of the government program, while others received a contract teacher under the coordination of a local NGO. The evaluation showed how the implementation by the NGO improved students' test scores across diverse contexts, while government implementation had no effect at all (Bold and others 2013).

The series of experiments on commitment devices for farmers discussed earlier also illustrates how experimental implementation can be used iteratively to learn how to adapt policies before scaling them up.

A first set of field experiments in Kenya showed that investment in fertilizer is surprisingly low despite high returns (Duflo, Kremer, and Robinson 2007, 2008, 2011). Diagnosis suggested that several factors, some psychological and social and some market related, could help explain this puzzle: credit constraints, information constraints, absent-mindedness, and intention-action divides. A second set of experiments tested these proposed theories by implementing several different interventions simultaneously and found that interventions that provided a way for farmers to commit to fertilizer purchases (by paying for them when they had cash on hand) were the most successful. Similar commitment products were tested in Malawi with tobacco producers with large positive effects (Brune and others 2013). The findings were then taken to scale and evaluated by the World Bank in Rwanda in the context of a government intervention with a typical population of subsistence farmers (Kondylis, Jones, and Stein 2013).

As these examples show, a more adaptive, empirically agile approach to the intervention cycle can help identify effective ways to improve development outcomes. Has anyone succeeded in systematically implementing this more psychological and socially informed and experimental approach at a scale beyond field experiments with NGOs? The Behavioural Insights Team in the United Kingdom has dedicated itself to bringing psychological and social insights into government policy and service delivery and tests policy alternatives through experimentation (Haynes, Goldacre, and Torgerson 2012) (box 11.3).

Box 11.3 Using psychological and social insights and active experimentation in the United Kingdom

The Behavioural Insights Team (BIT, also known as the “Nudge Unit”) was created in 2010 with the objective of applying insights from academic research in behavioral economics and psychology to public policy and services. It was created at a time of economic and financial crisis and resource scarcity, when psychological and socially informed interventions seemed a viable alternative to legislation.

BIT uses a four-part methodology to identify what works and can be scaled up and what does not: (1) define the desired outcome; (2) use ethnography to understand better how individuals experience the service or situation in question; (3) build new interventions to improve outcomes; and (4) test and try out the interventions, often using randomized controlled trials.

The unit tried to harness the power of social norms to encourage timely tax payments. They tested various messages in letters sent to taxpayers, which either invoked no social norm or contained messages like “9 out of 10 people in [Britain/your postcode/your town] pay their

tax on time.” Citing social norms that referred to others in the taxpayer’s own town led to a 15 percentage point increase in the fraction of taxpayers responding with payment in the following three months (BIT 2012).

In another experiment, team members from BIT embedded themselves in an unemployment center to see what obstacles the unemployed faced in moving off unemployment benefits and into a job. They identified a cumbersome process that involved considerable paperwork and that failed to motivate job seekers. They then designed a pilot program that asked job seekers to make commitments for future job search activities (as opposed to reporting on past activities) and to identify their personal strengths. These changes increased transitions away from benefits by nearly 20 percent (Bennhold 2013).

Until January 2014, BIT was funded by the public. It is now a company owned by its employees, the U.K. government, and Nesta (the leading innovation charity in the United Kingdom).

Conclusion: Learning and adapting

As these and countless other examples throughout the Report have demonstrated, finding effective solutions requires continual research and development (R&D). Although time and resource constraints might interfere with efforts to adopt more systematic diagnoses and experimental implementation, the biggest challenge may be overcoming the psychological and social obstacles within development organizations themselves. Measures are needed to ensure that development practitioners account for their own automatic thinking, mental models, and the social influences on their own choices. To do so, they may need to rethink the process of research and development.

First, R&D is not meant to yield immediate profits or immediate improvements. It delivers uncertain benefits in the future. Time-pressed and risk-averse organizations in search of immediate but certain results might thus underinvest in R&D. They may require commitment devices or risk-mitigating measures that can help them set aside the time and resources required for adequate diagnosis and experimental implementation. Over the years, development practitioners have become familiar with the risks related to financing and political economy and to technical uncertainty. But they need to pay more attention to another set of risks: those associated with the development and implementation of new products, services, and modes of delivery. What matters are not just the high-level policies that governments and development actors adopt but also how those policies are implemented and delivered. Just as a science of designing development interventions may be emerging, so too might an art and science of service delivery in development.

Second, R&D entails failure. Tendencies to continue paying sunk costs or to pay attention only to evidence that confirms their own biases (confirmation bias) can prevent development practitioners from admitting to and learning from failures. However, it is often through the process of experimenting, failing, and learning from those failures that effective, evidence-based diagnoses and intervention strategies emerge.

Creating the mechanisms for accommodating the optimal levels of diagnosis, risk taking, and failure is an organizational challenge. Ideally, development practitioners would be accountable for the quality of the learning and experimentation strategy all along the intervention cycle rather than just for compliance with the plans defined before the start of the intervention—a situation that cripples innovation, hampers midcourse corrections, and stirs fears of failure.

The findings in this Report bring another very large and complex source of uncertainty to development projects: the role of psychological and social factors in the decision making and behavior of end users, implementers, and development practitioners themselves.

To account for the fact that development practitioners themselves face cognitive constraints, abide by social norms, and use mental models in their work, development organizations may need to change their incentive structure, budget processes, and institutional culture.

This uncertainty is not insurmountable for development practice. Indeed, one purpose of this Report has been to synthesize some of the most compelling scientific research on the topic. It is hoped that this Report can inspire development practitioners who are ready to take up the challenge.

References

- Abdul Latif Jameel Poverty Action Lab. 2012. "Cleaner Water at the Source." *J-PAL Policy Briefcase* (September). <http://www.povertyactionlab.org/publication/cleaner-water-source>.
- Ahuja, Amrita, Michael Kremer, and Alix Peterson Zwane. 2010. "Providing Safe Water: Evidence from Randomized Evaluations." *Annual Review of Resource Economics* 2 (1): 237–56.
- Allcott, Hunt. 2011. "Social Norms and Energy Conservation." *Journal of Public Economics* 95 (9): 1082–95.
- Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack. Forthcoming. "No Margin, No Mission? A Field Experiment on Incentives for Public Service Delivery." *Journal of Public Economics*. <http://dx.doi.org/10.1016/j.jpubeco.2014.06.014>.
- Banaji, Mahzarin R. 2001. "Implicit Attitudes Can Be Measured." In *The Nature of Remembering: Essays in Honor of Robert G. Crowder*, edited by James S. Nairne, Henry L. Roediger III, Ian Neath, and Aimee M. Surprenant, chap. 7. Washington, DC: American Psychological Association.
- Banaji, Mahzarin R., and Anthony G. Greenwald. 2013. *Blindspot: Hidden Biases of Good People*. New York: Delacorte.

- Beaman, Lori, Raghavendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *Quarterly Journal of Economics* 124 (4): 1497–1540.
- Bennhold, Katrin. 2013. "Britain's Ministry of Nudges." *New York Times*, December 7. <http://www.nytimes.com/2013/12/08/business/international/britains-ministry-of-nudges.html?pagewanted=all>.
- Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment." *Quarterly Journal of Economics* 127 (3): 1205–42.
- BIT (Behavioural Insights Team). 2012. *Applying Behavioural Insights to Reduce Fraud, Error and Debt*. London: Cabinet Office, Behavioural Insights Team.
- Blair, Graeme, and Kosuke Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20 (1): 47–77.
- Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. 2013. "Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education." Working Paper 321, Center for Global Development, Washington, DC.
- Brune, Lasse, Xavier Giné, Jessica Goldberg, and Dean Yang. 2013. "Commitments to Save: A Field Experiment in Rural Malawi." Policy Research Working Paper 5748, World Bank, Washington, DC.
- Bryan, Gharad, Dean Karlan, and Scott Nelson. 2010. "Commitment Devices." *Annual Review of Economics* 2 (1): 671–98.
- CFED and ideas42. 2013. "Small Changes, Real Impact: Applying Behavioral Economics in Asset-Building Programs." <http://www.ideas42.org/publication/view/small-changes-real-impact-applying-behavioral-economics-in-asset-building-programs/>.
- Datta, Saugato, and Sendhil Mullainathan. 2014. "Behavioral Design: A New Approach to Development Policy." *Review of Income and Wealth* 60 (1): 7–35.
- Droitcour, Judith, Rachel A. Caspar, Michael L. Hubbard, and Trena M. Ezzati. 2011. "The Item Count Technique as a Method of Indirect Questioning: A Review of Its Development and a Case Study Application." In *Measurement Errors in Surveys*, edited by P. B. Beimer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman, 185–210. Hoboken, NJ: John Wiley & Sons.
- Duflo, Esther, and Michael Kremer. 2005. "Use of Randomization in the Evaluation of Development Effectiveness." In *Evaluating Development Effectiveness*, Vol. 7 of World Bank Series on Evaluation and Development, edited by Osvaldo N. Feinstein, George Keith Pitman, and Gregory K. Ingram. New Brunswick, NJ: Transaction Publishers.
- Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2007. "Why Don't Farmers Use Fertilizer? Experimental Evidence from Kenya." Working Paper, Massachusetts Institute of Technology and Harvard University, Cambridge, MA.
- . 2008. "How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya." *American Economic Review* 98 (2): 482–88.
- . 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review* 101 (6): 2350–90.
- Fryer, Jr., Roland G. 2010. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." Working Paper 15898, National Bureau of Economic Research, Cambridge, MA.
- Gerber, Alan S., and Todd Rogers. 2009. "Descriptive Social Norms and Motivation to Vote: Everybody's Voting and So Should You." *Journal of Politics* 71 (1): 178–91.
- Glennster, Rachel, and Kudzai Takavarasha. 2013. *Running Randomized Evaluations: A Practical Guide*. Princeton, NJ: Princeton University Press.
- Greenwald, Anthony G., Debbie E. McGhee, and Jordan L. K. Schwartz. 1998. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology* 74 (6): 1464–80.
- Grynbaum, Michael. 2009. "New York's Cabbies Like Credit Cards? Go Figure." *New York Times*, November 7, A1.
- Haynes, Laura, Ben Goldacre, and David Torgerson. 2012. "Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials." Policy Paper, Cabinet Office, Behavioural Insights Team, London.
- Holbrook, Allyson L., and Jon A. Krosnick. 2009. "Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique." *Public Opinion Quarterly* 74 (1): 37–67.
- IDInsight. 2014a. "Measuring the Impact of Mama Kits on Facility Service Delivery Rates in Rural Chadize and Serenje Districts in Zambia." Technical Report for the Government of Zambia. Available at <http://idinsight.org/wp-content/uploads/2014/10/Mama-Kits-Report-Aug-2014.pdf>.
- IDInsight. 2014b. "Measuring Community-Level Point ITN Distribution Dynamics and the Impact of CHW Hang-Up in Rural Rufunsa District in Zambia." Technical Report for the Government of Zambia. Available at <http://idinsight.org/wp-content/uploads/2014/10/ITN-in-progress-report-October-2014.pdf>.
- iNudgeyou. 2014. "When Prescription Works Better than Proscription." Blog, February 10. <http://inudgeyou.com/when-prescription-works-better-than-proscription/>.
- Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy* 98 (6): 1325–48.
- Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. 2010. "Getting to the Top of Mind: How Reminders Increase Saving." Working Paper 16205, National Bureau of Economic Research, Cambridge, MA.
- Karlan, Dean, and Jonathan Zinman. 2012. "List Randomization for Sensitive Behavior: An Application for Measuring Use of Loan Proceeds." *Journal of Development Economics* 98(1): 71–75.

- Kondylis, Florence, Maria Jones, and Daniel Stein. 2013. "Rwanda: Program of Impact Evaluations in Agriculture." Development Impact Evaluation Initiative (DIME) Concept Note, March. World Bank, Washington, DC. https://www.gafspfund.org/sites/gafspfund.org/files/Documents/LWH%20Impact%20Evaluation%20Concept%20Note_final.pdf.
- Kremer, Michael, Jessica Leino, Edward Miguel, and Alix Peterson Zwane. 2011. "Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions." *Quarterly Journal of Economics* 126 (1): 145–205.
- Ludwig, Jens, Jeffrey R. Kling, and Sendhil Mullainathan. 2011. "Mechanism Experiments and Policy Evaluations." *Journal of Economic Perspectives* 25 (3): 17–38.
- Pop-Eleches, Cristian, Harsha Thirumurthy, James Habyarimana, Joshua Graff Zivin, Markus Goldstein, Damien de Walque, Leslie MacKeen, Jessica Haberer, John Sidle, and Duncan Ngare. 2011. "Mobile Phone Technologies Improve Adherence to Antiretroviral Treatment in Resource-Limited Settings: A Randomized Controlled Trial of Text Message Reminders." *AIDS* 25 (6): 825–34.
- Pritchett, Lant, Salimah Samji, and Jeffrey Hammer. 2013. "It's All About MeE: Using Structured Experiential Learning ('e) to Crawl the Design Space." Working Paper 322, Center for Global Development, Washington, DC.
- Richburg-Hayes, Lashawn, Caitlin Anzelone, Nadine Dechausay, Saugato Datta, Alexandra Fiorillo, Louis Potok, Matthew Darling, and John Balz. 2014. "Behavioral Economics and Social Policy: Designing Innovative Solutions to Programs Supported by the Administration for Children and Families." OPRE Report No. 2014 16a, Office of Planning, Research, and Evaluation (OPRE), U.S. Department of Health and Human Services, Washington, DC.
- Schultz, P. Wesley. 1999. "Changing Behavior with Normative Feedback Interventions: A Field Experiment on Curbside Recycling." *Basic and Applied Social Psychology* 21 (1): 25–36.
- Shih, Margaret, Todd L. Pittinsky, and Nalini Ambady. 1999. "Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance." *Psychological Science* 10 (1): 80–83.
- Thaler, Richard H., and Shlomo Benartzi. 2004. "Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving." *Journal of Political Economy* 112 (S1): S164–S87.
- Wansink, Brian, and Koert van Ittersum. 2003. "Bottoms Up! The Influence of Elongation on Pouring and Consumption Volume." *Journal of Consumer Research* 30 (3): 455–63.