Yeung, Catherine & Sharon Tham (2021).
Behavioral Science in Policy and Government: A Roadmap. in the Behaviorally Informed Organization.
Edited by Dilip Soman & Catherine Yeung. University of Toronto Press.

CHAPTER SIXTEEN

Behavioral Science in Policy and Government: A Roadmap

Catherine Yeung and Sharon Tham

Behavioral science has provided solutions to many business and policy problems. Indeed, this book is replete with success stories from for-profit and government and welfare organizations alike. In chapter 1, however, Soman cautions readers against believing that the application of BI is straightforward. Behind many of the success stories are struggles, and significant effort has been needed to overcome challenges in bringing behavioral-science projects to fruition. In this chapter, we look past the success and zoom in on these challenges.

Our partners have shared stories of obstacles faced and lessons learnt. Some of them went through a long process to get to the root cause of a problem, while others expended tremendous effort to understand how behavioral principles could be contextualized. For example, in an initiative to increase tax-filing among low-income Canadians, it was unclear why low-income citizens were not filing tax returns despite the fact that many income-tested benefits are tied to tax returns. Only after extensive research did community groups in charge of the initiative realize that low-income citizens wanted to avoid being stigmatized as being in need of government support (see chapter 12). Identifying this cause was a breakthrough, because designing interventions to address the stigma issue was key to truly helping the low-income group. If the community groups had not dug deep into understanding the root cause but had used a standard, off-the-shelf solution that had been shown to work elsewhere (e.g., a social-norm message), the initiative would have been much less successful.

The stories of eMBeD illustrate the team's daunting challenges of contextualizing and adapting behavioral-science principles in different countries (see chapter 14). The team could not implement the same interventions in countries where people and their behaviors differed, and where infrastructures, political landscapes, and resource constraints of the countries were also different. An education program that worked in the United States needed to be modified before it could be conducted in Peru, and modified yet again for Indonesia. Each adaptation was a brand-new project with its very own behavioral issues to tackle and constraints to overcome.

In this chapter, we build on our partners' experiences and insights to propose four key points that can guide policymakers and practitioners through their challenges and help them discover the best way of using behavioral insights for their own policymaking. Although we focus on the applications of behavioral science to policy and governments, we note that the insights apply to organizations more generally.

We start by first digging deeper into the fundamental differences between BI scientists and policymakers that were first identified in chapter 1. This will help us better understand the chasm between research-based knowledge and policy practice and will inform us on how to bridge it and work toward building coherence in designing behaviorally compatible public policy.

THE IMPERFECT MATCH

The two major players in the ecosystem of behavioral public policy are academic scientists and policymakers. Scientists are producers of knowledge, and policymakers consume this knowledge and embed it into policies and programs. Ideally, the two parties pursue the common goal of using behavioral insights to benefit society and build on one another's expertise by focusing on what they are best at. Scientists design and test intervention ideas, often in small-scale, "proof-of-concept" studies; policymakers make use of the research findings and decide on adoption and adaptation based on their domain knowledge and frontline experience.

However, the incentives of both players are not aligned. Scientists' careers are predominantly advanced through publications. Papers that show large effects for solving important policy problems are more likely to be published than non-novel results or small-effect sizes.[1] Therefore, in conducting a proof-of-concept study, scientists are motivated to give an intervention idea the best shot at success by selecting a sample or situation that might yield a large treatment effect. In reporting the findings, scientists tend to leave out details that expose the limitations of the applicability and generalizability of their findings. Scientists may further be tempted to position an intervention as capable of solving a grand policy problem when in reality it might only tackle the problem in some very specific contexts. Therefore, academic research findings are often "premature" – not ready to be embedded into policies *in full scale* – until they have been tested or replicated in field studies representative of a specific context.

Unfortunately, policymakers could easily accept a premature evaluation of a solution because of pressure to reach a solution to a problem. Solution-minded policymakers may be inclined to seize on a seemingly adequate "off-the-shelf" solution[2] before achieving a sufficient understanding of the true nature of the problem and the suitability of the proposed solution. They may read results reported in the academic literature without attending to research details that are consequential to implementation success. If BI is presented as a powerful tool to design public policy, behavioral scientists could be likened to passionate salespersons who promote the tool as powerful, economical, and user-friendly, while policymakers are the excited and enthusiastic consumers who begin to use the tool without going through the details of the user manual. Indeed, the seemingly good match between solution-hungry policymakers and scientists offering attractive solutions can create blind spots that lead to frustrations and ineffective use of BI, from the beginning of a BI project when a policy problem is defined all the way through to the later stages when contextualization and adaptation of BI principles take place.

THE PROBLEM WITH PROBLEMS

"If I had an hour to solve a problem, I'd spend 55 minutes thinking about the problem and 5 minutes thinking about solutions." This famous quotation, attributed to Albert Einstein, underlines that accurate problem definition is critical to finding an appropriate solution. While there is no way for us to verify whether the quotation actually comes from Einstein, we do know that many behavioral-science practitioners – ideas42, eMBeD, BEworks, to name a few – stick to the discipline of starting BI projects with a comprehensive diagnosis of a problem based on fieldwork and qualitative and quantitative research. By sticking to this discipline, BI practitioners guard against the tendency to make assumptions about a problem and propose a solution early on (see chapter 14, for example, for eMBeD's approach). We believe this insight is of particular relevance to policymakers who solve complex problems – what at first seems to be the problem is often a symptom of a deeper problem. When a problem is complex, finding an effective solution entails breaking it up into smaller problems and tackling each (or part) of them in targeted ways. This is the fundamental idea of problem definition; while the idea is deceptively simple, it is often quite challenging in practice.

Imagine a supervisor who gives their team a problem to solve. What would happen if the team tells the supervisor that there are actually five problems in it? It is not inconceivable for the supervisor to lament that the team is creating more problems rather than solving the one pressing problem. Unfortunately, teams will learn from such a pushback that they need to find quick fixes to problems. In a nutshell, the challenge of putting problem definition into practice is that there is a psychological force that sets us in motion toward solving the problem at hand; breaking it up into several problems slows things down and sends the signal that we are not making progress toward finding a solution. This tendency, known as solution-mindedness,[3] stimulates the eagerness to reach a solution to a problem "as given" without due diligence being done to uncover the true character of the problem. Solution-mindedness can be further entrenched by scientists who supply policymakers with

Figure 16.1 Behavioral science and public policy: A nudge-centric view

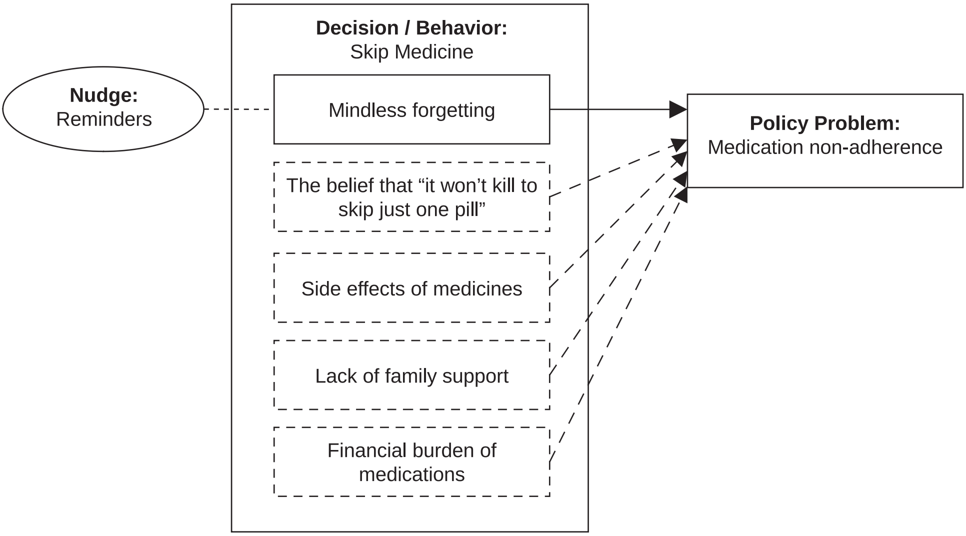
the quick solutions they ask for. We illustrate how this happens using one of the most widely adopted nudges – the use of reminders.

Reminders have been used in various behavioral domains across different countries. In the United Kingdom, sending text messages that remind patients of their upcoming medical appointments, the phone number to call for cancellation and appointment rearrangement, and the cost of missing appointments ("Not attending costs NHS £160 approx.") was found to reduce the incidence of missed appointments from 11.1 percent to 8.4 percent.[4] In Singapore, reminding patients that "missed appointments keep others waiting" reduced the incidence of missed appointments by 6.9 percent.[5] In medical care, reminders have also been found to improve disease-combating behaviors. A number of studies have shown that reminders (through SMS messages or other electronic reminder devices) improve patients' adherence to medication among patients with chronic diseases such as HIV, hypertension, asthma, and glaucoma.[6]

Now imagine that a member of a diabetes care team is identifying ways to improve medication adherence among patients. After surveying the literature on the positive effects of reminders (and knowing that it is inexpensive to do so), they might start thinking about nudging medication adherence by sending patients text messages three times a day to remind them to take their medicines. This sounds like a reasonable judgment that leads to the decision to nudge medication adherence using reminders, as depicted in figure 16.1.

However, a missing step in this process is an analysis of the factors that contribute to medication nonadherence. A closer look at these contributing factors reveals that while some people simply mindlessly forget to take their medicines, others might skip doses deliberatively – for example, diabetic patients may think it is fine to skip when they do not "feel" sick. For patients who hold this belief, the inconvenience and discomfort of taking medicine might outweigh the small perceived benefits of taking the medicine, and so

Figure 16.2 Behavioral science and public policy: A problem-centric view



they intentionally skip some doses.[7] Reminders can reduce mindless nonadherence but not deliberative nonadherence. The adoption of this problem-centric view of medication nonadherence shows clearly that the use of reminders could at best solve part of the problem (figure 16.2). So, back to a healthcare provider's question of whether reminders can solve the medication (non)adherence problem – the answer is "yes" only if the problem is primarily one of mindless nonadherence (i.e., when a significant number of the patients know the importance of taking medicines and intend to take the medicine as prescribed but simply forget to do so).

Taking a problem-centric view of a policy problem from the outset allows the true nature of the problem to show, thereby resulting in a more robust process of finding a truly effective solution. Healthcare providers will come to realize that they must find out which type of patient (the mindless or the deliberative) is more representative of their nonadherent patient population. If a significant number of patients know the importance of taking medicines as prescribed but simply forget to do so, reminders will work very well. But if a significant number of patients hold wrong beliefs about how diabetic medicines work, an informational approach that aims at changing beliefs and raising awareness of the benefits of adherence would be more suitable (see chapter 1 for a discussion of the cocktail of approaches that could be used to solve different shades of problems).

There are hurdles that policymakers must overcome before they gain access to the problem-centric view. Their solution-mindedness is a major hurdle, but the publication strategy adopted by many scientists could aggravate the problem. As we have mentioned, scientists who publish their proof-of-concept studies might position their nudge ideas almost as a silver bullet to a grand policy problem. A meta-analysis that examined thirteen studies on the topic of electronic medication reminders found that none of the studies differentiated between mindless and deliberative nonadherence, despite the theoretical and practical relevance of doing so.[8] If scientists do not tie an intervention idea to a specific contributing factor but instead frame it as a solution to a general policy problem, we can imagine how tempting it is for a solution-minded policymaker to adopt the intervention without going through a thorough problem-definition process.

Our suggestion to policymakers is therefore to resist this temptation, accept that a given intervention can at best address only part of the broader policy problem, and commit to a thorough problem-definition stage of a BI project, one that deconstructs every policy problem in order to drill down to the very specific behavioral issue that needs to be tackled. Although a solution sounds narrow when it can only address part of a problem, policymakers must learn to embrace the "narrowness" of a behavioral solution, because specificity is the key to unlocking the full potential of behavioral public policy.

SAMPLE REPRESENTATIVENESS: DEMOGRAPHICS AND BEYOND

One of the most common questions policymakers ask when they attempt to use BI research findings for policymaking is, "Can the research participants be considered representative of my audience?" Indeed, humans can potentially vary along many different dimensions, and a policymaker's audience may not react the same way to a nudge as those in the original study. It is rare for a nudge to have the same effect on every type of person. Chapter 7 illustrates that some people are more rational than others; some are more nudgeable; while the rest are not ready to change. As many of our partners have pointed out, there is no hard-and-fast rule to finding out whether the participants in one trial are comparable to the targeted population of another; the key is to estimate and manage the uncertainties.

Unfortunately, policymakers may start off assuming that findings are generalizable at some level and that the participants are representative of the population unless they see obvious signs that suggest otherwise. As a result, policymakers incur substantial costs in implementing policies with little prospect of success. For example, after iron-fortified salt was found to reduce anemia among adolescents in several trials, India's National Institute of Nutrition and the Indian government granted licenses to produce iron-fortified salt and issued a policy to encourage broad public adoption of iron-fortified salt. Back then, the assumption underlying this policy was that iron-fortified salt is effective in reducing anemia in all sectors of the population. However, a large-scale experiment[9] subsequently found an increase in hemoglobin only among adolescents but not in any other age groups, so the fortified salt had no effect on the policy goal of reducing anemia in the general public. This finding was very disappointing, considering that vast amounts of money and resources had been allocated to the formulation, production, and public-wide distribution of iron-fortified salt.

In the fortified-salt case, the policy did not consider the representativeness of the research sample in relation to the broader population. Sample unrepresentativeness is one of the major causes of the problem of "voltage drop" – the phenomenon that when a treatment is implemented at scale in a community, the size of the measured treatment effect diminishes significantly relative to that found in the original research study, greatly undermining policy success.[10]

There are two major factors that contribute to sample unrepresentativeness. The first is the scientist's biased choice of subject samples. Scientists might seek out a specific research sample that will benefit significantly from an intervention in order to yield a more promising treatment effect (the "let's give the idea its best shot of working" approach).[11] The reason, as we discussed, is that findings showing large treatment effects often increase publication success. The bias could also be a consequence of constraints on the ability to access representative samples. For example, in Volpp and colleagues' (2008)[12] ground-breaking study examining the use of financial incentives to promote weight loss, 94.7 percent of the participants were males, because the study was conducted at a veterans' affairs medical center. Yet another example comes from the domain of pharmaceutical testing; women of child-bearing potential are often excluded from participation because of concerns about potential adverse reproductive effects. Therefore, historically, women's involvement in clinical trials for prescription drugs has been limited, and prescription drugs have been mostly tested on males. Not until recently have pharmaceutical scientists pointed out systematic gender differences in drug response, and hence the importance of accounting for gender differences in pharmaceutical research.[13] Whether the selection of a subject sample is driven by convenience or interest in securing a promising result, this selection bias may yield a particular group of subjects with different characteristics from the policy-relevant population.

The second contributing factor to sample unrepresentativeness is known as "volunteer bias." People are fundamentally not interested in participating in scientific research unless they can derive meaningful benefit from their participation – the benefit could be money (in the form of a participation fee), but more often it is the expected positive outcomes of the treatment. Therefore, people who volunteer to participate in a research project could be more concerned with the issue being studied than the rest of the population.[14] For example, people who volunteer to sign up for a randomized controlled trial (RCT) that tests the effect of incentives on exercise are likely to be already convinced of the benefits of exercise but need help to get started; they represent only a subset of the wider population that also includes people who think "exercise is not for me."

What can policymakers do to ensure sample representativeness? Currently, the CONSORT (Consolidated Standards of Reporting Trials) guidelines require scientists to report subject characteristics in their research papers, including information on gender, age, income, race, and other socioeconomic information about the sample. It is also mandatory for scientists to provide a detailed diagram showing the number of participants who received and declined the invitation to join a study, and the number of those who quit mid-way, thereby providing policymakers with some (very) basic information with regard to the unrepresentativeness bias.

However, even with such information at hand, it would be challenging for policymakers to determine the nature and extent of sample unrepresentativeness. Consider again the example from the pharmaceutical domain. For a long time, knowing that participants in pharmaceutical trials are predominately males has not led practitioners to question the generalizability of the findings to females, except when there has been broad awareness of the possible connection between gender and medication effects. In addition, several other individual characteristics – such as age, ethnicity, family orientation, culture, and worldview – might also play a role.

To truly tackle the issue, economists and public-policy experts strongly encourage policymakers to subject a behavioral-intervention idea to a "stage-two trial" – a trial conducted by the organization in the community that is representative of where an intervention would eventually be implemented. The aim of conducting this trial is to confront the problems that an intervention would have met with when implemented at scale. Many of the studies conducted at the Privy Council Office of Canada, the ESDC Innovation Lab, eMBeD, and Ontario BIU serve this purpose, at least in part. Also, machine-learning techniques could allow us to identify heterogeneity in response to interventions and hence allow the practitioner to customize interventions as a function of observable features of the respondent.[15]

THE TRICKY SITUATION OF SITUATION REPRESENTATIVENESS

The features of a situation in which an intervention is launched have an effect on intervention effectiveness. A situation encapsulates a collection of variables, so it is challenging for policymakers to anticipate whether any of those variables will interfere with an intervention.

We draw on an example[16] to illustrate the role of situation change on the efficacy of an intervention. The intervention was a pedagogical approach known as "Teaching at the Right Level" (TaRL) designed by Pratham (a nongovernmental organization in India). The basic idea was to organize children by their level of knowledge (but not according to their age) and match the teaching to the knowledge level of the students. In the proof-of-concept study, students identified as "lagging behind" were withdrawn from regular classes for two hours every day and were instead taught by paid community members trained by Pratham. The pedagogical approach was simple and the result promising. However, when TaRL was integrated with the government school system, its overall effect size was much smaller. One interesting observation was that it did not work when it was made part of the in-school effort but did work well when implemented *outside* school hours (e.g., in summer camp). It turns out that in the regular school year, there was strong resistance against TaRL from teachers and parents because of their emphasis on covering the grade-level curriculum during formal school hours. Teachers and parents were, however, more receptive to the idea in summer camps, when there was no formal curriculum to follow.

In the TaRL example, the results differed dramatically when the situation changed from the regular school term to summer camp, even though the same students were involved. Given the complexity of real-life settings, it is almost impossible to fully anticipate the influence of situation characteristics. In academia, scientists are advised to handle this uncertainty by replicating findings in different situations, so that they can explore how treatment effects vary across those situations.[17] Similarly, we encourage policymakers to adopt this strategy in experimentation.

The value of replicating an experiment is evident at the Civil Service College (hereafter, CSC) Singapore, which has been experimenting with blended learning (blending online content with in-class training) to deliver its training programs. Blending has the benefit of letting course participants cover some online reading assignments prior to the course so that the in-class sessions are used more productively to answer questions and facilitate discussion. The challenge, however, is to help participants overcome procrastination and complete the pre-course assignments prior to the in-class sessions. The completion rate was as low as 11–13 percent. This meant that a large majority of the participants would not be ready to dive into deeper discussions in class as intended.

In 2019, a team at CSC redesigned the pre-course e-mail notification to include a nudge with two key features.[18] First, the team made it salient that each reading assignment was bite-sized – they indicated the time required to complete each reading item (e.g., five minutes for Article 1; seven minutes for Article 2). This was designed to make the pre-course reading seem less daunting and hence to reduce procrastination. The CSC team also set the deadline for the online reading assignment to be on the Friday of the same week as the notification was sent out. Prior research has shown that people tend to procrastinate less when the deadline is in the same time period (e.g., the same week) than when it is in a later time period (e.g., next week).[19] Nevertheless, the decision to use the "this-week intervention" was not taken trivially, because the task and situation studied in the original article were clearly different from the blended learning task. So the CSC conducted a small trial to test the idea; the "this-week intervention" group was tested against a control group, which was given two weeks to complete the same assignment. The initial results showed that 50 percent of the "this-week intervention" group completed the online reading assignment by the suggested deadline, compared to only 12.5 percent in the control group, even though the latter had more time to finish the assignment.

In the second step of putting the idea of replication into practice, the CSC team sought to find out whether the same nudge would work as effectively in a course covering different content or one with heavier content. While the trial is still ongoing, initial findings already show differences in the effectiveness of the "same-week intervention" among courses of similar duration but different content, even though the broad objective of improving the completion of the reading assignment has still been met. The findings of these replication efforts will direct the team to the best use of the "this-week intervention." This is where and how replicating the same intervention is key to understanding the generalizability of the nudge.

FORMING AN ECOSYSTEM OF LEARNING

Making BI work in solving various policy problems is almost a moving target in the complex and fast-changing environment we operate in. There is a plethora of target groups and contexts such that interventions that work in one context may not in others, or even if they did, could evolve over time. Therefore, in our last point, we want to emphasize the importance of having a platform for sharing and learning in order to save resources and promote the generation of new ideas.

As mentioned in chapter 15, often there are only a handful of BI experts serving an organization. Therefore, the capacity for BI experts to support "non-BI" colleagues is often limited, and collaborations on BI projects may only last for a limited time. To keep the BI momentum going after a project is over, it is important for organizations to build organization-wide BI capacity. Having access to resources for upgrading BI sciences will facilitate learning and keep people engaged. A platform for people to learn about one another's BI initiatives will serve this purpose.

The Ecosystem of Learning: Aspirations of the Singapore BI Community. In Singapore, BI application to public policies is decentralized across ministries and statutory boards, with the Innovation Lab housed in the Public Service Division (PSD) working with agencies through an innovation process that incorporates BI (among other tools) in interagency projects.[20] In such a landscape, the Civil Service College (CSC) plays a unique role in curating and facilitating an ecosystem of learning opportunities across the public service. This ecosystem goes beyond large-scale conferences and workshops to include more informal settings such as the BI and Design Community of Practice within the government, where agencies get to share their projects and to hear from others in the community. The next phase of work aims to increase the vibrancy of the ecosystem by boosting learning opportunities via "LEARN," an online learning platform and an internal repository of BI projects aimed at more organic sharing and at forming networks across the public service more nimbly.

However, it is not just the platforms for learning that matter but also the nature of the content that is shared. It is easy to celebrate successes (experiments that yield positive and statistically significant results) but much harder to be open about what can be learnt from experiments with nil or negative results. This is because, in the policymaking environment, the actions of bureaucrats and politicians are greatly affected by the need both to claim credit and to avoid blame. Often, policymakers will prefer inaction (i.e., status-quo bias) because of the fear of making mistakes and attracting blame.[21] Frequently, not making mistakes (rather than gaining praise for policies done right) is sufficient for career progression.

Bearing this in mind, experimentation that includes creating a fail-safe environment is crucial to the successful integration of BI in organizations,[22] because not all interventions will be spot-on no matter how well they have been designed or how much deliberation has gone into the process. In other words, learning to using BI effectively in policy design involves learning from experiments that yield both positive and negative results. In an RCT that the Singapore Ministry of Manpower ran to test messaging that reminded self-employed people to make mandatory Medisave (health insurance) contributions, the use of infographics (in the form of cartoons) was found to reduce contributions.[23] What was thought to be an effective way to explain difficult government policies in public communications did not work in this context. The team hypothesized that using cartoons may have trivialized the subject rather than encouraging compliance.

Experiments like this that yield nil or negative results could have been readily seen as a failure, not to be shared or mentioned further. However, it is possible to frame such findings as ways to prevent potentially large investments in programs that would not have succeeded, or worse, that would have produced unintended negative results. Sharing and learning from such experiments is thus just as valuable as building on those that yield positive results, because they help policymakers to avoid making mistakes on a larger scale. Therefore, a repository of results (or a “what-works” database, as

Feng, Kim, and Soman call it in chapter 2) is actually an excellent way to bring down the costs of experimentation and to increase the chances of success.

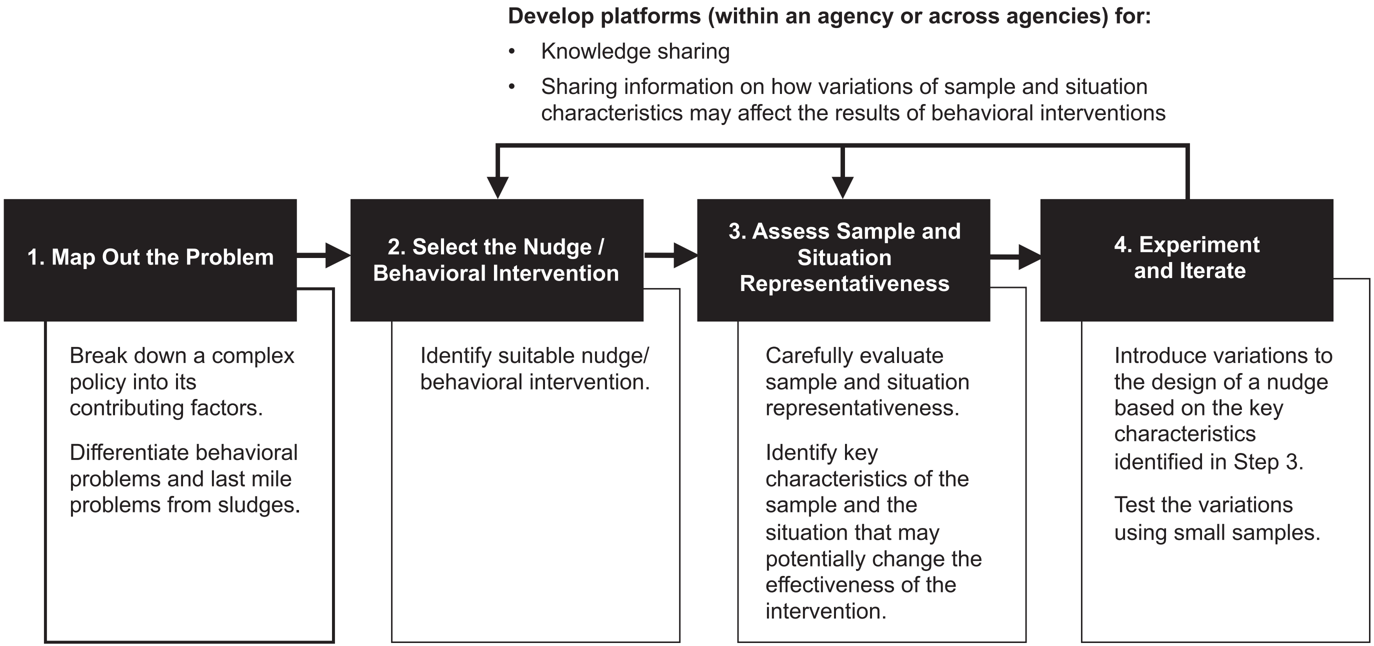
In short, to create and facilitate an ecosystem of learning opportunities to make BI work effectively in public policies, attention needs to be paid to what is being shared (both successes and failures), as much as to the availability of channels and platforms for such sharing and learning.

MAKING IT WORK

"Making it work" is the theme of the last two chapters of this book. Interestingly, we often talk about "making it work" when "it" is something challenging. Indeed, the implementation of BI in policies can be challenging. But these challenges can be overcome. In this chapter, we discuss a sequence of things to think about that could guide policymakers through a more fruitful implementation of BI. The sequence is summarized in figure 16.3. We have gone beyond discussing the importance of making the right choice of intervention (something that our readers are well aware of) to explain how the other steps – problem definition, assessment of sample and situation representativeness, and experimentations and iterations – determine the success (or failure) of the intervention.

A roadmap of how to make things work often entails not only "what to do" but also "when to do it." Healthcare providers may order a CT scan of the brain for people who have head injuries (what to do), but most of the time CT scans are ordered only when specific signs are observed, such as reduced vision, repeated vomiting, tenderness over the skull, and so on (when to do). This is so because unnecessary tests and scans use up scarce resources. In an ideal world, more testing – assuming that the testing has high validity and reliability – can only be a good thing. However, too much testing comes at a cost. Hence, it would be helpful for practitioners to have a checklist to help them assess when in-situ testing (step 4, figure 16.3) is relatively more valuable, and when it might be skipped in the face of cost and other constraints.

Figure 16.3 **Our roadmap**



Here, we put together a list of questions that practitioners can ask. The greater the number of *Yes* responses, the higher the marginal value of running in-situ testing of an intervention:

- Does past evidence of the intervention come primarily from laboratory studies rather than field studies?
- Does past evidence come from a discipline or domain (finance, health, education, tax payment, charity, environment conservation, energy saving, etc.) that is different from the one in which your intervention would be implemented?
- Compared with the RCT in which the intervention was previously tested, is the audience significantly different in terms of the dimensions listed in the CONSORT table of participant characteristics?
- Would intervention failure harm your audience? (Yes = there will be harm; No = simply no effects or status quo).

This checklist is by no means intended to replace in-situ testing, nor does it replace the judgment of professionals. It provides preliminary guidance for practitioners who are under significant resource constraints for running in-situ testing, face a high cost of experimentation, or need to conserve resources for testing interventions in other more complex and dynamic environments (see Soman's discussion under "Identify the Value of Being Evidence Based"in chapter 1, and the call for future research on this topic). Use of these pragmatic guidelines will increase the feasibility of using BI in policymaking, thereby facilitating a positive impact on the public and in organizations from the application of behavioral science.

NOTES

1 Al-Ubaydli, O., Lee, M.S., List, J.A., Mackevicius, C., & Suskind, D. (2019). How can experiments play a greater role in public policy? 12 proposals from an economic model of scaling. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2019-131); Al-Ubaydli, O., List,

J.A., & Suskind, D. (2019). The science of using science: Towards an understanding of the threats to scaling experiments. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2019-73); National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. Washington, DC: The National Academies Press.

2 Soman, D., & Hossain, T. (2020). Successfully scaled interventions need not be homogeneous. *Behavioral Public Policy*. In press.

3 Solem, A.R. (1958). An evaluation of two attitudinal approaches to delegation. *Journal of Applied Psychology, 42*(1), 36–9; Maier, N.R., & Solem, A.R. (1962). Improving solutions by turning choice situations into problems. *Personnel Psychology*, *15*(2), 151–7.

4 Hallsworth, M., Berry, D., Sanders, M., Sallis, A., King, D., Vlaev, I., et al. (2015). Stating appointment costs in SMS reminders reduces missed hospital appointments: Findings from two randomised controlled trials. *PLOS One*, *10*(9), e0137306.

5 Chen, L. (2017). Using behavioral insights to improve service delivery. *Ethos*, *17*, 17–24. Civil Service College Singapore.

6 Vervloet, M., Linn, A.J., van Weert, J.C., De Bakker, D.H., Bouvy, M.L., & Van Dijk, L. (2012). The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: A systematic review of the literature. *Journal of the American Medical Informatics Association*, *19*(5), 696–704.

7 Wroe, A. (2002). Intentional and unintentional nonadherence: A study of decision making. *Journal of Behavioral Medicine*, *25*(4), 355–72.

8 Vervloet, M., Linn, A.J., van Weert, J.C., De Bakker, D.H., Bouvy, M.L., & Van Dijk, L. (2012). The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: A systematic review of the literature. *Journal of the American Medical Informatics Association*, *19*(5), 696–704.

9 Banerjee, A., Barnhardt, S., & Duflo, E. (2018). Can iron-fortified salt control anemia? Evidence from two experiments in rural Bihar. *Journal of Development Economics*, *133*, 127–46.

10 Al-Ubaydli, O., Lee, M.S., List, J.A., Mackevicius, C., & Suskind, D. (2019). How can experiments play a greater role in public policy? 12 proposals from an economic model of scaling. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2019-131); Al-Ubaydli, O., List, J.A., & Suskind, D. (2019). The science of using science: Towards an understanding of the threats to scaling experiments. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2019-73); Banerjee, A., Barnhardt, S., & Duflo, E. (2018). Can iron-fortified salt control anemia? Evidence from two experiments in rural Bihar. *Journal of Development Economics*, *133,* 127–46. Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., et al. (2017). From proof of concept to scalable

policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, *31*(4), 73–102.
11 Ibid.
12 Volpp, K., John, L., Troxel, A., Norton, L., Fassbender, J., & Loewenstein, G. (2008). Financial incentive-based approaches for weight loss: A randomized trial. *JAMA*, *300*(22), 2631–7.
13 Liu, K.A., & Dipietro Mager, N.A. (2016). Women's involvement in clinical trials: Historical perspective and future implications. *Pharmacy Practice (Granada)*, *14*(1), 1–9.
14 National Research Council. (2013). *Nonresponse in social science surveys: A research agenda*. Washington, DC: National Academies Press.
15 Shah, A., Osborne, M., Lefkowitz, J., Fishbane, A., & Soman, D. (2019, 25 September). Can making family salient increase financial savings? Quantifying heterogeneous treatment effects in voluntary retirement contributions using a field experiment in Mexico. http://dx.doi.org/10.2139/ssrn.3460722.
16 Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., et al. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, *31*(4): 73–102.
17 Al-Ubaydli, O., Lee, M.S., List, J.A., Mackevicius, C., & Suskind, D. (2019). How can experiments play a greater role in public policy? 12 proposals from an economic model of scaling. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2019-131)
18 Lim, C. (2020). Nudging a higher completion of pre-course online assignments for blended learning. *Institute of Governance and Policy, Civil Service College*. Ongoing work.
19 Tu, Y., & Soman, D. (2014). The categorization of time and its impact on task initiation. *Journal of Consumer Research*, *41*(3), 810–22. doi:10.1086/677840.
20 Afif, Z., Islan, W.W., Calvo-Gonzalez, O., & Dalton, A.G. (2018). *Behavioral Science around the World: Profiles of 10 Countries* (English). eMBeD brief. Washington, DC: World Bank Group.
21 Ho, T-H., Leong, C., & Yeung C. (2020). Success at scale: Six suggestions from implementation and policy sciences. *Behavioral Public Policy*. In press.
22 Kok, P-S. (2017). Nudges: Why, how and what next. *Ethos, 17*. Civil Service College, Singapore.
23 Ibid.