Getting Practical with Causal Mechanisms: The application of Process-Tracing Under Real-World Evaluation Constraints

Estelle Raimondo (Independent Evaluation Group)[1]

Abstract

Over the past decade, the field of development evaluation has seen a renewed interest in methodological approaches that can answer compelling causal questions about what works, for whom, and why. Development evaluators have notably started to experiment with Bayesian Process Tracing to unpack, test, and enhance their comprehension of causal mechanisms triggered by development interventions. This paper conveys one such experience of applying Bayesian Process Tracing to the study of citizen engagement interventions within a conditional cash transfer program under real-world evaluation conditions. The paper builds on this experience to discuss the benefits, challenges, and potential for the applicability of this approach under real-world evaluation conditions of time, money, and political constraints.

Keywords: evaluation methods, process tracing, citizen engagement

[1] Eraimondo@worldbank.org Estelle Raimondo, PhD is a Senior Evaluation Specialist at the Independent Evaluation Group. I would like to thank, with no implications, Eduardo Fernandez Maldonado, Elena Bardasi, Jos Vaessen, Barbara Befani, Helene Grandvoinnet and the whole team implementing the "Reportes Comunitarios" in the *Progresando con Solidaridad* Program of the Dominican Republic, in particular Evangelista Cornelio The author's views, findings, or interpretations do not necessarily coincide or reflect the views of the World Bank, the Independent Evaluation Group, its affiliated institutions, or its executive directors.

Introduction

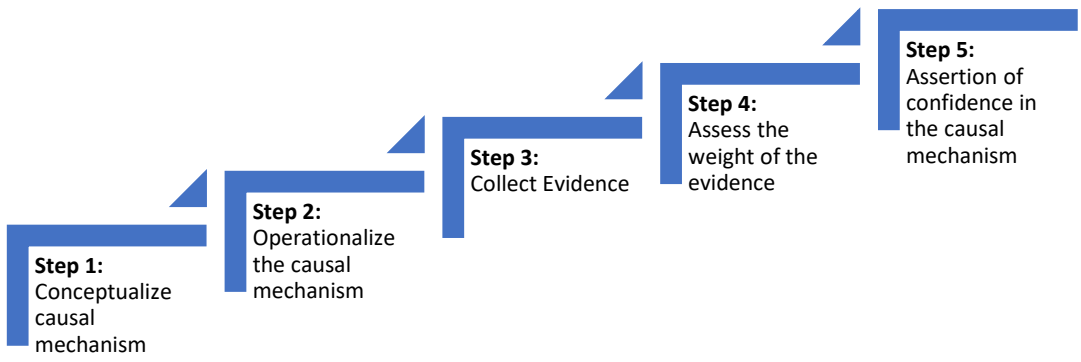
The necessity to open the "black-box" of development interventions resonates well with most development evaluators nowadays and materializes into a growing appetite for mechanism-based explanations. Three concurrent patterns may explain the increased popularity of causal mechanisms in development evaluations. First, after a period of active methodological warfare in the mid-2000s, a consensus on the value of understanding the inner-working of programs—in addition to their net-effect on specific outcomes of interest—seems to have arisen. Moreover, there has been an increased awareness that crucial development interventions, worthy of causal analysis, did not lend themselves to (quasi)experimental testing. Finally, there has been a recognition that conventional evaluation approaches have failed to comprehend the complexity inherent in the development enterprise (e.g., Bamberger, Vaessen, & Raimondo, 2015; Byrne, 2013; Forss et al., 2011; Befani, 2012). The groundbreaking DFID report (Stern, Stame, & Mayne, 2012) on expanding the range of impact evaluation methods in international development epitomizes this new appetite. It borrows from other social science disciplines to propose methodological avenues for empirically studying causal mechanisms, including Bayesian Process-Tracing (BPT) (Beach & Pedersen; 2013; George & Bennett, 2005; Bennett, 2008).

Over the past 5 years or so, the field of development evaluation has seen a slow but enthusiastic take-up of various methodologies. Evaluators and commissioners of evaluations are increasingly willing to experiment with alternative approaches, but most operate within real-world evaluation constraints, of limited time, money, and the need to reconcile multiple demands (Bamberger & Mabry, 2019). Many are thus wondering if it is worthwhile investing in studying causal mechanisms and are seeking "proofs-of-concept" to better understand the feasibility, costs, and benefits of such application. Yet, the collective experience of applying mechanism- based approaches, notably Process Tracing, to real-world evaluation contexts remains scarce.

Can Bayesian Process Tracing (BPT) provide relevant insight into com- plex causal-mechanism, and is it applicable under real-world evaluation challenges? Two additional lines of inquiry

motivate this article. First, to what extent can BPT bolster our capacity to establish a causal link between development interventions and outcomes? Second, in so doing, does BPT also enhance the policy relevance of the evaluative knowledge produced? This article seeks to provide some elements of responses to this practical conundrum. In short, BPT consists of looking for observable manifestations of hypothesized causal mechanisms within a single case and weighing the strength (probative value) of the evidence collected to make inferences about whether the causal mechanism was present in the case and whether it functioned as expected (Beach & Pedersen, 2013, p. 81). BPT uses Bayesian logic of inference by asking how we should update our confidence in the causal explanation in light of new evidence collected. This article proposes a demonstration of the application of Process Tracing to study the mechanism of "citizen action" in the context of international development interventions. This research was embedded in a larger evaluation of the World Bank's support of Citizen Engagement conducted by the Independent Evaluation Group (IEG) in 2017 (World Bank, 2018).

Figure 3.1. The five steps of Process Tracing.



Source: Punton, M., & Welle, K. (2015). Adapted with permission.

After providing some background on the context of the evaluation, the paper explains how the team applied the five steps of Process-Tracing, illustrated in Figure 3.1. In the methodology section, we describe step 1 (conceptualizing the causal mechanism) and step 2

(operationalizing the causal mechanism). In the data collection and analysis section, we go through step 3 (collecting evidence) and step 4 (weighting the evidence). The final section presents the fifth step (forming conclusion) and draws lessons from this pilot by discussing the specific challenges in applicability and the added-value of studying causal mechanisms for the quality and use of the evaluation.

Choice of Methodological Approach

In recent years, we have observed an intensification of aid agencies' efforts to put citizens front and center in defining their development agenda. The notion that citizens are "change agents" and can improve development outcomes resonates increasingly with development discourse and practice, especially in the Sustainable Development Goals era. The World Bank decided in 2014 to mainstream citizen engagement activities in all of its projects where direct beneficiaries could be identified. In making this policy commitment, the World Bank claimed that engaging citizens was not only the "right" thing to do, but it was also going to improve the effectiveness of its projects (World Bank, 2018).

The choice of adopting BPT as the methodological approach to investigate the inner working of citizen engagement mechanisms was motivated by three considerations: gaps in the literature, practical knowledge needs, and feasibility.

First, the literature to date on the impact of social accountability and citizen engagement interventions on development outcomes is quite mixed. A few studies have reviewed the existing evidence on whether citizen engagement improves development outcomes (Fox, 2015; Gaventa & Barrett, 2010; Grandvoinnet, Aslam, & Raha, 2015; Mansuri & Rao, 2013) pointing to a wide range of possible effects—including unintended negative effects such as elite capture—depending on the context. Scope conditions, such as the country development stage, the client government's buy-in, and capacity, and the level of societal inequality, conflict, and fragmentation were thought to have a strong influence on the process and its results. So far, establishing a causal link between citizen engagement activities and development outcomes has been fraught with methodological challenges, given the heterogeneity of both the intervention and the outcome space (Joshi, 2013; Ringold, Boyd, Landers, & Weber, 2013). So

far, the body of empirical knowledge on the topic mostly consists of descriptive case studies, surveys, or correlation analysis.

Second, little attention had been paid to the causal mechanisms that underlie such change processes both at the behavioral and at the macro (institutional levels). Answering the question—what causal mechanisms are at play to move from a situation of little voice for citizens to active citizen's actions, and responsive state's action?—had the potential to provide practical insights to a very wide portfolio of interventions within the World Bank. We opted for a causal mechanism approach that could elicit the process mechanisms, but also allowed us to investigate the behavioral mechanisms at the individual citizen level. We expected, BPT to yield useful, practical insight because of the following reasons: (1) it allows unpacking and testing the hypothesized causal mechanism through a rigorous and trans- parent analytical process; (2) it enables identifying whose action through which activities is a constitutive part of the change processes; (3) it can identify what scope conditions enables the mechanism to be triggered both at the local and institutional levels.

Third, we considered the feasibility of the approach within a broader thematic evaluation conducted by IEG as well as the potential for the transferability of the findings to other operations. We selected a "typical case" of the application of citizen engagement activities, that had been replicated in a number of operations and zeroed in on the case of participatory monitoring in the national Conditional Cash transfer of the Dominican Republic called "Reportes Comunitarios" (Aston and Cavatore 2016; World Bank, 2018).

Methodological Design

In Process Tracing, the causal mechanism consists of a process made up of multiple interlocking parts. In our application, we followed Beach and Pedersen (2013; 2016) approach and conceptualized the mechanism as parts composed of entities engaging in activities that together form a system transmitting causal forces from the intervention to the outcome (Machamer, Darden, & Craver, 2000). Specifically, the evaluation sought to assess how citizen engagement intervention (X) could help improve the quality of service delivery at the local level (Y).

The first step of the Process Tracing methodology consists of formulating the causal mechanisms so that it provides a full explanatory account of the process that links the intervention with the outcome (Beach & Pedersen, 2016, p. 465). We proceeded through a two-pronged analytical process. First, the evaluation team reviewed the empirical literature on citizens' engagement in three different development sectors (public service delivery, natural resource management, and social inclusion) to identify alternative or similar conceptualization of the causal mechanism. Second, the team conducted a number of interviews and focus group discussions with the World Bank and outside experts on citizens' engagement to corroborate or recalibrate specific parts of the causal mechanism.

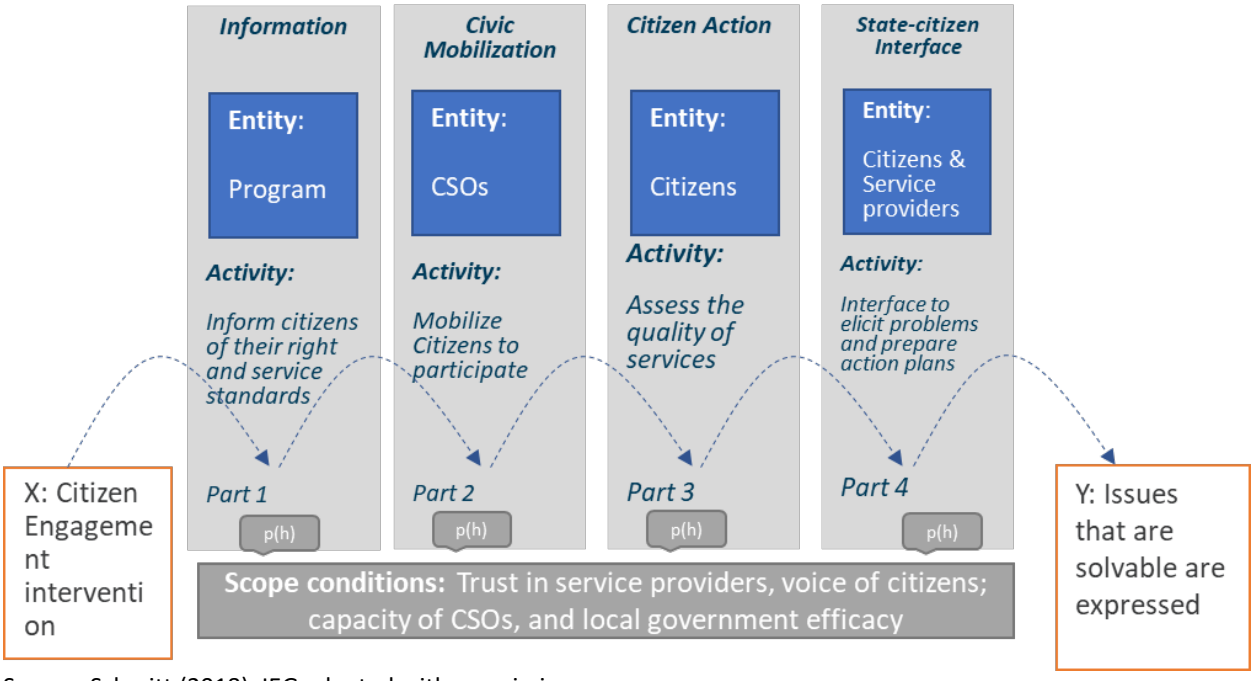
The social accountability literature (Grandvoinnet et al., 2015) was the evaluators' starting point to develop the hypothetic causal mechanisms to be tested via process tracing. There are two main intersecting causal mechanisms that need to be triggered for the outcome to come about: "citizen action" and "state action." The first causal mechanism consists of triggering human motivation to voice grievances to service providers in a context of low initial trust and low incentives. The second intersecting causal mechanism consists of triggering a response from local state actors to these grievances, and ultimately initiate a virtuous cycle of trust and voice to improve service delivery. This second part of the mechanism was also assessed in the evaluation, but we will not detail it in this paper. The causal logic of the "citizen action" mechanism hypothesizes a causal mechanism in four parts (Information; Civic Mobilization; Citizens Action; and State-Citizens interface). In addition, there are several scope conditions that can either support or hinder the citizen action that needs to be taken into account. These comprise the existing level of trust in service providers, the existing level of the voice of citizens; capacity of local civil society organizations; and local government efficacy. The causal mechanism is described in Figure 3.2.

The second step of the methodology is the operationalization of the various parts of the causal mechanism into a series of tests. The logic of empirical testing in process-tracing is well summed-up by Beach and Pedersen (2013, p. 101):

"…if we expected X to cause Y, each part of the mechanism between X and Y would leave the predicted empirical manifestations that can be observed in empirical material. Detecting these manifestations, or fingerprints, requires the development of carefully formulated case-specific predictions of what evidence we should expect to see if the hypothesized part of the mechanism exists."

In this step, we started to apply Bayesian logic to guide our evidence- gathering efforts. In practice, this step consists of thinking about the evidence we should be able to observe if the causal theory is valid. We brainstormed with experts in citizen engagement and in the program about the likelihood of finding specific pieces of evidence if the program truly triggered citizen action and state action to enhance service delivery. We listed the type of evidence that would give us the greatest confidence in the mechanism. To avoid confirmation bias, we also listed the type of evidence that would significantly decrease our confidence in the causal mechanism, and we planned our data collection in the field to look for both.

Figure 3.2. Operationalization of the causal mechanism of Citizen Action.



Source: Schmitt (2018). IEG adapted with permission.

Data Collection

The task of data collection corresponds to step 3 of the BPT sequence. The application of BPT helped us to efficiently organize data collection and focus on gathering evidence that would have the highest probative value. We will illustrate our approach by focusing on how we tested the "interface" part of the citizen action mechanism, which is a critical part of the theory.

First, we strategically gathered "account evidence" through a number of interviews and focus groups with key informants. In selecting which informant to interview, we applied Bayesian thinking to balance evidence of confirmatory power with evidence of disconfirmatory power. We also made sure that respondents were as independent of each other as possible, meaning that they had not discussed the topic of the interviews together beforehand and that they had different stakes and interests in the intervention and its outcomes. For example, if the "interface" part of the citizen action mechanism was true, we expected that staff and community liaisons in charge of facilitating the interface meeting would confirm that the interaction produced agreements between citizens and service providers on action plans. If they disagreed, our confidence in the claim would have diminished significantly. We made sure to cover these informants, but without oversampling them, as their responses did not have confirmatory power.

On the other hand, if members of the community who did not take part in the interface meetings—and thus did not have a stake in providing socially desirable answers to our questions—also vouched for the transformative nature of the interface meetings, then our confidence in that part of the mechanism would be higher. Similarly, we spent more time gathering account evidence from independent civil society actors who were neutral observant of the process, than with officials of the programs.

As is often the case in real-world evaluation settings—of time, money, and political constraints—everything did not go as planned. Specifically, this pro-poor program is highly sensitive and publicized, often used in political campaigning. It was thus natural for the program to organize a press delegation to follow the World Bank evaluators as they learned and talked to program beneficiaries and local actors. However, for applying BPT, this was highly

problematic. The level of bias that creeps into interviews conducted while surrounded by a camera crew does not pass the evidence muster. We thus had to discard some evidence collected during our first day of the mission.

Second, we focused on collecting “trace evidence.” For instance, we set out to gather and analyze meeting records of the “interface meetings,” as well as the audited database of action plans and the status of the follow- up for each action maintained by the ministry in charge of the program. Having access to this database helped us confirm that the meeting took place, that action plans were produced, and that the progress on realizing these action plans was tracked and escalated to the appropriate entities in the decentralized government units. We were able to observe whether the grievances of citizens, which had been translated into action plans, appeared in the database as “solved” were actually resolved. For example, we stopped at several schools and community health centers to check on the facilities, opening hours, presence of teachers, and health workers. We also attended district coordination meetings, unannounced. We made sure that our spot- checking was not guided by the program staff to diminish the risks of biases and boost our confidence.

Third, we gathered “pattern evidence,” such as data from a survey that had been conducted for the purpose of another study but contained useful information on the level of engagement of citizens in the intervention and their patterns of behaviors. We were able to obtain the raw data and exploit it for our own purpose. For example, we were able to gather evidence that contributed to enhancing our confidence in the first and second parts of the “citizen action” causal mechanism. Notably, we found that before the implementation of the “reportes comunitarios” beneficiaries of the cash transfer had limited avenues to voice their complaints and that in villages where the intervention was implemented, there was improved civic mobilization, willingness to voice concerns and take part in the solution.

Finally, we collected “sequence” evidence, reconstructing the chrono- logical sequence that underpinned the improvement of specific services to the community and testing whether the sequence of activities and events confirmed or not the causal mechanisms. For example, if a particular set of school latrines had been fixed prior to an “interface” taking place, this would

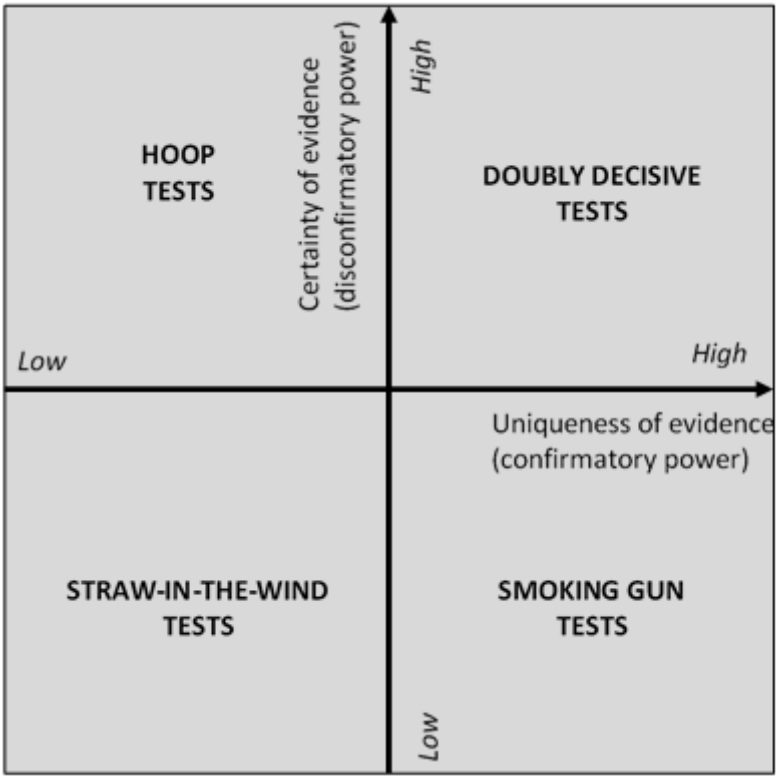
have weakened our confidence in that part of the causal mechanism playing out within our case.

Data Analysis

In steps four and five, we analyzed the evidence collected, assessed the probative value of each piece of evidence, and weighted their relative contribution to updating our confidence in the causal mechanism in our case (Beach & Pedersen, 2013, p. 83). Simply put, for each piece of evidence gathered, we asked ourselves two questions: First, what type of inference can we make based on having found the predicted evidence and to what extent does it imply updating (increase or decrease) our confidence in the existence of each part of the causal mechanism? The second question we asked ourselves was, what type of inference could we make if we did not find the predicted evidence even after searching for it? In other words, if we do not find a piece of evidence even after having thoroughly looked for it, does it mean that the part of the mechanism does not exist or work as expected? Here, evidence that we really expected to find in the case, if not found, leads us to decrease the confidence in the part of the causal mechanism by larger margins than evidence that we rated less likely to be found in the field.

Each piece of evidence can thus pass or fail four types of Process- Tracing tests, depending on its combined level of uniqueness to the theory and certainty, as illustrated in Figure 3.3. The weakest test is a "straw- in-the-wind test," pieces of evidence that are neither unique nor certain. In our case, for instance, the interviews conducted during the first day under the projectors of the communication team would all qualify as "straw-in-the-wind." Failing a "hoop test" reduces our confidence in the causal mechanism; but finding it does not help us confirm the evidence. For example, if program actors, citizens and service providers who participated in the interface meetings did not provide account evidence that the interaction had been constructive, that the action plans were meaningful and that the feedback was taken seriously, then our confidence in the interface part of the citizen action mechanism would have been highly diminished. However, a positive appreciation from these participants, who have a stake in the process, is not unique enough to help us confirm the claim.

Figure 3.3. Four Process Tracing tests.

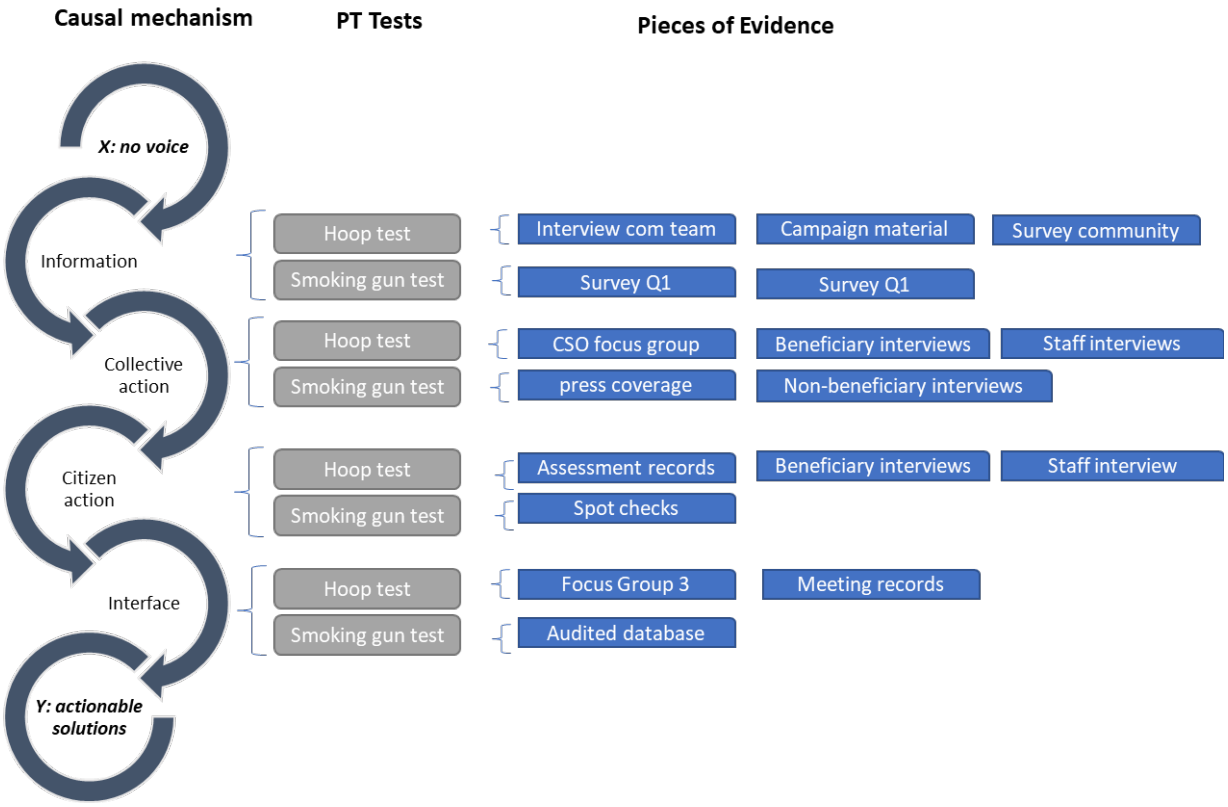


Conversely, “smoking gun tests” are quite unique but have low certainty. Finding a “smoking gun” evidence helps increase our confidence in the causal mechanism; but failing to observe it does not disconfirm the claim. In our case, gaining access to trace and sequence evidence in the form of an audited database of information about interface meetings taking place and generating action plans combined with evidence that these action plans are monitored and followed-up passed the smoking-gun test. The strongest test is the “doubly decisive” as it combines both certainty and uniqueness. However, finding pieces of evidence that pass this type of test is very unlikely in real-world evaluation.

What is more feasible is to find multiple pieces of evidence balancing hoop and smoking gun tests on each part of the causal mechanism to augment the probative value of the combined pieces of evidence. In practical evaluation situations, there is hardly ever a single piece of evidence with sufficient probative value. Instead, we aim to collect a cluster of pieces of

evidence that together are high in probative value and allow updating the hypothetic causal mechanism in the given case. Figure 3.4. illustrates the structure of the empirical evidence.

Figure 3.4. Structure of the empirical evidence.



Source: Schmitt (2018). IEG adapted with permission.

Step five consisted of updating our confidence in each part of the causal mechanism. In addition to the evidence itself, we needed three parameters to be able to weight the probative value of the evidence to help us strengthen or weaken our belief about each part of the causal mechanism. First, we needed a "prior," which is our initial confidence in the validity of each part of the causal mechanism based on prior knowledge and before collecting the evidence. To formulate this prior, we reviewed the existing empirical literature and consulted with experts on social accountability. For parts of the causal mechanism where there was little existing evidence to inform our initial confidence level, we used "uninformed" priors of 0.5. Second,

we needed to specify "sensitivity," which is the likelihood that we would find a particular type of evidence in the case if the causal mechanism was true, and "type 1 error," which is the likelihood that we would find the same evidence if the causal mechanism was false. Pieces of evidence that have low type I error are considered unique to the theory and have high confirmatory power. We used "subjective" likelihood ratios, as described by Bennett (2019). We consulted with experts on social accountability and the local context to mobilize their knowledge and asked them to "bet" on the strength of evidence, using a qualitative rubric for different levels of confidence that we adapted from Befani and Stedman-Bryce (2017). The rubric had six levels of strength, from "practical certainty" (which is equivalent to a 0.99+ level of confidence) to "no information" (which represents a 0.50 level of confidence).

Our approach to Bayesian updating can be qualified as "informal," instead of aiming for precision in each parameter we used a qualitative grid to assess the extent to which the confidence in each part of the claim had increased or decreased based on the evidence (Bennett & Checkel, 2014). We used the Bayesian formula described in Befani and D'Errico (this issue) as a heuristic tool for each part of the causal mechanism. Given that the internal validity of the causal mechanism is only as strong as its weakest part, we proceeded with testing each part of the causal mechanism before asserting our confidence in the entire mechanism. As discussed below, we were able to significantly and transparently increase our confidence in the effectiveness of the citizen engagement intervention in improving service delivery at the local level.

Findings and Discussion

Piloting process-tracing under real-world evaluation conditions was meant to generate lessons on the applicability, value-added, and potential for replicability of the approach. In this section, we reflect on our experience and seek to answer two questions: Did the use of process tracing to test the causal mechanisms of citizen and state action enhance the internal validity of our findings? My answer to this question is a wholehearted yes. Beach (2019) lays out three core criteria that determine the degree of internal validity of mechanistic findings: (i) the quality of the mechanistic explanation (an unbroken causal mechanism); (ii) the uniqueness and

confirmatory power of the expected empirical fingerprint left by the causal mechanism in action; (iii) the trustworthiness of the sources of evidence used in practice. By making the process of evidence gathering and weighting the probative value of the evidence for each part of the causal claim, the team achieved a level of transparency and rigor rarely achieved in typical case study approaches. The application of the methodology also bolstered the defensibility of the findings.

Note that we also considered assigning numerical probabilities to the various pieces of evidence to compute the likelihood ratio and apply the Bayesian updater. We tried it for one part of the causal mechanism. The extent to which this bolstered the internal validity of the finding remains questionable. It was tedious and sometimes counter-productive to assign the probabilities. More importantly, it was very difficult to communicate with program staff and other evaluation stakeholders. We thus decided to go for the informal variant of BPT. However, the application of the Bayesian logic as a heuristic was well understood and appreciated by the main audience to the evaluation.

Now, did unpacking and testing the causal mechanism underlying citizen engagement activities fulfill its promise of enhancing policy relevance? Here my answer is a "qualified yes." It certainly enhanced the evaluation team's understanding of the behavioral, operational, and institutional inner workings of the operation and the conditions under which citizen engagement could transmit causal power to change the quality of services. For example, we were able to recommend to the program that the interface meeting be facilitated by civil society institutions, as opposed to program staff, to enhance the confidence of citizens in the integrity of the feedback process. The policy relevance of our findings for the program, which is implemented across regions, also hinges on the generalizability of the findings from the case that we studied in other cases. In the case of citizen engagement interventions, there is extreme heterogeneity in the scope conditions that would enable the causal mechanism to generate an outcome, many of which are cultural, or steeped in the history of a place. In order to enhance the generalizability of our findings, we paired process-tracing with a comparative case study design, sampling interventions across a wide range of contexts (Beach, 2019). The process- tracing

study alone would not have been sufficiently robust to communicate policy-relevant findings to our main audiences.

Moreover, communicating the findings of the study to the program staff and the government counterparts required a significant amount of knowledge, translation and brokering. Reaching the right balance between conveying the rigor of the exercise—which requires keeping a certain level of technicity—and communicating clear and concise messages about the findings and their policy implication was a challenge and required multiple iterations in writing and oral presentation. Given the piloting nature of this exercise, the evaluation team also had to demonstrate to the leadership team the uniqueness and "value for money" of the exercise compared to more conventional case studies approaches. This was particularly challenging in the absence of well-known standards of quality and rigor in case-based methods.

Finally, to what extent is this approach applicable and replicable in other real-world evaluation settings? Here I am more pessimistic. On the one hand, I believe that some core methodological principles underlying process tracing can be adopted widely to bolster the transparency of the evidence gathering and protect against confirmation bias. Additionally, process tracing principles can help distinguish between absence of evidence and evidence of absence, which are very common in evaluation (D'Errico, Befani, Booker, & Guiliani, 2017, p. 39). On the other hand, a robust application of process-tracing requires very thorough and intense preparation, data collection, and data analysis process in a single case. Many real-world evaluations are required to look at many evaluation questions, with multi-level evaluand, across a multiplicity of evaluation criteria and dimensions. In this context, finding opportunities for rigorous within-case causal analysis is challenging and shares some of the issues encountered by rigorous experimental designs, insofar as it answers only a limited set of questions, provides high internal validity, but requires relatively lengthy preparation and analytical phases, and requires specialized expertise.

References

Aston, T., & Cavatore, M. (2015). Sistematización general del proceso de implementación a escala nacional del monitoreo participativo de servicios sociales mediante “Reportes Comunitarios” en el Programa Progresando con Solidaridad. London: Care International UK.

Bamberger, M., & Mabry, L. (2019). Real world evaluation: Working under budget, time, data, and political constraints. Thousand Oaks, CA: Sage.

Bamberger, M., Vaessen, J., & Raimondo, E. (2015). Dealing with complexity in development evaluation. Thousand Oaks, CA: Sage.

Beach, D. (2019, August). Developing better standards for evaluating how policy interventions work in practice – a new evidence hierarchy. Paper presented at the Annual Meeting of the American Political Science Association, Washington, DC.

Beach, D., & Pedersen, R. (2013). Process-tracing methods: Foundations and guidelines. Ann Arbor, MI: University of Michigan Press.

Beach, D., & Pedersen, R. (2016). Causal case study methods. Ann Arbor, MI: University of Michigan Press.

Befani, B. (2012). Models of causality and causal inference. In E. Stern, N. Stame, J. Mayne, K. Forss, R. Davies, & B. Befani, Broadening the range of design and methods for impact evaluation (DFID Working Paper no. 38). London: DFID. Retrieved from http://www.dfid.gov.uk/r4d/Output/189575/Default.aspx

Befani, B., & Stedman-Bryce G. (2017). Process tracing and Bayesian updating for impact evaluation. Evaluation, 23(1), 42–60.

Bennett, A. (2008). Process tracing: A Bayesian perspective. In J. Box-Steffensmeier,

H. Brady, & D. Collier (Eds.), The Oxford handbook of political methodology. Oxford: Oxford University Press.

Bennett, A. (2019). Process tracing for development evaluation. Training to Independent Evaluation Group delivered March 18, 2019.

Bennett, A., & Checkel, J. (2014). Process tracing: From metaphor to analytic tool. Cam- bridge: Cambridge University Press

Byrne, D. (2013). Evaluating complex social interventions in a complex world. Evaluation, 19, 217–228. https://doi.org/10.1177/1356389013495617

D’Errico, S., Befani, B., Booker, F., & Guiliani, A. (2017). Influencing policy change in Uganda: An impact evaluation of the Uganda Poverty and Conservation Learning Group’s

work. Retrieved from https://pubs.iied.org/pdfs/G04157.pdf

Forss, K., Marra, M., & Schwarts, R. (2011). Evaluating the complex: Attribution, contribution and beyond. New Brunswick, NJ: Transaction.

Fox, J. A. (2015). Social accountability: What does the evidence really say? World Development, 72, 346–361. https://doi.org/10.1016/j.worlddev.2015.03.011

Gaventa, J., & Barrett, G. (2010). So what difference does it make? Mapping the out- comes of citizen engagement. IDS Working Papers, 2010(347), 1–72. https://doi.org/ 10.1111/j.2040-0209.2010.00347_2.x

George, A., & Bennett, A. (2005). Case studies and theory development in the social sciences. Cambridge, MA: MIT Press.

Grandvoinnet, H., Aslam, G., & Raha, S. (2015). Opening the black box: The contextual drivers of social accountability. Washington, DC: World Bank.

Joshi, A. (2013). Do they Work? Assessing the impact of transparency and accountability initiatives in service delivery. Development Policy Review, 31(1), 29–48. https://doi. org/10.1111/dpr.12018

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. Philosophy of Science, 67(1), 1–25. https://doi.org/10.1086/392759

Mansuri, G., & Rao, V. (2013). Localizing development: Does participation work? Washington, DC: World Bank.

Punton, M., & Welle, K. (2015). Straws-in-the-wind, Hoops and Smoking Guns: What can Process Tracing Offer to Impact Evaluation?, CDI Prabrctice Paper 10, Brighton: IDS Ringold, P. L., Boyd, J., Landers, D., & Weber, M. (2013). What data should we collect? A framework for identifying indicators of ecosystem contributions to human well-being. Frontiers in Ecology and the Environment, 11(2), 98–105. https://doi.org/ 10.1890/110156

Schmitt, J. (2018). Testing causal mechanisms with Bayesian process-tracing: Strengthening explanatory power of case study evaluations, The Evaluator, 2018, 16–18.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). Broadening the range of designs and methods for impact evaluations. Brighton, UK: Institute for Development Studies. Retrieved from https://doi.org/10.22163/fteval.2012.100

World Bank (2018). Engaging citizens for better development results: An independent evaluation. Retrieved from https://ieg.worldbankgroup.org/sites/default/files/Data/ Evaluation/files/Engaging_Citizens_for_Better_Development_Results_FullReport.pdf