Testing causal mechanisms with Bayesian process-tracing: Strengthening explanatory power of case study evaluations

Dr Johannes Schmitt, Evaluator at the German Institute for Development Evaluation (DEval) outlines two significant developments which have improved the rigour and quality of process tracing in evaluation, including clear examples of how these have been used.

CAUSAL MECHANISMS take a key role in debates about evaluation methodology. Motivated to strengthen rigour in small-n evaluations, scholars and practitioners intensely debate how to open the black box of causation and increase internal validity. The core of working with causal mechanisms is to systematically analyse interrelations between an intervention and its expected outcomes.

Although they are also relevant in statistical designs like mediation analysis, mechanisms are prominently discussed to improve small-n research, particularly through the method of process tracing. In recent years, social science scholars have introduced two major innovations in process tracing: (i) an explicit theorisation of the causal mechanism before conducting empirical research and (ii) critical assessment of the probative value of evidence according to Bayesian logic of causal inference (Beach and Pedersen, 2013; Schmitt and Beach, 2015).

These innovations have started to push through to the evaluation arena but clear guidance is missing. At a conceptual level, scholars have transferred these innovations to the evaluation arena and its first applications have tapped the potential of the Bayesian inspired within-case method (e.g. Bamanyaki and Holvoet, 2016; Befani and Stedman-Bryce, 2016). However, practical guidance on how to apply Bayesian process tracing (BPT) in evaluation is limited. In particular, there is little information on how to assess the probative value and do the causal inference.

This article attempts to give a glimpse of how BPT can be applied in evaluation case studies based on experiences made in a recent DEval-evaluation. Building on earlier evidence about the effectiveness of the aid modality of budget support (Orth et al., 2017), the evaluation asks, among other questions, whether effects on reforms in public finance management (PFM) were sustained after the donors' exit from budget support and whether there is a causal link between ending budget support and the robust reform dynamics in PFM in the four case studies (Orth et al., 2018).

Explicit theorisation of the causal mechanism

The first exercise in PT is to develop the hypothetic causal mechanism based on prior knowledge. The mechanism takes the form of a causal process that consists of multiple, interlocking parts leading from the intervention (X) to the outcome (Y). For each part, entities (real actors) engage in activities (traceable action). For the mechanism to provide a solid basis for causal inference, it needs to be complete and explicit. The evaluation team needs to make sure that it contains all the parts necessary to provide a convincing and sufficient explanation for the occurrence of the observed outcome.

To operationalise this concept, we first defined X and Y, and then described how entities engage in activities for each part. Furthermore, we defined case specific scope conditions for the causal mechanism to function (see Figure 1).

FIGURE 1: HYPOTHETIC CAUSAL MECHANISM

CM Part	Entity	Activity
X		**Suspension of budget support (BS)**
Part 1.1	Donors	signal willingness to relaunch BS or similar financing modalities
Part 1.2	Donors	...conditional to relaunch of (IMF) structural reform program
Part 2	Gov't of Malawi	addresses corruption scandal issues through PFM Action plan
Part 3	IMF	provides suggestions on further PFM reforms and monitors performance of IMF structural benchmarks regarding PFM
Part 4.1	Gov't of Malawi	implements PFM Action Plan successfully and exerts higher budget discipline
Part 4.2	Gov't of Malawi	improves fiscal discipline
Part 5	Audit Office	receives more support and produces more output
Y		**The dynamic of PFM reform is robust**
Scope conditions		Malawi: Poor and aid dependent country, weak state capacities, history of embezzlement of funds

Source: Adjusted from Orth et al., 2018.

Before studying the hypothetical causal mechanism in the field, we further operationalised its individual parts. First, we determined the *prior*, i.e. the confidence in our hypothesis for each part to be true based on prior knowledge from earlier case studies and country reports. Second, the literature on process tracing recommends thinking of possible types of evidence for each part. Like detectives, we asked ourselves what we would expect to find if our theory were true. Where would we need to look for evidence? Whom should we talk to? What should we ask? This step was crucial to preparing a search strategy and questionnaires for qualitative interviews with different stakeholder groups.

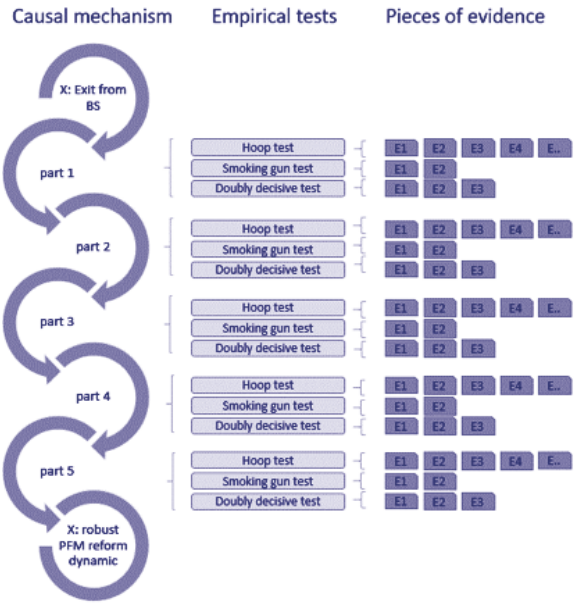
During one week of data collection in Malawi, a team of four researchers conducted 29 interviews. We first talked to country experts and independent consultants in order to readjust our hypothetic causal mechanism. After this first round, we interviewed representatives from civil society, Government and donor agencies.

Critical assessment of the probative value of evidence and Bayesian updating

In order to draw causal conclusions and infer into the existence of the causal mechanism, BPT requires that evaluators structure empirical evidence, assess its probative value and then conduct Bayesian updating.

Using qualitative data analysis software, we coded interview notes and reports and assigned the pieces of evidence (i.e. a coded interview statement or a section in a report) to one or more empirical tests for each part of the causal mechanism (see Figure 2).

FIGURE 2: LINKING EVIDENCE, TESTS AND THEORY



Source: Author's own

> Like detectives, we asked ourselves what we would expect to find if our theory were true.

We determined the probative value for each empirical test by quantifying its respective *theoretical certainty* and *uniqueness* (see figure 3). We assessed *certainty* by discussing: 'What is the probability to observe this piece of information if the hypothesis was true?' For theoretical *uniqueness* we thought of our evidence asking: 'What is the probability that, despite observing this evidence, the hypothesis is not true?' High scores in certainty mean that a test has a strong disconfirmatory power. If we find contradictory evidence for a test of high *certainty*, we need to downgrade our confidence in our theory. On the other hand, if a test is *unique* (low scores!), it has a high confirmatory power. Finding evidence for a unique test lets us increase our confidence in the theory.

Based on the principles of theoretical certainty and uniqueness, four test types can be distinguished. A `hoop test´ is what we would `expect-to-find´ if our theory was true. A `smoking gun test´, on the other side, is what we would `love-to-find´ as it is sufficient evidence to confirm our hypothesis. The `doubly-decisive test´ is strong to confirm but also to disconfirm parts of our hypothetic causal mechanism. `Straw-in-the-wind tests´ are neither necessary nor sufficient and therefore low in probative value (Van Evera, 1997; see also Beach and Pedersen, 2013; Punton and Welle, 2015).

FIGURE 3: ASSESSING THE PROBATIVE VALUE FOR EMPIRICAL TESTS

Mechanism part 1.1	Test	Prior p(h)	Certainty p(e|h)	Uniqueness p(e|~h)
Donors signal willingness to relaunch BS or similar financing modalities	Donors earmark funds for further disbursement (hoop test)	0.6	0.7	0.4
	Donors show willingness to relaunch payments if conditions are met (doubly-decisive test)		0.8	0.1

Source: Adjusted from Orth et al., 2018

Having identified the prior for each part and the probative value for each empirical test, we calculate posterior values using the Bayesian formula.

> Even more promising, also for other theory-based approaches, is the critical assessment of the probative value of evidence and Bayesian updating put forward by BPT.

THE BAYESIAN FORMULA

$$\text{posterior} = \frac{\text{prior}}{\text{prior} + \frac{\text{uniqueness*}}{\text{certainty}}\,(1\text{-prior})}$$

Source: Beach and Pedersen, 2013

First, we calculated posteriors for each empirical test. Second, we aggregated the posterior values of individual tests for each part of the causal mechanism. Third, we determined the direction and degree of updating for each part by comparing the posterior value to the prior. If the posterior was higher than the prior, we updated positively. If vice versa, we updated our confidence downwards. Finally, causal inference was done in two steps: A mechanism part was existent if the posterior value ranged above .5. We find that all five parts of the causal mechanisms were present in the case. In a second step, we inferred into the presence of the entire causal mechanism. Given that all parts were present in the case, we inferred that there is a causal connection between the donors' exit from budget support (X) and the robust PFM reform dynamic (Y) in Malawi.

Outlook

Against the backdrop of our own application, I conclude that BPT is a viable method to test the causal mechanism between intervention and outcome and thus increase rigor in qualitative case study evaluations. Explicit theorisation of the causal mechanism and thinking of different types of evidence helps to make data collection more efficient. Even more promising, also for other theory-based approaches, is the critical assessment of the probative value of evidence and Bayesian updating put forward by BPT. Reaping the full benefits of BPT is quite challenging. Given the method is still in its infancy, more applications and detailed guidance on how to go about the technical steps in BPT are needed.

This work is part of the methodological research project on 'The use of Causal Mechanisms in complex Evaluations' at the German Institute for Development Evaluation (DEval).

Johannes Schmitt
johannes.schmitt@deval.org

References

Bamanyaki, P.A. and N. Holvoet (2016), *Integrating Theory-Based Evaluation and Process Tracing in the Evaluation of Civil Society Gender Budget Initiatives,* Evaluation, Vol. 22, No. 1, pp. 72–90.

Beach, D. and R.B. Pedersen (2013), *Process-Tracing Methods: Foundations and Guidelines,* University of Michigan Press, Ann Arbor.

Befani, B. and G. Stedman-Bryce (2016), *Process Tracing and Bayesian Updating for Impact Evaluation,* Evaluation, Vol. 23, No. 1, pp. 1–19.

Orth, M. et al. (2018), *The future of integrated policy-based development cooperation,* German Institute for Development Evaluation (DEval), Bonn.

Orth, M. et al. (2017), *What We Know about the Effectiveness of Budget Support. Evaluation Synthesis*, German Institute for Development Evaluation (DEval), Bonn.

Punton, M. and K. Welle (2015), *Applying Process Tracing in Five Steps*, Practice Paper, No. 10 Annex, Centre for Development Impact (CDI), Institute of Development Studies, Brighton.

Schmitt, J. and D. Beach (2015), *The Contribution of Process Tracing to Theory-Based Evaluations of Complex Aid Instruments*, Evaluation, Vol. 21, No. 4, pp. 429–447.

Van Evera, S. (1997), *Guide to Methods for Students of Political Science*, Cornell University Press, Ithaca, NY.