CDI Centre for Development Impact
PRACTICE PAPER
Innovation and learning in impact evaluation

Process Tracing Innovations in Practice: Finding the Middle Path

Abstract Evaluation practitioners in the international development sector have given considerable attention in recent years to process tracing as a method for evaluating impact, including discussion of how to assess the relative importance of causal factors. Despite the increasing interest, there is a relative dearth of examples of practical learning and evidence of applying process tracing in practice. This CDI Practice Paper draws on comparative learning from applying three different types of process tracing in international development initiatives. It argues in favour of a 'middle path' of applying evidence tests and rubrics to structure evaluative judgements rather than formal Bayesian updating or looser forms of process tracing. It also calls attention to the potential added value of taking a participatory approach, offering practical recommendations for how to do this effectively.

1 Introduction
This paper builds on CDI Practice Paper 10 (Punton and Welle 2015) which discusses what process tracing can offer to impact evaluation of international development initiatives. It expands upon this by drawing on three evaluations through which we assessed initiatives that sought to influence decision makers to improve consultation and inclusion practices, and to reform public policy and resource allocation. We review the application of three types of process tracing: process tracing through causal process observations; process tracing with formal Bayesian updating (contribution tracing), and process tracing with evidence tests and rubrics. We first provide an overview of process tracing fundamentals and emerging variations. We then unpack the three evaluation cases to illustrate the relative strengths and weaknesses of these forms in practice. We focus on aspects of causal claims and chains, and evidence assessment, discussing how a participatory approach can add rigour and other benefits to this methodology. We conclude with a summary of our main reflections.

2 Process tracing fundamentals and innovations
Process tracing can be understood as a theory-based, case-study method for evaluating impact. It emerged from political science decades ago but has only been applied as an evaluation method in the last decade (Stedman-Bryce 2013; Punton and Welle 2015; Befani and Stedman-Bryce 2016; Befani et al. 2016). Its approach to causation is generative, meaning that the focus is on unpacking complex causal 'mechanisms', explaining how and why a cause (or set of causes) led to a particular outcome (Beach and Pedersen 2019). Some forms of process tracing also contain specific techniques and tools for data collection and analysis (Collier 2011; Bennett and Checkel 2014; Beach and Pedersen 2019). The starting point is identifying an observable outcome. A theory is then defined and broken down in a series of causal mechanisms, which together are deemed to be a sufficient causal package for achieving the outcome.

Beach and Pedersen (2019: 31) identify four types of process tracing: (1) as descriptive narratives of a sequence of events in between the cause and outcome; (2) as intervening variables (i.e. factors) between cause and outcome; (3) as causal process observations or diagnostic evidence which serve as empirical fingerprints; and (4) as systems of interlocking parts transmitting causal forces between cause(s) and outcome(s). This paper looks at the latter two.

Causal process observations, or what Beach and Pedersen (2019) call 'minimalist mechanisms' process tracing, go beyond a descriptive narrative of events in a temporal and logical sequence (Bennett and Checkel 2014), and specifically look for evidence where the causal process can be observed. Yet, this type of process tracing tends to be quite abstract, as the relationships between cause and outcome are not unpacked in detail. The systems form of process tracing is more fine-grained at a lower level of abstraction. In this form, mechanisms are considered to have several parts that are interlinked in the causal process, composed of entities (for example, people or organisations – *nouns*) that engage in activities (for example, arguing, supporting – *verbs*).

Particularly in the systems form of process tracing, causal mechanisms can consist of various complementary causal chains which may also contribute to an outcome. It can even include rival (or alternative) causal chains which could help to potentially rule out an intervention's contributory role to an outcome. To verify causal mechanisms, it is necessary to consider these other contributing factors and configurations (Hay 2016).

Process tracing assesses the strength of one's proof for the causal relationship through four evidence tests. These tests assess the 'probative value' of evidence (i.e. strength of proof), either helping to confirm or disconfirm a contribution claim. The more unique and certain the evidence is, the stronger its probative value. The two most useful tests are 'hoop' tests and 'smoking gun' tests. These either help to rule out one's explanations (hoop tests) or to confirm the hypothesised causal relationship (smoking gun). Evidence tests allow evaluators to assess how well causal claims can stand up to scrutiny, to show convincing evidence for a specific explanation, and to reject rival explanations. These tests can be applied at each step within a causal mechanism. The more specific the causal mechanism, the more precise the tests can be. Defining these tests – that is, skilfully identifying what evidence can help to confirm or disconfirm a claim – requires strong contextual knowledge (Punton and Welle 2015).

Various scholars have argued that process tracing is underpinned by Bayesian reasoning (Mahoney 2016; Fairfield and Charman 2017). Bayesian reasoning is a means of updating one's views about which hypothesis best explains the outcomes of interest as additional information is gained. Inferences that are made relate to the level of confidence in the validity of a theory and causal claim. Evaluators first establish a 'prior' confidence in the claim, and then update it during data collection. Different items of evidence are graded based on their supposed inferential power or 'probative value'.

There has been increased interest in recent years to move to the formal application of Bayes' theorem to update probabilities in a numerical form. Various researchers oppose this quantification and plea for a more informal application of the logic. They argue that the main aim of process tracing is to compare rival hypotheses; thus, they see the formal classification of evidence as either unnecessary or inappropriate (Bennett 2008, 2014; Fairfield and Charman 2017; Beach and Pedersen 2019; Bennett, Charman and Fairfield 2022; Zaks 2021). Some evaluators have nonetheless seen value in both applying evidence tests directly and using Bayesian logic through the formal quantification of confidence levels for an intervention's contribution to an outcome (Befani and Stedman-Bryce 2016; Befani *et al.* 2016; Mayne 2019). In this paper, we discuss the relative merits of both the formal and informal applications and argue for a 'middle path' between these.

3 Key learning from process tracing innovations: Three evaluation cases

The following three evaluations represent different variations of process tracing: (i) process tracing with formal Bayesian updating (contribution tracing); (ii) process tracing through causal process observations; and (iii) what we call 'contribution rubrics'. These three cases and their main features are presented below.

3.1 Causal claims and chains
Making a clear contribution claim and developing explicit causal chains based on a detailed theory of change is essential to good process tracing (see Bennett and Checkel 2014; Punton and Welle 2015; Befani and Stedman-Bryce 2016). However, as we will illustrate, different forms of process tracing can take different approaches to articulating contribution claims, theories of change, and specific causal chains. We now present the explicit contribution claims and causal chains developed for each case and explain the rationale.

Table 1 **Selected evaluation cases**

Initiative/Agency	Sectors	Overview	Process tracing features
G7/G20/COP Advocacy (Oxfam America (OUS))	Climate change policy and advocacy	Advocacy efforts on international climate policies, mainly the Paris Agreement and US backsliding under the Trump administration (2016–18)	▪ Minimalist mechanisms via explicit causal chains ▪ Causal process observations ▪ Evidence tests ▪ Summative external evaluation
Journey for Advancement in Transparency Responsiveness and Accountability – JATRA (CARE International)	Infrastructure and social accountability	Promotion of social accountability to improve the responsiveness of public financial management and participatory planning at the lowest tiers of local government in Bangladesh (2014–17)	▪ Systems process tracing ▪ Explicit causal chains ▪ Evidence tests ▪ Formal Bayesian updating ▪ Formative, partner-led evaluation
G7CSO Policy advocacy (Coalition of World Vision Canada, Plan International, Save the Children Canada, Right to Play, Results Canada, and UNICEF)	Girls' education policy and advocacy	Influencing of a strong policy and funding commitment for girls' education in crisis settings via the 2018 G7	▪ Systems process tracing ▪ Explicit causal chains ▪ Evidence tests ▪ Explicit testing of rival claims ▪ Contribution rubrics ▪ Summative, partner-led evaluation

1 Oxfam causal chains development

The OUS evaluation employed process tracing to explore a case study and specific outcome within the organisation's overall climate change and energy advocacy portfolio. The contribution claim evaluated was:

> Linked to its US influencing strategy, in the lead-up to the G7 and G20 summits, OUS played a leadership role, both within Oxfam International and in broader civil society networks, on the strategy and actions that successfully influenced governments to uphold their commitment to the Paris Agreement and associated actions, in the face of US backsliding.
> (OUS evaluation report)

The claim directly reflects the logic and strategic pathways represented in the theory of change, developed collectively as part of the evaluation process.

The evaluation team identified the following five causal chains, considered as fundamental to explaining the OUS contribution to the outcome:

1 OUS played a leading role in relevant NGO networks and advocacy bodies to develop common strategies to respond to and counter US withdrawal from the Paris Agreement.

2 OUS succeeded in making sure that the defence of the Paris Agreement was prioritised within Oxfam International Confederation (OI) and shaped OI's response.

3 OUS mobilised insiders (policymakers) to influence the Trump administration's decision-making process regarding the potential withdrawal from the Paris Agreement.

4 OI (and/or OI-influenced) key messages reached G7 and G20 leaders, particularly those from France, Italy, Germany, and the UK.

5 OUS, OI and/or OI-influenced public outreach efforts helped to create public pressure, also through media visibility, on G7 and G20 leaders to counter US backsliding.

These were conceived as 'minimalist mechanisms'. Two similar and competing rival claims were identified. Rather than being specific to one of the causal chains, they were relevant to the observed outcome overall. In sum, the first rival claim was that the relevant messaging was largely independent from OUS and the second that other civil society actor(s) were the leading entity driving unification around the Paris Agreement defence. They could not be evidenced or tested directly due to available time and resources but were assessed using expert judgements based on the available information. This highlights a common practical limitation we have found for resourcing of process tracing evaluations, often requiring compromise based on available resources and the level of effort required to assess rival claims.

2 JATRA causal chain development

As a formative evaluation, the JATRA team was given complete freedom to choose the contribution claims they wanted to evaluate. The project did not have a theory of change before the evaluation. Partly as a result, there were long discussions on how ambitious the claims should be before agreeing on the following claim: 'JATRA's facilitation of poor citizens' engagement has led to greater budget allocation of their demands in Union Parishad annual budgets (the lowest tier of government administration)' (JATRA evaluation report). The team developed six causal chains which had a total of 23 components (i.e. steps). The team did not test the rival explanations, largely because confirmatory evidence for the 23 components was already a lot to assess. The dominant causal chain had 12 steps. Below are two of the most important ones to illustrate the level of detail:

■ **Part A:** Citizen Forum members co-facilitate Ward Shava pre- and post-budget meetings and increase the number of issues raised by participants from poor *Para* (hamlets).

■ **Part B:** With the assistance of citizen forum members, Union Parishad organises an open budget meeting to raise awareness on resource commitments, declares the draft budget and opens for public discussion.

Developing the causal chain revealed that several steps fed into the Ward Shava meeting: formal Ward Shava planning meetings; courtyard meetings, and folk song events which were designed to sensitise and mobilise poor *Para* to attend and present their demands at Ward Shava meetings. These efforts contributed to a higher number and proportion of poor peoples' demands being prioritised in Ward Shava open budget meetings and the final Union Parishad budget declaration, from which connections to poorer parts of communities could be traced.

3 G7CSO Coalition causal chain development

The contribution claim in this third example was:

> The G7CSO Coalition's policy influencing efforts through coalition building, direct government engagement and youth and public advocacy, secured the Government of Canada's commitment to girls' education in crisis contexts, as identified through the G7 Charlevoix Declaration on Quality Education and Canada's financial pledge of CAD$400 million.
> (G7CSO Coalition evaluation report)

G7CSO Coalition members were very confident in a high level of contribution to this observed outcome. As various stakeholders brought their own values, goals, and biases into the process, the evaluation team had to negotiate to ensure that the contribution claim was feasible to assess. The Coalition agreed to review one complementary and two rival claims to enhance credibility. As Aston *et al.* (2021) note, these can be thought of as logical (rather than experimental) counterfactuals which fit within a generative approach. We identified three causal chains, of which one was evaluated, called 'insider advocacy'. The final chain included five components – here is a summarised version of the main three components:

■ **Part A:** Upon the request of Canada's Sous-Sherpa via Global Affairs Canada, the G7CSO Coalition developed a draft policy implementation plan outlining its policy and funding aims for girls' education in crisis (GEiC).

■ **Part B:** Based on this plan, Canada's G7 Sherpa advocates to the Prime Minister (PM) for a Canadian commitment to a GEiC declaration and the PM provides approval, agreeing to work on other G7 countries for financial contributions.

■ **Part C:** The G7CSO Coalition communicates directly with key ministerial and PM Office staffers and delivers the final policy content and financial ask to Canada's Sherpa, prior to the Charlevoix Declaration on 9 June 2018.

Generally, we found that teams often want to showcase more than one outcome, and the selection of the focus of the process tracing exercise comes with political dynamics related to this desire. So, it is important to manage expectations early on, and to agree on a clear outcome and plausible claim, considering available resources. Developing more specific theories of change and outcome statements can help teams to focus, make claims more testable, and identify other contributory factors.

3.2 Evidence assessment

Evidence assessment is a critical aspect of different process tracing approaches. As discussed, some question whether to explicitly test rival hypotheses, to conduct process tracing tests, or to take a formal Bayesian approach to quantification of confidence (Befani and Stedman-Bryce 2016; Fairfield and Charman 2017; Bennett *et al.* 2022). Our cases shed some light on the relative merits and demerits of these options.

In each case, we took a different approach to evidence assessment. Starting with OUS, we identified and conducted brief evidence tests for each of the five identified chains overall, but not for each component in their chains nor the rival claims. Table 2 provides examples of how evidence tests were employed at the causal chain level.

Table 2 Examples of tests and evidence assessment for two Oxfam causal chains

Causal chain	Expected evidence	Test	Collected evidence	Strength of analysed evidence and test outcome
OUS played a leading role in relevant NGO networks and advocacy bodies to develop common strategies to respond to and counter US withdrawal from the Paris Agreement.	OUS staff held leadership positions within relevant networks.	Hoop	Meeting notes from relevant NGO coalitions.

Interviews with relevant OUS staff and NGO coalitions' representatives.

Both are relevant for the two tests. | Strong and specific enough to be sufficient; but it was weaker in relation to OUS' research; the value of its contribution was more focused on its savvy ability to gather intelligence and navigate the political environment.

Both tests passed. |
| | OUS knowledge products shaped common strategies, positions, and/or messages. | Smoking gun | | |
| OUS mobilised insiders (policymakers) to influence the Trump administration's decision-making process regarding the potential withdrawal from the Paris Agreement. | OUS and OI had unique access to some of the most relevant inside influencers in the lead-up to main relevant events (G7, G20, COP 23). | Hoop | Meeting notes with insiders.

Interviews with key insiders (policymakers) and bellwethers. | Weak. Access was confirmed but the evidence was not specific enough to differentiate between OUS and OI entities, making it difficult to verify the contribution of OUS for this chain.

Test failed. |
| | NGO leaders credit OUS and/or OI for reaching insiders in the lead-up to key events (G7, G20, COP 23). | Smoking gun | Interviews with NGO coalitions' representatives/ bellwethers. | Weak. The test sheds doubt on whether the main insiders targeted by OUS had significant influence on the decision to withdraw from the Paris Agreement. The evaluators did not find any evidence of OUS reaching those officials with higher influence.

Test failed. |

Despite the limitations discussed, the approach taken was sufficient for the purposes of this evaluation. Looser forms of process tracing, in which each component in a chain and rival claims are not tested, can still provide practical value, but come with trade-offs. For example, testing rival claims can help increase the confidence in one's contribution claim while the granular details provided by testing each step in a chain increases the clarity on how an outcome happened and the specific activities and entities directly involved at each step.

For JATRA, a much more intensive approach was taken to evidence tests. All 23 components and 80 items of evidence were assessed. This process took several months because the project team had to be trained and coached to conduct formal Bayesian process tracing. Staff were supported to: identify potentially relevant evidence for each step in the causal chains; assess the probability of finding evidence *ex ante* if their contribution claim were true or false; and provide a numerical judgement which represented that probability (Befani and Stedman-Bryce 2016). The team used a spreadsheet which automatically calculated the Bayes formula when probability estimates were inputted. After collecting the data, the team assessed their posterior confidence in the contribution claim after observing evidence. This provided a final number to classify the confidence level. There was a trade-off between the ambition/boldness of contribution claims and the confidence in them. In practice, this meant that the claim had around 70 per cent confidence with an ambitious claim, rather than 90 per cent confidence with a less ambitious claim.

The JATRA team initially struggled with the formal Bayesian updating process. It was onerous and time-consuming to provide a probability assessment for each potential item of evidence. Yet, explicitly appraising the probative value of evidence meant that the team could determine which evidence was worth collecting and which was not. Perhaps because of these sunk costs, the team became strong defenders of formal Bayesian updating, hailing its merits in helping them to defend their evidence claims to donors. We also saw the downside. Once users understand how the updating is calculated, it is possible to game the system to get closer to a favoured (higher or lower) number. These instruments are only as reliable as the judgements of users themselves. So, precise quantification seems to do little to increase accuracy or our level of confidence in contribution claims.

The G7CSO Coalition took a 'middle path' to balancing time, cost, and value added of available process tracing tools. They selected one causal chain for the contribution claim (insider advocacy) and three alternative claims. The evaluation team held a two-day workshop with selected Coalition members who had the most contextual and historical knowledge about the events to review the theory of change and develop causal chains, starting with a timeline of key events. This helped them to develop the final contribution claim and key components of the selected chain. It resulted in five causal steps. The team was also introduced to evidence tests, identifying 'need to find' (necessary) and 'love to find' (unique) items of evidence for all steps. This was sufficient for them to grasp the logic of the tests without using too much of their time, after which the evaluators applied the tests once data was collected for each component. The same was done for the alternative explanations. The team used rubrics against five colour-coded standards (no confidence, low, ambiguous, medium, and high) with descriptions for each of these levels. These rubrics reflected the logic of process tracing evidence tests, assessing the level of confidence in each component of the causal chain and the whole contribution claim overall.

We found that developing rubrics for evidence tests made the process more rigorous. Some tests are more precise than others. Both over and under-precision can compromise the credibility of evaluative judgements. Establishing clear criteria, levels, and descriptions assessing confidence at each step helped to clarify the probative value of evidence. Although the JATRA team saw benefits in formal Bayesian updating, using rubrics offered a less stringent but equally credible framework for structuring evaluative judgements derived from evidence tests.

It was worth the effort to test alternative claims explicitly, as we did in the G7CSO Coalition case, with three potential alternative claims (one complementary and two rival). Demonstrating careful consideration of other influences increases the credibility of findings as it helps to eliminate other potential causal explanations for an outcome. However, it is more challenging to collect evidence for rival claims than it is for an initiative's own claim because this often requires access to external actors' data. Additionally, interviewing potentially important key informants for rival claims may not always be possible. Therefore, rival claim analysis relies more on secondary data which is publicly available, potentially leading to greater confirmation and self-serving bias. Finding credible and knowledgeable key informants is often crucial for assessing rival claims. In the G7CSO case, we were fortunate to interview several key informants who did not have clear incentives to support the Coalition's claim as they were from 'rival' CSO Coalitions that advocated for different G7 2018 government commitments. So, obtaining their validating testimony about the Coalition's influence on the observed outcome was high-quality evidence that helped to rule out rival claims.

Assessing the probative value of evidence for overall contribution, causal chain components, and alternative claims is highly context-specific. For example, in the G7CSO evaluation, email evidence from a high-level policy official was deemed to be doubly decisive due to the specific nuances in the language used and knowledge of the CSO relationships with key government stakeholders in the Canadian policy landscape. Overall, we found that emails can be surprisingly powerful sources of evidence. Akin to interviews, their language and tone provide important insights into the nature and quality of stakeholder relationships. They also provide a time-stamped interaction that typically identifies key actors and actions, and they are difficult to fake. A series of email threads between different stakeholders provides a reliable paper trail in the chain of events within a decision-making process. However, the assessment of these contextual details will almost always come from initiative stakeholders and depend on how well they can explain the evidence in context.

Overall, our three cases demonstrated the relative added value of making theories of change more testable, conducting process tracing tests, assessing rival claims, and developing rubrics for assessing levels of confidence in claims. We found that formal Bayesian updating was unnecessarily onerous and did not offer any clear value added over more flexible rubrics.

3.3 Participatory approaches
Participatory approaches are often considered as critically flawed because they can generate small-n biases (White and Phillips 2012). However, our experience demonstrates that participatory approaches also have the potential to **enhance** rigour (Chambers 2015; Aston *et al.* 2021). We found that ensuring multiple and diverse stakeholder perspectives (e.g., staff and organisations) can help improve

the clarity and credibility of claims and how well these can be tested. This section discusses the merits as well as the trade-offs that evaluators may face in practice.

In the G7CSO Coalition case, it was helpful to adopt a flexible approach to maximise the potential benefits of stakeholder participation, while remaining aware of the different voices inside and outside the room and of how evaluations can affect stakeholder interests and reputations. The inclusion of additional perspectives is often worthwhile but requires extra time and further negotiation of contribution claims.

With JATRA, the project team was much more engaged throughout the whole evaluation, while for OUS, staff had much less involvement. From these three approaches, we found that the most appropriate and useful moments to involve teams were in developing theories of change, contribution claims, and testable causal chains, and for the identification of evidence and appraising its potential probative value. As in the G7CSO Coalition, it proved beneficial to engage project stakeholders at different points during a process tracing evaluation while still limiting their involvement to a degree that balances potential bias, time, and resources. Furthermore, discussions with project teams about which causal steps were required for the claim to be valid forced the team to recognise the complexity of their own programming and how the efforts of other actors must be included as part of their explanations.

Like JATRA, soliciting diverse perspectives from partners in a participatory manner allowed the G7CSO Coalition evaluators to prioritise time and resources. The Coalition agreed that only one of the three causal chains identified (insider advocacy) was worth focusing on, given available resources, and because they believed it was the most influential chain for their contribution. In both evaluations, we found a need to coach project stakeholders to ensure that claims are sufficiently clear and testable, also advising them on the trade-offs of choosing more and less ambitious ones. The risk of making theories of change, claims, and causal chains too ambitious is that there is a high risk that the necessary evidence (i.e. hoop test evidence) will not be found, and this would potentially invalidate the contribution claim. Process tracing evaluations with testable claims may also find that the intervention was only a small contributory (but not necessary) factor towards an outcome. So, project teams should be made aware of both the benefits of finding confirmatory evidence and the risks of potential disconfirmation. Risk aversion therefore comes into play.

Interestingly, project managers, field staff, and monitoring and evaluation staff had quite different views regarding what was causally relevant to a particular claim. Through being prompted to justify explicitly how and why particular strategies were directly relevant for the outcome of interest,

the project teams' evaluative thinking increased significantly. However, the capacity to prompt this kind of thinking also depends upon the role and influence of evaluators, and how much they can persuade teams to be self-critical about their interventions and contribution scope.

Process tracing has various idiosyncratic metaphors and complicated technical language, but it is possible to navigate this challenge. As found in the G7CSO Coalition evaluation, it is easier to ask:

- ■ 'What evidence do we need to see if X is true?' This represents our minimum expectations to pass the hoop test.

- ■ 'What evidence would we like to see if X is true?' This represents the evidence that will convince us the most and help to cast doubt on (or rule out) alternatives to pass the smoking gun test.

Whether project stakeholders should be involved in primary data collection is a more contentious question. If they take part in primary data collection, providing guidance on interviewing techniques is critical to reducing selection, courtesy, and confirmation biases. If guided appropriately, they can be a good choice for data collection; in both JATRA and G7CSO evaluations, many key items of evidence were already available from project teams themselves. In the OUS evaluation, the project team played a much smaller role in evidence identification and data collection, limited to sending requested documents and linking the evaluators to key informants. We think that this enabled a more efficient process with reduced bias risks, but it also potentially resulted in a less in-depth exercise.

In sum, we found that participatory processes can help to elevate evaluative thinking. Soliciting multiple perspectives can improve triangulation, which in turn, can counter various small-n biases and actually enhance rigour. However, this approach needs to be carefully managed, especially if project teams are engaged in primary data collection with project stakeholders.

4 Conclusions

This paper focused attention on three critical areas of debate in recent process tracing innovations: causal chain development, evidence assessment, and stakeholder participation. It also provided arguments for taking a middle path among the options for process tracing evaluation, using three different evaluation cases for illustration.

Firstly, we found that well-developed and testable theories of change are a prerequisite for effective process tracing. If these are not in place, it is worth allocating the time up front to develop them because then causal chain development is much more efficient. Contribution claim-making is a deeply political process with clear stakes

for organisational reputations. So, evaluators must be sensitive to the potential trade-offs between the level of ambition in claims and levels of confidence, and negotiate this tension with project teams. Causal chains should be developed in as much detail as possible, but evaluators also need to be cognisant of feasibility limitations in evaluating many claims with several steps on a modest evaluation budget. It is possible that a 'causal hotspot' approach to contribution analysis offers a means to help decide which parts of a theory of change and which potential claims may be worth testing (Apgar and Ton 2021).

In considering different approaches to evidence assessment, contrary to Fairfield and Charman's appraisal (2017), we found that formal process tracing tests are worth the effort because they enhance evaluative thinking and prompt deeper critical analysis on the probative value of evidence than can be achieved without them. Probative value is context-specific, and thus stakeholder perspectives and positionality have a considerable bearing on evidence value. We judged that assessing rival claims explicitly is useful because this helps to explain contextual influence, and this enhances credibility of findings. We also found that while formal Bayesian updating can enhance the evaluative thinking of evaluation teams regarding the

probative value of evidence, the process is unnecessarily onerous and, in our view, precise probability calculations can be potentially misleading. So, they add relatively limited value. Instead, we found that rubrics provide a more flexible and suitable means to approach evaluative judgements regarding the quality of evidence.

Perhaps most importantly, we found that process tracing lends itself to stakeholder participation more than is commonly recognised. While employing a participatory approach requires more time and effort, it also adds considerable value. Developing plausible causal chains requires strong contextual and tacit knowledge. Therefore, supporting project stakeholders to develop contribution claims, to create causal chains, and to assess the probative value of evidence can enhance rigour; and it ensures that they are engaged at moments where their diverse perspectives are the most useful. Participation can also enhance evaluative reasoning, ownership, and potential utilisation of evaluation findings. However, especially if project stakeholders play a role in primary data collection, it is important to provide capacity building and coaching on interviewing to diminish potential biases. Participation is not a panacea, but it can play a valuable role in process tracing evaluation.

References

Apgar, M. and Ton, G. (2021) 'Learning Through and About Contribution Analysis for Impact Evaluation', *IDS Opinion*, 10 September (accessed 4 May 2022)

Aston, T.; Roche, C.; Schaaf, M. and Cant, S. (2021) 'Monitoring and Evaluation for Thinking and Working Politically', *Evaluation* 28.1: 36–57, DOI: 10.1177/13563890211053028

Beach, D. and Pedersen, R.B. (2019) *Process-Tracing Methods: Foundations and Guidelines*, Ann Arbor MI: University of Michigan Press

Befani, B. and Stedman-Bryce, G. (2016) 'Process Tracing and Bayesian Updating for Impact Evaluation', *Evaluation* 23.1: 42–60, DOI: 10.1177/1356389016654584

Befani, B.; D'Errico, S.; Booker, F. and Giuliani, A. (2016) *Clearing the Fog: New Tools for Improving the Credibility of Impact Claims*, IIED Briefing, London: International Institute for Environment and Development

Bennett, A. (2014) 'Disciplining Our Conjectures: Systematizing Process Tracing with Bayesian Analysis', in A. Bennett and J.T. Checkel (eds), *Process Tracing: From Metaphor to Analytic Tool*, Cambridge: Cambridge University Press

Bennett, A. (2008) 'Process Tracing: a Bayesian Perspective', in J.M. Box-Steffensmeier, H.E. Brady, and D. Collier (eds), *The Oxford Handbook of Political Methodology*, Oxford: Oxford Academic

Bennett, A. and Checkel, J.T. (2014) 'Introduction: Process tracing: From philosophical roots to best practices', in A. Bennett and J.T. Checkel (eds), *Process Tracing: From Metaphor to Analytic Tool*, Cambridge: Cambridge University Press

Bennett, A.; Charman, A. and Fairfield, T. (2022) 'Understanding Bayesianism: Fundamentals for Process Tracers', *Political Analysis* 30.2: 298–305, DOI: 10.1017/pan.2021.23

Chambers, R. (2015) 'Inclusive Rigour for Complexity', *Journal of Development Effectiveness* 7.3: 327–35, DOI: 10.1080/19439342.2015.1068356

Collier, D. (2011) 'Understanding Process Tracing', *PS: Political Science & Politics* 44.4: 823–30, DOI: 10.1017/S1049096511001429

Fairfield, T. and Charman, A.E. (2017) 'Explicit Bayesian Analysis for Process Tracing: Guidelines, Opportunities, and Caveats', *Political Analysis* 25.3: 363–80, DOI: 10.1017/pan.2017.14

Hay, C. (2016) Process Tracing: A Laudable Aim or a High-tariff Methodology?, *New Political Economy* 21.5: 500–4, DOI: 10.1080/13563467.2016.1201806

Mahoney, J. (2016) 'Mechanisms, Bayesianism, and Process Tracing', *New Political Economy* 21.5: 493–99, DOI: 10.1080/13563467.2016.1201803

Mayne, J. (2019) *Assessing the Relative Importance of Causal Factors*, CDI Practice Paper 21, Brighton: Institute of Development Studies

Punton, M. and Welle, K. (2015) *Straws-in-the-wind, Hoops and Smoking Guns: What can Process Tracing Offer to Impact Evaluation?*, CDI Practice Paper 10, Brighton: Institute of Development Studies

Stedman-Bryce, G. (2013) *Health for All: Towards Free Universal Health Care in Ghana*, End of Campaign Evaluation Report, Oxford: Oxfam GB

Wadeson, A.; Monzani, B. and Aston, T. (2020) *Process Tracing as a Practical Evaluation Method: Comparative Learning from Six Evaluations, Monitoring and Evaluation News* (accessed 8 February 2023)

White, H. and Phillips, D. (2012) *Addressing Attribution of Cause and Effect in Small N Impact Evaluations: Towards an Integrated Framework*, 3ie Working Paper 15, New Delhi: International Initiative for Impact Evaluation

Zaks, S. (2021) 'Updating Bayesian(s): A Critical Evaluation of Bayesian Process Tracing', *Political Analysis* 29.1: 58–74, DOI: 10.1017/pan.2020.10

> " Well-developed and testable theories of change are a prerequisite for effective process tracing. "

Centre for Development Impact (CDI)

The Centre is a collaboration between **IDS**, **Itad** and the **University of East Anglia**.

The Centre aims to contribute to innovation and excellence in the areas of impact assessment, evaluation and learning in development. The Centre's work is presently focused on:

(1) Exploring a broader range of evaluation designs and methods, and approaches to causal inference.
(2) Designing appropriate ways to assess the impact of complex interventions in challenging contexts.
(3) Better understanding the political dynamics and other factors in the evaluation process, including the use of evaluation evidence.

This CDI Practice Paper was written by **Tom Aston** and **Alix Wadeson**.

The opinions expressed are those of the authors and do not necessarily reflect the views of IDS or any of the institutions involved.