

---

## 16. Exploring causal pathways amid complexity

*Jewlya Lynn and Marina Apgar*

---

Historically, impact evaluation has focused on measuring the effects of interventions, using experimental and quasi-experimental designs to measure health outcomes, educational outcomes, outcomes of social service interventions, and even environmental outcomes, such as increased air quality and improved river health – outcomes and net effects that can often be measured through careful selection of a few quantitative indicators. In these contexts, impact evaluations tend to address the question “what works” (an approach advanced by the “What Works” movement in the UK and US; see White, 2019). However, these evaluations are less likely to answer questions about how an intervention worked, why it worked, and for whom.

This type of net effect measurement is frequently not enough. Evaluators are *increasingly being asked to make sense of what works in interventions being implemented in complex settings*. Additionally, as pressure has increased to answer other questions like “why,” “how,” and “for whom” (see Stern et al., 2012, for a well-known call to action on this issue), evaluation methods like randomized controlled trials (RCTs) have shown they are not as relevant or effective. No longer are our impact evaluations primarily focusing on individual-level outcomes and interventions with, in many cases, controllable variables and repeatable outcomes. Instead, interventions are now often seeking to address the complex causes of problems that manifest across systems (e.g., by tackling root causes like poverty, seeking to change narratives underlying many different structural issues, and advocating for policy changes).

The evaluation response has differed by issue area, country, and funding source. In the United States’ philanthropic context, *impact evaluation with causal thinking but without causal analysis became a default*, rather than replacing RCTs with more appropriate causal methods. These descriptive “impact” evaluations are grounded in causal thinking (e.g., predicting causality via a unidirectional and simplified theory of change), and often include findings that imply causality by measuring outcomes of programs and strategies that were included in the original predictive causal thinking. However, the analytical processes used do not generate causal inferences; they do not answer the question of “how” the change occurred. Rather, they assume causality, based on the original theories of change or similar tools developed to make program assumptions visible (which often do not reflect on existing evidence), or even on unarticulated beliefs. This absence of causal inferences is not unique to evaluations that engage in predictive thinking about causality; assumptions about causal links are also sometimes made in evaluations that discover emergent changes, yet do not interrogate deeply why the change happened.

In part, the use of impact evaluations without causal analysis may have been embraced due to the ways in which more descriptive evaluations can support the initial assumptions held by funders and/or implementers (known as confirmation bias). Descriptive impact evaluations may also be embraced because assumptions about causal relationships (sometimes articulated through theories of change and other similar tools) are accepted as sufficient to make causal inferences when outputs, outcomes, and context are observed and measured. What descriptive

impact evaluations fail to do is collect data on and interrogate the causal links within the theory of change (Lynn et al., 2022b).

In the international development context, the rejection by some evaluators of RCTs led to a different outcome: a variety of methods has flourished, and new ones have been developed to be implemented in more complex settings, even as other evaluators and commissioners have continued to try to make RCTs and quasi-experimental designs work. For example, Ray Pawson and other Realist evaluators have called for more attention to be given to the ways that causal mechanisms work in different contexts for different intended beneficiaries (Pawson & Tilley, 1997).

Depending on one's disciplinary training and sectoral evaluation experience, engagement with multiple forms of causality and different ways of developing causal explanations may feel more or less comfortable. We agree with Gates and Dyson (2017), who have made the case that evaluators and commissioners need to become more literate about multiple ways of understanding causality to feel confident in their work developing causal explanations.

Knowledge and evidence hierarchies that come from medical science and put experimental designs at the top (dismissing other methods as not valid) remain influential in methodological debates, but fail to engage with the more important question of how causal relationships are understood through them. They have fueled the momentum behind the “four waves of the evidence revolution” (White, 2019), which has led to institutionalizing the “What Works” movement in the UK and US, counter to the parallel trend of broadening evaluation designs and choice of causal methodologies.

In this chapter, we are not arguing for or against any particular approach or method, but we do suggest that a deeper exploration of causality is necessary to evaluate rigorously amid complexity. We will show through cases that mixing whole methods (Ton, 2012). and combining parts of high-quality methods (known as bricolage; Aston & Apgar, 2022), can reveal causal pathways in useful ways.

This chapter *assumes there is a need to continue to explore and interrogate causal relationships*, even amid complexity. In an integrated fashion throughout this chapter, we will:

- Briefly explore the concept of causality in the context of evaluation;
- Describe the steps and methods that can be helpful as an evaluator designs a study intended to explore causality amid complexity;
- Lightly introduce methods, provide resources to go deeper, and explore how to mitigate threats to the quality of implementation of an evaluation and an evaluator's ability to make causal claims; and
- Share examples of evaluations to bring these methods to life.

Our work is grounded in a value we, the authors, hold central: *evaluation should not be a tool of the funder or evaluator, but rather a partnership between those most affected by the intervention and those seeking to intervene.*<sup>1</sup> For this reason, we will center participatory approaches and reflect on power dynamics throughout the chapter.

## BROADENING AND DEEPENING HOW WE UNDERSTAND CAUSALITY

We approach the exploration of causality in the context of evaluations that need not only to describe changes that occur, but also to examine how the change happens, inclusive of intentional interventions and contextual factors. Box 16.1 provides our definitions of two terms central to the contexts in which such interventions and their evaluations operate: “systems change” and “complexity.”

### BOX 16.1 WHAT WE MEAN BY SYSTEMS CHANGE AND COMPLEXITY

Throughout this chapter, we use two key terms to describe the social change work that can be evaluated using causal methodologies:

- *Systems change*: Strategies that focus on shifting and tapping into visible dynamics (e.g., power, structures, resources, information) and hidden dynamics (e.g., norms, narratives, mental models) in systems to improve outcomes.
- *Complexity*: The inherent non-linearity and unpredictability evident in how change happens, particularly in systems change work. A complex system can be understood as an environment always in flux, with many different sources of influence, and unpredictable and often unexpected pathways to change.

When seeking to change systems, including complex adaptive systems (see Chapter 18 of this *Handbook*), it is natural to make assumptions about causality – that is, assumptions that help us explain why something we observe is happening. We make assumptions about how and where to act, and how to make sense of our environments. Instead of making implicit assumptions, we should first make them explicit, to then explore, investigate, and make sense of how change is actually happening.

For clarity, we are not talking about finding root causes, or the upstream drivers of social and environmental problems (e.g., poverty as a root cause of children ending up in harmful work). Rather, our focus is on examining cause-and-effect relationships between intentional interventions and the outcomes to which they contribute. Understanding causal relationships is fundamental to being able to make a causal claim in an evaluation that can support actionable recommendations.

We are also not talking about simply using causal *thinking* (often codified through a predictive theory of change or other prospective thinking; see Chapters 14 and 33), but rather about the interrogation of how change actually occurred – the causal pathway. We work to discover how our programs and strategies produce the outcomes we observe, rather than seeking evidence only of the strategies and outcomes (desired or otherwise).

We do not have space here to fully describe the extensive and overlapping debates about causal inference within both research and evaluation communities. Their historical longevity and vibrancy are evidence of causation’s being a highly contested topic across different disciplinary and epistemological domains. Disagreements are found both in terms of the main features of causal relations that exist in the world, as well as the methods through

Table 16.1 Different approaches to causality in evaluation

	Counterfactual	Regularity	Configurational	Generative
Basis for making causal inferences	Difference between two otherwise identical cases	Frequency of association between cause and effect	Combinations of conditions together causing an effect	Identifying the “mechanisms” that explain effects
Related evaluation approaches	Experimental and quasi-experimental	Statistical	Comparative case-based	“Theory-based” and Realist

Sources: Gates & Dyson (2017); Jenal & Liesner (2017); Stern et al. (2012, pp. 16–17).

which we can discover and account for them (Cartwright, 2007). Zooming into the field of evaluation of development and social change interventions, the Stern Review (Stern et al., 2012), commissioned by the UK Department for International Development (now the Foreign and Commonwealth Development Office), was pivotal in supporting movement beyond the until-then commonly held view of the “gold standard” for evaluation, that is, promoting the experimental design as the best design to test causality. Deepening evaluation discussions to engage with the underlying frameworks through which causal inference (the basis for a causal claim) can be made was a necessary step in this evolution. Table 16.1 summarizes four approaches to causality that are now commonly employed within the evaluation field, and are the basis for causal inference in a range of approaches to evaluation.

To better understand the assumptions behind the arguments for using experimental approaches to measure causality, as well as the alternatives (including qualitative research), we encourage you to read Maxwell’s (2004) text, in addition to Chapter 20 of this *Handbook*. In summary, the journey has brought us to the current understanding that causal explanations can be made through many different evaluation approaches, and by employing different types of methods.

Table 16.1 shows two alternatives to the counterfactual (the basis for experimental and quasi-experimental designs) and regularity (the basis for statistical designs) causal frameworks, which open opportunities for qualitative approaches to causal inference. A key feature of both the configurational and generative causal frameworks is their recognition of multiple factors working together to create change. These frameworks are particularly useful when evaluation is designed to understand complex causal relationships produced by interventions within social, cultural, environmental, and political contexts that influence and are influenced by these interventions.

As Maxwell (2004) argues, inferring causality does not have to depend on a high volume of cases, but rather can be explored by looking at the events and processes that connect them, or by using case-based approaches (e.g., see Ragin, 1987), both of which are useful when one is interested in exploring the complexity of causation. Similarly, using a generative logic, Pawson and Tilley (1997, p. 69) argue that “the relationship between causal mechanisms and their effects is not fixed, but contingent” upon many different factors which cannot be controlled for in the larger context. Yin (2003) similarly emphasizes that case studies are useful to answer “how” and “why” questions, and to provide rich explanations of change in context. He adds that exploration of “rival” explanations is important to support rigor within case study research and evaluation designs.

### Whose Knowledge and Experience Matter?

Increasingly, evaluators and evaluation commissioners are shifting their questions from asking simply “what works” (seeking to measure only the net effect of an intervention), to asking how and why change happens, in what conditions, and for whom? Reframing causal evaluation questions invites us to reconnect with long-standing questions in evaluation and research about “whose knowledge” and “whose experience” count in what is considered success, and more importantly, in the causal explanations of how change happens (e.g., Estrella & Gaventa, 1998). Ultimately, the systems that interventions are designed to influence are made up of many actors with diverse positions, experiences, and values, which inform their perceptions of what constitutes a desirable pathway to change. Interventions and their evaluations are not value-free; indeed, they are often shaped by what commissioners and evaluators value, and how their choice of methods attend to their needs (see Chapter 4).

There are participatory approaches and specific methods designed to develop causal explanations through acknowledging power and centering the perspectives of excluded and marginalized groups whom much philanthropic strategy aims to reach (Apgar & Allen, 2021). Participatory evaluations often entail developing narratives of change and employing creative practices, allowing the full lived experiences of different actors and their own sensemaking of change in systems to inform our understanding of causality (see the case study in Box 16.2 on how the CLARISSA program is embedding causal analysis in participatory action research with children working in the worst forms of child labor). As we suggest in this chapter, causal explanations are more robust, and therefore more useful, when we attend to power and participation throughout the evaluation process.

It is also important to acknowledge that while participatory evaluation methods do not have the same problematic history as many quasi-experimental and experimental design approaches in terms of often being imposed on communities with little input from them, they also do not automatically address equity or handle race and other ways people are marginalized in appropriate and culturally respectful ways. As is true with most research methods, the cultural relevance of the method stems from how it is used and by whom, with attention to facilitation as central to quality in the process (Dick, 2021). It is critical that the evaluator examine the context of the system and how the system is potentially driving oppression, and take care to avoid contributing to that oppression.

#### BOX 16.2 CASE STUDY OF CLARISSA: DEEP PARTICIPATORY AND ACTION- ORIENTED CAUSAL LEARNING

The Child Labour Action Research in South and Southeast Asia (CLARISSA) program was a five-year systemic action research program funded by the UK Foreign and Commonwealth Development Office (UK Aid), led by the Institute of Development Studies, and implemented through a consortium of research and development partners experienced in child participation. CLARISSA was designed to: (1) generate new evidence on the dynamics of the worst forms of child labor (WFCL) in supply chains and in urban neighborhoods in Nepal and Bangladesh; (2) generate innovative responses to WFCL through participatory processes; and (3) explore how to scale those innovations. The program employed a participatory, child-centered, and adaptive approach to programming, such that evidence was

co-produced with working children and other stakeholders (such as employers, and parents and guardians), centering their lived experiences. Participatory learning was used to co-generate innovative responses and evaluate their effectiveness.

Recognizing the limited existing evidence on what drives children into WFCL, from the outset, the consortium partners, together with the donor, developed a learning-oriented and complexity-aware monitoring, evaluation, and learning system. Further, all partners were committed to not only focusing on what was achieved through the program, but also to learning how outcomes were achieved for extremely marginalized children through the program's innovative participatory approach. Consequently, the program's evaluation was embedded within its participatory processes, and employed contribution analysis as the overarching approach to answering specific evaluation questions as they related to identified causal hotspots within three main causal pathways.<sup>2</sup>

The foundation for understanding system dynamics and causal relationships that lead children into WFCL consisted of participatory life story collection and analysis processes involving the working children (Karki et al., 2022; Sayem et al., 2022). This level of involvement of marginalized people in participatory analysis illustrates the potential for combining participatory and causal methods to build ownership as we navigate complexity in systemic interventions (see Burns, 2021, for more on the evolution of this analysis process).

In urban neighborhoods with high proportions of children engaged in hazardous work in Kathmandu and Dhaka, 400 life stories of working children were collected by adults and children. Through a series of participatory analysis workshops, children were supported to undertake causal analysis to identify salient themes for initiating action research groups. The analyses revealed hidden dynamics driving children into WFCL, including family financial struggles leading to the breakdown of family relations, and family health problems and death (often caused by alcoholism) connected to these financial struggles that lead to children's dropping out of school and girls' being married off at a young age (see Hacker & Sharma, 2022, for full details of causal analysis in Nepal).

Children clarified causal dynamics through engagement in action research groups. They deepened their specific understandings of the macro-level issues, and began to formulate their own collective actions to stimulate change (e.g., engaging with adults in their families and neighborhoods through performing dramas about their experiences). Their micro-level interventions were evaluated by the groups through developing their own micro-level theories of change and monitoring indicators to evidence if and through what causal pathways their actions were creating desired change in the systems dynamics. The participatory causal analysis undertaken within the action research groups used multiple data sources to reconstruct the causal links as the participants envisioned them.

The data generated through these participatory processes were used by the CLARISSA evaluation team at the program level to evidence if and how the action research processes generated effective interventions in response to WFCL. Combined with outcome harvesting, the evaluation explored from the children's perspectives the multiple causal pathways to improving working conditions and wellbeing. This example illustrates how rigor can be built through engagement of system change agents in making sense of causal pathways, thus building credibility into the evaluation process.

*Sources:* Apgar et al. (2022); Burns (2021); Hacker & Sharma (2022).

## THE ITERATIVE PROCESS OF DESIGNING EVALUATION TO EXPLORE CAUSAL PATHWAYS

Designing evaluations to explore causal pathways amid complexity should be understood as an iterative, reflexive process that centers action learning throughout. In Figure 16.1, we list the core elements of an iterative, collaborative design process, and how to select appropriate methods, noting in particular the feedback loops between evaluation design and implementation. When working in a participatory and deliberative process, experimentation and learning as you go are central to achieving quality in the evaluation process and outcomes.

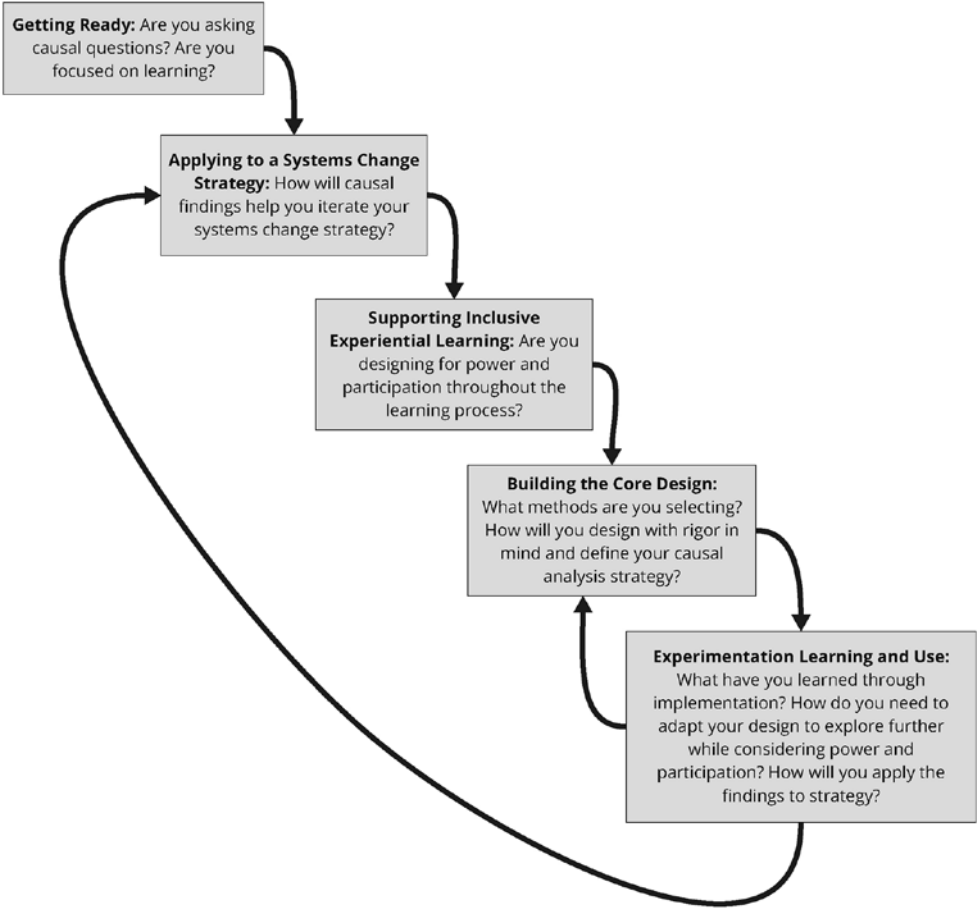


Figure 16.1 *The iterative codesign process for causal analysis*

## Getting Ready

Evaluations that seek to rigorously answer causal questions are not needed in all settings. Sometimes a description of what occurred in relation to a predictive theory of change is sufficient, though in those cases it is important to be clear about what the evaluation can and cannot tell you. It might feel appropriate to make a causal claim when causal thinking was used to design the evaluation at the outset, but these claims should really be reserved for when an explicit causal analysis is included.

Evaluations that can make visible causal pathways through causal analysis bring the greatest value when there is a readiness to engage in this type of evaluation, including:

- Stakeholders are asking questions about how, why, and in what context the interventions are contributing to intended or unintended outcomes;
- It is clear how the causal information can inform strategy (current strategy or a planned future strategy), and there is openness to revising current assumptions about how change happens;
- There is interest in and openness to discovering emergent outcomes, positive and negative, intended and unintended, and understanding how and why they emerged;
- There is not much prior evidence demonstrating that the change strategy leads to the desired outcomes in a similar context – in other words, uncertainty is sufficiently high to justify investment in deep causal analysis;
- The evaluators have technical knowledge of the methods and how to investigate causal relationships, including understanding when to use which method and how to combine them; and
- The funder and implementers are able and excited to engage in a participatory evaluation, and have the resources, time, and access to expertise and facilitative skills to implement a quality participatory evaluation.

Finally, part of getting ready to engage in a causal evaluation is identifying whether you are asking questions that require causal information to answer. Do you primarily need to test a theory of how change happens after the fact, track how change happens as it is emerging, or investigate a specific outcome by looking back to discern what caused it? The methods you pick will differ depending on the focus. The Think Tank Initiative case study described in Box 16.3 illustrates that, in spite of differences in the specific interests of stakeholders involved, the common objective of learning deeply about whether and how change was produced in the face of complexity created the conditions to pursue a causal evaluation.

### BOX 16.3 CASE STUDY OF THE THINK TANK INITIATIVE: COMING TO THE REALIZATION THAT CAUSAL ANALYSIS IS NEEDED

The Think Tank Initiative (TTI) was a large-scale, ten-year, multi-donor-funded program of institutional research capacity-strengthening for think tanks in Africa, Asia, and Latin America. It offered flexible, long-term funding, combined with technical support, to help over 50 organizations move along a pathway to sustainability, through two five-year phases. Overall, TTI was supported through an investment of £100 million by a con-



sortium of funders, including the International Development Research Centre (IDRC), Foreign, Commonwealth & Development Office (FCDO), Hewlett Foundation, Bill & Melinda Gates Foundation, Norwegian Agency for Development Cooperation (Norad), and Directorate General for International Cooperation (DGIS).

Given the scale and complexity of TTI, funders were committed to the idea of making resources available for evaluation as a tool to support learning from the first phase to inform the second phase. Consequently, funders invested in establishing a culture of evaluative thinking, building an extensive monitoring system, and convening learning events. Initially, some think tanks involved in the program were concerned that the evaluation process would lead to their performance being judged unfairly in the early phases of work, while the funders were largely concerned with learning from the process. Upon agreement of all stakeholders, the Phase 1 evaluation used existing research about how and why think tanks can influence policy, and collected 65 “stories of change” using the rapid Outcome Mapping technique. The stories helped validate findings about the “impact pathways” of think tanks, and led to the development of a detailed theory of change. Recommendations from the Phase 1 evaluation created the foundations for building Phase 2.

For Phase 2 of the work, while some individual funders were interested in how funded think tanks had performed, overall, the funding partners were more concerned with learning about how think tanks work in different contexts. Building on the successful evaluation in the first phase, a new phase of learning-oriented evaluation was implemented by a commissioned external evaluation team using the agreed-upon theory of change (developed in Phase 1) as the starting point. The push to a causal emphasis in the second phase was in part responding to learning from Phase 1 about the complexity affecting how think tanks worked in different contexts. The evaluators recognized that multiple causal influences were at work, and so they did not impose a simple, linear attribution of results to the program interventions. Rather, they proposed a Realist evaluation design to understand the causal pathways of each think tank in context.<sup>3</sup>

During Phase 2, the evaluators sought to answer questions about *how* the support offered by TTI led, or failed to lead, to stronger and more sustainable think tanks. There was an explicit intent to unpack the causal pathways of both successful and unsuccessful think tanks. Using a sample of 13 organizations, the evaluators collected qualitative baseline data, and followed up with in-depth monitoring interviews to map out the causal pathways of each think tank. They also produced annual interim reports that were shared in reflection sessions with all grantees, which facilitated actionable learning during the program itself. The support modalities were adjusted in real time to meet the needs of think tanks individually and as a community. The evaluators developed a typology of trajectories (or causal pathways) for think tanks that other funders are now considering as they explore opportunities to support these organizations. The evaluation findings were also of practical value for think tanks, for example, informing think tank leadership about business models and stakeholder engagement practices (Cristoplos et al., 2019).

*Sources:* Cristoplos et al. (2019); Taylor (2022); Young et al. (2013).

## **Applying to a Systems Change Strategy**

Notice that the criteria for readiness for evaluations that explore causal pathways do not include whether change is expected to be produced in simple or complex, interconnected systems. Causal analysis is relevant for evaluating any sort of systems change strategy, though when and how one undertakes causal analysis are affected by the underlying systems change approach being used. Table 16.2 lists the relevance and different uses of causal analysis across three common approaches to systems change employed by funders: ecological systems theory, complexity theory, and focusing on predictable pathways in discrete parts of the system.

Regardless of the conceptual model and entry point into systems change, causal analysis can reveal expected and unexpected outcomes, how and why they are occurring, harms that may need to be mitigated, and ultimately, new pathways to change. Causality may not be unidirectional in systems, and the analytical work to understand causality can help navigate complexity to produce actionable learning.

## **Supporting Inclusive Experiential Learning**

As with all evaluation approaches that acknowledge complexity, such as developmental evaluation (Patton, 1994), it is helpful to ground an ongoing and often embedded evaluation in the practice of experiential learning. For example, identification of emergent outcomes through a method like outcome harvesting may result in efforts to validate outcomes, broaden data collection to discover outcomes, or expand who is involved in the interpretation process.<sup>4</sup> Initial ideas about evaluation design, thus, are revisited as the needs become clearer and change. As noted already, intentionality in pursuing learning is a necessary precondition for causal pathways analysis.

Even when learning is valued, evaluation design may not be participatory, and commissioners and evaluators alone may define the questions and determine what methods to employ. When aiming to explore causal pathways in ways that support inclusion of diverse perspectives, we suggest that the design process itself should be participatory and deliberative. The case study in Box 16.1 is an example of how the design process was participatory, and began with hearing the lived experiences of children engaged in the worst forms of child labor. And the case study in Box 16.2 illustrates how a group of funders and grantees benefited from an extensive initial engagement to clarify their information needs. Acknowledging power dynamics upfront within the evaluation process can help move towards achieving the much harder goal of navigating these power dynamics to support full inclusion (Hanberger, 2022).

## **Building the Core Design**

Building a participatory evaluation design requires addressing three key sets of questions:

- What is our purpose with this evaluation? What are we trying to achieve? What will the evaluation inform?
- Who needs to be involved from the outset and throughout? Whose questions matter? Whose experience of causality matters?
- What specific causal questions are we asking?

*Table 16.2 Examples of causal analysis given different conceptual models and approaches to systems change*

	Systems dynamics approaches and theories (e.g., systems maps and leverage points)	Complexity theory approaches (e.g., using emergent strategy and assuming a highly dynamic system)	Approaches focused on influencing discrete pieces of the system with predictable pathways to change
Type of systems change strategy	Strategy discovers where to act through systematic mapping of the system, identifying leverage points, and considering their interconnected nature with other parts of the system.	Strategy seeks to find experiments worth trying amid complexity, assuming a dynamic and largely unpredictable system.	Strategy commits to a specific, more predictable pathway to change, based on prior knowledge and evidence (e.g., continuing a known strategy or learning from other settings).
Value and focus of causal analysis	Causal analysis can help understand HOW, WHY, and to WHAT EXTENT the strategy is contributing to intended outcomes, what other outcomes are emerging, what else is contributing to those outcomes, and how the change is rippling through the system. In doing so, it helps test theories of change and assumptions embedded in the systems mapping.	Causal analysis can offer rapid feedback loops that help understand HOW, WHY, and to WHAT EXTENT an intervention is contributing to direct and indirect outcomes, including emergent outcomes, and what else is contributing to those outcomes. Often it can help build an understanding of the system behaviors, which prepares the strategy to either take greater risks or prototype new ideas.	Causal analysis can confirm or disconfirm the EXTENT, HOW, and WHY the specific strategy is contributing to the intended outcomes (or other unintended outcomes), and how the larger context is contributing. It can help understand how the outcomes ripple out to affect other parts of the system. This can help clarify the need to continue investment in this discrete systemic action.
Examples	A strategy might seek to address government corruption by using a multi-audience, multi-lever approach to changing the dominant narratives that have normalized corruption. After a deep investment, a causal analysis might find the strategy has resulted in mindset shifts, but the changes in mindsets are not leading to changes in behavior, in part because the predicted increased accountability did not occur. This might result in a new leverage point being prioritized that is also assumed to have potential to influence accountability in the system.	A strategy might experiment with addressing government corruption by trying a mix of different bottom-up interventions. Early outcomes might suggest that some approaches are promising for shifting mindsets, and others are leading to new behaviors. Yet, unexpected harms are being produced, as those with less power are facing retribution for their bold actions due to an opaque environment with little accountability. These findings may lead to a rapid shift and new interventions that specifically seek to create a safer, more accountable environment, where the participants can be protected from retribution.	A strategy seeking to address government corruption might focus on building agency, skills, and power among women, based on research showing the impact in other settings. A participatory causal analysis might focus on questions of interest for the women involved, including asking whether the women’s increased sense of agency is translating to influence. From this, evaluators might discover specific aspects of the context that are more disabling than expected, with complex, multifaceted roadblocks. Findings may lead to a decision to expand the strategy to more directly seek to address the disabling conditions.

Source: Lynn & Coffman, 2024.

There are multiple potential starting points for embarking upon a causal evaluation, which relate to different purposes and imply different methodological needs. When evaluating the impact of a program after it was completed, or summative evaluation, it is possible to start with an observed change (an outcome achieved) and work backwards to identify the cause, thus developing “causes of effects” explanations (Goertz & Mahoney, 2012). The outcome

that is the starting point could be intended or unintended, but if a plausible contribution claim is suspected, then a case-based method presents a valuable way to explore if and how the intervention contributed to the change and for whom.

When conducting a formative, or process-oriented evaluation, the starting point may be more exploratory, and could aim to follow how change is emerging going forward, which could include monitoring the process as it unfolds. Such prospective approaches are often theory-based. There is risk, however, that an evaluation that narrowly focuses only on intended and expected change may miss emergent changes that are likely more important for evaluating systems change strategies (see Table 16.2). Mitigating this risk, we believe, is not about throwing out the prospective tools all together, but rather ensuring their use is iterative and reflexive. For example, contribution analysis, a theory-based approach, can also be used to inform adaptive management (Apgar et al., 2022).

As others have noted, employing both retrospective and prospective approaches likely constitutes the most robust approach, but is only possible with sufficient time and space to embed a causal evaluation from the outset (Jenal & Liesner, 2017). In addition, monitoring provides a tool to detect emergent and unexpected (positive or negative) outcomes.

Deciding who needs to be involved in the evaluation requires careful consideration of all relevant stakeholders, including commissioners or funders, program implementation teams, and actors within the systems where outcomes are sought (e.g., change agents on the ground and intended beneficiaries). Indeed, the evaluation community (especially in the context of international development and humanitarian work) has largely moved beyond the dichotomous view of learning versus accountability to embrace a more representative form of accountability. We can distinguish between managerial accountability, or sending information “up” the chain from the field to the funders, and representative accountability, where information is shared with all evaluation stakeholders (Guijt, 2020). When accountability is understood as relational and multi-directional, information-sharing becomes a vehicle for making the evaluation process more inclusive. Sharing information can build trust, a key ingredient for quality in the causal analysis and learning that follows (Bodorkós & Pataki, 2009; Wicks & Reason, 2009).

Diverse stakeholders will have different forms of power and needs. Commissioners or funders of evaluation may open up or shut down space for employing specific questions and methods. Program implementers, on the other hand, constitute the direct beneficiaries of learning, and have the power to directly feed learning from causal analysis back into strategy. Implementers may, however, prefer linear approaches to evaluation that measure indicators they can easily track, rather than answer harder questions around *how* and *why* their efforts support outcomes. Other stakeholders on the ground may not have a direct influence on the process of change, but have lived experiences which can provide vital information about the context shaping causal relationships.

The needs and values of different stakeholders are unlikely to all neatly align, calling attention to the critical role that facilitation plays in making key evaluation decisions. The role of evaluators needs to broaden from being solely the independent assessors of impact to also serving as facilitators of evaluative thinking and learning among stakeholders, which may raise ethical dilemmas (Barnett & Eager, 2022). When implementing participatory approaches, evaluators are enmeshed in the politics of evidence-building (Eyben, 2013; Parkhurst, 2017), and cannot fall back to play the less messy role of distant technical expert. The need for strong facilitation skills must be considered when constructing evaluation teams.

Table 16.3 *Non-experimental causal designs and methods*

Approach	Methods	Basis for making a causal claim	When to use it
Theory-based	Contribution analysis	In-depth theoretical analysis of causal processes or mechanisms	There is a strong theory of change.
	Process tracing	in context. Many of these	Differences in context are likely to matter.
	Realist evaluation	methods are grounded in	It is important to examine effects for specific groups.
	General elimination methodology	a generative causal framework, though configurational causality is also relevant to some.	
	Qualitative impact assessment protocol		
Participatory	Multiple lines and levels of evidence		
	Innovation history		
	Most significant change (MSC)	Validation by participants that their actions and experienced effects are “caused” by the intervention. Many of these methods are grounded in a generative causal framework.	To capture multiple understandings of change and unintended consequences.
	Outcome harvesting		
	Collaborative outcomes reporting		
Case-based	Collaborative yarning		
	Rapid outcome assessment		
Case-based	Within-case	Analysis of causal processes within a case or across multiple cases.	To identify causal factors within or across cases when effects are known.
	Across-case		
Systems-based	Causal link monitoring	Building a conceptual model of the causal relationships at work.	To illustrate multiple interdependent causal and nonlinear feedback processes.
	Causal loop diagramming		

Source: Authors.

### Selecting appropriate methods

Once specific causal questions have been developed and finalized, appropriate methods to answer the questions need to be selected. In the previous sections, we referenced some methods that can be employed to make causal pathways visible. Table 16.3 includes a longer (yet still not comprehensive) list of methods that, when implemented fully and with rigor, can help examine cause and effect in complex systems.

No one method shown in Table 16.3 will be universally better than the others – they all have a place and time when they can be particularly useful. See the case study in Box 16.4 for an example where one method was used, and yet other methods could have been used instead. Guidance already exists for selecting appropriate methods beyond experimental designs, such as the UK Government Magenta Book (HM Treasury, 2020), and a spreadsheet-based tool developed by the Centre for the Evaluation of Complexity Across the Nexus (Befani, 2020), among others. Underlying the guidance provided is a decision-support tree that requires consideration of the types of questions being asked, the types of causal chains being examined, and the context or attributes of the intervention or strategy being evaluated. Quadrant Conseil (2017) developed a very useful decision tree that includes many of the methods identified above; this trend of visualizing methodological choice through various forms of decision trees is only going to increase.

The evaluation community increasingly recognizes that employing a single method may not enable full exploration of all causal links, and recombination of two or more methods –sometimes called “bricolage” – offers greater potential for causal analysis (Aston & Apgar, 2022). For example, it is not uncommon to combine outcome harvesting, which begins with

describing observable outcomes, with case-based approaches such as process tracing or contribution analysis, to explore the causal links in greater depth. Or it may be that one function of a method is used to strengthen another method, such as the attention to verification with stakeholders in collaborative outcomes reporting enhancing the use of most significant change.

The methodological choices made must support the purpose of the causal analysis. And as noted above, we recommend that the design process overall, including the methodological choices, be undertaken in an iterative manner throughout the evaluation process, not simply by the evaluators, but rather within a shared and deliberative space including all stakeholders. Making methodological choices explicit through an inclusive process ensures that we open up the “black box” of evaluation design and acknowledge both strengths and weaknesses of the design choices made in relation to identified evaluation goals. This, we contend, will increase the probative (proof) value, as well as the use, of the causal analysis, as it responds to both managerial and representative accountability.

### **Designing and implementing with rigor**

Considering how to attain rigor when investigating causal pathways amid complexity entails going beyond adhering to traditional criteria for quantitative research. We need to consider additional criteria for ensuring a high-quality, rigorous design for causal analysis. We offer criteria here that build upon the work of Preskill and Lynn (2016) to guide evaluation aspiring to causal analysis amid complexity. We elaborate five principles to prioritize and balance when conducting evaluations of causal pathways amid complexity:

1. *Quality of the thinking*: The extent to which the evaluation’s design and implementation engage in deep analysis that focuses on patterns, themes, and values (drawing on systems thinking); seek alternative explanations and interpretations; are grounded in the research literature; and look for outliers that offer different perspectives.
2. *Credibility and legitimacy of the claims*: The extent to which the data are trustworthy, including confidence in the findings; the transferability of findings to other contexts; the consistency and repeatability of findings; and the extent to which findings are shaped by respondents, rather than evaluator bias, motivation, or interests.
3. *Cultural responsiveness and context*: The extent to which evaluation questions, methods, and analysis respect and reflect stakeholders’ values and context, their definitions of success, their experiences and perceptions, and their insights about what is happening.
4. *Quality and value of the learning process*: The extent to which the learning process engages the people who most need the information, in a way that allows for reflection, dialogue, testing assumptions, and asking new questions, directly contributing to making decisions that help improve the process and outcomes.
5. *Quality of the participatory process*: The extent to and ways in which the process design is power-aware and facilitated to create opportunities for diverse stakeholders to engage in a meaningful way throughout, as well as incorporate reflexivity on the part of evaluators and commissioners to explore their own bias and power.

### **Causal analysis data, methods, and analytical processes**

When we implement causal analysis, we often collect, code, and analyze data with a predefined theory or structure in mind. We may also present and interpret findings in the context of this same theory. The theory of change used may have been articulated by a funder (as seen in Box 16.4); be our own theory about how change is happening based on previous research;

reflect previous evaluation findings from work undertaken in this setting; or be an as yet barely articulated set of beliefs about how change should happen.

A variety of researchers using many different types of approaches have found that people will construct scenarios to make sense of discrete, related pieces of data, and essentially “discover” the causal pathway by assembling pieces of it, without necessarily interrogating the causal links themselves (Miles & Huberman, 1994). Such interpretative behavior applies both to the individuals from whom we capture cause and effect stories, and to ourselves as researchers or evaluators. As Miles and Huberman (1994, p. 144) note:

These fallibilities are crucial for us as qualitative researchers, who must traffic in meanings as well as actions; who must rely, in part, on the explanations that people give us; and who must live with the knowledge that because we too, are “people,” our own explanations are equally vulnerable.

Due to the heavy role of interpretation in causal analysis, we offer seven practical suggestions to guide data collection and analysis:

1. *Interrogate each step within a causal chain:* The presence of a theory (articulated by evaluators or participants) as to why one thing is leading to another is insufficient on its own to generate a causal claim. Rather, evidence (including contrary evidence about other possible causes) for each step in a causal chain is needed. Rarely in a complex intervention is there one action that alone leads to one desired outcome; rather, there may be a collection of actions, each of which have their own initial outcomes, and combined along with other actions can lead to the desired outcome. Each part of that chain of outcomes needs to be interrogated. When contrary evidence is included in the interrogation, developing an understanding of this evidence can help discover nuances in the causal pathway that are otherwise hidden.
2. *Triangulate sources:* To interrogate causality, you benefit from *multiple sources of data that are exploring the same causal pathways*, allowing you to look at one pathway through different perspectives and types of information. This is more specific than triangulating your sources for the overall evaluation, where the different sources of data collectively tell a larger story. Here, we’re talking about each link in your causal pathway being understood and validated using multiple sources of data (see Chapter 29 on process tracing). It ideally includes collecting data from sources less biased toward a particular theory of how change happened and sources who hold competing explanations. It can include attending carefully to the context of outcomes and what else may be contributing, and collecting data with that in mind, rather than aiming too carefully at the strategy being evaluated and its associated intended outcomes.
3. *Minimize data collection bias:* Evaluators know it’s critical to construct data collection tools in ways that minimize multiple forms of bias that favor a predefined set of outcomes or expectations. Yet, when we ask a series of questions to understand what happened, we can easily make the mistake of asking questions that make visible our existing theory of change and thus bound the data in ways that do not allow other pathways to change to become visible (referred to as intervention bias). It is important to review data collection tools and ask whether they encourage many different ways of understanding what happened to emerge. You can go further and review recordings of interviews to assess the extent to which informal prompting may have encouraged a specific causal pathway to emerge. In the Qualitative Impact Assessment Protocol, one of the methods for understanding

causality amid complexity, data collection bias is minimized by using local, independent researchers who are not given information about the context or purpose of the study (called “blindfolding”). This decreases the risk of confirmation bias toward the desired outcomes of the program (Copestake et al., 2019).

4. *Assess the quality of evidence:* The interrogation of the causal pathways can also include making structured judgment calls about the quality of the evidence underlying each step in a potential causal chain. Rubrics that define the quality of evidence for a causal claim can help weigh the strength of a causal chain step-by-step, rather than only looking at it overall (Lynn et al., 2022b). You must choose what criteria you define as having greater quality, and criteria like plausibility, triangulation, uniqueness, independence, and so on are commonly used in these types of rubrics (e.g., Aston, 2020). When applying criteria, you might find, for example, that one key causal link is understood primarily through a set of evidence that is lower-quality, perhaps based on a single key informant with a vested interest in the causal link looking a specific way. This would signal a need to identify other sources of data to validate that claim or identify alternative claims.
5. *Code for causality:* The interrogation of causal links can include coding data in ways that allow you to investigate causality – not just coding actions and outcomes, but also coding pathways between them and the context of those pathways. Exploring the full range of ways that coding data can prepare for causal analysis is beyond the scope of this chapter, but it is critical to take time to explore how a coding scheme is supporting discovering causality in the data, not just in the interpretation, and incorporating the larger context and its contributions.

As an example of one way to code for causality, in the Causal Map software program, this is accomplished by coding items as causes and effects, coding the link, and even adding temporal information (e.g., before intervention, after intervention; Causal Map, n.d.). It is important to note that tools like Causal Map are helpful for some causal coding needs, but not all, and it is useful to explore what different platforms can help with your causal coding (including to code the impact of context on the causal pathway). There is also movement towards opening up the coding of data and analysis processes to stakeholders outside the evaluation team through processes of collective analysis (see Burns, 2021 for one example); although this practice is still nascent, it is an opportunity to be pursued.

6. *Identify causal hypotheses:* Whether articulated or not, evaluation analysis processes often include the construction of hypotheses about what we believe happened. When interrogating a causal pathway, it can be helpful to clearly articulate multiple hypotheses about how change happened. Tools like Causal Map can help with visualizing these multiple stories about cause and effect. Methods like process tracing can guide the development of competing hypotheses as well. What is critical here is not just constructing hypotheses, but explicitly seeking and using evidence of alternatives to the pathway that is assumed at the beginning of the evaluation.
7. *Test causal relationships:* The interrogation can also include testing the strength of each link in a causal hypothesis or otherwise articulated causal pathway. For example, in process tracing, this is done by a series of tests that require you to have not just one hypothesis about how change happened based on your data, but multiple hypotheses, and then test them for their strength (Lynn et al., 2022a). This process helps weigh the strength



of different hypotheses about how change happened and build an argument for causal pathways emerging from the data.

Alternatively, you can examine causality through a lens of complexity, thinking about how multiple causes co-join and affect each other as well as the outcome. Conceptually, this is about making visible a network of causes and effects, and observing how this network shifts over time (i.e., systemic change). We can look for how these shifts in interrelated causes change the types of outcomes emerging, and how shifting one cause can change others (Miles & Huberman, 1994). Like a more linear cause-and-effect chain, as often investigated by methods like process tracing, this type of systems-aware causal pathway still requires multiple forms of evidence, not just about the presence of specific causes and effects, but also their relationships.

### **Implementation, Learning, and Use**

Evaluations are intended to generate usable results for one or more audiences, as described in the fourth rigor criteria we offered above, which emphasizes the value of the learning process. In evaluations designed to make visible causal pathways, sensemaking and use are closely intertwined, with one of the critical audiences being those who are in the system, experiencing the change.

#### **Sensemaking**

Sensemaking, or the process of using data to draw conclusions and make causal claims, can be facilitated among stakeholders involved in the evaluation to help them collectively make sense of what the data and analysis means for them. In the context of international development, such sensemaking may be embedded within program cycles, and is defined as “a process in which people jointly make sense of information and develop a shared understanding. It assumes that individuals have different interests and perspectives, and often see information in different ways” (Simister & O’Flynn, 2017, p. 1). Evaluators play an important role in bringing their expertise and findings into these facilitated processes. Learning workshops are common ways that all stakeholders of an evaluation (beyond just the commissioner and evaluators) can be afforded an opportunity to engage with data and findings in relation to their own learning questions. In the collaborative outcomes reporting approach, for example, a deliberative “outcomes panel” opens up the process of sensemaking to key stakeholders outside the evaluation team.

#### **Application to strategy or program design**

Causal findings can be a powerful input into strategy or program revision. Unlike descriptive findings, which focus on measuring implementation and outcomes, causal findings may challenge beliefs about whether the emerging outcomes are the result of the strategy or program, and what other outcomes are resulting. Evaluation commissioners and program implementers can use the results to: (1) refine their explicit or implicit theories about how change happens in the system; (2) refine their understanding of their own strategy in relationship to how change happens; (3) refine their plan for evaluation and learning moving forward; and (4) refine their engagement strategies, as emerging outcomes and pathways to change may suggest that additional stakeholders be engaged in the design or revision of a strategy or program.

Regardless of which method and set of analytical tools are used, the interrogation of causal pathways and generation of causal inferences should be appreciated as more than evaluation as usual. Causal analysis entails a deep examination of the relationships between causal factors, including more than the planned strategies and intended outcomes. And it brings the larger context into the evaluation, not as an accompanying story to help explain the findings, but as a set of critical variables within the causal pathway to produce intended outcomes. When well-implemented, causal analysis can challenge how implementers understand the system they are working in, and uncover where, how, and under what conditions they see opportunities to influence change.

### BOX 16.4 CASE STUDY OF COLLECTIVE IMPACT: SEEING THE METHODS AND ANALYSIS IN ACTION

In 2017, ORS Impact and the Spark Policy Institute embarked on an evaluation effort to understand the degree to which a specific systems change approach (called collective impact) can contribute to population-level change across many sites. The evaluation utilized a *process tracing* methodology, along with collection of rich descriptive data and storytelling-style focus groups, in sites with participants whose stories and approaches differed significantly from other sites. A causal story was created for each site, with the nine sites with the most significant population level effects also receiving detailed case reports to use in their own learning and communication work.

The evaluation sought to analyze population-level change with as much rigor as possible, without attempting to simplify the complexity of the context, the variability of implementation of collective impact efforts across sites, or the many interim changes needed to see impact at scale. Phase I of the study included holding interviews with key stakeholders, collection of secondary data from each initiative, and collection of independently-generated data regarding key population-level changes to which the initiative claimed contribution. The quality and core insights gleaned from each source of data were judged using three rubrics focused on the central concepts of interest. Use of the rubrics allowed for the application of clear and consistent criteria to make judgments within each site, across sites, and across the variety of types of data. A “change” memo was created for each site, laying out the causal pathway to the population-level change, including where there were different and non-linear pathways being articulated across data sources.

During Phase II of the study, in-person site visits were conducted with the nine sites whose causal stories were the strongest. This stage of the analysis specifically sought to examine whether the practices of the collective impact approach were necessary contributors to the population-level changes observed. To answer this question, one of the sessions during the site visits included a dialogue with participants who brought distinctly different experiences with the change process. Together, participants explored, shifted, and weighed different causal elements, and helped to explain the relationships between types of outcomes, including which ones were critical steps in the path to population-level change. A final stage of data collection was employed to validate additional causal claims identified during the in-person sessions.

The analytical process used in Phase II (site visits) and Phase III (validation and final analysis) drew upon process tracing methodology (see Chapter 29), where competing hy-

potheses were explored, each of which represents a plausible explanation of the causes leading to a given outcome. The exploration resulted in a rating of each hypothesis's level of inferential strength, using four tests that are part of the process tracing methodology, and lead to an assessment of the strength of the contribution to change (Lynn et al., 2022a). Process tracing was a highly effective method to employ for this study due to the focus on why and how the collective impact approach contributed to change, including how multiple interim outcomes and external factors contributed. Contribution analysis and Qualitative Comparative Analysis would also have worked well, and led to different priorities in how the causal pathways were interrogated. In any causal pathways analysis, there are likely multiple methods that are appropriate and others that are not. Outcome harvesting, for example, would not have been helpful for answering the questions posed by this study, as the study identified outcomes up-front (population-level change), and sought to investigate the associated causal pathways, rather than seeking to generate a variety of types of outcomes and associated pathways.

The findings did fully confirm all elements' contributions in the collective impact approach. For example, one element (called "Continuous Communication") was found to function more as a supportive activity of a strong "Backbone" (another element) than to have its own direct impact. The evaluation also found that initiatives focused on a wide range of policy changes (rather than looking for a few big wins), and which included policy changes to shift the allocation of resources in the system, had a more direct and significant contribution to improving population-level outcomes.

The findings were widely used, including by funders of collective impact, consultants who support its implementation, sites advancing the work, and even in the mainstream media to highlight the good work happening in systems change (Brooks, 2018). One of the more referenced aspects was the understanding of both how and why some sites provided a unique and necessary contribution to population-level change, while others offered a pathway to change that was important, but heavily supported by other enabling factors.

*Source:* Stachowiak et al. (2020).

## CONCLUSION: TAKING ACTION

The transition from causal description in impact evaluation to using causal analysis to make inferences that explain the how and why of change (including the impact of context) is not easy. It takes time to learn about new methods, and to learn to implement them with rigor and in participatory ways. It is also critical to set the evaluation up for interrogating causality from the beginning, recognizing that measurement of interventions and outcomes must be accompanied by data that helps to explain the unfolding causal pathways.

Making this effort is more than worth it. Collectively, humans are investing heavily in social and environmental change strategies throughout the globe. We are making decisions based on our best knowledge and assumptions, but absent tests of those assumptions, including the causal ones, we may continue to reinforce and repeat past decisions. Are we having the full impact we could have, and addressing the problems that cause inequities and harm? Could we go further and have greater impact if our collective set of assumptions about change were regularly examined and challenged? We call for program and strategy designers, commission-

ers and funders of evaluation, and evaluators themselves to prioritize deep, participatory, and high-quality causal analysis to make sense of what it takes to change the world for the better. We offer a call to action to use appropriate methods to examine causal pathways in ways that honor context, culture, and voice, to invite many ways of understanding change. We call for action to make evaluation findings meaningful, so that others act on them and share them, to build our collective understanding.

## NOTES

1. To go deeper into this idea, explore the work of the Big Push Forward, where the politics of evidence are explored by many leading European thinkers. In particular, Polonenko (2018) reviewed a book by leading thinkers in this initiative, offering insights about the politics of evidence-seeking, and their impact on development work.
2. Contribution analysis is a form of theory-based evaluation that iteratively maps available evidence against a causal theory of change, and identifies and then addresses challenges to causal inference through a six-step process (see Ton, 2021, for further information).
3. Realist evaluation (see Pawson & Tilley, 1997) is a form of theory-based evaluation that seeks to uncover underlying mechanisms that trigger outcomes, by asking what works, for whom, in what respects, to what extent, in what contexts, and how.
4. Outcome harvesting is an approach to evaluation that begins with the collection of observed outcomes and works backward to understand how these changes came about. It offers an opportunity for program implementers to be involved in making sense of outcomes, and supports feeding their learning from evaluation back into program implementation (see Wilson-Grau, 2019).

## REFERENCES

- Apgar, M., & Allen, W. (2021). Participatory monitoring evaluation and learning: Taking stock and breaking new ground. In D. Burns, J. Howard, & S. M. Ospina (Eds.), *The SAGE Handbook of Participatory Research and Inquiry* (pp. 831–844). Sage Publications.
- Apgar, M., Snijder, M., Prieto Martin, P., Ton, G., Macleod, S., Kakri, S., & Paul, S. (2022). Designing contribution analysis of participatory programming to tackle the worst forms of child labour. *CLARISSA Research and Evidence Paper 2*. Institute of Development Studies. <https://doi.org/10.19088/CLARISSA.2022.003>.
- Aston, T. (2020). Quality of evidence rubrics. [https://www.linkedin.com/posts/tom-aston-consulting\\_quality-of-evidence-rubrics-activity-6736598045133164544-mfkZ/](https://www.linkedin.com/posts/tom-aston-consulting_quality-of-evidence-rubrics-activity-6736598045133164544-mfkZ/).
- Aston, T., & Apgar, M. (2022). The art and craft of bricolage in evaluation. *CDI Practice Paper 24*. Institute of Development Studies. <https://doi.org/10.19088/IDS.2022.068>.
- Barnett, C., & Eager, R. (2022). Evidencing the impact of complex interventions: The ethics of achieving transformational change. In R. D. Van den Berg, P. Hawkins, & N. Stame (Eds.), *Ethics for Evaluation: Beyond “Doing No Harm” to “Tackling Bad” and “Doing Good”* (Chapter 6). Routledge.
- Befani, B. (2020). *Choosing Appropriate Evaluation Methods: A Tool for Assessment and Selection (Version Two)*. Centre for the Evaluation of Complexity Across the Nexus, Guildford.
- Bodorkós, B., & Pataki, G. (2009). Local communities empowered to plan? Applying PAR to establish democratic communicative spaces for sustainable rural development. *Action Research*, 7(3), 313–334.
- Brooks, D. (2018). A really good thing happening in America. *The New York Times*, October 8. <https://www.nytimes.com/2018/10/08/opinion/collective-impact-community-civic-architecture.html>.
- Burns, D. (2021). The role of collective analysis in generating ownership and action in systemic action research. In D. Burns, J. Howard, & S. M. Ospina (Eds.), *The SAGE Handbook of Participatory Research and Inquiry* (pp. 692–708). Sage Publications.
- Cartwright, N. (2007). *Hunting Causes and Using Them*. Cambridge University Press.
- Causal Map (n.d.). Guide to causal mapping. Causal Map. <https://guide.causalmap.app/>.

- Copestake, J., Morsink, M., & Remnant, F. (2019). *Attributing Development Impact: The Qualitative Impact Protocol Case Book*. Practical Action Publishing.
- Cristoplos, I., Pain, A., Kluykens, J., & Fruhling, P. (2019). *External Evaluation of the Think Tank Initiative (TTI) Phase Two, 2014–2019, Final Report*. NIRAS.
- Dick, B. (2021). Facilitating participatory research. In D. Burns, J. Howard, & S. M. Ospina (Eds.), *The SAGE Handbook of Participatory Research and Inquiry* (pp. 207–219). Sage Publications.
- Estrella, M., & Gaventa, J. (1998). Who counts reality? Participatory monitoring and evaluation: A literature review. IDS Working Paper 70. Institute of Development Studies.
- Eyben, R. (2013). Uncovering the politics of ‘evidence and ‘results’: A framing paper for development practitioners. Prepared for the Politics of Evidence Conference, Brighton, UK.
- Gates, E., & Dyson, L. (2017). Implications of the changing conversation about causality for evaluators. *American Journal of Evaluation*, 38(1), 29–46.
- Goertz, G., & Mahoney, J. (2012). *A Tale of Two Cultures*. Princeton University Press.
- Guijt, I. (2020). Accountability and learning: Exploding the myth of incompatibility between accountability and learning. In A. Fowler & C. Malunga (Eds.), *NGO Management* (pp. 339–352). Routledge.
- Hacker, E., & Sharma, R. (2022). Life stories from Kathmandu’s adult entertainment sector: Told and analysed by children and young people. *CLARISSA Research and Evidence Paper 4*. Institute of Development Studies. <https://www.ids.ac.uk/publications/life-stories-from-kathmandus-adult-entertainment-sector-told-and-analysed-by-children-and-young-people/>.
- Hanberger, A. (2022). Power in and of evaluation: A framework of analysis. *Evaluation*, 28(3), 265–283.
- HM Treasury (2020). *Magenta Book 2020: Supplementary Guide: Handling Complexity in Policy Evaluation*. HM Treasury.
- Jenal, M., & Liesner, M. (2017). Causality and attribution in market systems development. *The BEAM Exchange*. [www.beamexchange.org](http://www.beamexchange.org).
- Karki, S., Apgar, M., Snijder, M., & Sharma, R. (2022). Learning from life story collection and analysis with children who work in the worst forms of child labour in Nepal. *CLARISSA Learning Note 2*. Institute of Development Studies. <https://clarissa.global/resource/learning-from-life-story-collection-and-analysis-with-children-who-work-in-the-worst-forms-of-child-labour-in-nepal/>.
- Lynn, J., & Coffman, J. (2024). Passing in the dark: Making visible philanthropy’s hidden and conflicting mental models for systems change. *The Foundation Review*, 16(2).
- Lynn, J., Stachowiak, S., & Byers, J. (2022a). *How to Do Process Tracing: A Method for Testing “How Change Happened” in Complex and Dynamic Settings*. Center for Evaluation Innovation.
- Lynn, J., Stachowiak, S., & Coffman, J. (2022b). Lost causal: Debunking myths about causal analysis in philanthropy. *The Foundation Review*, 13(3), 6.
- Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, 33(2), 3–11.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook* (2nd ed.). Sage Publications.
- Parkhurst, J. (2017). *The Politics of Evidence: From Evidence-Based Policy to the Good Governance of Evidence*. Routledge.
- Patton, M. Q. (1994). Developmental evaluation. *Evaluation Practice*, 15(3), 311–319.
- Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. Sage Publications.
- Polonenko, L. M. (2018). Book review: *The Politics of Evidence and Results in International Development: Playing the Game to Change the Rules?* by E. Rosalin, I. Guijt, C. Roche, & C. Shutt. *Canadian Journal of Program Evaluation*, 33(2).
- Preskill, H., & Lynn, J. (2016). *Rethinking Rigor: Increasing Credibility and Use*. FSG. <https://www.fsg.org/resource/rethinking-rigor/>.
- Quadrant Conseil (2017). How can impact be evaluated? Which approaches can you use depending on the intervention’s characteristics and the available data? In response to which expectations? <https://www.quadrant-conseil.fr/ressources/ArbreImpact.html>.
- Ragin, C. C. (1987). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press.
- Sayem, M., Paul, S., Apgar, M., & Snijder, M. (2022). Learning from life story collection and analysis with children who work in the leather sector in Bangladesh. *CLARISSA Learning Note 1*. Institute

- of Development Studies. <https://www.ids.ac.uk/publications/life-stories-from-children-working-in-bangladeshs-leather-sector-and-its-neighbourhoods-told-and-analysed-by-children/>.
- Simister, N., & O'Flynn, M. (2017). Sensemaking. *INTRACT Brief*. <https://www.intrac.org/wpcoms/wp-content/uploads/2017/01/Sensemaking.pdf>.
- Stachowiak, S., Lynn, J., & Akey, T. (2020). Finding the impact: Methods for assessing the contribution of collective impact to systems and population change in a multi-site study. *New Directions for Evaluation*, 2020(165), 29–44.
- Stern, E., Stame, N., Mayne, J., Forss, K., & Davies, R., & Befani, B. (2012). Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development. Working Paper 38. Department for International Development.
- Taylor, P. (2022). Evaluating capacity-strengthening impact: A funder perspective. *CDI Practice Paper 23*. Institute of Development Studies. <https://www.ids.ac.uk/publications/evaluating-capacity-strengthening-impact-a-funder-perspective/>.
- Ton, G. (2012). The mixing of methods: A three-step process for improving rigour in impact evaluations. *Evaluation*, 18(1), 5–25.
- Ton, G. (2021). Development policy and impact evaluation: Learning and accountability in private sector development. In H. Zafarullah & A. S. Huque (Eds.), *Handbook of Development Policy* (pp. 378–390). Edward Elgar Publishing.
- White, H. (2019). The twenty-first century experimenting society: The four waves of the evidence revolution. *Palgrave Communications*, 5(1), 1–7.
- Wicks, P. G., & Reason, P. (2009). Initiating action research: Challenges and paradoxes of opening communicative space. *Action Research*, 7(3), 243–262.
- Wilson-Grau, R. (2019). *Outcome Harvesting: Principles, Steps, and Evaluation Applications*. Information Age Publishing.
- Yin, R. K. (2003). Designing case studies. *Qualitative Research Methods*, 5(14), 359–386.
- Young, J., Hauck, V., & Engel, P. (2013). *Final Report of the External Evaluation of the Think Tank Initiative, London and Maastricht: ODI and European Centre for Development Policy Management (ECDPM)*.