9-2021

# Lost Causal: Debunking Myths About Causal Analysis in Philanthropy

Jewlya Lynn
*PolicySolve*

Sarah Stachowiak
*ORS Impact*

Julia Coffman
*Center for Evaluation Innovation*

# Lost Causal: Debunking Myths About Causal Analysis in Philanthropy

Jewlya Lynn, Ph.D., PolicySolve; Sarah Stachowiak, M.P.A., ORS Impact; and Julia Coffman, M.S., Center for Evaluation Innovation

## Introduction

When philanthropy seeks to drive change — especially in messy, complex, and dynamic systems — it can feel like strategy development and implementation takes place in the proverbial "black box." We select ideas that are promising, have reasonably high confidence that positive short-term outcomes will occur, and then hold out hope that the strategy will eventually add up to more than the sum of its parts.

In our experience as evaluators in philanthropy, we observe that evaluations of complex philanthropic strategies often do little to unpack assumptions about what happens in the black box of change once strategies are unleashed. More often, we see evaluations that describe observed changes without investigating how or why they occurred, including the relationships between what was funded, what was implemented, and what resulted. Even more problematic, we see evaluations that simply assume, without investigation, that relationships exist between implemented strategies and observed outcomes.

What if, in addition to learning that change happened, we also learned how and why? What if our evaluations made change pathways more visible, tested assumptions and hypotheses, and generated new insights based on what happened in the black box of a systems-change strategy? What if we understood cause-and-effect relationships not just in more controlled programmatic work, but also in dynamic and emergent strategies that include network building, field building, advocacy, organizing, or movements? This type of learning comes from causal analysis — inquiry that explores cause-and-effect relationships.

## Key Points

- What if philanthropic evaluations told us that changes in the world had occurred, as well as how and why they occurred, including in what ways foundation funding and grantees contributed to those changes? What if evaluations made change pathways more visible, tested hypotheses and assumptions, and generated new insights based on what happened in the "black box" of systems change strategies? This type of learning comes from causal analysis — inquiry that explores cause-and-effect relationships.

- Yet currently in philanthropy, particularly for strategies and initiatives that feature high complexity, few evaluations use robust techniques for understanding causality. Instead, philanthropic evaluation tends to rely on descriptive measurement and analysis. These descriptions often are rich, meaningful, and in-depth, but they remain merely descriptions nonetheless. This article challenges the myths that hold us back from causal inquiry, allowing us to embrace curiosity, inquiry, and better knowing, even (or especially) if it means learning that our assumptions and theories do not hold up.

- We argue that philanthropy more frequently needs to examine causal relationships, using a growing suite of methodological approaches that make this possible in complex systems. Causal methodologies can challenge and strengthen the often uncontested beliefs that underlie philanthropic interventions, while offering evidence about enabling contexts and system drivers. Strong causal analysis considers not only the funder's model and assumptions, but also the beliefs others hold about how and why change occurs, opening the door to more equitable and less biased ways of understanding change.

Before diving in further, we want to be clear about our focus here — or rather, what our focus is not. First, by causal analysis we do not mean root cause analysis, or the process of discovering the root causes of problems in order to identify appropriate solutions. Second, we are not debating the utility of randomized control trials (RCTs) or quasi-experimental designs (QEDs), which is where conversations about causal analysis often go in philanthropy and among evaluators. While we do address the myth that RCTs and QEDs are the only way to do causal analysis, our purpose here is not to debate their merit or use. Finally, we do not rehash much of the existing literature in this space that covers philosophical theories associated with causal analysis, or its technical aspects, including relevant mathematical models, statistical equations, or machine learning challenges (e.g., Cartwright, 2004; Pearl & Mackenzie, 2018; Rubin, 2005).

Our purpose is to open the conversation about causal analysis so that we can better see what is possible. As we describe in this article, there are many ways to conduct rigorous nonexperimental causal analyses. Causal analysis is possible for strategies that play out in more controlled conditions and settings, as well as strategies situated in complex and dynamic systems, where causes and effects are interrelated and interdependent. We think shedding light on these possibilities will help philanthropy to see the value in causal analysis and when to use it, rather than avoiding it altogether or deploying it in ways that can do harm when the method does not fit the context or the intent.

In this article, we aim to dispel misconceptions about causal analysis that we regularly encounter in philanthropy, including among evaluators. We find these myths get in the way of evaluations that can lead to deep learning and support greater, more equitable impact. By naming and addressing these myths, we hope to increase demand for evaluations that use causal analysis in order to go beyond answering what change happened, and delve more deeply into how change happened. (See text box on page 19.)

## The Case for Causal Analysis

Philanthropic grantmaking requires that foundations make decisions based on assumptions about how to drive change. Grantee selection, the offering of additional supports, and general deployment of philanthropic resources are then based on these decisions. These decisions typically are documented in strategies and theories of change that lay out the short-, intermediate-, and long-term outcomes that are expected to occur as a result of choices made.

Some of these choices are better grounded in research and evidence than others. In complex systems-change work, they more often are based on a combination of experience and intuition. While this is understandable for complex problems and strategies where what you do in one setting does not always translate to another, it is rare that cause-and-effect assumptions for strategies situated in complex systems get tested through evaluation. More commonly, foundations commission descriptive studies about these strategies that feature rich depictions of the context, how strategies were implemented, and any observed changes, but contain little to no causal analysis. The consequence is that the philanthropic sector is not developing knowledge about how change happens (or does not) in systems as a result of different types of interventions.

Given the volume of philanthropic resources devoted to solving problems rooted in complex systems (e.g., criminal justice reform, democracy reform, climate change, education reform), we worry about the lack of causal insight that is emerging. Our concerns are based on the prevalence of:

- Strategies based on how people think change ought to occur rather than a clear-eyed look at how it actually occurs;

- Evaluations that look only for predefined outcomes without testing whether alternative pathways to change or different, equally important outcomes are emerging;

<div style="background:#8a1414; color:white; text-align:center; padding:8px;">

**Tales of Two Analyses: Imagining How Causal Analysis Can Change the Conversation**

</div>

Late on a Thursday afternoon, Jamie, a senior program officer at a community foundation, arrived at the community center where community members had been invited to discuss the results of an evaluation. The first participants were arriving alongside the two co-facilitators, a community leader, and the evaluator the community had selected. As she helped to set up the presentation and discussion materials, Jamie felt excitement. Something important was going to happen tonight.

With the dozen regular participants in the room, the meeting kicked off. After some grounding, everyone dug into a large visual on the wall — the story of how their community had tackled homelessness through an antiracist lens, ultimately increasing food and shelter access and the number of affordable housing units, and decreasing the number of people reporting housing insecurity.

### The Tale of Descriptive Analysis

There was some pride and joy in the room as a description of the years of work was explored, with a sense of "Yeah, you got that right" and an occasional, "That isn't how I remembered it." By the meeting's end, there was general agreement that the description resonated and would be useful to share with the foundation board, future funders, and the media.

It was a good story about how community members and advocates all worked hard to drive change, with community members taking many small actions they could advance right away while advocates pushed upstream change with local government that took longer to come to fruition. It was a story about how media advocacy mattered and how the work had impact.

By the time the meeting was done, people were tired and many were glad the evaluation had ended. It had used up a lot of their time, but it was important that their story was told in the right way, particularly to their primary funder.

### The Tale of Causal and Descriptive Analysis

There was some pride and joy in the room as the evaluation's causal and descriptive analysis was investigated, along with some pushback when interpretations differed, as well as acknowledgment of how much of the success depended on outside factors they appreciated, but could not control.

When the meeting was over, participants had come to some conclusions about how change had happened. A number of those conclusions challenged what Jamie believed coming into the work. The biggest was her assumption that investing deeply in nonprofit advocacy was foundational, given the government's large role in housing. Some evidence supported that advocacy had helped, but stronger evidence suggested the direct actions of local community members and organizations (and the media coverage they earned) had influenced policymaker actions.

In fact, it had been hard to find much evidence that supported advocacy's role, in part due to the credibility challenges many advocates had with policymakers. The broader racial justice movement that had gained visibility during the same time had also helped, bringing to light deep inequities in the housing system.

As the participants walked out the room, they agreed the evaluation had taken their time, but they also acknowledged it had led to insights and new ways of talking to the foundation about where resources were most needed. As one participant said, "I had a gut sense of what was happening, but it wasn't what everyone was saying, so I stayed quiet. It wasn't until I saw the story on the wall that I felt OK talking about my perceptions openly, and even then, I learned a lot about what was less visible to me in my work. We really have a full picture now."

- Foundations that expend all of their evaluative energy on landscape scans, descriptive narratives, or predefined metrics to explain what happened, without rigorously examining why it happened; and

- Descriptive stories about change that fail to contest philanthropic thinking, and that simultaneously fail to surface and test the assumptions, experiences, and beliefs of those closest to the problems that philanthropy seeks to solve.

*In our experience as evaluators in philanthropy over the last two decades, we have observed a growing number of foundations that are tackling highly complex problems rooted in deeply dysfunctional systems (e.g., improving the ability of Congress to function in spite of political polarization; supporting an economic paradigm shift away from neoliberalism).*

How did we get here, where descriptive evaluation is the norm? We think it has something to do with how philanthropic strategy and focus has shifted in the sector.

In our experience as evaluators in philanthropy over the last two decades, we have observed a growing number of foundations that are tackling highly complex problems rooted in deeply dysfunctional systems (e.g., improving the ability of Congress to function in spite of political polarization; supporting an economic paradigm shift away from neoliberalism). In doing so, foundations are embracing complexity principles and systems thinking, recognizing that many actors and factors interact in unpredictable and often invisible ways to create the problems that foundations seek to address (Kania et al., 2014). They are treating strategy as more dynamic and emergent and profoundly affected by context, rather than as a series of well-considered and predictable steps that can be forecast in a long-term plan (Coffman, 2016).

As strategy has become more emergent, and as the complexity of context has increased, the relevance of well-known causality evaluation methods has decreased. Instead of leaning into new causal methods that are appropriate for complexity, many foundations and evaluators have accepted descriptive designs as the best that they can do.

In addition, because constant adaptation must be an essential component of strategy in complex systems, evaluators have shifted to approaches like developmental evaluation (Patton, 2011) that are designed to support emergence. Many evaluators using these approaches, as well as evaluation users, assume that causal analysis is too retrospective and takes too long to be useful in evaluations that support real-time learning.

Complexity is becoming even more relevant now that philanthropy is wrestling explicitly with systemic racism, white supremacy, and how to advance racial equity (Daniels, 2020). Foundations are asking new questions about root causes, who is harmed by the status quo, and ways to produce change that do not reinforce existing inequities and injustices. These shifts have led to funding approaches that are highly dynamic and emergent and designed to tackle upstream drivers of systemic problems, such as advocacy, power building, networks, movements, and field building.

We believe this increasing movement toward complexity in philanthropy is precisely why we need more causal analysis in addition to the good descriptive work already happening. If we seek to advance equity and justice, understanding how change happens and the contribution of specific approaches are critical so that the status quo can be shifted and inequities truly addressed. (See text box on page 21.) In addition, if we don't get better at making our assumptions about change explicit and investigating them, we risk continuing to do harm by replicating processes and activities that allow inequities to persist.

When done well, causal analysis can lift up and leverage the power of stories, lived experiences, and multiple ways of knowing. It can generate powerful ways to create shared understanding across many people involved in the work. When

we do not use causal analysis, we have a more cursory understanding of what we did and what happened under a certain set of circumstances and we lose the ability to test our assumptions, create knowledge about effectiveness that can drive future work, and break through our cognitive and implicit biases.

## Myths About Causal Analysis

In our conversations with both philanthropists and evaluators, we find that misunderstandings and misinformation are at the root of much of the sector's hesitancy to focus evaluation on cause-and-effect relationships. These misunderstandings are repeated so often that they have become a powerful set of myths about causal analysis and its relationship to rigor, usefulness, and equity. It is past time to debunk these myths.

*When we do not use causal analysis, we have a more cursory understanding of what we did and what happened under a certain set of circumstances and we lose the ability to test our assumptions, create knowledge about effectiveness that can drive future work, and break through our cognitive and implicit biases.*

### Does Causal Analysis Matter When Our Focus Is Equity?

Philanthropy's increasing commitment to advancing racial equity will benefit from an increased use of causal analysis. This argument is being made by proponents of critical race theory, intersectionality theory, and evaluators working in real-world settings where equity is a focus. The need for causal evidence related to equity comes from multiple current gaps in our knowledge related to the programs and population-level work we fund, the systems in which we intervene, how we define problems and their potential solutions, and even how we understand pathways to change.

*Programmatic interventions:* While the evaluations of many programmatic interventions have used causal analysis, there is an absence of causal evaluations for programs designed to be culturally relevant, as has been documented in the education field (Dee & Penner, 2017). Lacking the same evidence base as other types of interventions, culturally relevant approaches are less likely to be funded and adopted.

*Population level interventions:* Often, when intervening to address inequities observed in a variety of social systems, "descriptive statistics highlight important outcome differences between groups, but they may do little to establish underlying causes or motivations that can guide policy change or the implementation of interventions" (Sablan, 2019, p. 185). The absence of causal findings means that the causal mechanisms that drive or address inequities remain grounded in theory and assumptions, but are not being tested. We lack evidence about which interventions truly work to advance equity at a population level, and the mechanisms by which they work.

*Pathways to change:* While philanthropic strategies rarely dig deeply into the academic literature on critical race theory, intersectionality, and other ways of deeply understanding inequities, many philanthropic frameworks, theories of change, and other tools are grounded in these concepts. The theorists behind them recognize that causal analysis is needed to move them from theory to evidence. A deeper understanding of the complexity of how multiple identities and needs intersect "is vital for understanding social injustice and intervening on behalf of oppressed groups" (Murphy et al., 2009, back cover). Similarly, systems dynamics that drive inequities and interventions that seek to disrupt them need to be tested in order to move from observing racial differences in opportunities and outcomes to testing theories and assumptions and building an evidence base about how change can happen (Sablan, 2019).

## Rigor and Causal Analysis: Myth 1

***RCTs or quasi-experiments are the only ways to test cause-and-effect relationships.*** Any time the phrase "cause and effect" or the word "causality" comes up in an evaluative context, instantly, the conversation tends to go to the (often polarizing) topic of experimental designs or RCTs. In fact, many evaluators have long argued that RCTs are the best method for assessing causality, with little attention to context (Gates & Dyson, 2017).

We see no need to argue the merits of RCTs or other QEDs. They are useful causal analysis tools in certain contexts. Rather, we want to draw attention to a set of nonexperimental methods that give us a wide range of rigorous options to choose from, many of which are a much better fit for complexity. (See text box below.)

Philanthropy's movement toward "contribution not attribution" signals an increasing understanding that attribution — definitively isolating whether an outcome would not have happened without a particular effort — is difficult to impossible to attain in complex and dynamic settings. The standard instead is on contribution and determining whether a credible and plausible case can be made, based on evidence, that causal connections exist. This has led to new thinking on the concept of causality itself. For example, evaluator John Mayne (2012) introduced the idea of a "causal package" as a useful way to think about how one organization's strategy must interact with a broader mix of complementary interventions, actors, events, and contextual factors to increase the probability that desired changes will occur. The core premise of a causal package is that multiple causal factors must work together in order to produce a change. Each cause alone is necessary but not sufficient. A "package" of necessary causal factors acting together, however, can be sufficient. Especially with ambitious long-term goals that require complex solutions,

---

### Three Types of Evaluation Designs

| Experimental | Quasi-experimental | Nonexperimental |
|---|---|---|
| Experimental designs (also called RCTs) have a defining characteristic: the random assignment of individuals or targets to intervention and control groups (also called the counterfactual, or the condition in which an intervention is absent). The intervention group participates in the program or intervention, while the control group does not. Random assignment results in intervention and control groups that do not systematically differ, creating a situation where any differences between the groups that are observed after the intervention takes place can be causally attributed to the intervention with a high degree of confidence. | Quasi-experimental designs are like experimental designs in that they aim to make cause-and-effect statements about an intervention or strategy's impacts, but they do not use random assignment. Most QEDs construct comparison groups or other types of counterfactuals to examine an intervention's impacts for those who do and do not participate. While attempts are made to make sure that intervention and comparison groups do not systematically differ, some differences may exist. | Nonexperimental designs, like experimental and quasi-experimental approaches, examine relationships between variables and draw inferences about the possible effects of an intervention, but they do not have counterfactuals that control subjects or conditions. They are most commonly used with interventions or strategies situated in complex systems. Nonexperimental designs that explore causality often incorporate validation, or checking back with key informants on the accuracy of data and reasonableness of interpretations; and counterfactual thinking, exploring whether alternative explanations could have caused or contributed to observed relationships or outcomes. |

**TABLE 1** Nonexperimental Causal Designs and Methods

| Approach | Methods | Basis for Making a Causal Claim | When and Why to Use It |
|---|---|---|---|
| **Theory-Based Approaches** | • Contribution analysis<br>• Process tracing<br>• Realist evaluation<br>• General elimination methodology<br>• Qualitative impact assessment protocol<br>• Multiple lines and levels of evidence<br>• Innovation history | In-depth theoretical analysis of causal processes or mechanisms in context | • When there is a strong theory of change<br>• When differences in context are likely to matter<br>• When it is important to examine effects for specific groups |
| **Participatory Approaches** | • Most significant change<br>• Outcome harvesting<br>• Collaborative outcomes reporting<br>• Collaborative yarning<br>• Rapid outcome assessment | Validation by participants that their actions and experienced effects are "caused" by the intervention | • To capture multiple understandings of change and unintended consequences<br>• More timely and affordable<br>• Sample size is small |
| **Case-Based Approaches** | • Within-case<br>• Across-case | Analysis of causal processes within a case or across multiple cases | To identify causal factors across cases when effects are known |
| **Systems-Based Approaches** | • Causal link monitoring<br>• Causal loop diagramming<br>• Statistically created counterfactual | Building a conceptual model of the causal relationships at work, and verifying it with empirical data for each variable, mathematical formula, or computer simulation | To example multiple interdependent causal and nonlinear feedback processes |

Source: Gates & Dyson, 2017

interventions or initiatives should be designed with an assessment of the full causal package thought to be necessary to effect change.

The widespread understanding in philanthropy that contribution is the appropriate standard for complexity paves the way for the use of nonexperimental designs and methods that test whether a strong enough contribution case can be made. But both a lack of awareness that these approaches exist and faulty assumptions about their rigor cause them to be rarely used in practice.

## Rigor and Causal Analysis: Myth 2

*There are no rigorous nonexperimental designs for examining causality.* There are many nonexperimental causal design options. Gates and Dyson (2017) place the array of nonexperimental options for assessing causal relationships (many of them developed in the last 10 to 15 years) into four main categories: (1) theory-based, (2) participatory, (3) case-based, and (4) systems-based. (See Table 1.)[1]

We think about rigor here not as a particular method or design, but as the practice of

---

[1] While some of the methods in Table 1 exclusively focus on examining cause-and-effect relationships, others can be implemented in ways that are more descriptive than causal. For example, case studies are commonly and usefully used in philanthropic evaluation. For case studies to be causal designs, they need to incorporate methodological approaches that examine cause-and-effect relationships (e.g., using contribution analysis within a single case study design; qualitative comparative analysis across multiple cases).

*While causal analyses in complex, systems-change examples do not seek to create replicable program models that can be implemented regardless of setting, they help to build a better body of knowledge about what has worked, when, and why than descriptive studies alone can achieve.*

embedding sound evaluation principles and practices into an evaluation. This means, for example, systematically collecting and analyzing data to make sure the conclusions drawn are accurate and credible, openly discussing and exploring possible alternative conclusions, and using participatory approaches for interpreting results and identifying their implications.

An example of a rigorous implementation of causal analysis is the study that ORS Impact and Spark Policy Institute conducted on collective impact, a form of cross-sector collaboration to address complex social and environmental challenges (Kania & Kramer, 2011). The evaluators used a nonexperimental theory-based approach for examining causality (Stachowiak et al., 2020). Their use of process-tracing methodology sought to answer a fundamental causal question: To what extent and under what conditions does the collective impact approach contribute to systems and population changes? The study examined 25 collective impact initiatives via interviews and document review, examined eight initiatives via site visits and process tracing to rigorously test the approach's contribution to demonstrated population changes, and conducted virtual focus groups with three equity deep-dive sites to better understand equity work in the collective impact context.

## Usefulness of Causal Designs: Myth 3

***Causal designs focus on the past and do not help with future decision-making.*** In some ways, this myth has merit — to explore how something happened, we must observe a change that already has occurred. However, this does not mean that causal analysis is only useful in hindsight. Having better evidence about how change happened has a number of benefits for future decision-making.

Building a stronger base of knowledge about complex change can strengthen future strategy decisions in related work. While causal analyses in complex, systems-change examples do not seek to create replicable program models that can be implemented regardless of setting, they help to build a better body of knowledge about what has worked, when, and why than descriptive studies alone can achieve.

Having evidence of how change happened can also help other funders and public-sector actors lean into specific approaches to systems change with more confidence. Many evaluators who use causal analysis promote greater uptake precisely because the approaches increase the confidence of policymakers and funders about what has worked and why (Barrett, D'Errico et al., 2020).

An example of how causal analysis affected future decisions comes from the Agricultural Research Center for International Development, which sought to help winemakers achieve a geographical designation that they believed would help Brazilian producers increase their income (Blundo-Canto et al., 2020). An impact analysis alone would have shown that the geographical designation was achieved, and that the producers increased their sales and margins. However, a participatory causal analysis explained the mechanism by which income increased — efforts to support learning and motivation across producers led to a virtuous cycle of ongoing learning and increased professionalism, which improved the quality of the wine and increased their income. It was this set of mechanisms, not the geographic designation, that made the difference. The causal analysis also found that the changes experienced by producers made them

more resilient to fluctuations in political and institutional support. Identifying these mechanisms had significant implications for future programming and resources dedicated to building capacity for innovation and learning.

### Usefulness of Causal Designs: Myth 4

***Causal designs are summative only and cannot be used for real-time decision-making.*** If an evaluation's goal is to conduct real-time learning that uses data-informed feedback loops to explore how a system is changing, it is true that some causal methodologies cannot be used. Other approaches, however, are appropriate for ongoing implementation, capturing insights about specific interim outcomes and other types of systems changes as they emerge.

Outcome harvesting is an example of an approach that supports continuous learning. This method collects (harvests) evidence on what has changed and, then, working backwards, determines whether and how an intervention has contributed to these changes. It is useful with complexity, when it is not possible to define in advance with precision what an intervention will achieve over time. Outcome harvesting implemented with processes for validating causal pathways can be repeated over time, providing systems-sensing information combined with evaluative information about how change has happened. This is how Humanity United's Peacebuilding Portfolio is using the method, with biannual debriefings as it harvests and validates outcomes along the way. It allows the foundation team to observe steadily how and why the system is changing, both in response to their interventions and to other factors (Personal communication, Jen Heeg, May 13, 2021).

### Usefulness of Causal Designs: Myth 5

***Causal designs are not appropriate for complex settings.*** Because many people think about RCTs when the topic of causal designs is raised, it is easy to believe that causal approaches are appropriate only for situations in which RCTs are most commonly deployed — for programmatic efforts or models that provide replicable or semi standardized solutions for problems

that can be addressed using similar methods and procedures. Randomized control trials are less applicable to adaptive initiatives like systems-change efforts that offer flexible and often emergent strategies to address problems that require unique, context-based solutions (Britt & Coffman, 2012).

We recognize that for dynamic adaptive initiatives that take place in complex settings, it is difficult to discern cause and effect, and rarely does any one person hold the whole story of change. Many nonexperimental causal methods explicitly recognize this complexity, seeking to validate causal pathways through multiple perspectives and to understand the impact of larger system dynamics. (See Table 1 on page 23.)

In the study by ORS Impact and Spark Policy Institute referenced earlier, collective impact's contribution to population-level change was examined across many sites. Collective impact as an approach is complex and deeply context dependent. The sites in the study ranged widely across geographic scope, topical area, target populations, and more. The use of process tracing to understand the degree to which the approach plausibly contributed to population-level impacts across settings provided practical advice about which aspects of the approach mattered most. The causal findings helped to build confidence among implementers about the pathways to change they were assuming. The study also helped to illuminate where additional inquiry could further bolster the work, including around equity and the use of data. While the study did not prove collective impact would always create population-level change, it did find that it can contribute meaningfully impacts across different sites, settings, and topics (Stachowiak et al., 2020).

### Usefulness of Causal Designs: Myth 6

***Causal designs are too burdensome for participants.*** For the four main nonexperimental causal designs identified by Gates and Dyson (2017), deep understanding of how change occurs requires engagement with key stakeholders involved in the work. (See Table 1.) It is through this inquiry that different experiences,

*The solution here is not to pick simplified methods (any more than communities should simplify their context and culture), but rather to give space, time, and opportunity for learning across methods and culture.*

vantage points, and pieces of information can be brought together to create a nuanced understanding of change. While some secondary or extant data can be additive, strong implementation of causal methods in complex settings does require stakeholder participation.

Rightfully, philanthropy and evaluators are regularly mindful of the burden associated with nonprogrammatic asks of grantees and partners on the ground, including participation in evaluation. In our experience, engagement around questions of causality is additive to the work and worth the time invested. Burden should be assessed not just based on the time required for participation; it should also be judged on the value received from the output. Too often, work that merely describes what happened and which outcomes resulted does not lift up lessons that can inform future efforts. In these cases, the cost of participation can outweigh the value returned for the efforts participants put in.

### Causal Designs and Equity: Myth 7

***Causal designs cannot be implemented in ways appropriate for working with communities.***
Another assumption often made is that causal designs are too inherently complex in their design, implementation, interpretation of findings, and reporting of results. This complexity is a barrier, so the myth goes, for communities engaging in participatory evaluation processes.

This myth assumes that community partners lack the ability to engage with and understand

an evaluator's analytical approaches. At the same time, we often assume that evaluators have the ability to understand sufficiently the depth and complexity of a community's experiences, cultures, systems, and history. We argue that causal methods, being merely technically complicated, are less complex to understand in this equation.

The solution here is not to pick simplified methods (any more than communities should simplify their context and culture), but rather to give space, time, and opportunity for learning across methods and culture. Evaluators can use some of the same methods to bring causal designs to communities as communities use to bring their culture and experience to evaluation — through stories, metaphors, visuals, and shared dialogues. Evaluators can also listen for and explore how communities are testing their own causal assumptions, including their ways of measuring and telling stories about how and why change happens over time.

For example, Jeph Mathias' approach to outcome harvesting centers the lived experiences of those in the most marginalized parts of complex systems, not just as data sources, but as part of the study team. In Kenya, he engaged street youth as partners in collecting stories, supporting them to learn from their peers and to listen to other system actors. Independent of the external evaluation team, street youth listened to other street youth. They also accompanied evaluation team members to interviews with leaders at UNICEF and Kenya's Ministry of Youth Affairs, for example. As "insiders," the youth could hear stories that were inaccessible to the external evaluation team and they could act as "contextual experts." This enabled the evaluation team to see otherwise hidden parts of the system and to understand motives and meanings that were underneath the surface (Personal communication with Jeph Mathias, Oct 8, 2021).

Perhaps the myth here actually gets at a very different barrier than the inaccessibility of causal designs. Instead, the real cause of this myth may be the lack of evaluator skills, time, or resources

needed to engage in causal analysis in ways that are accessible and meaningful to communities.

### Causal Designs and Equity: Myth 8

***Causal designs are rooted in white supremacy.*** We recognize that some causal designs use a statistical research practice that is grounded in a history of white supremacy (Bonilla-Silva & Zuberi, 2008). There are legitimate, serious concerns with these analytical tools that seek to explain differences among humans and human experience through mathematical analyses that simplify and reduce people. Equally legitimate are concerns about research tools that assume there is one factual reality, as opposed to multiple truths that need to be understood. In addition, concerns abound about how these methods define concepts like rigor, objectivity, and validity (Dean-Coffey, 2018).

Fortunately, causal designs are not limited to methods with this historical (and contemporary) set of challenges. The nonexperimental causal designs that appear in Table 1 include methods designed by Indigenous researchers (Shay, 2019); qualitative methods that seek to explore the full story, not simplify it; and mixed methods that seek to understand multiple dimensions through different lenses. Additionally, many of these methods either explicitly acknowledge different ways of knowing or are designed so that the sources of evidence are not limited to one way of knowing.

Causal analysis done well also addresses a fundamental problem with many studies that seek to explain causality in relationship to race. Too many social scientists have described the "effect of race" in their findings, implying causal relationships between race and other outcomes (Bonilla-Silva & Zuberi, 2008). An evaluation that explores causal relationships to understand the drivers of inequity and the effectiveness of interventions is unlikely to make claims that race is a driver of specific outcomes. Instead, such a study is likely to find the ways in which race is a characteristic of who is affected and how by a system and an intervention.

*Causal analysis done well also addresses a fundamental problem with many studies that seek to explain causality in relationship to race. Too many social scientists have described the "effect of race" in their findings, implying causal relationships between race and other outcomes.*

### Causal Designs and Equity: Myth 9

***Causal designs center philanthropy's ideas about change.*** Evaluation, as practiced now in philanthropy, tends to take a positivist approach that accepts a foundation's strategy as is and looks at whether the foundation's strategy and related theory of change is playing out as expected. As with strategy, most foundations set or approve all evaluation terms for their strategies: what the questions are, who the evaluator is, what the scope of inquiry is, what the design should be, which data matter, and most importantly, what constitutes success.

While this practice of centering the foundation's ideas about change is typical, it is more about traditional power dynamics in philanthropy and long-standing foundation-and-evaluator routines and habits than it is about evaluation methods. The habit we have of centering foundation priorities in evaluation is changeable. While we acknowledge that foundations have centered their strategies and points of view in philanthropic evaluation, we also recognize that this is the result of the questions we ask (and who is asking them), not the result of the methods we choose. Causal analysis can center anyone's ideas about change, not just the foundation's (Beer, Patrizi, Coffman, 2021).

Techniques like process tracing or contribution analysis can focus on outcomes that emerged,

whether intended or not, and build an evidence base for a pathway to change that may or may not match the funder's assumptions, depending on what the evaluator discovers through data collection and analysis. Outcome harvesting, similarly, can be implemented with a focus on the types of outcomes in a funder's theory of change. Alternatively, it can be implemented the way Humanity United's Peacebuilding Portfolio is using it, where the adaptive theory of change offers a general sense of the universe of outcomes to be harvested, but does not limit that universe. The foundation actively seeks the harvesting of outcomes that emerge, regardless of whether it predicted them.

Causal analysis can also be a powerful way to challenge assumptions that there is a "right" pathway. It can demonstrate when these assumptions do not hold true or surface complementary pathways or conditions in the larger context that are necessary parts of the change process. In this sense, causal analysis methods that leave room for emergence can actually feel quite risky for some in philanthropy, as they become a means by which a philanthropic strategy or theory of change can lose its credibility. Yet, this decentering and challenging of the foundation's point of view is an important part of shifting power in philanthropy as well as increasing philanthropic impact.

## Call to Action

Repeated and reinforced often enough, myths are notoriously challenging to dispel. Given the number and variety of myths named in this article, it is no surprise that causal analysis tends to be the rare evaluative exception in philanthropy more than the rule.

We write this article during a time of profound disruption, when the opportunity to change and transform both philanthropic and evaluation practice feels possible. The COVID-19 pandemic and issues like climate change have brought into stark relief the interconnectedness of systems, while the racial reckoning taking place in the United States makes clear that continuing with the status quo will not lead to meaningfully different outcomes and futures. These issues, along with broader questions about power and privilege, are leading to shifts in philanthropic priorities, strategies, and approaches to both grantmaking and evaluation.

We believe that part of evaluation's role in supporting transformative change in this moment is to help the sector get smarter about how change happens in complex systems. Engaging in evaluative work in complex settings should include the use of causal analysis, alongside the active revision of our ideas about change as we learn from these analyses.

Imagine yourself as the program officer in the story above about two types of analysis (page 21). Do you want to facilitate dialogue and learning about change and what can drive it further, or do you want to describe what happened to the people who have already lived through it? We need to let go of the myths that hold us back from using every possible tool that we can, including causal analysis, to both understand and effect change.

## References

Barrett, S., D'Errico, S., Anderson, S., & Nebsu, B. (2020). Unpacking mechanisms in climate resilient agriculture interventions. *New Directions for Evaluation, 2020*(167), 115–130.

Beer, T., Patrizi, P., & Coffman, J. (2021). Holding foundations accountable for equity commitments. *The Foundation Review, 13*(2), 64-78. https://doi.org/10.9707/1944-5660.1565

Blundo-Canto, G., Devaux-Spatarakis, A., Mathé, S., Faure, G., & Cerdan, C. (2020). Using a participatory theory driven evaluation approach to identify causal mechanisms in innovation processes. *New Directions for Evaluation, 2000*(167), 59–72.

Bonilla-Silva, E., & Zuberi, T. (2008). Toward a definition of white logic and white methods. In E. Bonilla-Silva & T. Zuberi (Eds.), *White logic, white methods: Racism and methodology* (pp. 3–30). Rowman & Littlefield.

Britt, H., & Coffman, J. (2012). Evaluation for models and adaptive initiatives. *The Foundation Review, 4*(2), 44–58. https://doi.org/10.4087/FOUNDATIONREVIEW-D-12-00001.1

Cartwright, N. (2004). Causation: one word, many things. *Philosophy of Science, 71*(5), 805–819.

Coffman, J. (2016, April). Oh for the love of sticky notes! The changing role of evaluators who work with foundations. *Medium*. https://medium.com/@jcoffman/oh-for-the-love-of-sticky-notes-the-changing-role-of-evaluators-who-work-with-foundations-66ec1ffed2e4#.ifphr6o4a

Daniels, A. (2020, July 16). Foundations nationwide commit nearly half a billion dollars to racial justice. *Chronicle of Philanthropy*. https://www.philanthropy.com/article/Foundations-Nationwide-Commit/249187

Dean-Coffey, J. (2018). What's race got to do with it? Equity and philanthropic evaluation practice. *American Journal of Evaluation, 39*(4), 527–542.

Dee, T. S., & Penner, E. K. (2017). The causal effects of cultural relevance: Evidence from an ethnic studies curriculum. *American Educational Research Journal, 54*(1), 127–166.

Gates, E., & Dyson, L. (2017). Implications of the changing conversation about causality for evaluators. *American Journal of Evaluation, 38*(1), 29–46.

Kania, J., & Kramer, M. (2011, Winter). Collective impact. *Stanford Social Innovation Review*. https://ssir.org/articles/entry/collective_impact

Kania, J., Kramer, M., & Russell, P. (2014, Summer). Strategic philanthropy for a complex world. *Stanford Social Innovation Review*. https://ssir.org/up_for_debate/article/strategic_philanthropy

Mayne, J. (2012). *Making causal claims* (Brief No. 26). Institutional Learning and Change Initiative. https://cgspace.cgiar.org/bitstream/handle/10568/70211/ILAC_Brief26_Making%20causal%20claims.pdf?sequence=1 http://betterevaluation.org/sites/default/files

Murphy, Y., Hunt, V. H., Zajicek, A. M., Norris, A. N., & Hamilton, L. (2009). *Incorporating intersectionality in social work practice, research, policy, and education.* National Association of Social Workers Press.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use.* Guilford Press.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect.* Basic Books.

Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association, 100*(469), 322–331.

Sablan, J. R. (2019). Can you really measure that? Combining critical race theory and quantitative methods. *American Educational Research Journal, 56*(1), 178–203.

Shay, M. (2019, February). Extending the yarning yarn: Collaborative yarning methodology for ethical Indigenist education research. *Australian Journal of Indigenous Education, 50*(1), 1–9.

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.

*Julia Coffman, M.S.,* is co-executive director of the Center for Evaluation Innovation. Correspondence concerning this article should be addressed to Julia Coffman, Center for Evaluation Innovation, 1660 L Street NW, Suite 450, Washington, DC 20036 (email: jcoffman@evaluationinnovation.org).

*Jewlya Lynn, Ph.D.,* is founder of PolicySolve.

*Sarah Stachowiak, M.P.A.,* is chief executive officer of ORS Impact.