BEAM EXCHANGE

Causality and attribution in market systems development

Report

Marcus Jenal
Mollie Liesner

March 2017

Citation

Jenal, M., Liesner, M. (2017) *Causality and attribution in market systems development*. The BEAM Exchange, available at: www.beamexchange.org

Published by:

The BEAM Exchange
c/o PricewaterhouseCoopers LLP
7 More London Riverside
London SE1 2RT
United Kingdom
contact@beamexchange.org
www.beamexchange.org

The BEAM Exchange is a facility for knowledge exchange and learning about the role of market systems approaches in reducing poverty.

Acknowledgements:

We would like to thank all the key informants who were willing to engage with us in this research. Thanks to Ray Pawson for allowing us to cite from his article (Pawson, 2007), which is very insightful and accessible, but unfortunately was never formally published. Thanks also to Charley Clarke and Tim Ruffer of Itad for their quality assurance and critical feedback to make this paper as accessible as it is now.

UKaid
from the British people

The BEAM Exchange is a programme funded by the UK's Department for International Development (DFID) and the Swiss Agency for Development and Cooperation (SDC). It is administered by PricewaterhouseCoopers LLP, working with organisations including the Institute of Development Studies.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Agency for Development and Cooperation SDC

Executive summary

Attribution is the establishment of a causal link between (parts of) an observed change and a specific intervention. Attribution is a key requirement of any monitoring and results measurement (MRM) and evaluation activity. Donors require programmes to show whether, and to what extent, the results they are reporting are caused by their interventions. There are many opinions on how to attribute changes to interventions. This is particularly true for interventions in complex systems such as markets. Some practitioners and authors say that it's not possible to attribute change to market system interventions, while others say that there are well-established quantitative methods of doing so. This paper sheds light on the current theoretical/academic and practitioner understanding of attribution and causality. It then suggests a typology for attribution to create a better understanding of when one method is more appropriate than another.

First, the paper presents a view of current attribution practice in market systems development, drawing on several interviews with practitioners and experts as well as on practitioner oriented literature. Three key themes emerged from both interviews and literature, specifically there are:

1. Conflicting views on the extent to which attribution is possible for programmes using a market systems approach
2. Three common approaches for measuring attribution, and
3. Political and practical constraints to measuring attribution.

Practitioners have a great desire for solutions that are pragmatic rather than dogmatic. Most interviewed practitioners appreciate that there is no perfect way of establishing attribution. At the same time, they clearly wish to become better at attributing change to interventions so that they can use this as a tool for learning and improvement.

The paper then presents how causality is viewed in the theoretical evaluation literature. We introduce three types of causality: a successionist, a configurational and a generative model. After discussing and combining insights from both practice and theory, we suggests a pragmatic approach to attribution based on three central observations:

1. A generative approach is better able to explain how the causal mechanisms work (the "why it worked" question).
2. A successionist approach is better at achieving a higher level of confidence in showing co-occurrence of cause and effect, and the size of the net effect (the "if and to what extend it worked" question).
3. Perfect attribution in complex contexts is difficult (if not impossible).

The paper then suggests a typology of approaches based on different contexts and perspectives on change. It discusses how this typology can be applied in market systems development and presents two programme case examples. This should facilitate practitioners to choose the methods that are most appropriate to their context and are grounded in what is currently understood as "good practice" in the evaluation literature. How close any programme can come to what would be a 'best-fit' in its context depends on the programme's available resources for monitoring and evaluation as well as the technical capacity of the staff which a programme can attract. Therefore it is important to be transparent about the logic and decisions behind developing the attribution strategy, and limitations therein.

1. Introduction

This section introduces the paper's rationale and methodology and gives an overview of the contents.

1.1 Rationale

How do we as development actors know whether our efforts in development in general, and in influencing market systems in particular, have been successful? We of course need to monitor for changes, measure results and evaluate the programme. Observing and measuring changes alone, however, does not give us the confidence that the observed and measured changes are due to our interventions. To achieve confidence that our interventions indeed played a role in achieving the change, we need to attribute the change to the intervention. Attribution is a key requirement of any monitoring and results measurement (MRM) and evaluation activity. Donors require programmes to show whether, and to what extent the results they are reporting are caused by their interventions.

There are many opinions on how to attribute change to interventions. This is particularly true for interventions in **complex**[1] systems such as markets. Some authors say that it's not possible, while others say that there are well-established quantitative methods to attribute change.

For practitioners, this diversity of views can be confusing or unhelpful. This paper attempts to bring together current practice and theoretical perspectives on attribution and causality. It suggests a pragmatic way forward for practitioners by doing two things:

1. Shedding light on the current theoretical/academic and practitioner understanding and measurement of attribution and causality.
2. Suggesting a typology for attribution to create a better understanding of when one method is more appropriate than another. Or in other words, what we can find out using a particular method (or family of methods) and what it (or they) cannot tell us.

Who is this paper aimed at? This research paper is aimed at professionals involved in commissioning and implementing evaluations, as well as monitoring and results measurement (MRM) professionals who are involved in developing attribution strategies for programmes using market systems approaches.

What is the purpose of this paper? This paper is not a guide to attribution. It rather provides a basis on which to develop an attribution strategy. It does so by providing the reader with an understanding of the different perspectives on attribution, both from an academic as well as a practitioner point of view. It then attempts to synthesise these perspectives into a typology that helps MRM and evaluation practitioners to make an informed choice about their attribution strategy.

1.2 Methodology

We first studied relevant theoretical and practical literature on attribution and causality. We drew on academia and practitioner-oriented literature concerned with monitoring and evaluation.

The literature review was supplemented with 11 key informant interviews to verify the key findings and further refine the understanding of current practice. The interviewees included experienced MRM and evaluation professionals. We also interviewed evaluation scholars who apply a scientific understanding of attribution and causality in their work.

1.3 Overview of the paper

Section 2 of the paper frames the debates by clarifying what is meant by "attribution" and the related concept of "contribution" and explores attribution on different levels.

1 See Appendix 2: Complexity and emergence

Section 3 gathers key themes and measurement challenges that emerged from the key informant interviews. Section 4 explores three different theories of causality in the evaluation literature. In Sections 5 and 6 we use the key insights to suggest a pragmatic approach to attribution in market development programmes.

2. Framing attribution

Attributing observed changes at market systems level to interventions is arguably challenging. Before we tackle this challenge, we need to clarify some concepts. Firstly, what do we mean by "attribution" and the related concept of "contribution"? Secondly, what do we want to attribute to what?

2.1 Attribution or contribution?

The OECD Glossary of Key Terms in Evaluation and Results Based Management defines attribution as:

"The ascription of a causal link between observed (or expected to be observed) changes and a specific intervention. … Attribution refers to that which is to be credited for the observed changes or results achieved. It represents the extent to which observed development effects can be attributed to a specific intervention ..." (OECD, 2010:17).

Hence attribution only refers to the extent to which observed changes are caused by an intervention. This implies that the overall change might be larger than the portion of that change caused by the intervention.

This definition seems to be clear enough. Yet some authors suggest using the term "contribution" rather than attribution. Their reason for doing this is to help them describe the effects of interventions in complex systems, where there are multiple other potential or actual causal factors. Ruffer and Wach (2013), for example, describe interventions in market systems as operating "*as part of a wider system where, in nearly all cases, they interact in some way with other public and private activities to achieve their intended results. Interventions aim to catalyse change, inducing spill-over effects to indirectly scale up change. Most interventions can therefore be seen as a "contributory" cause – i.e. the intervention is a vital part of a "package" of causal factors that are together sufficient to produce the intended effect."* (Ruffer & Wach, 2013:17).

This distinction between attribution and contribution leads to some confusion. When asked about their own definitions, the key informants' answers varied greatly. Some interviewees understood attribution as only "sole attribution". For this group, attribution means that an intervention causes the whole effect. Others define attribution as the process by which the proportion of a change that an intervention caused is isolated and *quantified*. For this second group, both attribution and contribution signify that the intervention is one of a group of factors contributing to a change, the difference between the two being that attribution quantifies the part played by the intervention, whereas causality does not. The confusion in the definitions even led one person to talk about "attributable contribution"!

A few interviewees criticised the distinction between attribution and contribution altogether. They said that no one measuring attribution in complex systems can claim sole responsibility for a change. One key informant raised an interesting point: If we think of human beings and all the entwined factors that influence us, then it does not make sense to talk about attributing a change to one sole cause. The point she made was that there is no technical difference between the two terms, only a semantic one.

Some interviewees said that the distinction can become an "excuse" for lazy measurement practices. They shared experience of programmes claiming that measuring attribution is too difficult, and therefore assess contribution instead. Programmes then use methods lacking in rigor (or no methods at all), even where better-fit methods are available.

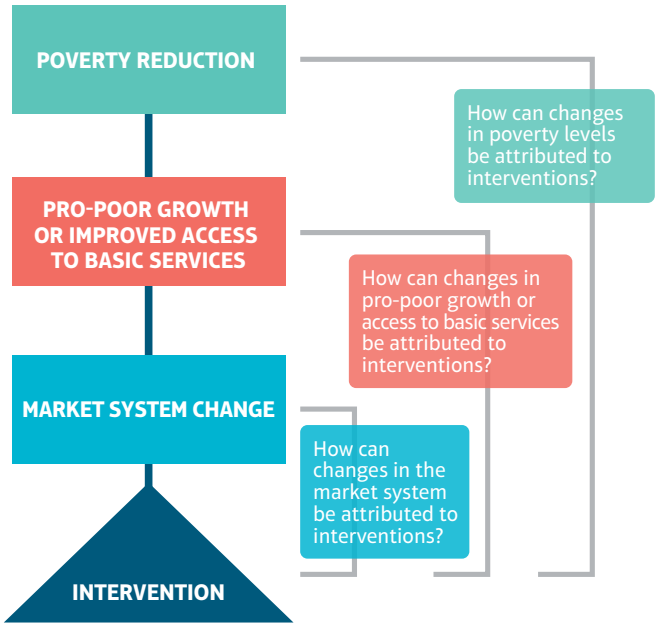
One of the key informants made another suggestion: We should retire this distinction altogether. Instead, he suggested differentiating between sufficient and necessary causes.[2]

So what definition are we going to use? In this paper, for reasons of simplicity and clarity, we use the original definition of attribution by the OECD: Attribution is the establishment of a causal link between (parts of) an observed change and a specific intervention. Whether we can quantify this link depends on the chosen view of causality (more on this in section 3).

2.2 Attribution on different levels

The generalised **Theory of Change**[3] for a market systems approach shown in Figure 1 illustrates the intended changes, represented as three "results levels": (1) market systems change, (2) change in pro-poor growth or improved access to basic services, and (3) poverty reduction. At each results level, we find a different operating context and an increasing number of influences external to the programme that contribute to any observed change. Therefore, different approaches and methods are required to measure and attribute change to interventions on these different levels. For example, different approaches are needed to attribute a change in market systems dynamics to a programme's intervention(s) or to attribute a change in household income or enterprise performance to a change in market system dynamics.

Figure 1: Generalised Theory of Change for programmes using a market systems approach with attribution questions



Source: The Springfield Centre (2014), questions added by authors

We will return to this discussion in Section 6 after looking at how attribution is done in practice and the challenges faced (Section 3) and then establishing a technical understanding of causality (Section 4).

2 For an explanation of the terms "necessary" and "sufficient" see Section 4.2.
3 For an introduction to Theories of Change on market systems development programmes see BEAM Guidance

3. Attribution in practice

Now that we have a better definition of attribution, we can explore the experiences and challenges of conceptualising and measuring attribution in practice, reported by practitioners.

This section summarises the insights gained from interviewing key informants supplemented by insights from practitioner oriented literature. The three key themes that emerged from both interviews and literature are:

- conflicting world views over to what extent attribution is possible for programmes using a market systems approach (Section 3.1)
- three common approaches for measuring attribution (Section 3.2) and political and practical
- constraints to measuring attribution (Section 3.3).

3.1 Conflicting world views and the middle ground

Varying views emerged from the key informant interviews over whether it is at all possible to attribute observed changes resulting from interventions in market systems.

On one end of the spectrum, interviewed practitioners commented that demonstrating causality in complex systems is near-on impossible. This is in line with Snowden (2013) who asserts that: "*you can't show actual attribution because it's dependent on other things that are going on at the same time."* This position is also echoed in the impact evaluation literature. Befani and Mayne (2014) argue for "*shifting the focus of impact evaluation from "assessing impact" to "assessing confidence" (about impact)."*

In opposition to this view, some interviewed practitioners are of the opinion that it is possible to attribute complex changes to interventions using a combination of qualitative and quantitative methods: a view that is also shared in the literature (White & Phillips, 2012; White, 2009b). One key informant claimed that the complexity argument is a "red herring". With **experimental methods,**[4] the statistical tools used can control for complexity. Furthermore, most experimental methods do not need to go into the "why" an intervention worked to tell whether a programme has made a change or not. Therefore, the informant claimed, the fact that causal relationships within complex systems cannot be predicted is irrelevant.

This debate, however, is not a dichotomy: there is a wide range of views between these two ends of the spectrum. The position of most of the key informants – the middle ground – was that programmes should be pragmatic in choosing their attribution methods to build up a realistic picture of what happened. There is a broad acceptance that due to the complexity of the systems that programmes work with, and the difficulty of assessing causality in complex systems, programmes cannot measure attribution with complete precision.

How then are programmes to know how much precision is "good enough"? A few interviewees suggested that guidelines should be developed on what is *not* good enough. This could steer programmes when developing their attribution strategies. Others cited existing guidance, in particular Posthumus and Wanitphon (2015). These authors however, favour counterfactual methods and do not suggest alternatives that could provide the same level of rigour. There are as we will introduce below, other approaches to attribution that might be a better fit for complex contexts. However, there is a lack of guidance for a pragmatic use of these methods.

This middle ground is characterised by three prominent features that were repeatedly mentioned by the interviewees:

1. Practitioners see it as **important to use a combination of methods to measure attribution**. The reason for this is to use complementary strengths to combat the limitations of any single method.

4 For an explanation of experimental and quasi experimental designs see http://bit.ly/2mbaO1K

2. Practitioners **emphasise participatory methods**. Any attribution measurement should involve participation from those within the system, in both the design of measurement frameworks and the analysis of data. Participatory methods also support capturing unintended changes.

3. Practitioners think we should **be more explicit in discussing the multitude of causal factors influencing the change** which the programme aspires to, including a recognition that there are causal factors we do not know about. This allows a more comprehensive understanding of how the changes take place and what the realistic role of an intervention was.

3.2 Three common approaches for measuring attribution in practice

The three types of approaches on measuring attribution used by interviewed practitioners and mentioned in the literature are: (1) comparative approaches, (2) prospective approaches, and (3) retrospective approaches. Examples and common tools used for each approach are provided later in the paper.

Comparative approaches

Comparative approaches take two or more cases that are very similar except for a few key characteristics and compares them. The cases generally differ over whether an intervention occurred or not, though they could also differ in the type or intensity of an intervention. The aim of the comparison is to observe what effects are visible and how they are similar or different, or what causal pathways can be discerned in each case.

The most prominent strategy used by practitioners for establishing attribution through comparison is the use of methods based on **counterfactuals** like experimental and quasi-experimental methods. Experimental and quasi-experimental methods are useful in that they allow one to understand what would have happened without an intervention. Furthermore, they are respected by many donors as being rigorous methodologies. They are, therefore, valuable for proving results and satisfying the demand for accountability.

Some of the key informants we interviewed and also parts of the literature, critique the use of experimental methods for this very reason: one interviewee argued that many experimental designs, such as **randomised controlled trials (RCT)**[5], do not allow programmes to assess *why* their interventions are creating the changes they are. This knowledge is crucial to improve and learn – for the programme itself, and for other programmes looking to try similar programme or intervention designs. This critique is also widely reflected in the evaluation literature. Stern et al. (2012) make the point that the "why" is not only important for learning, but also for attribution itself: "C*ounterfactuals associate a single cause with a given effect without providing information on what happens in-between: how the effect is produced. This information can be important for attribution purposes because knowing only the beginning and the end of the causal process might not be sufficiently fine-grained in some situations"* (Stern et al., 2012:7).

While not explicitly mentioned by interviewed practitioners, Stern et al. (2012) stress that there are comparative approaches that do not depend on counterfactuals. For example, configurational methods use different cases and compare their similarities and differences, and qualitative comparisons between different cases are also used to understand aspects of attribution.

Prospective approaches

Practitioners interviewed for this paper also reported using prospective approaches for establishing causality. Prospective approaches start off by developing a theory of how an intervention aims to achieve a change. This often takes the form of logic models (**results chains**[6], logframes) or Theories of Change. Once the intervention has started, changes at each step of the causal chain are validated to show the plausibility of the causal links between each step[7].

5 For more details on this method see: http://bit.ly/2moNWYE
6 For guidance on using results chains see the BEAM Guidance on Monitoring
7 This logic is codified in a widely applied method called Contribution Analysis, developed by Mayne (2008).

Interviewees reported that the value of prospective approaches is that they help programmes to think through what changes they expect to see. This is also useful for intervention design and management. Further, these approaches can be used to determine whether an intervention is on the wrong track, or if a change is not due to an intervention, by flagging which links within a causal chain have or have not been realised.

The key criticism of prospective approaches from interviewed practitioners was that they do not capture unexpected changes. Interventions in complex systems generally lead to unexpected effects, so this is a crucial omission. For this reason, practitioners suggested that prospective approaches should be paired with retrospective approaches that can capture these effects (see below). Prospective approaches, particularly if they apply a linear causal logic, are also criticised for not being equipped to deal with the **emergent**[8] nature of causality in complex systems. They are furthermore criticised because they give the impression that change is predictable.

In reaction to this criticism, the key informants argued that it is not the prospective tools that are problematic, but rather the way in which programmes use them. Theories should not be fixed blueprints, but should rather be continuously evaluated and adjusted using an iterative and adaptive approach.

Retrospective approaches and assessment of external causes

Retrospective attribution approaches look at changes that have occurred and trace back their causal factors to identify which factors contributed to the change. An example of a tool used by interviewees that uses retrospective logic is **outcome harvesting**[9]. When using this tool one starts by scanning the market system for the intended or unintended consequences of an intervention. Then one attempts to verify whether these changes were caused by a programme intervention.

The practitioners interviewed highlighted the value of retrospective approaches in capturing external factors that contribute to the changes being measured. This is critical for programmes given the multiplicity of causal influences in the complex market systems they operate in.

Interviewees, however, reflected that retrospective approaches are often primarily qualitative. They are also susceptible to **bias from overconfidence**[10], and it can be very difficult to link changes back to the intervention. Where these approaches are quantitative, they often involve very large surveys and can be difficult and expensive to apply.

3.3 Political and practical constraints to measuring attribution

Political economy

> *"There is a gap between what [some practitioners] think is true – that attribution is fanciful – and what the politicians and donors want to hear."*[11]

The political economy surrounding market systems approaches – and development in general – has a large effect on how programmes tackle the challenges of measuring attribution. The opinion of one practitioner interviewed was that the "hangover" from previous direct delivery approaches means that there is both a lack of understanding of how causality works in a market system as well as unrealistic expectations of how attribution should be approached.

A recurrent theme from interviewed practitioners was how the challenges of measuring attribution are directly influenced by donors' accountability requirements. It was pointed out that accepted industry standards on measuring attribution are influenced by the expectations of those to whom programmes report, whether they be government donors, philanthropists or taxpayers. One interviewee commented that it would be very difficult to go to the US Congress and explain to them that it is not possible to measure attribution in some contexts.

8 For an explanation of emergence, see Appendix 2: Complexity and emergence
9 For more details on outcome harvesting see: http://bit.ly/2ludeqZ
10 For details on this bias see: http://bit.ly/2mblcqt
11 Key informant interviewed for this paper

As he put it: “Their eyes would just glaze over.” Some interviewed practitioners see this restriction as a constraint to developing best practice methods for measuring systemic change.

They view it as restricting movement away from methods that are currently deemed to be rigorous in a generic sense – mainly quantitative methods using counterfactuals – towards new methods that better fit market development programmes and their contexts, while still being technically rigorous.

An additional point raised was how the resource restrictions on programmes’ monitoring and evaluating components constrain adequate attribution measurement. One key informant complained that some donors, “just want numbers, but do not want to spend the money.”

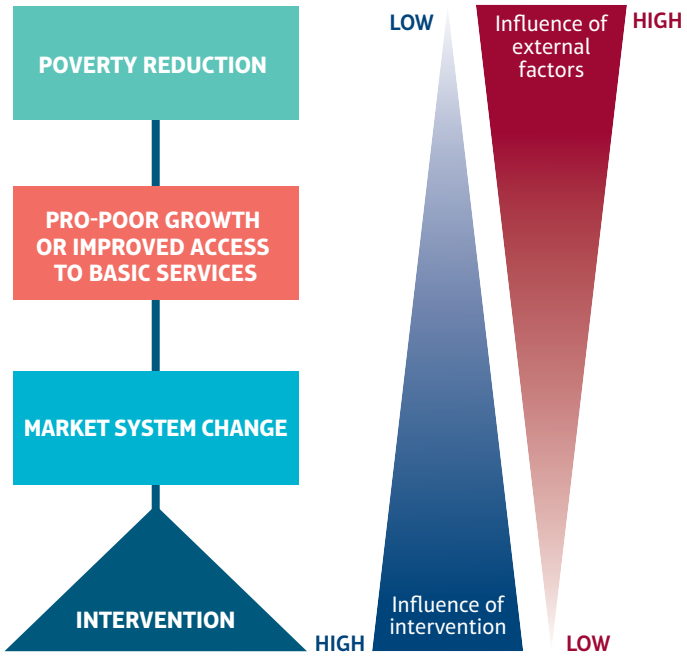
Interviewees raised the same concerns with regard to implementing agencies. One informant commented that, “actually, just as much resistance to rigorous attribution methodologies comes from implementers, who do not want to spend the resources and prefer light touch approaches.” He related this to the perverse incentives created by contracting structures such as payment-by-results contracts, which can disincentivise investments in progressively developing attribution strategies.

Another key informant, however, made the point that it is irresponsible to claim that market systems are too complex to measure attribution and that programmes should be designed in a way that the effects can be measured using counterfactual methods. He argued that it should not be a hard sell to convince donors to give programmes the necessary resources to rigorously test interventions.

A multitude of causal factors

Many factors influence the outcomes which programmes using market systems approaches are interested in, such as behavioural changes or changes in income and employment. With the increase in causal distance, i.e. the number of causal steps from an intervention to the observed effect, and scale of an anticipated outcome, the multiplicity of other factors influencing that outcome also increases. At the same time, the relative influence of the intervention on the outcome decreases. This is illustrated in *Figure 2*.

Figure 2: Causality between interventions and objectives and external influences



Source: Adapted from Ruffer and Wach (2013), The Springfield Centre (2014)

Both the academic literature and interviewed practitioners highlight causal density, i.e. the high number of other factors that have a causal influence on an outcome, as one of the core reasons for the challenges faced in establishing causality. This is particularly true for interventions that aim at catalysing wider system change. One key informant put it this way: “Life is not a lab.” Measuring changes in a market does not involve a finite number of causes that we can control.

Inability to predict causal chains

A common view revealed from the interviews was that causal chains in complex contexts are unpredictable. One practitioner said that we are incapable of either knowing or understanding all the causal factors in a market system. Even having complete knowledge of all the factors influencing an outcome would not be enough to build a clear causal chain. This is because causality in complex systems is emergent and therefore change does not occur along a chain of causally connected events (see Appendix 2: Complexity and emergence).

Consequently, the causal models that are developed at the beginning of a programme hardly ever match reality. Interviewees explained that this means that not all measurement can be planned in advance. This makes establishing control groups difficult (see below). Furthermore, results chains and Theories of Change can often over-state causal links between interventions and anticipated changes because they do not show other influences besides the programme interventions. We do not have the knowledge to predict all the other causal factors that contribute to the change.

One key informant takes a pragmatic view: he argues that it is less the tools (results chains and Theories of Change) that are the problem, but more how they are used. If used well, a results chain helps to provide a “best guess” prediction of anticipated changes that is not set in stone, but is continuously revised. If used this way, these models still have management and measurement value.

Constructing control groups

In efforts to apply counterfactual approaches to measuring attribution, many interviewed practitioners reported they find the construction of control groups in market systems challenging. Challenges include contamination, self-selection bias, changes in programme strategies that can make group set-ups useless, and the time frames of market system changes. The fact that many practitioners see this challenge as central in establishing attribution shows that counterfactual methods play a central role in their practice.

Proponents of counterfactual designs resist these positions. One key informant argued that there are statistical methods to overcome some of these challenges. Where there are no mechanisms to limit access to the treatment he suggests using methods such as **encouragement designs**[12] to prompt random differential take-up. He also suggests using **cluster analysis**[13] to overcome **self-selection bias**[14] and multiplicity of causal factors. Where effects are non-linear, for example because of diminishing returns or threshold effects, he suggests using statistical methods to compensate for the challenges.

12 For information on and an application of encouragement designs see Bradlow (1998)
13 For information on cluster analysis see: http://bit.ly/2lOGae7
14 For an explanation of self-selection bias see: http://bit.ly/2mbyu6r

4. Causality in the evaluation literature

In Section 2 we defined what we mean by attribution. Attribution depends on establishing causality, so how we understand causality determines how we approach attribution. In this section, we explore three different theories of causality from the evaluation literature.

Pawson (2007)[15] differentiates between three types of causal explanation**:**

- **Successionist causality:** Causality as statistical regularity, independent variables affecting dependent variables.
- **Configurational causality:** Causality as a combination of factors working together in a cumulative way.
- **Generative causality:** Causality as part of the whole system that is irreducible to the parts.

We will look at each of these in turn.

4.1 Successionist causality

> Successionists locate and identify vital causal agents as "variables" or "treatments". Research seeks to observe the association between such variables by means of surveys or experimental trials. Explanation is a matter of distinguishing between associations that are real or direct (as opposed to spurious and indirect) and of providing an estimate of the size and significance of the observed causal influence(s). (Pawson, 2007:1)

In a successionist perspective of causality, **variables** do the explanatory work. Each variable describes a single cause that has a specific effect. The effect that is observed or intended is defined as the "dependent variable" (Y). The research is about finding "independent variables" (X) that have an effect on the dependent variable. Multiple independent variables (X_1, X_2, X_3, …) can have independent contributions to an effect.

There are two ways to infer causality in successionist models. Causality is inferred either by:

1. How regularly cause and effect occur together, holding all other factors constant (correlational logic), or
2. The difference between two events, where everything but the cause is the same (counterfactual logic) (Befani, 2012).

In successionist models it is implied that more of the independent variable automatically leads to more of the dependent variable. However, as White (2009a) points out, this relationship does not have to be linear. Indeed, statistical methods can cope with various relationships between the independent and dependent variables. This includes diminishing returns, threshold effects as well as multi-variate causality.

In the real world, multiple factors usually influence one effect. To isolate the specific effect of one variable on another, researchers use either experimental or quasi-experimental designs or statistical means (causal modelling).

Researchers who use successionist approaches develop hypotheses on how variables could be connected. They then collect data using appropriate measurement tools and field work techniques to prove or disprove the hypotheses.

Experimental methods require **ex-ante designs** while quasi-experimental methods and

15 Quoted with kind permission of the author.

statistical modeling can be applied **ex-post.**[16] Data quality is essential and depends on the appropriateness of measurement tools and field work.

4.2 Configurational causality

> Configurationists begin with a number of "cases" of a particular family of social phenomenon [sic], which have some similarities and some differences. They locate causal powers in the "combination" of attributes of these cases, with a particular grouping of attributes leading to one outcome and a further grouping linked to another. The goal of research is to unravel the key configurational clusters of properties underpinning the cases and which thus are able to explain variations in outcomes across the family. (Pawson, 2007:1)

Instead of variables, the basic unit of research in configurational analysis is the **condition**. Combinations, or so-called configurations, of conditions have causal power. Like variables, conditions are identifiable, observable and measurable. Yet they interact with each other and with the context, rather than having their own independent causal power as in the case of successionist variables.

Befani (2012:11) differentiates between two different types of condition: "ground-preparing causes" that build a background against which "immediate causes" can be effective. Neither the ground-preparing causes nor the immediate causes are effective without the other. Befani uses the example of combustible fumes as ground-preparing causes and the lighting of a cigarette as an immediate cause: only when both are present would the effect of a fire occur. In technical language, both conditions are *necessary,* but on their own they are not *sufficient* causes. Hence configurationist research is about finding the specific combination of conditions that are *jointly sufficient* to cause an effect.

To make a causal claim, the cause needs to satisfy various conditions of necessity and sufficiency, in particular (Befani, 2016:13):

- whether an intervention, or other factors like for example specific contextual or historical conditions, are required to achieve a certain effect or if the effect can be achieved without them; and
- whether the intervention is good enough by itself to produce the outcome or needs other factors and components.

Configurationist research is based on identifying cases that are equivalent but not identical and observing the similarities and differences between their conditions. The goal of the analysis is to figure out which combination of conditions has causal powers over the outcome that is observed in all cases.

4.3 Generative causality

> Generativists, too, begin with measurable patterns and uniformities. It is assumed that these are brought about by the action of some underlying "mechanism." Mechanisms are not variables or attributes and thus not always directly measurable. They are processes describing the human actions that have led to the uniformity. Because they depend on this choice making capacity of individuals and groups, the emergence of social uniformities is always highly conditional. Causal explanation is thus a matter of producing theories of the mechanisms that explain both the presence and absence of the "uniformity." (Pawson, 2007:1)

16 For an explanation of ex-ante and ex-post designs see page 20 of http://bit.ly/2luA5m9

A generative approach to causality does not rely on identifying one variable or a set of conditions. It recognises that causal powers can be emergent and may not be reducible to a single cause or even a combination of causes. Instead, it tries to describe the overall association of three things; a specific context (C), a specific mechanism (M) and an observed outcome pattern (O) in that context.

Causal explanations in generative approaches take the form of propositions of what mechanisms (M) in what contexts (C) lead to what outcome patterns (O). These “CMO propositions” are first hypothesised and then empirically tested and revised in an iterative way to reflect observations. To revise and fine tune CMO propositions, evaluators use empirical observations, both quantitative and qualitative. These can include comparisons to strengthen a hypothesised proposition.

Contexts (C): Contexts are the institutional, organisational and social conditions. There are usually multiple mechanisms present in a context. Mechanisms often need to be triggered by an event or an intervention. This means that in any system, some mechanisms are “active” (the ones that have been triggered), others are only potential (the ones that are yet to be triggered). Multiple mechanisms can operate at the same time, and even compete, leading to combined effects.

Mechanisms (M): In social systems, mechanisms are constituted by the choices people make, individually and collectively. These choices depend on what is possible, what resources are available to implement them, and what decision-making mechanisms are in place if choices are contested. This implies that mechanisms are often not visible, yet they can be investigated.

Choices, however, are always implemented in a certain context. This context strongly influences the shape of the outcome patterns. The form of the outcome pattern is therefore different to each combination of a specific mechanism and a specific context.

In contrast to successionist logic, proponents of generative methods assert that, “what causes something to happen has nothing to do with the number of times we observe it happening” (Pawson, 2007:14). Rather than searching for regularities and repeat occurrences of variables, generative approaches look for associations of context, mechanism and outcome in each case.

4.4 Comparison of the three types of causal explanation

Table 1 brings these three theoretical views on causality together. It compares philosophical and theoretical underpinnings, typical causal questions the approach is best-fit to answer, overall approach to establishing causality, extent to which the causal mechanism is described by the approach, and typical methods used to determine attribution.

Table 1: Overview of approaches to causality and attribution.

	Successionist	Configurational	Generative
Theoretical perspective on causality	Theoretical perspective: Association between single cause and single effect	Theoretical perspective: Association between configurations of conditions and effects	Theoretical perspective: Description of causal mechanism
Causal questions	*To what extent did the intervention make a difference?* Did the intervention cause the effect? How much is the net effect of the intervention?	*Did the intervention make a difference and through which patterns?* What configurations of factors are necessary and/or sufficient to cause the effect?	*How did the intervention make a difference?* How was the effect produced? How did it come about?
Approach to establish causality	*Correlational*: establishing the frequency of associations between an intervention and an outcome, holding all other factors constant *Counterfactual*: establishing the difference between the case and another case where all other factors except the intervention were the same	Examining which conditions combine and interact to produce an outcome, which also depends on the context	Investigating the different mechanisms at play in a particular context which produce an outcome
Description of causal mechanism	None	Only basic ingredients are described: conditions, their combinations and disjunctions	In-depth
Methods for determining attribution	Statistical approaches, experiments, quasi-experiments	Qualitative Comparative Analysis (QCA) Decision tree learning	Realist approaches Process tracing Contribution analysis Outcome harvesting

Source: Authors' own compilation, drawing among others from Schatz and Welle (2016) and Befani and Mayne (2014)

5. Discussion

Through key informant interviews, we explored the experiences and challenges of measuring attribution in practice, which gave us some insight into how the different concepts are applied and what real world constraints practitioners face. The literature review provided an understanding of different views on causality and attribution. This section reflects on these two perspectives together, and looks at implications for future practice.

The successionist paradigm still strongly influences the thinking on attribution in monitoring and evaluation as can be seen from the key informant interviews summarised in Section 3. Counterfactual methods are, implicitly or explicitly, seen as the 'gold standard' to establishing attribution. Practitioners said they experience continuous pressure to use successionist causality and the related counterfactual models, particularly experimental methods. Establishing control groups is generally required to apply successionist approaches to measuring attribution.

However many interviewed practitioners expressed concern that these methods do not fit the realities of the market systems they work with. This view is also reflected in recent reviews of evaluation methods for market systems approaches (Fowler & Dunn, 2014; Ruffer & Wach, 2013), as well as in discussions in the evaluation community (Befani *et al.*, 2015; Befani *et al.*, 2014; Stern *et al.*, 2012). The evaluation scholars also make it clear that there are alternatives to successionist methods that can achieve the same level of rigour.

While there is pressure to use successionist methods for attribution, interviewed practitioners continue to explore approaches to both monitoring and evaluation based on a generative understanding of causality. This has, for example, led to making **theory-based approaches**[17] using Theories of Change, the norm in monitoring and evaluation in market systems development. Theories of Change are an inherently generative tool as they try to explain the "how" of the change process in a step-by-step manner (Pawson, 2007).

Hence the case for moving beyond a purely successionist approach to attribution, both in theory and practice, has clearly been made, though has not as-yet been explicitly stated in guidance and practice.

The different approaches applied to measuring causality, however, are sometimes seen as representing fundamentally different world views, which are difficult to reconcile. The successionist and the generative paradigms often view the other as unnecessary, or of very limited use. The configurational approach somehow occupies a position in-between. It is used to complement a successionist view and make it more adaptive to complexity, or it is used to make generative accounts more quantitative and amenable to statistical analysis.

Essentially, there is no ultimate correct answer to the question of what the best way is to attribute cause and effect. Too much depends on the world view that is adopted by the people involved, particularly those with the power to mandate certain approaches.

It is not the aim of this paper to revisit the arguments of both sides and take a stance on either. Rather, we take a pragmatic approach that will lead us to useful and practicable advice for practitioners who do not have the capacities, resources, or indeed, the liberty to be dogmatic about the adopted approach.

From that pragmatic point of view, we can make three observations:

- **A generative approach is better able to explain how the causal mechanisms work** (the "why it worked" question).

17 More information can be found on these approaches at: http://bit.ly/2mbpiP5

- **A successionist approach is better at achieving a higher level of confidence in showing co-occurrence of cause and effect, and the size of the net effect** (the "if and to what extend it worked" question). Although, the validity obviously depends on the ability of the researchers to establish the necessary conditions for the methods to work properly.
- **Perfect attribution in complex contexts is difficult (if not impossible).** Referring to Befani *et al.* (2014), we suggest that we need to shift from "assessing attribution" to "assessing confidence about attribution" (or contribution if one prefers that term). Or, to use the words of some of the key informants: "we should not let the perfect be the enemy of the good."

6. A pragmatic approach to attribution in market systems: recomendations

Building on the discussion above, this section provides a typology of attribution strategies adapted to different contexts and using different perspectives. This typology aims to help practitioners to make a pragmatic choice of attribution strategies for their various interventions and achieve a reasonable degree of credibility and validity in their monitoring and evaluation practice.

6.1 A typology of attribution methods

The first dimension of the suggested typology reflects the three dominant ways of understanding causality; successionist, configurational, and generative causality. At the risk of over-simplification, we can say that these world views reveal a spectrum rather than a dichotomy. On one side, cause-and-effect relationships are ***ordered*** and each cause and its specific effect can be isolated and quantified. On the other side, causal influence occurs at various emergent levels inside the system. To differentiate from the ordered side, we will call this side ***complex***[18] (Figure 3).

Figure 3: The spectrum from ordered to complex extends the different world views on causality

The successionist view of the world is located on the ordered side. In ordered systems, the whole equals the sum of the parts. All causal powers of individual variables can be added up to show the whole causal effect. In contrast, the generative view sees the world as complex, with change being emergent and unpredictable. In an emergent system, the whole is greater than the sum of the parts. Interactions between variables can lead to results that are greater than the sum of the causal powers of the individual variables. The configurational view sits somewhere in-between, requiring enough order to compare various configurations of conditions, while also considering that there is interaction between conditions.

In reality, market systems development programmes and their interventions can be situated across the spectrum. Not everything is complex, even in a market system. Some aspects are ordered and predictable for example procedures for storing produce in shared storage facilities. Others are unpredictable and emergent, such as behaviours of market players.

18 There are many definitions of complex. We use this term as it is easily recognised, and understand complex as a situation where cause-and-effect relationships are emergent and the system is dynamic and evolving. Kurtz and Snowden (2003) also use the term 'emergent order' for the second side.

It is therefore necessary for practitioners to understand where on the spectrum their interventions are situated, and what method can be used to answer the relevant causal questions. A particular programme can, therefore, be required to consider causality from a range of perspectives, depending on the part of the programme being considered.

The second dimension of the typology is based on the different approaches to attribution described in Section 3.2. Monitoring and evaluation frameworks that can establish credible attribution claims should comprise aspects of the following four perspectives:

- A **prospective** perspective that looks forward and describes what an intervention intends to achieve and how. Along the way, theories need to be updated based on new insights and evidence.
- A **retrospective** perspective that is based on observed changes and attempts to reconcile these with the interventions as well as other influencing factors.
- A **comparative** perspective of cases that are similar enough to give insights into the size of the effect and the functioning of the causal mechanism, ideally both.
- An **integrative** perspective that brings different perspectives together and adds **alternative causes** that were observed or assumed to have potentially influenced/ contributed to the outcome and when they occurred in relation to the programme interventions.

In Table 2 we map out these four elements and the context they are most suitable for. This table is intended to help practitioners chose attribution strategies that are adapted to their interventions. It provides guidance on selecting methods which match the programme's context. It can also be used as a basis to discuss such strategies with donors.

Table 2: Typology of attribution approaches along the two dimensions of ordered-complex and the four perspectives

	Complex ←----------→ Ordered		
	Generative	**Configurational**	**Successionist**
When does it work?	• Dynamic contexts • High number of interacting causal factors ("high causal density") • Long causal chains linking interventions with intended outcomes and impacts • Causal models need to be adaptive and change over time • Emergence and dispositionality (see appendices 1 and 2) do not allow for linear causal chains • Low certainty about working solutions and low levels of agreement among stakeholders • Repeated occurrences of a factor or combination of factors do not always create the same effect	• Fairly stable contexts • Causal influences not independently sufficient • Causal links between specific groupings of causal agents and anticipated effects • At least 3-5 reasonably similar cases to be compared	• Highly ordered contexts with stable cause-effect relations • Direct causal relation or reasonably short and non-controversial causal chain • Interventions focused on changing one or a few discrete aspects that can be measured • Control group set up or alternative counterfactual feasible • There may be multiple causal influences, but they are all independent of one another and can be controlled for

Example	Alliances Lesser Caucuses Programme (ALCP) application of Outcome Harvesting to identify unintended outcomes of ALCP's dairy interventions	Use of a configurational analysis as part of the Participatory Impact Assessment & Learning Approach (PIALA) for evaluating commodity chain development in Ghana	BIF RCT measuring how dehumanisation and industrial engineering skills cause changes in garment factory profits and employee welfare
Prospective	Use of Theory of Change, causal mechanisms or **CMO propositions** (see Section 4.3) to describe and subsequently refine potential mechanisms that could explain the change	Use of hypotheses to collect conditions to compare	Use of hypotheses about causal relations to establish relevant influences and variables to test or control for
Retrospective	Iterative approach to updating the theory based on data collection and analysis based on **Bayesian probability logic** (see Appendix 3)	Iterative approach of updating the theory based on data collection and analysis; refining, adding or taking away conditions and cases, depending on new insights that emerge during the analysis	Use of quasi-experimental methods applying different statistical mechanisms like **propensity score matching** to establish comparison groups ex-post
Comparison	Pragmatic use of comparison data where available and useful to add to the understanding of the causal mechanisms	Comparison between multiple selected cases, comparing different combinations of conditions	Different ways of constructing a counterfactual. Most common is the comparison between a treatment group and a control group in experimental and quasi-experimental approaches
Alternative causes	Generative enquiries generate rich narratives of mechanisms and contextual factors that work together to create a specific outcome, which includes possible alternative causes	Rejecting alternative causes by showing that they are not necessarily part of any jointly sufficient set of conditions	Rejecting alternative causes by statistically controlling for their influence

The choice of which of these elements should be applied in measuring attribution depends on a number of factors relating to the programme (and components within a programme) being considered. These include:

- type of **context** their interventions are implemented in
- monitoring or evaluation **questions** that need to be answered
- human and financial **resources** that are available for monitoring or evaluation activities and staffing
- **capacity of the staff** that will implement the activities
- **approaches and methods** that are appropriate to answer these questions in the given context with the given resources
- **time available** before reporting on results is required

6.2 Applying the typology to market systems development

In this section, we discuss how the typology discussed above could be applied in market systems approaches. We base this section on the programme examples given in Table 2.

Ordered contexts

In **ordered contexts**, causal chains are **relatively short** and the context is **stable enough** so the causal relations are likely to remain unchanged over the duration of the intervention. An example of a short causal chain is given in Figure 4. In market systems approaches this is usually only the case when direct effects of interventions are observed on the lowest level of a programme's Theory of Change, even before expecting market systems change to occur.

Figure 4: An example of ordered cause and effect relations

In these situations, a **successionist approach** can prove to be effective. The attribution strategy could be developed as follows:

Prospective: A results chain is developed as shown in Figure 4. Assumptions are added as to why these causal links are believed to be valid. For example, it can be assumed that businesses that receive direct advice on working conditions will implement this advice, since they understand that this could improve their revenues. A way to measure whether the hypothesised results chain holds true is to collect quantitative data from the companies on productivity. Qualitative data could comprise information about company regulations on working conditions and the perspective of the company's staff on what has changed in their working reality. The attribution claim is supported if this data collection shows that the working conditions and the productivity in those companies that were given advice had improved.

Retrospective: A retrospective approach would involve scanning the market for companies that seem to perform better than others and assess why. If the scan shows that many of the companies that perform better have better working conditions, this supports the causal claim between working conditions and improved productivity. If these companies are the ones that received advice from the programme, this supports the attribution claim by the programme. If the number of companies is high enough, the collected data can be collected in a quasi-experimental way by retrospectively assigning companies to a treatment and a comparison group.

Comparative: Comparison can be made between different companies. In this situation, this can be done by using a quasi-experimental or experimental method, depending on the way the intervention is designed. If the companies that receive the treatment are chosen by the programme, an experimental method may be possible. If they self-select into receiving the advice, a quasi-experimental approach works better – which is then the same approach as described above in the retrospective perspective.

Alternative causes: In this context, the number of alternative causes is relatively easy to assess. The people who make decisions about changes in the company, including working conditions, is small. These people can be questioned about what changes were implemented and where the ideas for these changes come from. If there are a sufficient number of companies, alternative causes can be controlled for statistically. If it is shown that the alternative causes did not have a significant impact, this strengthens the attribution claim.

By integrating both qualitative and quantitative elements from these different perspectives of causality and attribution strategy design, we can develop a rich picture of change and attribution that tells the story from what the situation looked like before the intervention started to the picture at the end. This story can include when and what was done, what changes were observed and why these changes are likely to have been caused by the intervention.

Case Study: The Business Innovation Facility (BIF) programme implemented a Randomised Controlled Trial (RCT) in the Myanmar garments market to study the effect of their interventions. BIF is working with 15 garment factories, providing HR and productivity training to support the businesses to improve their practices, increase efficiency and improve their positioning for supplying larger buyers in the international market. The RCT seeks to identify exactly how training affects worker welfare and productivity, and whether there is strong enough evidence to prove a link between the two.

Complex contexts

In complex contexts, **causal density is high** and **causal chains are long** and intricate. In market systems approaches, we find this is virtually always the case when attempting to attribute change at the market systems change level or 'higher level' changes (such as growth or access to basic services, and poverty reduction) – to an intervention. In market systems approaches, we find this is virtually always the case when attempting to attribute change at the market systems change results level, and 'above' to changes in growth or access to basic services, and poverty reduction.

In these situations, **a generative approach is most appropriate**. The different elements of the attribution strategy could look as follows:

Prospective: The programme will develop a Theory of Change (ToC) and results chains. From the generative world view, ToCs are used in a different way to successionist approaches. ToCs do not comprise fixed hypotheses that are tested over the implementation period of an intervention. Instead, they are living documents that are used by the programme team to make sense of what is going on and to keep track of their growing understanding. This understanding is generated by continuous learning through multiple channels. These can include informal channels, such as day-to-day observations and conversations, or more formal channels, such as monitoring efforts to measure different aspects of the ToC. The measurement framework that is applied could look quite similar to a framework applied in a successionist model. The use of a mix of qualitative and quantitative methods is, however, more critical. The emphasis of the measurement efforts is on learning, and updating the understanding about the context and effects.

Retrospective: Compared to the successionist world view, the retrospective element becomes more important in this strategy. The reason for this is that in complex contexts a ToC cannot capture all aspects of a situation. The predicted causal chain is only one among many possible causal chains, and there are many other influencing factors at play. Appropriate measurement will comprise mixed methods, and some methods will most likely be used for both prospective and retrospective analysis, such as field observations and diaries. More formal approaches to retrospective assessment can be used, such as Outcome Harvesting.

Comparative: Comparison in a generative context does not have the same central role as it does in successionist or configurational settings. Having said that, data generated by counterfactual or configurational methods can of course also be used in a situation where a generative attribution strategy is most appropriate, to add to the overall picture. In addition to systematic comparisons, more qualitative comparisons can also be used, for example through observation of different cases.

Alternative causes: Generative enquiries generate rich narratives of mechanisms and contextual factors that work together to create a specific outcome. The assessment of alternative causes is therefore inherent in generative enquiries.

As with alternative causes, Bayesian logic is inherent in generative enquiries. Each observation and all additional data are continuously used to update and refine one's understanding of a situation to enable it be described in as much detail as possible.

Case Study: The Alliances Lesser Caucasus Programme (ALCP) is a programme using a market systems approach. ALCP's monitoring and results measurement framework is built on and audited against the DCED Standard for Results Measurement. Its attribution strategy was initially based on a successionist world view on causality stipulated by the standard. The main attribution strategy is to validate causal links along a linear results chain. The programme leadership team has recognised that this approach was limiting. For example, it makes it difficult for the project to recognise unintended consequences and make a strong attribution claim for changes that were not originally foreseen in its results chains. Thus, the team has has recently added new components to this strategy, using generative logic to strengthen its attribution claims. The programme has trialled Outcome Harvesting. By employing this retrospective method, the programme was able to find a number of systemic changes in the region in which the programme is active and to link these to programme interventions. This information enhanced the findings from the DCED Standard-audited MRM system, leading to a richer picture of ALCP's achievements.

6.3 Choosing the most appropriate methods and increasing confidence

While arguing for a pragmatic approach, we still think that there should be a minimum level of quality for attribution strategies. This plays out on two levels.

Firstly, an agreement on a minimum level of quality in monitoring, evaluation and attribution work needs to be generally agreed for the market systems community. There are different ways of developing such an agreement. Two examples are the use of auditable standards or broader conventions. The widely used DCED Standard provides guidance on estimating attributable changes in monitoring and results measurement (Sen, 2016). This guidance, however, narrowly advocates an approach to measuring attribution that is inspired by a successionist world view. An alternative to a standard would be a broader convention which provides options for best-fit practices, such as Better Evaluation[19] for the evaluation field. Better Evaluation was developed by a group of universities, think tanks, and an accreditation body. It is an interactive website combining experience in applying best-fit and learning from less suitable practice, with a case-based approach to facilitate practitioners to decide on what is accepted as "good enough" for a particular programme's operating context and design.

Standards are very good instruments to improve general practice from a low level to an acceptable level. The DCED Standard is a good example for this as it has clearly improved MRM practice in market systems development. The challenge with standards, however, can be that they tend to drive the dominant world view and discount alternative approaches, and are often slow to adapt to new developments in our understanding. Conventions, on the other hand, are more flexible and respond better to emerging ideas and contested world views.

Secondly, field practitioners need to make pragmatic choices of methods they will employ to attribute change to interventions of their programmes. Implementing a methodological framework that best-fits the demands of the context requires investment in the right resources. Programmes need access to staff and advisers who have a sufficiently high level of technical capacity to design, plan and implement measurement frameworks, analyse data and report on results.

Additionally, not all programmes can implement the most rigorous of all methods. Programmes usually find that they have limited financial resources (budget) available for monitoring and evaluation. In addition, staff capacity is often limited as most programmes cannot hire academics to do rigorous research, or cannot find access to the right advice at the right time. Hence, it is important to understand the compromises made, and the limitations of an attribution strategy, while being transparent about the choices made.

19 http://betterevaluation.org/

To reach an exact answer as to whether an intervention contributed to a change, and the size of the net change it created, is very resource intensive in market systems development (if it is at all possible). By using several different methods instead, including participatory methods, we can develop a rich picture of the scenario and the changes observed over time. Instead of attempting to prove attribution beyond any doubt, each additional measurement or observation increases our confidence in the attribution claim. This logic of continuously increasing the confidence in a causal claim is called Bayesian updating.

It is also important to keep in mind that not everything is complex, even in a market system. Certain aspects of market systems can be fairly ordered, for example procedures for storing produce in shared storage facilities. An assessment of the context and the selection of the relevant method needs to be made for each intervention.

7. Conclusion

In this paper, we have attempted to shed light on the concept of attribution and its application in market systems approaches. Attribution is the establishment of a causal link between an observed effect and a specific intervention. We looked at how attribution is done in practice and what challenges practitioner faced. We also looked at the evaluation literature and provided a theoretical perspective on causality and attribution.

Bringing the practical and theoretical perspectives together, we have found that there is a legitimate spectrum of views on causality, spanning from successionist to generative positions. Still, successionist views dominate the discussion among practitioners. Practitioners report that donors continue to ask for methods that are rooted in a successionist view, such as experimental or quasi experimental methods. However, we show in this paper that the case for moving beyond a purely successionist approach to attribution is strong and has been made, both in theory and in practice.

We developed a typology based on different views on causality and requirements of different contexts to support practitioners in developing their attribution strategy. When developing an attribution strategy, practitioners need to be aware that perfect attribution in complex contexts is very difficult, if not impossible. We therefore advocate for a pragmatic approach of “best-fit” over “best practice”. Selecting attribution strategies in any context depends on available resources for monitoring and evaluation as well as the technical capacity of the staff. To still achieve a high level of confidence in results and attribution claims, it is important to be transparent about the logic and decisions that led to the development of an attribution strategy. Both standards and broader conventions which provide options for best-fit practices can be used to improve the credibility of these claims.

Programmes using market systems approaches should aim to increase the credibility of their attribution claims by consciously and transparently selecting best-fit methods that are in accordance with their specific context. There is no one-size fits all recipe for attribution.

Bibliography

BEFANI, B. 2012. Models of Causality and Causal Inference.

BEFANI, B. 2016. Pathways to Change: Evaluating Development Interventions With Qualitative Comparative Analysis. University of Surrey & University of East Anglia.

BEFANI, B., BARNETT, C. & STERN, E. 2014. Introduction - Rethinking Impact Evaluation for Development. *IDS Bulletin,* Vol. 45(6), pp. 1-5.

BEFANI, B. & MAYNE, J. 2014. Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation. *IDS Bulletin,* Vol. 45(6), pp. 17-36.

BEFANI, B., RAMALINGAM, B. & STERN, E. 2015. Introduction - Towards Systemic Approaches to Evaluation and Impact. *IDS Bulletin,* Vol. 46(1), pp. 1-6.

CUNNINGHAM, S. & JENAL, M. 2016. Rethinking systemic change: economic evolution and institution. Technical paper. The BEAM Exchange.

FOWLER, B. & DUNN, E. 2014. Evaluating Systems and Systemic Change for Inclusive Market Development. LEO Report: USAID.

JENAL, M. & CUNNINGHAM, S. 2015. Explore, Scale Up, Move Out: Three Phases to Managing Change under Conditions of Uncertainty. *IDS Bulletin,* Vol. 46(3), pp. 81-92.

JONES, H. 2011. Taking responsibility for complexity. How implementation can achieve results in the face of complex problems. Working Paper 330. Overseas Development Institute.

JUARRERO, A. 1999. Dynamics in Action: Intentional Behavior as a Complex System. Cambridge, Massachusetts; London, England: MIT Press.

KURTZ, C.F. & SNOWDEN, D.J. 2003. The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM Systems Journal,* Vol. 42(3), pp. 462-483.

MAYNE, J. 2008. Contribution analysis: An approach to exploring cause and effect. ILAC Brief No 16.

MUMFORD, S. & ANJUM, R.L. 2013. Causation: a very short introduction. First edition. Oxford; New York: Oxford University Press.

OECD. 2010. Glossary of Key Terms in Evaluation and Results Based Management. OECD. https://www.oecd.org/dac/2754804.pdf

PAWSON, R. 2007. Causality for Beginners. Unpublished Work.

POSTHUMUS, H. & WANITPHON, P. 2015. Measuring attribution: a practical framework to select appropriate attribution methods, with cases from ALCP Georgia, MDF in East Timor, Propcom Mai-Karfi in Nigeria and Samarth-NMDP in Nepal. DCED.

RAMALINGAM, B., JONES, H., REBA, T. & YOUNG, J. 2008. Exploring the science of complexity: Ideas and implications for development and humanitarian efforts. Working Paper 285. Overseas Development Institute.

RUFFER, T. & WACH, E. 2013. Review of M4P Evaluation Methods. Itad.

SCHATZ, F. & WELLE, K. 2016. Qualitative Comparative Analysis: A Valuable Approach to add to the Evaluator's Toolbox? Lessons from Recent Applications. CDI Practice Paper Number 13. Centre for Development Impact.

SEN, N. 2016. Guideline to the DCED Standard for Results Measurement: Estimating Attributable Changes. Donor Committee for Enterprise Development.

SNOWDEN, D.J. 2011. Good fences make good neighbors. *Information Knowledge Systems Management,* Vol. 10(1-4), pp. 135-150.

SNOWDEN, D.J. 2013. Systemic M&E Interview. Episode 1: Dave Snowden. In Jenal, M.Podcast. Systemic M&E Interviews. The SEEP Network.

STERN, E., STAME, N., MAYNE, J., FORSS, K., DAVIES, R. & BEFANI, B. 2012. Broadening the Range of Designs and Methods for Impact Evaluations. DFID Working Paper 38. Department for International Development. http://www.dfid.gov.uk/Documents/publications1/design-method-impact-eval.pdf

THE SPRINGFIELD CENTRE. 2014. The Operational Guide for the Making Markets Work for the Poor (M4P) Approach.

WHITE, H. 2009a. Some Reflections on Current Debates in Impact Evaluation. Working Paper 1. International Initiative for Impact Evaluation (3ie).

WHITE, H. 2009b. Theory-Based Impact Evaluation: Principles and Practice. Working Paper 3. International Initiative for Impact Evaluation (3ie).

WHITE, H. & PHILLIPS, D. 2012. Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework. Working Paper 15. International Initiative for Impact Evaluation (3ie).

Appendix 1: Different philosophies of causality

As we have seen, attribution depends on our ability to establish causality between an intervention and an observed effect. This paper first explores different philosophical views on causality, because "*we have to know what causation is before we can start to look for it"* (Mumford & Anjum, 2013:7).

One might ask why philosophy is important. Mumford and Anjum (2013:9) clarify: *"There is something elusive about causation that makes it a particularly difficult matter to know … the supposed causal connection is not itself part of our experience … we have to infer its presence from other factors of the situation*." Hence causality is rather a matter of belief, not of fact. There are thinkers, including Bertrand Russell for example, who wholly deny the existence of causality. Nevertheless, for this exploration of attribution we assume that causality is indeed a real thing with real implications for our work in market systems development.

In the following, we briefly introduce different philosophical views on causality[20].

Causality as regularity

David Hume was the philosopher who introduced the notion that you cannot actually see causal mechanisms. He postulated that only through observation of regularities can we capture causality. In its simplest form, his theory says that if one type of event is always followed by another, this allows us to believe that the first type of event caused the second. This view is called a constant conjunction account.

An important caveat to this view is that a regularity is only a constant conjunction if it occurs throughout all time and space. This of course makes it impossible for any human being to decide whether an observation indeed involves genuine causality as we cannot possibly observe all of time and space. This means that we cannot in effect differentiate between genuinely causally connected events and coincidentally increased frequencies of co-occurrences of two events. Nevertheless, regularity remains a central aspect of our everyday approach to causality.

Necessity, contingency or probability

A different view of causality assumes that for an event to cause another event, the second event must be more than just a mere possibility among others but a necessary consequence of the cause. This view is called necessitarianism. Necessitarianism does not allow any contingency in causal relations, and in its purest says that the world is entirely deterministic. This is the opposite to Hume's view, which allows complete contingency.

There is a variation on the necessitarian view that allows for more complex causal relations, namely that almost every effect could have several causes. We can, for example, find conditions that are necessary but not sufficient, which means that they could not cause a change on their own but are essential as part of a package to make the change occur. Conditions can be seen as non-necessary but sufficient if they can cause something on their own but also if other conditions exist that could cause the change.

As a reaction to necessitarianism, the notion of indeterministic causality emerged. This view sees a cause as something that increases the probability that something will happen while not ensuring it. This puts causality somewhere between necessitarianism and complete contingency.

Counterfactual dependence

Some scholars state that to know what caused something we need to find a "difference maker" and understand what would have happened if this difference maker had not been there – this is called a counterfactual dependence test.

From a philosophical point of view, this makes causality dependent not on something that is there, but on something that is not, which can be problematic. The question is whether the counter-fact depends purely on our imagination (called fictionalism), which is a rather weak basis

20 This overview draws heavily on the very short but invaluable introduction to causation by Mumford and Anjum (2013).

for an understanding of causality in the real world; or whether there are, as some philosophers postulate, an infinite number of parallel worlds which all differ in only one respect from each other.

Instead of relying on mere fictions or parallel worlds, there are instances in the real world where we can actually produce two situations that only differ by one fact. This idea is used for example in randomised controlled trials. It is, however, as Mumford and Anjum put it, "*one thing to use some sort of counterfactual dependence as a test of causation and another thing to say that causation consists in such dependence"* (2013:59). Nevertheless, the use of counterfactual dependence to test for causality is justified in principle.

Reductionism and emergence

Physicalicism sees the essence of causality in the transfer of a conserved physical quantity such as energy, momentum or electrical charge between cause and effect. While this notion of causality seems, attractive and is easily accessible, it has difficulties explaining complex phenomena like economies or human behaviour. There are two reactions to this.

Firstly, reductionism postulates that all higher-level phenomena can be explained in terms of lower-level phenomena. Hence according to this view, complex phenomena should also be reducible to some basic physical phenomena. This view is contrary to real-world observations. For example, while biochemistry can explain in very fundamental terms how the human body works, it fails to explain things like consciousness.

Secondly, the view called emergentism says that certain higher-level phenomena "emerge" at that level and are not merely the sum of their lower-level parts. This entails that certain causal characteristics only occur at that higher level. To use the same example, this view sees consciousness as emerging from interactions in the human body while not being reducible to the parts of the human body. At the same time, it allows consciousness to assume causal agency, for example over human behaviour, that the individual parts of the body could not assume.

Inherent causal powers

Instead of looking for causality as something separate from the involved entities or events, dispositionalism postulates that individual objects contain their own dispositional properties or "causal powers". These causal powers can exist as both manifested (e.g. the shape of an object) or unmanifested (e.g. if something is fragile it only manifests itself when it hits the floor). These powers can vary in strength and are highly context sensitive. A slight change in context can lead to a huge difference in outcome. This means that the effect of the causal powers varies both in the extent to which they manifest and whether they manifest at all. Or to quote Mumford and Anjum again: *"Causes tend or dispose towards their effects, where this is short of necessity but much more than pure contingency"* (2013:102).

Mumford and Anjum use smoking as an example. Smoking has causal power towards developing cancer, which means that those who smoke manifest a disposition towards cancer. This does not mean that everyone who smokes necessarily develops cancer. Cancer is a very complex causal phenomenon and just a small change in circumstances can make a difference to whether somebody in fact gets it or not.

There are two different views on how dispositions are manifested. Some think that causal powers need to be triggered by a "stimulus condition". Others instead think that dispositions form "mutual manifestation partnerships", producing together what neither could have produced alone.

In this section we briefly introduced several relevant philosophical views on causality. We can distil the following insights from these different views:

- Causality is more than pure regularity.
- Causes do not necessitate their effects nor are the effects completely contingent. Causes can be seen as propensities, increasing the probability of an effect to occur.
- Counterfactual dependence can be used to test for causality.

- A causal mechanism cannot be found in a reductionist way; indeed, some causal powers only reside in the whole that emerges from the parts.
- Causality does not reside outside interacting elements but is inherent in them in the form of dispositional properties or causal powers, which might only manifest when being stimulated or when combined with other causal powers.

Appendix 2: Complexity and emergence

The argument that social change is complex and that this has specific implications for attributing change is commonly expressed both in the academic literature and by practitioners. Aspects that are often mentioned when talking about complex systems are that they consist of a high number of interconnected and interdependent agents; that there are feedback processes that lead to non-linear dynamics; or that the agents are adaptive, i.e. that they adapt their strategy based on the feedback that they gather locally (Jenal & Cunningham, 2015; Jones, 2011; Ramalingam *et al.*, 2008). Two aspects that are mentioned most frequently in connection with attribution are emergence and the unpredictability of causal chains. While the latter is self-explanatory, the former needs more explanation.

Emergence is a process of the elements in a system self-organising into a qualitatively novel state of interrelation, and hence it is a higher-level order. Emergence occurs when previously uncorrelated elements or processes in a system suddenly become coordinated and interconnected (Juarrero, 1999). Juarrero (1999) uses the so-called Rayleigh–Bénard convection in a liquid that is heated as an example of emergence taken from a physical system. In the process of increasing heat from below, the molecules of the liquid suddenly and spontaneously self-organise and a regular macro-pattern of convection cells emerges. Also in human systems, spontaneous order emerges, leading to structures that enable more effective interactions between individuals, such as habits, social norms or the rule of law (Cunningham & Jenal, 2016).

As the example of social norms and the rule of law show, emergent structures can in and of themselves have causal powers. A social norm can cause somebody to do something (or not do something). This “internal” causal structure in complex human systems is the reason some scholars talk about complex human systems being dispositional (Snowden, 2011; Juarrero, 1999), i.e. having inherent causal powers that cannot be known in advance but influence the outcome of each intervention. These causal influences can also not be captured by variables or conditions.

Appendix 3: Paradigms for inferring causal connections

In Appendix 2 we described different philosophies of causality and different modes of causal explanation. In addition, there are different ways to make causal claims from data and to show the strength of inferred causal connections. The two most common ones are *frequentist* inference and *Bayesian* inference.

Frequentist logic

Frequentist inference statistically analyses the evidence that has been collected with regard to a hypothesis (for example that the independent variable X influences the dependent variable Y). The analysis of the data either leads to a “true or false” conclusion from a significance test (i.e. that the effect of X on Y is significant), or to a conclusion in the form of a given sample-derived confidence interval that covers the true value (i.e. the calculated net impact of X on Y that lies within a specified confidence interval).

Frequentist logic of inference is employed in successionist models of causal explanation, where large amounts of sample data are statistically analysed either to find similarity or differences within the data.

Bayesian logic

Rather than producing “true or false” statements about causal connections between variables, Bayesian inference aims to improve the confidence of the researcher in the validity of the inference claim. Or in other words, the goal of evidence gathering is to update our confidence in a hypothesis. This process is also called “Bayesian updating”. This updating always occurs as a matter of degree; we can never absolutely confirm or disconfirm a hypothesis. Applied in monitoring and evaluation, it helps practitioners to improve the confidence in the validity of their hypothesised causal mechanisms.

Bayesian updating is often used in evaluations that use a generative approach to causality. Bayesian logic allows them to continuously refine the underlying hypotheses, Theories of Change or causal mechanisms and systematically increase the confidence in the postulated theory.