Contribution Analysis: Theoretical and Practical Challenges and Prospects for Evaluators

Suman Budhwani
University of Toronto

James C. McDavid
University of Victoria

Abstract: Contribution analysis (CA) is a theory-based approach that has become widely used in recent years to conduct defensible evaluations of interventions for which determining attribution using existing methodologies can be problematic. This critical review of the literature explores contribution analysis in detail, discussing its methods, the evolution in its epistemological underpinnings to establishing causality, and some methodological challenges that are presented when CA is applied in practice. The study highlights potential adaptations to CA that can improve rigour, and describes areas where further work can strengthen this useful evaluation approach.

Keywords: contribution analysis, program evaluation, theory-based evaluation

INTRODUCTION

Interventions, including programs, policies, initiatives, and projects, can often operate within environments where complex contextual factors affect processes and results (Government of Canada, 2012). Interventions are often designed so that
there are multiple components required to achieve results using multiple pathways of causation. For such complicated and complex interventions, more innovative methods of measuring outcomes and assessing attribution can be more appropriate than experimental or quasi-experimental evaluation designs (Government of Canada, 2012; Mayne, 2011; Rogers, 2011). Theory-based evaluation approaches, such as realist evaluation (Pawson & Tilley, 1997) and developmental evaluation (Patton, 2010), can be particularly effective in this regard (Government of Canada, 2012). More recently, contribution analysis (CA) has become widely known as a theory-based evaluation approach by which evaluators can conduct defensible assessments of interventions for which assessing cause-effect relationships using standard methods can be problematical. However, despite its popularity, questions remain with respect to its applications in practice and its approach to assessing causality. This article aims to address these questions. It presents a literature-based critical review of the CA approach to evaluation that situates it in the wider realm of theory-based approaches, discusses its evolution to a fourth generation, presents options for how evaluators can conduct CA, and presents opportunities to clarify and further develop conceptual thought and practice on CA.

WHAT IS CA AND WHAT IS ITS UNDERLYING APPROACH TO CAUSALITY?

CA is a theory-based approach to assessing the contribution of a program to observed results (Mayne, 1999; Mayne, 2008; Mayne, 2012; Mayne, 2015a). It is highly relevant in today’s growing environment of results-based management where programs are expected to produce impacts and where evaluations of these impacts have moved away from the assessment of inputs-processes-outputs to delineating and measuring outcomes and impacts resulting from the program and how and why they occur (Global Affairs Canada, 2015; Mayne, 2004; Wimbush & Beeston, 2010). CA thus presents a heuristic framework, by which an analysis can be conducted to determine and reasonably conclude the degree of contribution based on an explicitly defined theory of change (BetterEvaluation, 2013; Mayne, 2012; Vaessen & Raimondo, 2012).

Although considerable conceptual diversity exists with respect to theory-based evaluation approaches, Weiss (1997a, 1997b), one of the founders of the theory-based evaluation approach, differentiates among program theories, implementation theories, and theories of change. She defines implementation theory as an explication of how an intervention is carried out, emphasizing implementation failure and/or success based on quality, intensity, and fidelity of the intervention activities. Program theory, on the other hand, is defined as the collective mechanisms that enable the activities of the program to translate into observed results. Theories of change combine implementation theories and program theories. Analysis for each vary and require different methods and forms of evidence (Stocks-Rankin, 2014). Like other theory-based approaches, CA aims to explore the underlying theory of change by which an intervention is expected to
produce impacts, including, but not limited to, understanding the activities of the intervention itself, the underlying assumptions based on which the intervention is proposed to work, and the relative influences of external factors on the intervention in achieving observed outcomes (Mayne, 2008, 2012). However, it is favoured in the spectrum of theory-based approaches due to the examination of both implementation and program theory components in its systematic process-based framework by which claims of contribution can be made and iteratively strengthened, particularly with the collection of various lines of evidence over time (Mayne, 2008; Wimbush & Beeston, 2010). By gathering evidence to justify or reevaluate a suggested theory of change that describes how an intervention is expected to work (implementation theory) and cause the expected impact through its proposed framework (program theory), CA attempts to provide comprehensive insights on the contribution made by an intervention to observed results (Mayne, 2008, 2015a, 2015b). Consequently, CA addresses some of the limitations of other theory-based approaches by moving away from focusing solely on program implementation and making an attempt to comprehensively understand all of the necessary components within a theory of change, thus making it an important evaluation approach for evaluators (Dybdal, Nielsen, Lemire, & Ramboll Management Consulting, 2011).

The CA approach is considered to be particularly relevant when determining the attribution of complicated and complex interventions to observed impact and results (American Evaluation Association, 2015; Dybdal, Nielsen, Lemire, & Ramboll Management Consulting, 2011). Due to the nature of the multiple emergent constructs (complex programs) and multiple pathways to causation inherent within these interventions, it becomes difficult to establish an intervention’s attribution to any observed changes and results (Mayne, 2011). In addition, effects of competing interventions or other contextual factors may diminish the ability of an evaluator to establish straightforward linear cause-and-effect relationships through experimental or quasi-experimental research designs, particularly if results take a long time to emerge (Kotvojs & Shrimpton, 2007). As such, evaluation methods using counterfactual designs from the post-positivist tradition (Shadish, Cook, & Campbell, 2002) may not necessarily be suitable due to methodological constraints such as the evaluation being conducted after the interventions have already been implemented; practical constraints including the nature of the interventions themselves, the possible resources required, and the inability to adjust the implementation of an intervention; and ethical constraints, including the implications of withholding treatment of control groups, may also be deterrents (American Evaluation Association, 2015; Befani & Mayne, 2014; Dybdal, Nielsen, Lemire, & Ramboll Management Consulting, 2011; Mayne, 2008).

In such situations, CA can provide an attractive alternative evaluation approach as it does not require the use of counterfactual-based evaluation designs nor promises unequivocal understandings of the direct cause-and-effect relationships between an intervention and the observed results (Befani & Mayne, 2014; Dybdal, Nielsen, Lemire, & Ramboll Management Consulting, 2011; Mayne,
By seeking to provide evidence by which one can “reasonably” assert the contribution of an intervention to observed results, it relies on tests of alternative explanations to act as substitute candidates in place of counterfactuals to determine the plausibility of a proposed theory of change and the significance and, accordingly, the influence of the intervention in producing observed changes (Befani & Mayne, 2014; Lemire, 2010; Wimbush, Montague, & Mulherin, 2012).

Due to these advantages, CA has been increasingly applied in areas such as the field of international development evaluation where intended, and actual causal relationships between interventions and results are heavily dependent on the situational context, including the considerable amount of time that may be required before results are observed (Kotvojs & Shrimpton, 2007; Noltze, Gaisbauer, Schwedersky, & Krapp, 2014). Other examples of CA in practice include various government-driven interventions with complicated or complex program characteristics (including causal chains) and undefined time frames for expected results (Biggs, Farrell, Lawrence, & Johnson, 2014; Delahais & Toulemonde, 2012; Wimbush, Montague, & Mulherin, 2012). With increasing requirements from government and funding agencies to move away from evaluations based solely on implementation performance metrics and toward measurable impacts, CA is intended to provide evaluators with the ability to conclude with a reasonable degree of confidence that an intervention is contributing to observed results and to provide an evidence-based justification of how this is probably occurring (Kotvojs & Shrimpton, 2007).

HOW IS CA EXPECTED TO BE CONDUCTED?

The overall goal of CA is to examine a theory of change for an intervention with evidence to determine with relative confidence whether the intended results were achieved and whether this intervention made a contribution to the observed results. Mayne (2008, 2011, 2012, 2015a; see also Befani & Mayne, 2014) highlights six integral steps to conducting CA analysis (with an additional seventh step for complex interventions). We will summarize these steps and illustrate them by including a summary of a published evaluation of a public health intervention conducted by Biggs, Farrell, Lawrence, & Johnson (2014). The Crunch&Sip® nutrition primary prevention program is a public health intervention administered to primary school-aged children to improve fruit, vegetable, and water consumption in New South Wales, Australia, on a voluntary basis, as part of the New South Wales Healthy Children Initiative. The primary goals of the evaluation were to “determine whether there was a reasonable theory-of-change driving the program” (Biggs, Farrell, Lawrence, & Johnson, 2014, p. 216) and also “whether the program was contributing to the intended goal” (p. 216) using the CA approach. Mayne’s original steps were configured based on the evaluation requirements and program context and augmented with tools from more recently published literature advancing the approach (Dybdal, Nielsen, Lemire, & Ramboll Management Consulting, 2011; Kotvojs & Shrimpton, 2007; Lemire, Nielsen, & Dybdal,
We will present this evaluation as an exemplar of the CA evaluation approach, followed by a discussion of other potential areas of clarification and improvement.

**The steps in Mayne’s general approach**

**Step 1: Set out the cause-effect issue to be addressed**

This step involves recognizing the attribution problem; scoping out the problem including determining the specific cause-and-effect questions to be addressed, including the level of confidence required to answer these questions; determining and establishing the standards of proof required to conclude contribution confidently; and identifying the type of contribution expected and other key influencing factors that might influence results. Finally, this step involves assessing the plausibility of the expected contribution based on an intervention’s size and reach (Dybdal, Nielsen, Lemire, & Ramboll Management Consulting, 2011).

**Step 2: Develop the postulated theory of change and risks to it, including rival explanations**

The primary aim of this step is the development of a theory of change for the intervention and causal links (see Figure 1). The causal link is the basic unit of analysis in CA comprising a proposed causal mechanism, created by the linkage of two or more elements within the hypothesized results chain (Mayne, 2012; Noltze, Gaisbauer, Schwedersky, & Krapp, 2014). Development of a complete theory of change includes not only causal links but also the identification of any risks and assumptions within the theory of change itself as well as the identification and statement of the roles of the other influencing factors and rival explanations to the observed results. This development of the theory of change is based on available terms of reference and discussions with the program team and any prior and existing research in the field (Befani & Mayne, 2014; Noltze, Gaisbauer, Schwedersky, & Krapp, 2014; Organisation for Economic Co-operation and Development, 2014). Lastly, this step involves determining how contested the proposed theory of change is among stakeholders. If a single theory of change cannot be decided upon following discussion, then multiple theories of change should be tested for goodness of fit to the program and the observed results. According to Mayne (2011), the soundness of a theory can be established through four criteria namely (1) the plausibility of the theory; (2) the agreeability of the relevancy and appropriateness of the theory of change; (3) the embeddedness of the theory of change to reflect social and economic contexts; and (4) the testability to ensure measurement in order to establish the theory of change’s credibility.

**Step 3: Gather the existing evidence on the theory of change**

This step includes an assessment of the postulated theory of change against the logic and gathered evidence. It includes collecting evidence from existing or previous (performance) measures and past evaluations or research to date to validate the probability of occurrence for: (1) proposed results, assumptions, and risks;
(2) each causal link in the results chain; and (3) other influencing factors that have an impact aside from the intervention. Evidence collection is limited to the use of secondary data, including, but not limited to, previous studies and evaluations, existing research, and data from stakeholders used in developing the postulated theory of change in Step 2 (Befani & Mayne, 2014; Mayne, 2011).

**Application of Steps 1–3**

The evaluation team of Biggs, Farrell, Lawrence, & Johnson (2014) developed cause-effect questions to understand (1) the factors affecting the implementation of the program and (2) the factors contributing to observed outcomes of increased fruit and vegetable consumption by primary school-aged children. They also constructed

---

**Figure 1. A Basic Theory of Change** (adapted from Mayne, 2015b)
a retrospective logic model. The authors next used existing and secondary data, including a review of evaluations of the program implemented by other jurisdictions and interview data from stakeholders collected for a related evaluation of the program in New South Wales. Both quantitative and qualitative data were collected.

**Step 4: Assemble and assess the contribution claim and challenges to it**

This step includes assembling and assessing the causal claim of a postulated contribution story or the narrative presentation of one or more causal claims (subcomponents or the full theory of change) and its supporting evidence, based on the evidence collected this far (Mayne, 2012; Noltze, Gaisbauer, Schwedersky, & Krapp, 2014). It includes analysis of the strengths and weaknesses of the postulated theory of change and its individual causal links, including the analysis of the relevancy of influencing factors. This step also includes updating the postulated theory of change based on the analysis so far.

**Step 5: Gather new evidence from the implementation of the intervention**

In this step, based on the assessed robustness of the postulated theory of change in Step 4, additional evidence is gathered to further ensure its credibility, including any assumptions and risks inherent within the causal links and related influencing factors. Additional evidence can be gathered using a broad array of approaches, including conducting primary data collection, speaking to subject matter experts, conducting case studies, tracking differences in the implementation of intervention activities and associated differences in results, undertaking program evaluations of specific components of the theory of change requiring further clarification, and the use of pre-existing research (Mayne, 2011).

**Application of Steps 4–5**

Biggs, Farrell, Lawrence, & Johnson (2014) adapted the relevant explanation finder (REF), a framework for systematically identifying the most salient and influencing factors and alternative explanations when conducting CA (Lemire, Nielsen, & Dybdal, 2012). The REF helps with categorizing influencing factors and alternative explanations by type of mechanisms/factors within a heuristic framework, contextual level of influence, identifiers offering proof or disproof, and the degree of influence, with the overall aim of making an informed evidence-based decision on the contribution of these to observed outcomes. Within the Crunch&Sip® program evaluation, the evaluation team first modified the logic model to include a list of assumptions developed in corroboration with program practitioners. This list of assumptions was then analyzed guided by the REF to understand each one’s potential influence on achieving program outcomes. The REF was also used to delineate the mechanisms underlying the suggested pathway to program outcomes and influencing contextual factors. Once all of the influencing factors and alternative explanations had been generated, the program team evaluated each one’s potential impact on the outcomes until consensus was reached among the team categorizing each as a high, medium, or low level of influence. The two data sources described in Steps 1–3 were used for this analysis (Biggs,
Farrell, Lawrence, & Johnson, 2014). The analysis findings were used to answer the evaluation questions.

**Step 6: Revise and strengthen the contribution story**

Using the new data gathered in Step 5, this step involves assembling a more credible contribution story by strengthening the contribution claim. It includes reassessing strengths and weaknesses within the theory of change by examining whether the proposed results, assumptions/risks, and other influencing factors actually occurred. If further evidence to support the theory of change is required, Step 5 can be revisited. However, at this stage, a conclusion can be reached on the credibility of the contribution claim based on the strength of the postulated theory of change and the hypothesized and accounted for role of other influencing factors affecting the observed results.

**Application of Steps 5–6**

While the need for Step 6 was identified by Biggs, Farrell, Lawrence, & Johnson (2014) to further confirm or refute theories of change, the step was not conducted due to the unavailability of appropriate evidence at the time of the evaluation.

**Step 7: For complex interventions, assess and assemble the complex contribution story**

This includes developing the contribution story for any sub or nested theories of change and for the overall theory of change (Mayne, 2011, 2015a). This step was not applied by Biggs, Farrell, Lawrence, & Johnson (2014), possibly due to the assumption that their intervention was not complex and, as such, did not require this additional analysis.

**METHODOLOGICAL CHALLENGES ASSOCIATED WITH CA: ASSESSING CONTRIBUTION, DETERMINING CAUSALITY**

Although the steps described in the previous section are the standard analytical framework by which CA is intended to be conducted, program evaluators have been encouraged by Mayne (2012) to adapt the steps as required in evaluating interventions using this approach. The program evaluation of the Crunch&Sip® public health intervention conducted by Biggs, Farrell, Lawrence, & Johnson (2014) is an example of how Mayne’s guidelines have been adapted in practice. The absence of a standard method with associated benchmarks and quality criteria recognized by evaluation practitioners as key features of a sound CA-based evaluation has led to differing understandings and adaptations of the approach—the Crunch&Sip® program evaluation is an example (Biggs, Farrell, Lawrence, & Johnson, 2014; Delahais & Toulemonde, 2012; Government of Canada, 2015; Mentzer, Czerniak & Struble, 2014; Noltze, Gaisbauer, Schwedersky, & Krapp, 2014; Patton, 2012; Srivastava & Enriquez, 2013; Wimbush, Montague & Mulherin, 2012). This presents a potential challenge for evaluators using CA as an evaluation approach. There are additional methodological questions about CA that need to be addressed.
Will a contribution always be found (positive bias)?

Some critics suggest that CA contains an inherent positive bias due to its focus on degree of contribution, suggesting that the approach would lead to a contribution always being found (Patton, 2012). This problem of positive bias has also been identified more broadly for the evaluation field itself (Scriven, 1993; Scriven & McNulty, 2012). In CA specifically, this notion may arise due to the absence of standards or a level of evidence by which one is able to confidently validate the contribution claim through the CA steps presented earlier, such as the potential subjective bias involved in the identification and assessment of the importance of each link in a theory of change, or by missing key alternative explanations altogether due to the application of “narrow logic” during identification (Delahais & Toulemonde, 2012; Tangata Whenua Community & Voluntary Sector Research Centre, 2015). These challenges can be inherent in all theory-based evaluations such as CA (Delahais & Toulemonde, 2012) but may present themselves more obviously within CA, as noted by Biggs, Farrell, Lawrence, & Johnson (2014) within their program evaluation.

One solution to this challenge is a standard proposed by Mayne (2011, 2012; see also Befani & Mayne, 2014), by which attribution can be determined when conducting CA. He suggests using four minimum conditions to infer causality. The first condition is plausibility—the postulated theory of change for the intervention in question is credible if the chain of results and underlying stipulation of how the intervention leads to results are plausible, sound, and based on, and validated by, existing research and at least some key stakeholders. The second condition is fidelity—the activities of the intervention are implemented as proposed in the theory of change. The third condition is a verified theory of change—the theory of change has been verified by collected evidence, including that the chain of expected results was actually observed and that all causal assumptions were confirmed. Finally, the fourth condition of accounting for other influencing factors ensures that all identified contextual and other factors did not have an influence on the theory of change or, if they did, then their relative contributions were included within the theory of change.

Meeting these four conditions is expected to provide justification for the theory of change that passes the reasonable person test in concluding that an intervention contributed to the observed results—namely, that it is based on the presented logic and collected evidence if a reasonable person would conclude that an intervention contributed to the observed results and thereby suggesting “plausible association” (Hendricks, 1996, as cited in Mayne, 2011, 2012). However, the reasonable person criterion is not elaborated in detail by Mayne and remains largely unspecific, though one could assert that, unless we assume some universal model of “reasonableness,” different persons with their own criteria for soundness, even based on experience, could differ among themselves as to what is reasonable. Moreover, use of this criterion suggests that the theory of change can never be the best it could be, as there can always be further improvements that can be made to make it more “reasonable.” In effect, this criterion is problematical if one’s aim
of defensible evaluations is to use a methodology that is transparent (Bannister & O’Sullivan, 2013).

**Identification and validation of theory of change components (bias from a narrow focus)**

The systematic identification and investigation of evidence for a postulated theory of change and related influencing factors may be subject to bias from a narrow focus. Paying unbalanced attention to some causal links over others or some influencing factors over others can create biases in the interpretation of the results, thereby producing conclusions that are based on a partial or narrow view of the theory of change without an accounting of all relevant and influential factors (Delahais & Toulemonde, 2012). Challenges can also occur when there are more causal links to be tested within a single evaluation than there are lines of evidence to test them, making the selection of some causal links over others within the analysis a necessity to meet evaluation scope and resource requirements (Noltze, Gaisbauer, Schwedersky, & Krapp, 2014). Evaluators of the Crunch&Sip® program were also faced with the challenge of an “interminable” number of possible influencing factors and resorted to focusing on “key statewide issues, and attempted to capture the diversity of successes and limitations of the program” (Biggs, Farrell, Lawrence, & Johnson, 2014, p. 222).

Moreover, when examining evidence, judgment of its merit by evaluators in asserting whether a result occurred—what it is due to (causal linkages, rival hypotheses, or both)—implies invoking epistemological assumptions that can influence how evidence is weighted and whether or not some lines of evidence are credible. No framework exists that addresses the criteria by which these judgments should be made, leaving it largely up to the interpretation of individual evaluators and analysts (Lemire, 2010; Wimbush & Beeston, 2010). This includes the inability of the CA approach not only to systematically identify causal links and influencing factors but also to understand the extent to which these factors play a role in the theory of change hypothesized, thereby reducing the ability to infer contribution (Lemire, 2010). Although the reasonable person test mentioned previously is a good theoretical notion for determining the credibility of a contribution claim, further quality criteria and benchmarks need to be developed to reduce subjective bias, particularly as this approach becomes more well known and more frequently used (Lemire, 2010). The reasonable person test may be too general and vague to provide a systematic framework by which to judge and assess causality and may also be impacted if sufficient resources do not exist to execute it at the depth required to validate contribution-related conclusions (Bannister & O’Sullivan, 2013; Delahais & Toulemonde, 2012; Noltze, Gaisbauer, Schwedersky, & Krapp, 2014).

**Resource intensiveness of CA and fidelity to the approach**

A final challenge is the resource-intensive nature of the steps required to conduct a CA, a point emphasized by Mayne himself (2015a; see also Delahais & Toulemonde, 2012). Building and then elaborating the theory of change suggests the
possibility of several cycles of evidence gathering and contribution story testing, a process that may not be practical in most applied settings when faced with limited resources, as also encountered by Biggs, Farrell, Lawrence, & Johnson (2014) during the evaluation of the Crunch&Sip® program (Bannister & O’Sullivan, 2013; Dybdal, Nielsen, Lemire, & Ramboll Management Consulting, 2011). Limited resources may thus impact the ability of evaluators to conduct a full six- or seven-step contribution analysis, thereby opening the evaluation to questions of integrity for any contribution claims.

Mayne (2011) and others (Bannister & O’Sullivan, 2013; Dybdal, Nielsen, Lemire, & Ramboll Management Consulting, 2011) seem to recognize this problem and have suggested a way (conceptually) of differentiating among different kinds of CA used, depending on the requirements of the evaluation, the strength of the available evidence, and the degree of rigour required to establish a contribution claim. Minimalist CA develops a theory of change and confirms the delivery of expected outputs: (1) direct influence CA includes building on minimalist CA by gathering further evidence that indicates that results expected under the direct influence of the theory of change were observed especially within the context of any influencing factors and (2) indirect influence CA seeks to build on the first two levels through the measurement of intermediate and some final outcomes, while collecting evidence that corroborates the assumptions in areas of indirect influence within the theory of change (Bannister & O’Sullivan, 2013; Dybdal, Nielsen, Lemire, & Ramboll Management Consulting, 2011; Mayne, 2011).

More recently, Mayne (2015a) also suggests the use of light CA, where the usual process of developing a theory of change is followed by a confirmation of expected outputs and early outcomes and is based on the overall strength of the theory of change, allowing claims of contribution by the intervention to be made. However, this light CA requires more methodological detail and guidance to address the methodological issues discussed. Moreover, having a hierarchy of the types of CA that can be conducted suggests that not all kinds of CA will account for all of the relevant factors necessary in order for a contribution claim to be sound. We will come back to the issue of resources to conduct CA and how fidelity to the suggested approach may affect contribution claims in our conclusions.

APPLICATIONS AND ADAPTATIONS TO CA TO REDUCE METHODOLOGICAL CHALLENGES

The previously mentioned methodological challenges for adequately establishing attribution have led to a variety of adaptations within the CA approach to establish a contribution claim. Since the primary aim of the CA approach is to present a credible theory of change and to provide sufficient evidence that allows one to have confidence in its attribution with observed results, many of these adaptations have targeted this particular area to establish more systematic processes to reduce two kinds of bias: the challenge of positive bias—that is, the inevitable claim that some level of contribution will always be detected when using this approach—and
also a narrow focus that can bias how the theory of change is established and how influencing factors and alternative hypotheses are accounted for, thereby reducing the strength of the contribution claims (Delahais & Toulemonde, 2012). To reduce the potential for these biases, Delahais and Toulemonde (2012) suggest executing a systematic search for any potential alternative hypotheses and influencing factors to improve confidence in establishing a contribution claim. Such a systematic search should take place before gathering any evidence, instead of during the data collection stage and should include not only the identification of alternative hypotheses and influencing factors but also an assessment of the degree of their influence on the observed results (Delahais & Toulemonde, 2012; Lemire, Nielsen, & Dybdal, 2012). This systematic search can be aided by standardized frameworks for the identification and assessment of influencing factors and alternative explanations such as the REF developed by Lemire, Nielsen, & Dybdal (2012) and as implemented within the Crunch&Sip® program evaluation (Biggs, Farrell, Lawrence, & Johnson, 2014) (see Table 1). Even using this approach, identification of the most salient factors and explanations continues to be subjectively determined based on the evaluator’s judgment, but the REF presents a first step in systematizing the judgment of the relevancy of these factors (Lemire, Nielsen, & Dybdal, 2012).

However, identification of the most salient factors and explanations does not have to be entirely subjective. Focusing on the weakest or most contested causal links within the theory of change (identified by the most plausible alternative explanations or the greatest risks) is recommended in an effort to improve confidence in establishing contribution claims (Nakrošis, 2014; Noltze, Gaisbauer, Schwedersky, & Krapp; Mayne, 2015a). In effect, a stance is suggested wherein the default position is that the program has not made a contribution. If the weakest links are corroborated by the lines of evidence, then confidence in the overall theory of change is increased. Moreover, an iterative process of analysis, as suggested within Mayné’s (2012; see

**Table 1.** The Relevant Explanation Finder

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Adapted from Lemire, Nielsen, & Dybdal, 2012
also Wimbush, Montague, & Mulherin, 2012) stepwise analysis framework, may be useful in strengthening contribution stories as the understanding of the role of the intervention in contributing to the observed results improves. Embedded case studies that construct specific instances of how the theory of change actually operates *in situ* may assist in this iterative process to ensure evidence is collected for the entire theory of change. This involves using a case study approach to collect detailed information on the first logical causal link within the theory of change and then selecting and exploring sub-cases within each case to validate other further logical causal links until examination of the entire theory of change is completed (Delahais & Toulemonde, 2012). Process tracing, or a probabilistic method using Bayesian principles to determine causality, can also be used under the overarching CA framework to provide guidance on what evidence is needed and how to judge the evidence’s merit based on cumulative observations to confirm contribution (Befani & Mayne, 2014). Bayesian inference relies on the estimation of probabilities based on prior probabilities and new evidence, which is different from traditional methods of statistical inference based on relative frequencies (Befani & Mayne, 2014; International Society for Bayesian Analysis, 2009).

The triangulation of methods and evidence is considered important in CA when trying to verify the theory of change and discount alternative explanations and the role of influencing factors in observed results (Nakrošis, 2014; Noltze, Gaisbauer, Schwedersky, & Krapp, 2014; Rotem, 2008; Ton, Vellema, & Ge, 2014). Delahais & Toulemonde (2012) go one step further and suggest that evidence triangulation takes place at each level of judged reasoning, not just for each overarching evaluation question or even the whole theory of change. This may include, for example, triangulation at each step and sub-step of Mayne’s analysis framework. Triangulation can involve using both qualitative and quantitative lines of evidence as required or of evidence that can include collecting existing and new evidence (Mayne, 2012; Nakrošis, 2014). However, when faced with tight timelines and time budgets, the feasibility of using triangulation will be affected, thereby presenting the question of how elaborate an approach to CA should be in order to demonstrate a credible contribution claim (Vaessen & Raimondo, 2012).

Stakeholders can provide a wealth of data at the outset of an evaluation to narrow down the scope and focus of the evaluation and evaluation questions (Noltze, Gaisbauer, Schwedersky, & Krapp, 2014). Moreover, stakeholders can be used to collectively build a theory of change (Government of Scotland, 2011; Mayne, 2012; Srivastava & Enriquez, 2013). Once a theory of change is constructed, stakeholders can be used to participate in collecting evidence to assess the validity of the proposed theory of change and address alternative explanations and other influencing factors (Delahais & Toulemonde, 2012). In accounting for this evidence, stakeholders can act as a jury not only to assess the reliability of the evidence in relation to the contribution story but also to validate the contribution story itself based on the evidence presented (Rotem, 2008).

Patton (2012) is an advocate for the use of stakeholders in this way by suggesting the relevance of the use of CA within a utilization-focused evaluation approach.
He suggests that the quality of evidence acquired in the analysis of a contribution story needs to be judged and that this responsibility should fall to the primary intended users of the evaluation results, with the evaluator acting as a facilitator of these conversations. This ensures user participation from the outset in both determining the type of evidence required to claim contribution and in the assessment of the degree or level of contribution determined within the end results.

Patton (2012) further underlines the importance of rigorous thinking over rigorous methods when using CA. He presents a framework to establish a rigour metric that consists of eight criteria, namely: (1) exploring multiple hypotheses using multiple perspectives; (2) searching for information until available avenues have been exhausted; (3) validating using a systematic approach to verify, corroborate, and cross-check sources; (4) stance analyzing sources to fully understand visible biases in sources of evidence; (5) analyzing sensitivity using a strategic approach to understand assumptions and limitations of the analysis; (6) collaborating with specialists to extract input from field subject matter experts; (7) synthesizing information with the inclusion of diverse interpretations in the compilation of data; and, finally, (8) conducting an explanation critique to understand how many perspectives have been included, which inferences are stronger, and which are weaker. This type of rigour is intended to detect and remedy biases that can occur within the CA approach; however, the steps presented may again prove to be too onerous and resource intensive for evaluators when faced with limited scope and resources.

A FOURTH GENERATION OF CA: CUMULATIVE EFFECTS OF INCREMENTAL CHANGES

Since its introduction in 1999, Dybdal, Nielsen, Lemire, & Ramboll Management Consulting (2011) have suggested that CA has evolved as an evaluation approach, from the first to the third generation (see Table 2). Originally conceptualized as a means of creating a performance story to elaborate on observed performance results by eliciting existing data from performance measurement and monitoring systems, and other more limited lines of available evidence, it has since evolved to focus more on the attribution question for a wide range of programs and evaluation requirements (Dybdal, Nielsen, Lemire, & Ramboll Management Consulting, 2011; Mayne, 1999, 2012, 2015b). This includes a clear demarcation of the components of a theory of change to understand how and why an intervention is intended to work in a given setting and what refinements are necessary to the analytical strategy to gather relevant evidence that could indicate that an intervention has contributed to the observed results (Dybdal, Nielsen, Lemire, & Ramboll Management Consulting, 2011; Mayne, 2011, 2015a).

However, with more recent clarifications from Mayne and adaptations of the approach by other users of this approach (discussed in the preceding section), we suggest that CA has further evolved into its fourth generation. The primary reason for this is because of evidence that suggests that the scope
Table 2. Overview of the Three Generations of Contribution Analysis

<table>
<thead>
<tr>
<th>Article</th>
<th>First generation CA</th>
<th>Second generation CA</th>
<th>Third generation CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope</td>
<td>How monitoring data can be used to address the attribution question through CA (p. 5)</td>
<td>How monitoring data can be used to address the attribution question through CA (p. 5)</td>
<td>How CA can be used in evaluation and/or performance measurement (p. 1)</td>
</tr>
<tr>
<td>Problems addressed</td>
<td>Sketches elements in CA</td>
<td>Elaborates operational CA design</td>
<td>Elaborates operational CA design and considers implications in terms of causal claims</td>
</tr>
<tr>
<td>Operational approach</td>
<td>CA structured around nine elements</td>
<td>CA’s nine elements consolidated into six steps</td>
<td>CA’s six steps changed slightly. New Step 1 and Steps 3 and 4 consolidated as new Step 4. Step 7 added.</td>
</tr>
<tr>
<td>Level of detail for each step</td>
<td>Outline guidelines</td>
<td>General guidelines</td>
<td>Elaborate guidelines</td>
</tr>
<tr>
<td>On attribution</td>
<td>Acknowledges experimental design most appropriate for addressing attribution (p. 5)</td>
<td>Acknowledges experimental design most appropriate for addressing attribution (pp. 4–5)</td>
<td>Argues that another approach to attribution can be valid (pp. 5–6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Argues that “plausible association” can be inferred (p. 8)</td>
<td>Argues that probabilistic causation can be inferred (p. 7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Contribution can be inferred when four conditions are met (pp. 21–22)</td>
<td>Contribution can be inferred when four conditions are met (p. 7)</td>
</tr>
<tr>
<td>On magnitude of contribution</td>
<td>None offered</td>
<td>None offered</td>
<td>Based on evidence available and causal links demonstrated three levels of contribution can be determined (pp. 25–26)</td>
</tr>
<tr>
<td>Methodology</td>
<td>Advocates mix of quantitative and qualitative methods. Use of existing data and additional data collected (p. 5)</td>
<td>Same (p. 18)</td>
<td>Same (pp. 17–21)</td>
</tr>
<tr>
<td>Epistemology</td>
<td>None offered</td>
<td>None offered</td>
<td>None offered</td>
</tr>
<tr>
<td>Ontology</td>
<td>None offered</td>
<td>None offered</td>
<td>None offered</td>
</tr>
</tbody>
</table>

Source: Adapted from Dybdal, Nielsen, Lemire, & Ramboll Management Consulting, 2011
and epistemological underpinnings of the approach have evolved. First, versions of CA have appeared to align with successionist models of causality that are focused on causal links, the basic unit of analysis in CA (Mayne, 2012; Noltze, Gaisbauer, Schwedersky, & Krapp, 2014). Under the successionist model of causality, CA entails corroborating the causal links in a program’s theory through observed co-variation between and among variables, controlling for rival hypotheses using appropriate experimental or quasi-experimental research designs and clear specification of independent and dependent variables (Dybald, Nielsen, Lemire, & Ramboll Management Consulting, 2011; Mayne, 1999; Pawson, 2007). Early versions of CA had the scope to address how monitoring data collected via existing performance measurement systems could be used and elaborated on to shed light on attribution, while acknowledging that experimental or quasi-experimental designs based in a successionist model of causality were still the most appropriate for addressing attribution (Dybald, Nielsen, Lemire, & Ramboll Management Consulting, 2011; Mayne, 1999). In this first version of CA, Mayne (1999) suggested that when more definitive evaluative information was required, beyond constructing and elaborating performance stories, a program evaluation would be more appropriate.

In contrast, more recent iterations of the scope of CA appear to be more aligned with a generative causal model that focuses on understanding the relationships between context, mechanisms, and outcomes through situationally sensitive theories of change (Befani & Mayne, 2014; Mayne, 2015a; Pawson, 2007). Situated in this epistemology, CA is concerned with the relative confidence of observers of impacts produced for groups of participants due to an intervention, through the provision of evidence that reduces uncertainty about the intervention’s contribution to the observed results (Befani & Mayne, 2014; Kotvojs & Shrimpton, 2007; Mayne, 2015a). The implied model of causality is based on the successful identification and validation of the causal package that consists of intervention components as sufficient factors to obtain results as well as contextual and process factors as necessary conditions required to produce these factors or as sufficient factors themselves (Delahais & Toulemonde, 2012; Mayne, 2012; Nakrošis, 2014). As such, the results are based on collecting evidence that can support the hypothesized relationships between all of the internal intervention and external contributing factors, thereby making this a causal approach based on relational causation (Steinberg, 2007, as cited in Nakrošis, 2014). To further demarcate contribution from non-contribution, Mayne (2012) converts the three conditions of attribution for successionist-based causality to four conditions (previously discussed) that must be met to claim a contribution and that are examined through an assessment of the logical validity of an intervention’s suggested theory of change, the time sequence between activities and results, and the evidence that confirms the theory of change and discounts competing explanations for why the observed results occurred, usually by collecting and analyzing a mix of qualitative and quantitative data (American Evaluation Association, 2015; Dybald, Nielsen, Lemire, & Ramboll Management Consulting, 2011).
Using a generative model of causality within CA can be useful where a deeper understanding of the mechanisms and contexts for group-level change is required to assess what factors contributed to the observed results and how an intervention may hold up among competing explanations. Most recently, Mayne (2015a) himself accepts the usefulness of generative frameworks for demonstrating causal attribution and suggests that CA, as a theory-based approach, situates well within this framework. He also suggests that the scope of CA is to demonstrate that an intervention is a contributory cause, with the end goal of delivering a credible contribution claim based on systematic identification and the analysis of causal links and mechanisms pertaining to the intervention in question. These indications suggest an evolution of both the scope and epistemological underpinnings of CA to a fourth generation.

Other than in epistemology, the work of other evaluators has also led to further development and refinement of the set of steps to conduct CA that can also suggest an evolution to a fourth generation (Dybdal, Nielsen, Lemire, & Ramboll Management Consulting, 2011). These modifications have occurred through the detailed work of Mayne (2015a, 2015b) on what constitutes a theory of change itself, including pathway components, assumptions, and external influences for which evidence must be collected within CA that further clarifies the operational approach of CA. Other than developments in the operational approach to CA provided by Mayne, clarifications on how to apply CA in practice have emerged because of CA being applied by practitioners. Suggestions such as the triangulation of methods (Delahais & Toulemonde, 2012), the REF framework (Lemire, Nielsen, & Dybdal, 2012), and contribution tracing (Befani & Mayne, 2014) have helped to provide the level of detail for each step and to fill in gaps on methodology. As CA is used more frequently by evaluators to fulfil various evaluation needs, it is likely that further clarifications of, and enhancements to, the approach will occur (such as the adaptations by Biggs, Farrell, Lawrence, & Johnson (2014) discussed earlier), which will lead to the further evolution of CA as an evaluation approach.

CONCLUSIONS AND IMPLICATIONS FOR EVALUATION PRACTICE

While the above adaptations provide some suggestions on how to systematically apply CA in practice, the question for practitioners is whether there is a “gold standard” approach to be followed when conducting evaluations using CA? Unfortunately, there is no clear answer to this question due to the multiple adaptations of the approach present in the literature. The evaluation of the Crunch&Sip® nutrition primary prevention program is arguably a mature example of how to conduct CA, but it clearly differs from the six-step process specified by Mayne. Aside from questions of whether variations and adaptations compromise the fidelity of applications of CA, resource requirements alone suggest that applying the full framework is a daunting process.

The stepwise process proposed by Mayne can be considered the skeleton when using the CA evaluation approach, and, in this article, we have pointed out
ways by which this process could be improved. However, these suggestions may not be applicable or relevant for all forms of program evaluations using CA. For example, on the face of it, the suggestion that data triangulation take place at each step in assessing the theory of change has merit (Delahais & Toulemonde, 2012). But adding this requirement makes the CA process, which is arguably already demanding in terms of effort, even more demanding and resource intensive, particularly in determining the validity of causal links in the theory of change when the evidence is not strong, convergent, or unable to be triangulated and where strong professional expertise by evaluators may be required (Delahais & Toulemonde, 2012; GIZ Denmark, 2015; Stocks-Rankin, 2014; Wimbush & Beeston, 2010). What is suggested is a direct relationship between the scope of work required to demonstrate a contribution and the credibility of evidence supporting a contribution claim, which Mayne (2015a) also recognizes. In many cases, this relationship may translate itself as a trade-off when practitioners are faced with limited resources.

Involving stakeholders in constructing and validating the theory of change has also been suggested as a means by which CA can be made more robust (Delahais & Toulemonde, 2012; Leeuw & Vaessen, 2009; Rotem, 2008; Wimbush & Beeston, 2010). When, and to what extent, stakeholders should be engaged within CA leads to an overall lack of clarity in whether and how CA can be melded with existing participatory evaluation approaches (Leeuw & Vaessen, 2009; Wimbush, Montague, & Mulherin, 2012). Evaluators differ in their degree and extensiveness of involvement of stakeholders within CA, using stakeholders to assist in better understanding the intervention in question, designing the theory of change, deciding the specific evaluation questions and metrics, and collecting, analyzing, and interpreting data (Delahais & Toulemonde, 2012; Mayne, 2012; Rotem, 2008; Wimbush, Montague, & Mulherin, 2012). Patton (2012) has suggested subsuming the CA approach within his utilization-focused framework to ensure that evaluation results are of importance to stakeholders. While the application of CA to these various approaches provides evidence of its versatility, the absence of consistency and clarity regarding stakeholder involvement suggests uncertainty whether CA is a distinct evaluation approach or an innovative strategy for addressing the core methodological challenge in program evaluation—that is, whether an intervention has caused observed results.

Additional challenges arise from the need for clarity surrounding what it means to construct and test a theory of change. The standards for a fully developed theory of change, including “adequate” levels of detail, remain points of discussion in the literature, with little consensus (Sridharan & Nakaima, 2012). Moreover, there is ambiguity on what constitutes a theory of change, although initial attempts have been made by Mayne (2015a) to define the various components of a theory of change. Some evaluators identify theory of change as the program theory, others think of it as implementation, whereas some consider it to be both.

Within CA, Mayne (2015a, 2015b) defines theory of change to include both of these elements by describing intervention activities in the results chain to signify
implementation theory and stating causal assumptions to account for program theory (see Figure 1). However, more thinking and clarification is needed to understand how to merge these two forms of analysis for both program theory and implementation theory so that they complement each other in arriving at conclusions about an intervention’s contribution (Stocks-Rankin, 2014). Ambiguities in how theories of change are understood and applied more generally in evaluation practice might affect the quality of a constructed theory of change and affect the integrity of any resulting contribution claims derived from CA.

Questions also arise on the ability of the CA approach to handle contribution questions for complex interventions. While some authors advocate for the use of the CA approach for complex interventions (GIZ Denmark, 2015; Lemire, Nielsen, & Dybdal, 2012; Noltze, Gaisbauer, Schwedersky, & Krapp; Stocks-Rankin, 2014; Wimbush & Beeston, 2010), including Mayne (2011, 2015a, 2015b) himself, others are hesitant to support this claim (Bannister & O’Sullivan, 2013; Nakrošis, 2014). This might be due to the often common visualization and conceptualization of the theory of change as a single path approach to causation in an effort to tell an evaluation story, when what may be required for complex interventions are more emergent program structures that include recursivity (Bannister & O’Sullivan, 2013; Funnell & Rogers, 2011). Thus, there may be programs that the CA approach may not be able to evaluate effectively. Mayne (2011, 2015a) proposes solutions to this issue with the suggestion of a seventh step in the CA process that includes developing an impact story for each nested theory of change and then developing the contribution story for the overall theory of change. However, according to Rogers (2008, as cited in Bannister & O’Sullivan, 2013), this may be difficult to achieve, and Bannister & O’Sullivan (2013) provide examples of interventions where little success was achieved when attempting to construct theories of change for more complex interventions.

Finally, external validity or the ability to generalize findings to other settings, other people, outcome variations, or other intervention variations is also important. This means verifying the theory of change following CA by comparing results across similar interventions to verify if the theory of change still holds (Srivastava & Enriquez, 2013) and to determine whether it contributes to the “cumulation of theoretical knowledge” (Weiss, 1997b, p. 52). However, generally it is difficult to produce generalizations using the CA approach due to its reliance on intervention-specific contextual and influencing factors, an implication of using a generative causation model to determine attribution as previously discussed (Lemire, 2010; Nakrošis, 2014). External validity related to CA is discussed less frequently in the literature suggesting an area where further work could take place.

Many of these challenges and questions for applications of CA can be traced to a lack of conceptual clarity about what CA represents. For example, clarifying and further delineating its epistemological roots to understand CA’s approach to causality would affect how we approach CA methodologically and how we view the generalizability of contribution claims in particular evaluation settings. Conceptual ambiguities can also be traced to the lack of clarity on how to conduct CA—what
types of research designs are best suited for CA and what types of methods are most suitable to ensure rigour in addressing causality (Lemire, 2010; Stocks-Rankin, 2014). Fundamentally, there is a question of what it means for an intervention to contribute to an outcome or outcomes. Mayne (2011, 2015b) suggests that it is possible to come up with the strength of a contribution that is dependent on the robustness and detailed nature of the underlying theory of change. This implies a metric for degree of contribution that, even if qualitatively expressed, has the effect of making CA similar in this respect to economic evaluation (cost-benefit analysis, cost-effectiveness analysis, or cost-utility analysis). If the degree of contribution is to be a meaningful phrase, then there needs to be more work to clarify how CA can yield this “bottom line.” In evaluation practice, to be able to sum up a program’s degree of contribution could be a powerful tool that would appeal to stakeholders.

CA is a multi-step and potentially complex approach to evaluation that is demanding of both evaluation practitioners and evaluands. How the six (or seven) steps translate into an evaluation work plan for consultants who are faced with limited time and resources is critical to the integrity and, ultimately, the utility of this approach. Whether it is possible to conduct a full CA is an issue in many evaluations, as was demonstrated within the evaluation of the Crunch&Sip® program, where a call for further evidence (not possible because of resources) was indicated to elicit a more complete understanding of the contribution story (Biggs, Farrell, Lawrence, & Johnson, 2014).

A working hypothesis, based on this literature-based assessment of CA is that although many evaluations declare their intention to do a contribution analysis, all the steps outlined in this article are rarely applied, particularly in large-scale evaluations of programs in developing countries. Instead, what consultants may do is “fill in the blanks” with their own judgments and adaptions relevant to each situation—an approach also taken by the evaluators of the Crunch&Sip® program (Biggs, Farrell, Lawrence, & Johnson, 2014). All program evaluations require professional judgment, but to substitute judgment for methodologies that are sufficiently transparent that they can be critically assessed can weaken the credibility of evaluations.

CA is now an important part of the theory and practice of program evaluation internationally. This review of the literature has suggested several issues that should be addressed as efforts to elaborate and strengthen this approach continue. It may well be time to conduct a review of existing contribution analyses similar to what Miller and Campbell (2006) did for empowerment evaluation. Such a meta-analytic review would help us to look at issues like fidelity to the approach and to better understand the variations and their implications that have evolved over the time that John Mayne and others have contributed to the theory and practice of CA. Such a review would help to better situate CA in evaluation theory and practice.

REFERENCES


Organisation for Economic Co-operation and Development. (2014). *Evaluation of the OECD Development Assistance Committee contributions with the Trade Committee to


doi: 10.3138/cjpe.31121


**AUTHOR INFORMATION**

Suman Budhwani is a PhD candidate at the Institute of Health Policy, Management and Evaluation at the University of Toronto. Her area of specialization is in outcomes and evaluations within health services research, focusing on the burden and utilization of advanced disease and end-of-life populations. She holds a Master’s degree in public policy and administration from Ryerson University and a graduate certificate in program evaluation from the University of Victoria. Suman has a keen interest and passion for program evaluation and performance measurement and applies these to address research questions in health services research in Ontario and in the not-for-profit sector.

James C. McDavid is a professor of public administration at the University of Victoria. He specializes in program evaluation, performance measurement, and performance management. He has conducted research and evaluations focusing on federal, state, provincial, and local governments in the United States and Canada. His published research includes articles and books on program evaluation, performance measurement, performance management, and comparisons of local government service performance across Canada. The second edition of his textbook, *Program Evaluation and Performance Measurement: An Introduction to Practice* was published by Sage Publications in 2013. He is a co-founder of the Consortium of Universities for Evaluation Education, a Canadian organization dedicated to increasing opportunities for students and practitioners to acquire graduate education in evaluation.