Theory of Change Analysis: Building Robust Theories of Change

John Mayne
Ottawa, Ontario

Abstract: *Models for theories of change vary widely as do how they are used. What constitutes a good or robust theory of change has not been discussed much. This article sets out and discusses criteria for robust theories of change. As well, it discusses how these criteria can be used to undertake a vigorous assessment of a theory of change. A solid analysis of a theory of change can be extremely useful, both for designing or assessing the designs of an intervention as well as for the design of monitoring regimes and evaluations. The article concludes with a discussion about carrying out a theory of change analysis and an example.*

Keywords: *analysis of theories of change, criteria for good theories of change, impact pathways, theory of change*

Résumé : *L'utilisation qui est faite de modèles de théories du changement varie grandement. Par ailleurs, il y a peu de discussion sur ce qui constitue une bonne ou solide théorie du changement. Le présent article décrit et analyse les critères de détermination de la robustesse d'une telle théorie. De plus, il discute de la façon dont ces critères peuvent servir à l'évaluation rigoureuse d'une théorie du changement. Une analyse approfondie d'une théorie du changement peut être extrêmement utile, autant pour concevoir ou évaluer la conception d'une intervention, que pour concevoir des évaluations et systèmes de monitorage. L'article se termine avec une discussion sur l'analyse d'une théorie du changement et un exemple.*

Mots clés : *analyse de théories du changement, caractéristiques d'une bonne théorie du changement, chaine des résultats, théorie du changement*

INTRODUCTION

Theories of change (ToCs) are now widely used in evaluations. They are the basis of theory-based evaluations (Coryn, Noakes, Westine, & Schroter 2011; Donaldson, 2007; Funnell & Rogers 2011; Rogers, 2007). As many have noted, the specific models used vary greatly (James, 2011; Valters, 2014; Vogel, 2012) and there is no overall agreement on just what comprises a ToC. Funnell and Rogers (2011, pp. 15–34) discuss the range of terms used and their histories, as does Patton (2008, pp. 336–340). Further, what constitutes a good or solid ToC is not at all clear; the characteristics or criteria of a robust ToC have not been widely discussed.

Corresponding author: John Mayne, john.mayne@rogers.com

 Canadian Journal of Program Evaluation / La Revue canadienne d'évaluation de programme
32.2 (Fall / automne), 155–173 doi: 10.3138/cjpe.31122

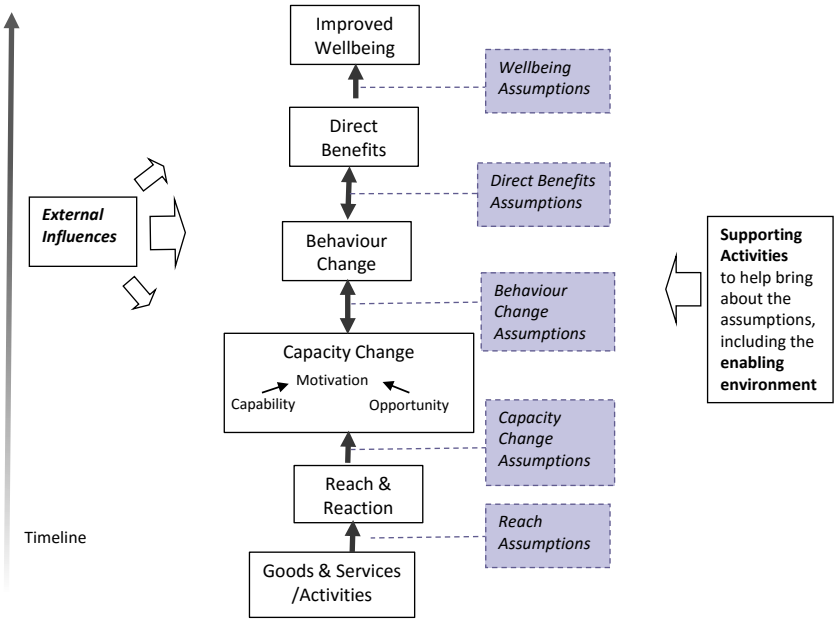

Figure 1. The COM-B Based Theory of Change

This article discusses criteria for a robust ToC and a tool for carrying out analysis of ToCs, namely Theory of Change Analysis (ToCA) to assess and strengthen ToCs.

When discussing specific aspects of ToCs and presenting examples, I will be using the behaviour change-based ToC model shown in Figure 1. Behaviour change-based ToCs are discussed in Mayne (2015) and the COM-B model in Mayne (2016a).[1] However, the steps and principles discussed apply to theories of change generally.

SOME TERMS

Given the diversity of how terms around ToCs and results are used, let me first a review the terms being used here:

- *Results* is used to include outputs, outcomes, and impacts, where impacts are the final outcomes affecting well-being. A *result statement* is the exact text used to describe the result. The term *intervention* is used here to describe specific activities undertaken to make a positive difference in outcomes and impacts of interest. It covers policies, programs, and projects.

1 The COM-B model postulates that behaviour (B) occurs as the result of interaction between three *necessary* conditions, capabilities (C), opportunities (O), and motivation (M).

- *Impact pathways* describe causal pathways showing the linkages between a sequence of steps in getting from activities to impact. An intervention may have several pathways to impact.
- A *theory of change* (ToC) adds to an impact pathway by describing the causal assumptions behind the links in the pathway—what has to happen for the causal linkages to be realized. Theories of change are models of how change is expected to happen (*ex ante* case) or how change has happened (*ex post* case).
- *Rationale assumptions* identify the underlying hypotheses or premise(s) on which the intervention is founded.
- *Causal link assumptions* are the salient events or conditions necessary (or likely necessary) for a particular causal link in a ToC to be realized; if the assumption doesn't hold, then the expected effect from that link will not occur. This can be a very demanding requirement, if interpreted literally. We can rather think in probabilistic terms, whereby causal link assumptions can be thought of as *likely necessary* assumptions, events, and conditions that almost always have to occur for the causal link to work.

Further discussion of these terms and other alternative terms used such as *logic models* and *program theory* can be found in Mayne (2015).

Because they are necessary or likely necessary, causal link assumptions also represent risks to the causal link occurring—the risk being that the assumption does not occur, that is, is not realized. For example, if an assumption is that local government takes some action, the risk is that it does not take the action. Consequently, rather than listing assumptions and risks, one can just identify assumptions.

Typically, some assumptions are less likely to be realized than others. For example, if an assumption is that some party, perhaps a local government, will take some action that has not been taken before, and nothing is being done to encourage the government to do so, then that assumption is quite likely at risk—and indeed may not be plausible. If an assumption is that a market will emerge for a new product and nothing is being done to encourage such a market, then that assumption is at risk. In addition, an assumption may be at risk because of counter pressures trying to ensure the assumption is not realized. An assumption that monitoring will be done by a third party may be at risk if there are other powerful parties who do not want the monitoring to be effective.

For theory of change analysis, I will call these *at-risk assumptions*. In an *ex ante* situation, at-risk assumptions represent potential gaps in the design of the intervention and likely serious threats to the intervention working. As a result, one may want to identify possible *confirming actions* that could be taken early on to give assurances that the assumption is likely to be realized, or *corrective actions* that might be taken to mitigate the at-risk assumptions. In an *ex post* situation, these are areas that need special attention in evaluations to see if in fact anything was done to address the risk.

In *ex ante* situations, it is important to keep the timeline in mind. At-risk assumptions for causal links well in the future may be less of a problem—realizing

the risk, actions could be taken later to address the issue. Many assumptions would not be expected to be at risk, such as when the assumption can be expected to be realized based on past experience and/or research, if the intervention design is solid, or even if it is agreed by stakeholders that it is likely to occur.

In a theory of change model, at-risk assumptions could be identified by bolding the assumption and discussing it in the accompanying text.

ISSUES IN ANALYZING A THEORY OF CHANGE

In addition to their use in evaluations, ToCs have also been found useful in designing interventions or assessing the designs of interventions (Leeuw, 2012; Rey, Brousselle, & Dedobbeleer, 2012; Tremblay, Brousselle, Richard, & Beaudet 2013). Mayne (2015) and Mayne and Johnson (2015) discuss a variety of uses of theories of change. Mayne (2015) and Johnson, Mayne, Grace, and Wyatt (2015) discuss some forms of analysis of ToCs. However, no structured approach for such analysis has yet been proposed.

Those developing theories of change use forms of analysis both during development and after. However, given the numerous elements of a ToC and the various possible purposes, it is useful to undertake a structured analysis with specific aims in sight. The *theory of change analysis* (ToCA) discussed here aims at addressing two questions:

1. Does the intervention ToC appear robust? That is, is the ToC structurally sound and plausible?
2. What are the implications for monitoring and evaluating the intervention?

ToCA is done on a proposed ToC, one that has been developed to reflect how an intervention is working or was expected to work; hence the "appear" term in the question. Reality might suggest that the intervention and its ToC were not in fact that robust! But *a priori*, before undertaking extensive data collection, we would want to identify any evident shortcomings in the ToC and hence the intervention design. And *ex post*, if we find that a ToC that has been used to model an intervention is not very robust, we might find that helpful in explaining a less than successful intervention and/or identifying issues that an evaluation should explore.

Criteria for Robust Theories of Change

When a ToC is being developed, the expectation is that it is not just a bunch of ideas, but that it is well articulated, credible, plausible, and logical—that it is robust. A robust ToC is defensible, would support a well-designed plausible intervention design, and would provide a solid basis both for monitoring and for theory-based evaluations.

A related idea is that of a ToC being evaluable, for which Rick Davies (2012) has set out a list of criteria. Davies's criteria are quite broad in their coverage, meant to

include anything that is called a ToC. And indeed, as noted above, a wide range of models and representations have been used to depict theories of change, and several of Davies's criteria challenge what has been set out as to whether it is a ToC at all.

My starting point is a little different. In defining above what a ToC is, it is assumed that what is being examined sets out the pathways of change as a causal sequence of results, and assumptions behind the pathways.

Thus, several of Davies's criteria are "assumed," namely testable, explained, complete, and inclusive. How well those criteria are addressed in a ToC is part of the robust criteria discussed below. Most of Davies' other criteria are covered in the robust criteria as well. In addition, several other criteria are needed to assess the robustness of a ToC.

A robust ToC is one that is structurally sound and plausible. A robust ToC supports a solid and plausible intervention design: with this design, it is reasonable to expect that the intervention, if implemented as designed, will be able to contribute to the intended results. Criteria for a *robust theory of change* for an intervention would address the following questions:

For a structurally sound ToC:

1. Is the ToC *understandable*? Are there pathways of results, and are causal link assumptions set out? Is there a reasonable number of results?
2. Are the ToC results and assumptions *well defined*?
3. Is the *timing* sequence of results and assumptions *plausible*?
4. Is the ToC *logically coherent*? Do the results follow a logical sequence? Are the causal link assumptions pre-events and conditions for the subsequent effect? Is the sequence plausible or at least possible?
5. Are the causal link assumptions *necessary* or *likely necessary*?
6. Are the assumptions *independent* of each other (recognizing that some assumptions may apply for more than one causal link)?

For a structurally sound ToC that is plausible:

7. Is the ToC generally *agreed*?
8. Are the results and assumptions, or at least the key results and assumptions, *measurable*? What is the likely strength or status of evidence?
9. Are the causal link assumptions likely to be *realized*? Are at-risk assumptions mitigated through confirming or corrective actions?
10. Are the sets of assumptions for each causal link along with the prior causal factor plausibly *sufficient* to bring about the effect?
11. Is the *level of effort* (activities and outputs) commensurate with the expected results?
12. To what extent are the assumptions *sustainable*?

doi: 10.3138/cjpe.31122

A ToC that is reasonably robust would provide a solid basis for using the ToC (Mayne & Johnson, 2015) in (a) designing and planning an intervention, (b) managing an intervention, (c) assessing and evaluating an intervention, and (d) scaling up an intervention. Robustness, as imagined here, is not a 0-1 variable. Meeting all the criteria could be quite demanding. Rather, in most cases, one would be improving a ToC over time, moving toward a more robust version.

There is evident need for an intervention to be plausible. At the outset, clear gaps or flaws in the design will most probably lead to a less successful intervention. Evaluability assessments are now seen as exploring the plausibility of intervention design with a view to improving the design and/or to identifying if it makes sense to undertake an evaluation (Davies, 2013; Peersman, Guijt, & Pasanen, 2015; Trevisan & Walser, 2014). The criteria here for a robust ToC include those used in evaluability assessments. The criteria also include those set out for SMARTly describing outcomes (Smart, Measurable, Achieved, Relevant, Timely) in Outcome Harvesting (Wilson-Grau & Britt, 2013).

M&E Implications

One purpose of a ToC is to provide a framework for setting out monitoring and evaluation plans. In carrying out ToC Analysis, it becomes clear just what needs to be monitored and paid attention to in evaluations. Questions here would be:

1. What data on results and assumptions should be monitored?
2. What issues need attention in an evaluation?
3. What is the likely strength or current status of evidence for the various results and assumptions, and in particular for each causal link?

Table 1 pulls these criteria together for ToCA. Each criterion is then discussed. These criteria build on ones I suggested earlier (Mayne, 2011).

Ex Ante *and* Ex Post *Perspectives*

In carrying out ToCA, it is important to keep in mind the perspective being used, namely if the situation is *ex ante* or *ex post*. The analysis is similar in both cases, but the implication of the findings will differ.

The context for the *ex ante* perspective is where a ToC is being developed for an intervention that has yet to be implemented or is in the early stages of implementation. The intent would be to develop a robust ToC to match a plausible intervention design, so that at the outset it seems reasonable that the intervention would bring about the expected results. In this setting, ToCA can be used to

- facilitate agreement on a ToC
- identify possible gaps in the intervention design and what can be done

- identify results and assumptions that need monitoring
- identify issues that a future evaluation needs to address.

In an *ex post* setting, the intervention has been in operation for some time and an evaluation is to be undertaken to see the extent to which the intervention has actually worked. Some monitoring data may have been gathered and some changes in the intervention may have been made over time. There is a need to either build (reconstruct) a ToC or revise an earlier ToC to reflect how the intervention is now seen as working. *Ex post* ToCA can be used to

- facilitate agreement on a robust ToC, often a reconstructed ToC
- identify current intervention design weaknesses that may explain limited expected results being achieved
- identify results and assumptions data that an evaluation needs to collect or get from monitoring data
- identify evaluation questions that need addressing in an evaluation.

Note that in this *ex post* scenario, ToCA itself would not be assessing if the ToC was in fact realized. That would be done as part of the evaluation, using something like contribution analysis.

In either case then, ToC Analysis would seek to

- *strengthen the ToC*: identify and correct any structural weaknesses in proposed theories of change.
- *strengthen the intervention design*: identify weakness in intervention design and what could be, or should have been, done to strengthen the design
- *identify data needs*: identify monitoring and evaluation data that need to be collected for assessing performance of the intervention.

THE THEORY OF CHANGE ANALYSIS CRITERIA

The criteria in Table 1 can be used to assess the robustness of the ToC and the underlying intervention. However, as noted earlier, robustness is not a 0–1 rating. That is, because there can be different models for the ToC of an intervention with different levels of detail, the criteria need to be applied in a sensible manner. They might best be thought of as guidelines for assessing the strength of a ToC and the intervention it represents.

Overall Criteria

Understandable: The ToC and especially its pathways should be clearly evident so that readers understand the intervention in the same way. I have argued elsewhere that a complex ToC needs to be unpacked into several nested ToC models (Mayne, 2015). Further, in any one ToC, there should be a reasonable number of results statements, so that the ToC model is "readable" to others beyond those

Table 1. Criteria for Theory of Change Analysis

Overall Criteria	
Understandable	**Is the logic and structure of the ToC clear?**
Agreed	To what extent is the ToC agreed or contestable?
Level of effort	Are the activities and outputs of the intervention commensurate with the expected results?
Criteria for Each Result	
Well-defined	Is the results statement unambiguous?
Plausible timing	Is the time frame for the result reasonable?
Logical coherence	Does the result follow logically from the previous result? Is the sequence plausible or at least possible?
Measurable	Is there a need to measure the result? How can the results be measured? What is the likely strength or status of evidence for the result being realized?
M&E implications	What are the implications for monitoring and evaluation?
Criteria for Each Assumption	
Well-defined	Is the assumption unambiguous?
Logical coherence	Is the assumption a precondition or event for the effect sought?
Justified	What is the justification for the assumption as being necessary or likely necessary?
Realized	Is it plausible that the assumption will be realized? Are there at-risk assumptions that should be addressed?
Sustainable	Is the assumption sustainable?
Measurable	Is there a need to measure the assumption? How can the assumption be measured? What is the likely strength or status of evidence for the assumption being realized?
M&E implications	What are the implications for monitoring and evaluation?
Criteria for Each Causal Link	
Independence	Are the assumptions for the link independent from each other?
A sufficient set	Are the set of causal link assumptions along with the prior causal factor sufficient to bring about the effect? Is the link plausible?
Strength/Status of evidence	What is the strength or current status of evidence for the causal link being realized?

who developed it. A rule of thumb that I have used is that if you have more than 13 result "boxes," you may have a mess instead of a ToC.

Agreed: If the ToC, no matter how well constructed, is the product of just one person or a small group, it may not have much support or buy-in, and may not be

sustainable in the sense of not being used or not lasting long. Broad agreement among stakeholders would usually suggest a more robust ToC, often built up through a participatory approach to building the ToC. And there may be different views on how the intervention is supposed to work. In this case, one may need to build more than one robust ToC and check each against reality in due course, See, for example, Hansen and Vedung (2010).

Level of effort: This is a rough check on the plausibility of the intervention. Does it seem reasonable that the activities of the intervention and their outputs will be enough for the intervention to make a difference in the ways expected? Interventions sometimes have quite ambitious intentions that are expected to be realized from a quite modest level of effort.

Criteria for the Results

Well defined: The results need to be as *well defined* as practical as to their meaning and content, and their measurability. They should not be subject to different interpretations by different readers.

Plausible timing: There should be an indication of when the results are expected to come about, and the *time frame* set out should be realistic, that is, *plausible*. Setting out realistic timing for when results can be expected is frequently neglected in developing ToCs, indeed often completely absent. Unrealistic expectations about timing can point to quite unrealistic interventions. Even less attention is paid to the trajectory of the expected results, as Woolcock (2009) discusses.

Logical coherence: That is, the step-by-step model from activities/outputs to impact should make sense, based on plausible or at least possible logic and perhaps prior evidence. The distinction here between plausible and possible logic reflects the fact that different ToC models provide different levels of detail. A "possible" logic sequence implies that the causal step is possible but represents a large leap in logic, which may be due to the level of detail in the ToC or to a causal link at-risk. Remember, the ToC is a model of expectations, which may of course turn out otherwise. If the ToC is behaviour-based, such as discussed by Mayne (2015) or Morton (2015) and illustrated in Figure 1, this significantly strengthens the logical coherence of the model.

Measurable: The results, or at least key results, should be measurable—there are reliable and valid measures of the results, and the needed data can be (readily) collected. Depending on the use being made of the ToC, there may not be a need to measure all the results set out in a pathway. For example, in a behaviour-based theory of change model, measuring capacity change can sometimes be a challenge. On the other hand, measuring behaviour changes is usually much simpler, and may be all that is needed if the expected behaviour changes have occurred and other aspects of the model are verified. The ToC analysis should indicate if the result (a) needs to be measured, (b) might be useful to measure, or (c) do not really need to be measured. It is useful here to note the likely *strength of evidence* based on the measures.

Implications for monitoring and evaluation: As part of the analysis, one can also assess what the implications of each component of the ToC are for monitoring and evaluation. Implications could be identifying evaluation questions to be addressed;

issues that need to be carefully watched or explored; issues, results, and/or assumptions that should be monitored; and/or identifying data that should be collected.

The analysis would identify specific *M&E actions* that should be taken to strengthen monitoring and evaluation.

Criteria for the Assumptions

Well defined: The events and conditions set out in the assumptions need to be as well defined as practicable as to their meaning, content, and measurability. They should not be subject to different interpretations by different readers.

Logical coherence: Because the assumption should be needed for the effect to occur, it should be a logical precondition or event.

Justified: The assumptions are justified by a solid argument as necessary or likely necessary events or conditions for the causal link to work.

Realized: One should expect that the assumptions would be realized. That is, there is general agreement, strong logic, actions being taken, or prior evidence that make the assumption plausible. The analysis could identify the at-risk assumptions that exist to the intervention and *corrective or confirming actions* that could or need to be taken to mitigate the risk. *Ex ante*, this would identify weaknesses in the intervention design and what might be done to strengthen the design. *Ex post*, it would identify assumptions that need careful examination in an evaluation. In essence, here one is assessing the degree of control the intervention has over the assumption.

Sustainable: An assumption may be realized during the period of the intervention, but one normally would hope that the assumption is sustainable after the intervention is over. If not, then the assumption is at future risk, as would be the causal link, the result in question, and indeed the intervention. Where sustainability is an issue, the intervention might want to undertake some form of corrective action.

Measurable: The assumptions, or at least key ones, are measurable: there are reliable and valid indicators, the relevant data can be (readily) collected, and/or there is adequate prior evidence. The analysis should indicate if the assumption (a) needs to be measured, (b) might be useful to measure, or (c) does not really need to be measured. Again, it is useful here to note the likely *strength of evidence* based on the measures.

The *M&E Implications* criteria are discussed above.

Criteria for Each Causal Link

Independence: For each causal link, the assumptions should be independent of each other—that is, be separate events/conditions—and hence be a minimum set of assumptions, recognizing that the same assumption may be needed for more than one causal link.

A sufficient set: The set of the initial result plus assumptions for the causal link should be seen as sufficient for that link to work, that is, for the cause plus assumptions to contribute to the effect. The link should be plausible—the link causal package should be enough to likely bring about the effect.

Strength/status of evidence: This final criterion is about the likely strength of the evidence on the causal link occurring (*ex ante*), or the current status of evidence

about the link having been realized (*ex post*), classified as strong, medium, or weak. Again the analysis would seek to identify intervention design weaknesses or issues that need exploring in an evaluation. Where evidence appears weak, this might suggest the need for additional monitoring, research, and/or evaluation.

CARRYING OUT TOC ANALYSIS

The actual ToC Analysis needs to carried out in a step-by-step manner. Too often, a theory of change is developed on the basis of the ideas and beliefs of those involved without much challenge and analysis. Without structured analysis and challenge, it is unlikely that a robust theory of change and the implications for intervention design would emerge. ToCA entails a careful examination of each element in the theory of change, how the elements fit together, and an assessment of the ToC weaknesses, data needs, and their implications.

ToCA would use the criteria in Table 1 as the basis for analysis, roughly in the order set out. There would likely be some interplay among results, assumptions, and the pathway. The findings could then be summarized in terms of implications around the two questions noted earlier.

Step 1: Overall Criteria

The initial analysis is to determine if there is indeed an actual ToC model to work with, and if the intervention seems at all plausible.

Understandable: If the ToC is hard to understand, such as if pathways are unclear or there is a proliferation of results, then rethinking and redrafting are needed so that there is something resembling a ToC with impact pathways.

Agreed: If there are different views as to how the intervention is expected to work, more discussion on the ToC is probably warranted. If differences persist, then it may be necessary to build more than one ToC and analyze each of them.

Level of effort: If the expectations for results are quite out of line with the level and nature of the activities being undertaken, there may be a need to rethink the design of the intervention or to reduce expectations to a more realistic level.

Step 2: Detailed ToC Analysis

The detailed ToC Analysis is best done result level-by-result level in sequence. That is, using the behaviour-based ToC model, in order:

- *Getting to Reach*: Will the outputs delivered reach the intended target groups with the right reaction?
- *Getting from Reach* to *Capacity Change*: Will the outputs delivered and their reach lead to the intended capacity changes?
- *Getting from Capacity Change* to *Behaviour Change*: Will the capacity change lead to the intended Behaviour Changes?
- *Getting from Behaviour Change* to *Direct Benefits*: Will the behaviour changes lead to the intended Direct Benefits?

- *Getting from Direct Benefits* to *Well-being Changes*: Will the direct benefits lead to the intended Well-being Changes?

If another ToC model is being used, the steps are the same: getting from one level to the next. For each level, an analysis of results and an analysis of assumptions would be done and a summary of findings set out:

Analysis of Results

Definition: If not well defined, need to further define terms.

Timing: If not sensible, suggests a structural change to the ToC needed.

Logical coherence: If not OK, suggests a structural change to the ToC needed.

Measurement: Indicate if *needed*, *might be needed*, *not needed,* and through what means. If strength of evidence is weak and the measurement important, suggests an issue to be addressed in the M&E Implications.

M&E implications: Brings together the M&E issues. Need to remember that not all results may need to be measured.

Analysis of Assumptions

Definition: If not well defined, need to redefine terms.

Logical coherence: If not OK, suggests the need for structural changes in the ToC.

Justification: If not necessary or not likely necessary, then the assumption should be dropped.

Realization: If realization is in doubt, then need to identify assumption as at-risk and set out confirming or corrective actions.

Sustainability: Similarly, if the assumption is found not to be sustainable, a corrective action may be needed, or, in an *ex post* case, the issue noted as a lesson learned for future similar or follow-up interventions.

Measurable: Indicate how measures would be taken and if *needed*, *might be needed*, *not needed.* If strength of evidence is weak and measurement important, suggest an issue to be addressed in the M&E Implications.

M&E implications: Need to remember that not everything need be measured.

Assessing the Causal Link

Independence: If assumptions are not independent, consider merging assumptions.

A sufficient set: If not a sufficient set, additional assumption(s) or more of the prior result are needed.

Strength/status of evidence: Indicate level of evidence for the link being realized.

Summary of Findings for Getting to a Result

The summary of the analysis can depend on the specific purpose and context, but in general can highlight (a) the changes needed to enhance the

robustness—structural soundness and plausibility—of the ToC, (b) the level of evidence there is on results and assumptions, (c) the actions that are needed to enhance the robustness of the intervention design, and (d) the M&E implications.

Structural Changes Needed
: Where the structural criteria for a robust theory of change are not met, structural changes are needed to the ToC to enhance its robustness—that is, changes in descriptions used, result statements, coherence, assumptions, and/or causal links. After any structural changes, we would want to conclude that the ToC is reasonably sound.

Strength/Status of Evidence
: Summary analysis can indicate the strength of evidence for (a) the result in question, (b) the assumptions associated with the link, and (c) the link being realized.

Additional Intervention Effort Needed to Enhance Plausibility
: If plausibility or sustainability of the ToC/intervention design is questionable due to at-risk assumptions identified and/or sustainability being questioned, then confirming or corrective actions are likely needed. We would want the analysis to conclude that, with the confirming/corrective actions, the intervention design is (or *ex post* would be) robust. Where the ToC is seriously contested, more than one ToC may needed to be developed and analyzed.

M&E Actions
: To monitor how well implementation is going or to verify the ToC in an evaluation, it is important to identify what data need collecting and the likely strength of the resulting evidence.

Conclusion: Overall conclusions for the specific link (component) in the ToC on robustness, level of evidence, and sustainability.

Theories of change are best developed in a participatory manner involving those designing/implementing the intervention and the evaluator (Mayne, 2015, pp. 137–138). During this development, of course, the criteria for a robust theory of change can be kept in mind. In other cases, the ToC Analysis is done on a completed theory of change, probably (although not necessarily) by the evaluator. The findings of the analysis should then be discussed with intervention implementers. This discussion may bring to light issues that were not, but need to be, included in the ToC, identify issues about the intervention design that need addressing, and/or identify data that need to be monitored or that need to be addressed in a planned evaluation.

AN EXAMPLE

To illustrate issues and concepts in ToC Analysis, I examined a previously used case of an intervention aimed at improving the nutritional diets of children through providing trainintg and information to mothers (Mayne, 2015). The ToC used there

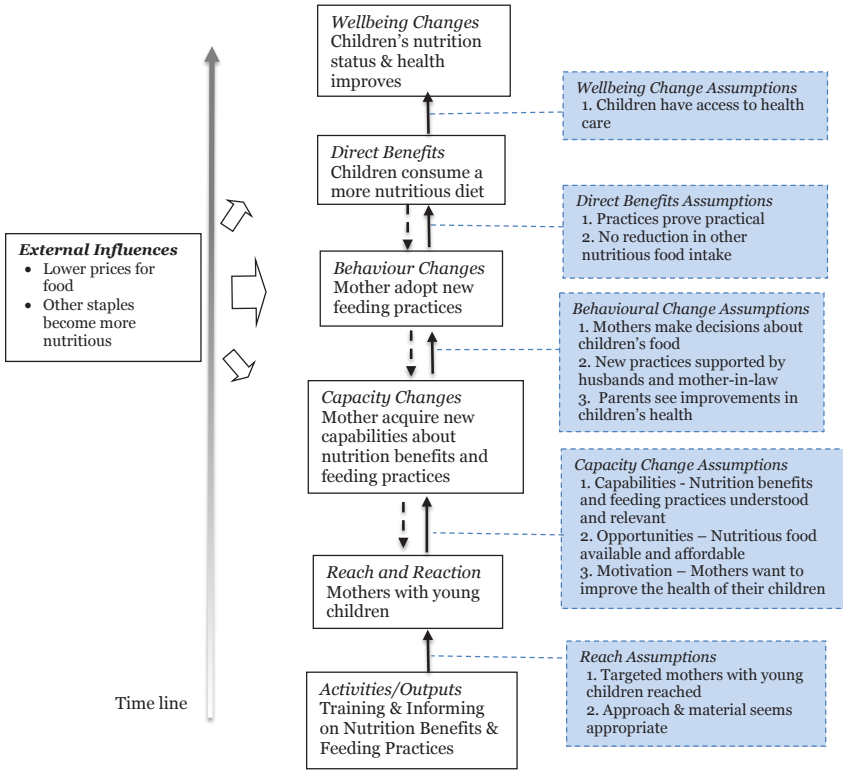

Figure 2. A Nutrition Intervention Theory of Change

is shown in Figure 2, with a few small changes to be consistent with the COM-B model: the motivation and food availability assumptions have been shown as capacity change assumptions rather than behaviour change assumptions.

The ToC in Figure 2 was used to carry out the ToCA. All the details are not provided here—but can be found in Mayne (2016b), as the analysis is quite lengthy. And that is worth a note. ToC Analysis is not a quick and dirty approach: it takes time, but not a lot, and patience to go through each criterion for each result and each assumption. But it can be worthwhile. Having developed the original example, I was not expecting many new insights, but I was wrong!

The findings from the ToCA are summarized below for each results level.

Getting to Reach and Reaction

Several needed *structural changes* were identified. The Reach result statement was not well-defined. What did "mothers with young children reached" mean? It could mean several things, such as mothers heard about the training, mothers were asked to participate, or mothers participated in at least the first session. I

assumed it was the latter case: reach and reaction was asking if mothers at least started the training and if they had a positive initial reaction. So the result statement needed to be changed.

Further, the first reach assumption was the same as the reach statement! Clearly there was a logical coherence problem. A new assumption, in fact two, were needed: targeted mothers are well identified, and targeted mothers can be communicated with. To get participation, the intervention needed to know who and where the targeted mothers were, and needed to be able to get the message to them about the nutrition training sessions.

If *sustainability* in the target area was an issue, then there would need to be a plan of how new mothers beyond the initial reach were to be reached, such as perhaps building into the training the need to spread the word within their communities.

There are two M&E implications: namely, the need to track the percentage of targeted population that initially participated, and to monitor initial reaction of participants.

Getting from Reach to Capacity Change

Several small structural changes were needed in the wording of the capacity result and assumptions (see Figure 3, where the changes are underlined).

Assumption 2 about the availability and affordability of nutritious food (opportunities) is possibly *at-risk* without more information. In Figure 3, at-risk assumptions are bolded. A useful *corrective action* would be to make local markets aware of the intervention and the expected increased demand for certain food products.

And a *confirming M&E action* is needed: the availability and affordability of nutritious food should be monitored during the life of the project.

Getting from Capacity Change to Behaviour Change

Behaviour change assumptions 1 and 2 overlap somewhat and may be *at-risk*. The intervention may need a better understanding about how decisions on food are made in households, and the sessions offered to households rather than only mothers.

M&E implications: Household surveys could track adoption of the new practices and general household support, and identify problems. Perhaps schedule a survey after 2 months and a follow-up 1 year later.

Getting from Behaviour Change to Direct Benefits

Assumption 2 about substituting other foods is *at-risk*. A *confirming action* could be to include this substitution issue in the nutrition training.

M&E implications: Follow-up household surveys could track children's dietary intake.

Overall, although implied by the timeline, Figure 2 did not set out a clear time frame for the intervention to have an impact. The level of evidence on realizing the ToC would be good, using the measures suggested.

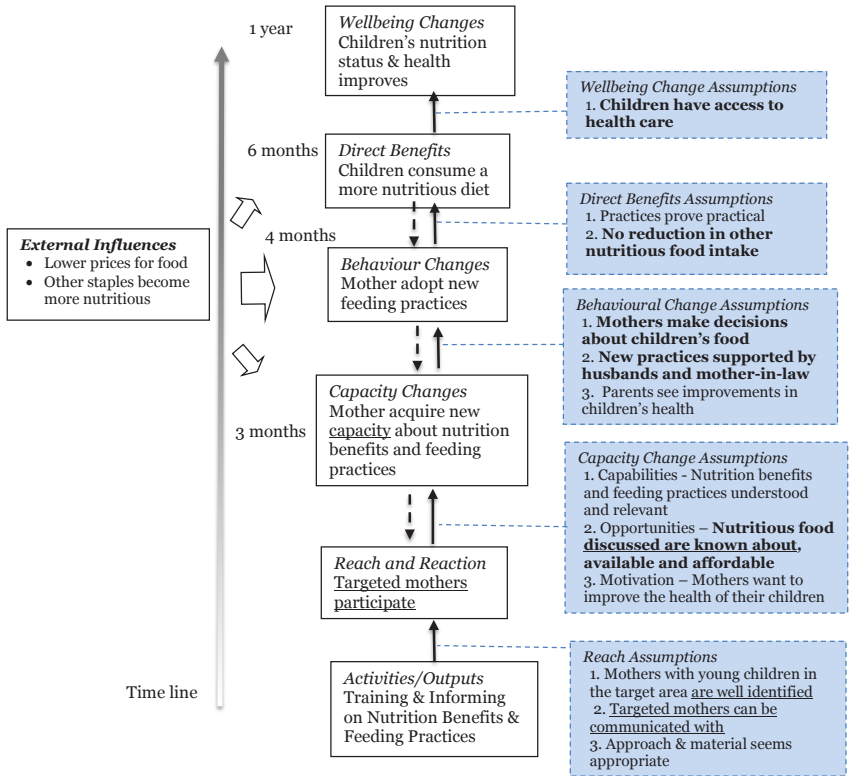

Figure 3. A Robust Nutrition Intervention Theory of Change

Based on this ToC Analysis, the revised and more robust ToC is shown in Figure 3. At-risk assumptions are shown in bold, and wording changes are underlined.

CONCLUDING REMARKS

Theories of change are the basis for theory-based evaluation approaches, such as logical analysis (Brousselle & Champagne, 2011; Rey et al., 2012), realist evaluation (Blamey & Mackenzie, 2007; Pawson, 2013), contribution analysis (Mayne, 2012), and process tracing (Schmitt & Beach, 2015). As such, the robustness of the theory of change used matters. A weak theory of change can only generate weak findings. For example, confirming a weak theory of change—one poorly structured with evident logical gaps—in contribution analysis cannot lead to credible causal contribution claims.

This article argues the usefulness of building robust theories of change and structured theory of change analysis, so that evaluation findings based on these theories of change are strengthened. ToC Analysis involves assessing a theory of

change against a set of criteria (Table 1) for each result, each assumption, and each causal link, challenging the structure and logic of the theory of change. The analysis takes some time and discipline to carry out. But it is mainly a desk review, and overall it entails hours rather than days of work. In my experience, it inevitably leads to improvements in the theory of change. The results are usually quite informative, leading to

- more robust ToCs,
- better intervention designs,
- useful M&E actions to help manage the intervention and support future evaluation, and
- *ex post*, more credible theory-based evaluations.

REFERENCES

Blamey, A., & Mackenzie, M. (2007). Theories of change and realistic evaluation: Peas in a pod or apples and oranges? *Evaluation*, *13*(4), 439–455. http://dx.doi.org/10.1177/1356389007082129

Brousselle, A., & Champagne, F. (2011). Program theory evaluation: Logic analysis. *Evaluation and Program Planning*, *34*(1), 69–78. http://dx.doi.org/10.1016/j.evalprogplan.2010.04.001

Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Schroter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*, *32*(2), 199–226. http://dx.doi.org/10.1177/1098214010389321

Davies, R. (2012). Criteria for assessing the evaluability of theories of change. http://mandenews.blogspot.com/2012/04/criteria-for-assessing-evaluablity-of.html.

Davies, R. (2013). *Planning evaluability assessments: A synthesis of the literature with recommendations* (Working Paper 40: DFID). Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/248656/wp40-planning-eval-assessments.pdf.

Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and application*. Mahwah, NJ: Lawrence Erlbaum. Retrieved from https://www.researchgate.net/publication/235930890_Program_theory-driven_evaluation_science_Strategies_and_applications

Funnell, S., & Rogers, P. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco, CA: Jossey-Bass.

Hansen, M. B., & Vedung, E. (2010). Theory-based stakeholder evaluation. *American Journal of Evaluation*, *31*(3), 295–313. http://dx.doi.org/10.1177/1098214010366174

James, C. (2011). *Theory of change review. A report commissioned by Comic Relief*. Available at http://mande.co.uk/blog/wp-content/uploads/2012/03/2012-Comic-Relief-Theory-of-Change-Review-FINAL.pdf.

Johnson, N., Mayne, J., Grace, D., & Wyatt, A. (2015). *How will training traders contribute to improved food safety in informal markets for meat and milk?: A theory of change analysis* (IFPRI Discussion Paper 1451). Washington, DC: International Food Policy

Research Institute (IFPRI). Available at http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/129293.

Leeuw, F. (2012). Linking theory-based evaluation and contribution analysis: Three problems and a few solutions. *Evaluation*, *18*(3), 348–363. http://dx.doi.org/10.1177/1356389012452051

Mayne, J. (2011). Contribution analysis: Addressing cause and effect. In R. Schwartz, K. Forss, & M. Marra (Eds.), *Evaluating the complex* (pp. 53–96). New Brunswick, NJ: Transaction Publishers.

Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, *18*(3), 270–280. http://dx.doi.org/10.1177/1356389012451663

Mayne, J. (2015). Useful theory of change models. *Canadian Journal of Program Evaluation*, *30*(2), 119–142. http://dx.doi.org/10.3138/cjpe.230

Mayne, J. (2016a). *The COM-B theory of change model: Working paper*. Retrieved from https://www.researchgate.net/publication/314086441_The_COM-B_Theory_of_Change_Model_V3.

Mayne, J. (2016b). *An example of ToC analysis*. Retrieved from https://www.researchgate.net/publication/305957815_An_Example_of_ToC_Analysis.

Mayne, J., & Johnson, N. (2015). Using theories of change in the CGIAR Research Program on Agriculture for Nutrition and Health. *Evaluation*, *21*(4), 407–428. http://dx.doi.org/10.1177/1356389015605198

Morton, S. (2015). Progressing research impact assessment: A "contributions" approach. *Research Evaluation*, *24*(4), 405–419. http://dx.doi.org/10.1093/reseval/rvv016

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.

Pawson, R. (2013). *The science of evaluation: A realist manifesto.* Atlanta, GA: Sage.

Peersman, G., Guijt, I., & Pasanen, T. (2015). *Evaluability assessment for impact evaluation: Guidance, checklists and decision support, A Methods Lab Publication*. London, UK: Overseas Development Institute.

Rey, L., Brousselle, A., & Dedobbeleer, N. (2012). Logic analysis: Testing program theory to better evaluate complex intervention. *Canadian Journal of Program Evaluation*, *26*(3), 61–89. Retrieved from https://evaluationcanada.ca/system/files/cjpe-entries/26-3-061.pdf

Rogers, P. (2007). Theory-based evaluations: Reflections ten years on. *New Directions for Evaluation*, *114*, 63–67. http://dx.doi.org/10.1002/ev.225

Schmitt, J., & Beach, D. (2015). The contribution of process tracing to theory-based evaluations of complex aid instruments. *Evaluation*, *21*(4), 429–447. http://dx.doi.org/10.1177/1356389015607739

Tremblay, M.-C., Brousselle, A., Richard, L., & Beaudet, N. (2013). Defining, illustrating and reflecting on logic analysis with an example from a professional development program. *Evaluation and Program Planning*, *40*, 64–73. http://dx.doi.org/10.1016/j.evalprogplan.2013.05.004

Trevisan, M., & Walser, T. (2014). *Evaluability assessment: Improving evaluation quality and use*. Thousand Oaks, CA: Sage.

Valters, C. (2014). *Theories of change in international development: Communication, learning, or accountability?* (JSRP Paper 17: The Asia Foundation). Retrieved from http://www.lse.ac.uk/internationalDevelopment/research/JSRP/downloads/JSRP17.Valters.pdf.

Vogel, I. (2012). *Review of the use of "Theory of Change" in international development:* Department for International Development (DFID). Retrieved from http://www.oxfamblogs.org/fp2p/wp-content/uploads/DFID-ToC-Review_VogelV4.pdf

Wilson-Grau, R., & Britt, H. (2013). *Outcome harvesting* (FordFoundation, MENA office). Retrieved from http://www.outcomemapping.ca/download/wilsongrau_en_Outome%20Harvesting%20Brief_revised%20Nov%202013.pdf

Woolcock, M. (2009). Toward a plurality of methods in project evaluation: A contextualized approach to understanding impact trajectories and efficacy. *Journal of Development Effectiveness*, *1*(1), 1–14. http://dx.doi.org/10.1080/19439340902727719

AUTHOR INFORMATION

John Mayne is an independent advisor on public sector performance. Over the past 13 years he has focused largely on international development evaluation and results-based management work. He has been working with several government, NGOs, and international organizations. He has authored numerous articles and reports, including several on contribution analysis, and co-edited eight books on program evaluation and performance monitoring. In 1989 and in 1995, he was awarded the Canadian Evaluation Society Award for Contribution to Evaluation in Canada. In 2006, he was made a Canadian Evaluation Society Fellow. Dr. Mayne's current research interests are on approaches for strengthening impact evaluation, useful theories of change, and dealing with attribution.