

# 3 My perspective on theories of change

*Markus Palenberg*

## Introduction

This essay lays out my personal understanding of theories of change and how I use them in practice. It highlights one aspect I consider important to render this a truly value-adding instrument to the evaluator's toolbox: the *perspective* taken when designing theories of change.

Before beginning, I must admit that I do not think the term *change* in the name *theories of change* is entirely appropriate. To me, it suggests that there must always be some form of change, which excludes the important possibility that a system may react by preserving its status quo—for example, when I pedal my bike and steer to avoid falling over. I will, however, continue using the term for lack of a simple alternative.

The essay is structured as follows. In the following section, I sketch my understanding of theories of change and introduce an example. I then describe two perspectives that can be taken when designing theories of change: the intervention and the system perspective. In the final section I argue in favour of the system perspective.

## How I understand theories of change

My way of thinking about—and making use of—theories of change is shaped by my education and research in physics, where models are used all the time. To me, theories of change are *qualitative conceptual tools* that can help *understanding*, systematically *analysing*, and *making predictions* about *cause-and-effect relationships* in the real world.

Theories of change do this by breaking down the cause-and-effect relationships of interest into smaller parts. They describe how these parts are causally related—to each other and to their context—and they explain what assumptions the model is based on.

Theories of change—like all models—have limitations in how well they can describe real-world phenomena. Even under the best of circumstances, a theory of change will only describe specific aspects of the real world. And it will only be able to do that for specific purposes. This limitation sets important boundaries

for the validity of any specific theory of change. In consequence, theories of change can only add value if their limitations are remembered when interpreting what they can tell us about the real world.

In more specific terms, I consider a theory of change to represent a simple *causal model* of the *reaction* of a *system* to *external influences*. Of course, these terms deserve further explanation:

- The *system* consists of selected parts and aspects of the real world that are of interest.
- With *model*, I refer to a systematic conceptual description that attempts to approximate the system at hand. Theories of change are *causal models* because they attempt to describe and explain cause-and-effect relationships within the system and—to some extent—also between the system and its surroundings.
- *Influences* are everything that affects the system. *External influences* originate from outside the system. These can be activities purposefully implemented with specific objectives in mind, such as the activities planned and implemented in projects, programmes, and institutional reform processes. External influences can also be any other activity, changing condition, or evolving situation originating from beyond the system boundary that affects the system in some way. In contrast, *internal influences* originate from within the system and describe causal connections between different parts of the system.
- Finally, in a theory of change, the *reactions* of the system to external influences are approximated by how the explicitly modelled states of affairs describing the system are affected (that is, how they change or remain stable under external influences).

Admittedly, this is still a rather generic and unspecific characterization of theories of change—partly on purpose. It reflects the flexibility I believe is required to make theories of change useful tools, so that they can best serve specific purposes at hand. The way I understand theory of change also draws on more rigorous and structured approaches, foremost on the causal calculus introduced by Judea Pearl (Pearl, 2009; Pearl & Mackenzie, 2018) but also on functional equation modelling and system dynamics.

As in those approaches, I usually summarize important elements of a theory of change by means of a diagram, as shown in examples below. I consider this an intuitive way to visually highlight the most important causal connections in the theory of change. In addition, I write up the theory of change as a narrative, referring to the diagram for illustration. When done, the diagram and the narrative should fully explain all the above-mentioned elements of the theory of change.

In my work to date, I have not made much use of theories of change prepared by others (that is, programme managers) prior to the commencement of

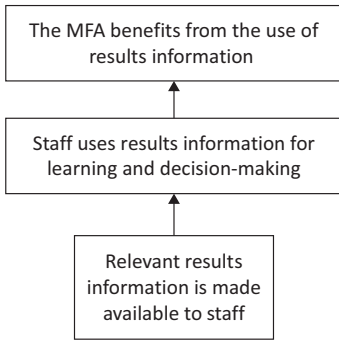


Figure 3.1 Initial theory of change

evaluations in which I have been involved. This may seem surprising because most of my evaluations are theory based. However, in most instances, there simply was no theory of change. In the few instances when a theory of change had already been developed, we (the evaluation team) usually did not find it particularly useful in relation to our evaluative information needs; hence, we ended up adapting it significantly or simply developed a new one, as explained in the following case.

#### *A theory of change for learning from results*

In 2019, I led an evaluation for the Ministry for Foreign Affairs of Finland (Palenberg et al., 2019). In this evaluation, the team was asked to assess how information about the results of Finnish development policy and cooperation had been used in the past and what could be done to strengthen evidence-informed learning and decision making in the future. There was no explicit theory of change. However, implicitly, the initial idea was that *results information*—the umbrella term we introduced for evidence of *what* results had been achieved and *how* they had been achieved—was what caused learning and decision making that, in turn, was thought to contribute to increased performance of the ministry towards its mandate (Figure 3.1).

However, once we understood the topic better it became evident that this was not the most important causal pathway. After a first round of interviews and literature review, we designed a theory of change around the basic causal relationships shown in Figure 3.2.<sup>1</sup> At its heart, this theory of change used an adapted version of the COM-B framework introduced by Michie et al. (2011).

1 This is a simplified version of the theory of change used in Palenberg et al. (2019).

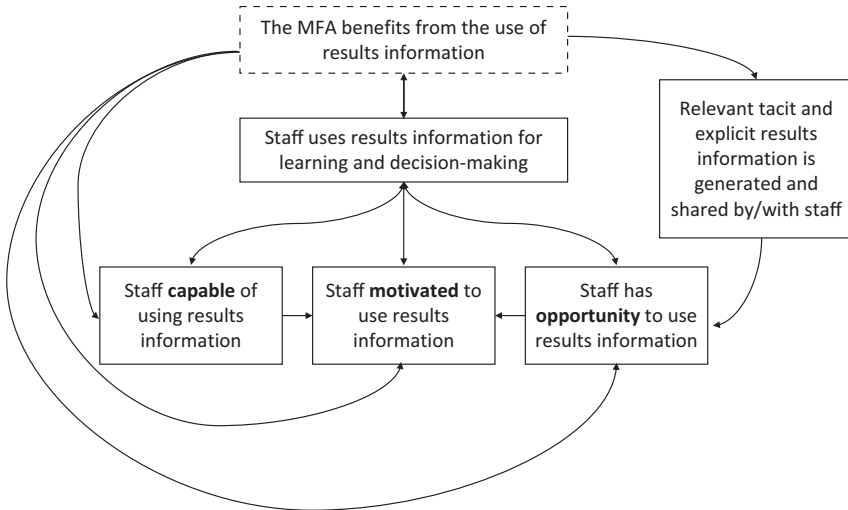


Figure 3.2 Theory of change including basic causal relationships

John Mayne, the evaluation team member leading the theory of change design, had also used this framework in other work (Mayne, 2019). It explains behaviour change—the ‘B’ in COM-B—by means of three groups of conditions: capability, opportunity, and motivation (COM), defined as follows (Michie et al., 2011):

- *Capability* is the individual’s psychological and physical capacity to engage in the activity concerned. It includes having the necessary knowledge and skills.
- *Motivation* is all those brain processes that energize and direct behaviour, not just goals and conscious decision making. It includes habitual processes, emotional responding, as well as analytical decision making.
- *Opportunity* is all the factors that lie outside the individual that make the behaviour possible or prompt it.

These three conditions for greater use of results information for learning and decision making were adopted as essential by the evaluation team. The initial starting point—the availability of results information—was reduced to one factor among several, feeding into the opportunity staff had for making use of it.

The full theory of change we used had more detail than what is shown in Figure 3.2. In addition to the diagram, it explained what we meant by what was in the boxes and what assumptions were behind the arrows connecting them. To

avoid overloading the diagram, we decided not to add detail on how all activities influence any of the boxes or the connections in the diagram (as in the case of the lowest box in Figure 3.1); instead, we listed them separately as ‘supporting activities’. We also decided that the causal step from use of results information (the middle box) to increased organizational performance (the top box) represented a hypothesis we would not be able to verify empirically (hence the dotted border).

We used this theory of change as the primary framework throughout the evaluation, adapting it repeatedly in light of what we learned. It guided our interviews and desk review, and it was used to structure our findings, synthesis, and sense making.

I bring up this example because it reflects two different perspectives that can be used when designing theories of change, as explained below.

## **Perspectives**

When designing theories of change, an obvious starting point is the specific ‘intervention’ at hand. This reflects typical evaluation questions, such as How effective was the intervention in contributing to intended results? or What was the impact of the intervention?

In such cases, the term *intervention* refers to a set of planned or already implemented activities that are part of a project, programme, or some other change process. During theory of change design one can then start with the intervention and explain how its activities cause subsequent effects. For this reason, I refer to this perspective as the ‘intervention perspective’.

In contrast, designing theories of change from a ‘system perspective’ entails first considering a change of interest, and then figuring out what may have caused it: the intervention but also other contributing factors.

In my experience, the intervention perspective is simpler, more often used, but often rather useless. The system perspective, on the other hand, is more complicated, less often used, but often very useful.

In what follows, I describe both perspectives in more detail.

### ***Intervention perspective***

One way to approach theory of change is what I call ‘forward explanation’. In this perspective, the intervention features as the root cause of subsequent change: starting from an intervention, the theory of change explains what immediate result(s) the intervention contributes to. Then, continuing along the causal chain, it explains what subsequent result(s) the immediate result(s) contribute to, and so forth (Figure 3.3). In an earlier publication (Belcher & Palenberg, 2018), Brian Belcher and I referred to this as ‘intervention perspective’.

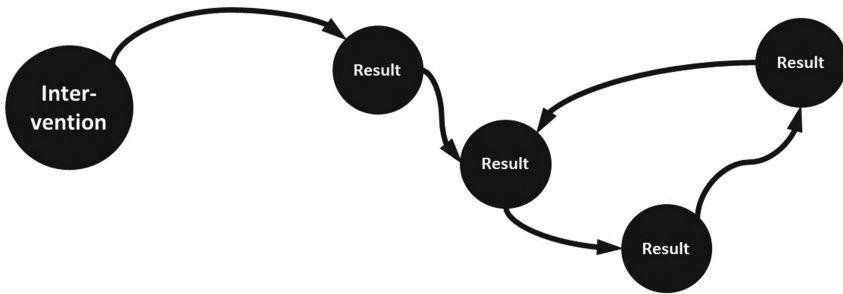


Figure 3.3 Example of a ‘forward explanation’ theory of change

The implicit, initial theory of change shown in Figure 3.1 is an example of an intervention perspective theory of change. It starts with the intervention of making results information available, which contributes to the intermediary result of staff using that information, which then helps the organization.

Theories of change designed from this perspective build on concepts like ‘results chains’ or ‘impact pathways’. In this perspective, possessive language (such as, ‘the results of the intervention’) is often used. This type of intervention-centric language is widely used in evaluation literature. For example, the OECD Development Assistance Committee’s Glossary of Key Terms in Evaluation and Results-Based Management defines results as the ‘output, outcome or impact of a development intervention’ (OECD, 2010), and the evaluation criterion of effectiveness is defined as the ‘extent to which the intervention achieved, or is expected to achieve, its objectives, and its results’ (OECD, 2021, italics added).

### *System perspective*

Another way to approach theory of change reflects a ‘system perspective’ (Belcher & Palenberg, 2018). It uses ‘backwards explaining’ of an observed or desired change by means of its contributing causes, as shown in Figure 3.4. The key difference to the intervention perspective is that the intervention has—*a priori*—no special status among the contributing causes.

In this perspective, all relevant causes that contribute to a specific change should be described. Here *relevant* means that these causes—taken together—explain most of the change.

The theory of change in Figure 3.2 is a system perspective theory of change. It models how a small part of the ministry’s evidence-based learning and decision-making system works and what factors contribute to its performance.

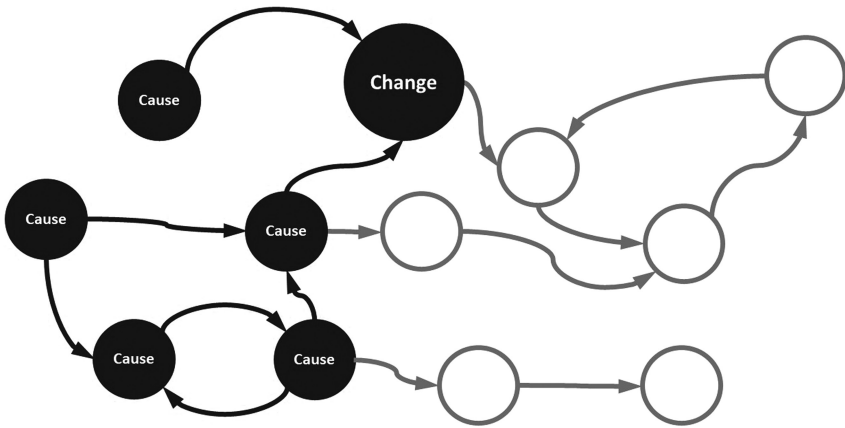


Figure 3.4 Example of a ‘systems perspective’ theory of change

**Which perspective works better?**

The intervention perspective is widely used. I believe this is because it reflects how we intuit simple cause-and-effect relationships in our daily lives. We say, for example, that we reduce our headache by taking an aspirin.

When we apply this type of intuitive monocausal thinking to more complex interventions and their effects, the result is an intervention perspective theory of change like the ones depicted in Figures 3.1 and 3.3.

Unfortunately, the validity range of intervention perspective theories of change is rather limited, making them ill-suited for answering the more complex questions of interest in the evaluations I have been involved in. Or, to keep it simple—but still beyond what an intervention perspective theory of change can handle—the question of how interesting you find reading this essay likely depends on many other causal factors beyond the fact that an aspirin reduced my headache while writing.

Strictly speaking, intervention-centric theories of change only serve their modelling purpose well when they describe cause-and-effect relationships that are indeed almost monocausal—that is, in which the intervention alone is the principal and dominant cause of the next result, which again is the principal and dominant cause of the next, and so forth.<sup>2</sup> However, in this special case, both perspectives will produce the same result.

2 This also applies if there are multiple root causes (for example, if different intervention activities are separately modelled) and if they contribute to more than one result at a time.

In all other cases, intervention perspective theories of change are prone to ignoring important contributing causes beyond the intervention. This is because they are guided by the question, How effective was my intervention? which risks putting blinkers on the evaluator's eyes.

For example, if we had conducted the entire evaluation referenced above of learning from results based on the initial understanding that the availability of results information represented the primary driver of all further change (Figure 3.1), then we could have easily missed out on the important elements of capability, opportunity, and motivation, which, in the end, turned out to be a good description of key enabling factors for the desired behaviour change. The risk of missing out on important causal factors would not go away if we had reflected them in assumptions on the side because the intervention-centric theory of change would still suggest that making results information available represents the principal driver of results-informed learning and decision making.

In contrast, system perspective theories of change are guided by the design question: What contributes to this change? They are therefore less likely to miss out on important contributing causes. We experienced this when we moved from the implicit, initial theory of change in Figure 3.1 to the more informed theory of change in Figure 3.2.

The intervention perspective theory of change in Figure 3.1 started with the silent assumption that the provision of results information would trigger its use. There might be other conditions and assumptions, but the causal relation between making the information available and using it would be the gist of it. When looking at what really contributes to results-informed learning and decision making, we quickly realized that this initial model was too limited and that we were missing out on other important factors. We therefore took a system perspective and started from the desired (behaviour) change: staff making use of results information. Then we checked the literature and found a simple model for explaining behaviour change. Suddenly, there were several additional candidates for contributing causes to that change. Importantly, the system perspective forced us to widen our focus from the single activity of making results information available to a wide range of other activities targeting any box and/or connection in the theory of change in Figure 3.2.

The relative importance of these other activities in contributing to results-informed learning and decision making was high. When we finalized our evaluation work, only one of seven conclusions and only two of eight recommendations concerned the availability of results information.

What does this tell us? For me, it primarily serves as a stark warning against oversimplifying theories of change.

In virtually all cases in my work, observed or desired change cannot adequately be described by a chain of monocausal cause-and-effect relationships. Instead, describing change as a consequence of multiple interacting causes—some of



which are related to the specific intervention at hand while others are not—often does the job.

Developing useful theories of change therefore requires using a system perspective. Despite much evaluation literature using intervention perspective language—such as ‘the intervention’s results’—we must consciously avoid ignoring other contributing causes because they often turn out to be more relevant and important than the ones we initially had in mind.

## References

- Belcher, B., & Palenberg, M. (2018). Outcomes and impacts of development interventions: Toward conceptual clarity. *American Journal of Evaluation*, 39(4), 478–495. <https://journals.sagepub.com/doi/10.1177/1098214018765698>
- Mayne, J. (2019). *The COM-B theory of change model*. Working paper. An earlier version of this paper can be retrieved from [www.researchgate.net/publication/323868561\\_The\\_COMB\\_ToC\\_Model4](http://www.researchgate.net/publication/323868561_The_COMB_ToC_Model4)
- Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6, 42. [www.implementationscience.com/content/6/1/42](http://www.implementationscience.com/content/6/1/42)
- OECD. (2010). *Glossary of key terms in evaluation and results-based management*. OECD Publications.
- OECD. (2021). *Applying evaluation criteria thoughtfully*. OECD Publishing. <https://doi.org/10.1787/543e84ed-en>
- Palenberg, M., Bartholomew, A., Mayne, J., Mäkelä, M., & Esche, L. (2019). *How do we learn, manage and make decisions in Finland’s development policy and cooperation: Management of results information and knowledge at the Ministry for Foreign Affairs of Finland*. Final Report. Ministry for Foreign Affairs of Finland. Retrieved from [https://um.fi/publications/-/asset\\_publisher/TVOLgBmLyZvu/content/evaluointi-tietojohdamisesta-miten-opimme-johdamme-ja-teemme-paatoksia-suomen-kehityspolitiikassa-ja-yhteistyossa-/384998](https://um.fi/publications/-/asset_publisher/TVOLgBmLyZvu/content/evaluointi-tietojohdamisesta-miten-opimme-johdamme-ja-teemme-paatoksia-suomen-kehityspolitiikassa-ja-yhteistyossa-/384998)
- Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books, Hachette Book Group.