Journal of MultiDisciplinary Evaluation
Volume 15, Issue 32, 2019

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
http://www.jmde.com

Theories of Change: Making Value Explicit

Steve Powell
ProMENTE Social Research, Sarajevo

Background: This article addresses two problems. The first is the Flexibility Problem: If we are to use a more flexible format for theories of change than for traditional logic models, one in which we can no longer assume that we only value things which are at the end of causal chains, nor that we intervene on all the things at the beginning of causal chains, how then can we show which things we value, and which things we intervene on? The second is the Definition Problem: What is the difference between a theory showing the causal influences within and around a project and, more specifically, a theory of change for the project?

Purpose: To solve the Flexibility Problem and the Definition Problem.

Setting: N/A

Intervention: N/A

Research Design: N/A

Data Collection and Analysis: N/A

Findings: A definition of "Theory of Change" is introduced, based upon a definition of "Theory" together with two symbols to mark variables we value ("❤", or any suitable alternative symbol) and variables we intervene on ("▶", or any suitable alternative symbol). These two definitions and the two symbols together answer both the Flexibility Problem and the Definition Problem, and have some interesting side-effects as follows. Firstly, they suggest that it is the task of evaluators to model how stakeholders value aspects of a project just as much as it is to model the causal chains within a project. Secondly, evaluators are able to model the fact that stakeholders may value variables which are not at the end of a causal chain, throwing a new light on the debate between results-based and principles-based programming. Thirdly, they provide a way to understand the behaviour of stakeholders and stakeholder groups in terms of their own theories of change – the way they view the world and how they can get what they want – rather than from the traditional behaviourist perspective more familiar to most evaluators.

Keywords: *Theories of Change; valuing; theory.*

The Flexibility Problem: Going Beyond Logic Models

In most traditional logic models and logical frameworks, the items are ordered into a neat system of layers ("inputs", "outputs" etc) in a strict hierarchical format (see for example Coleman, 1987; Department for International Development, 2011).

Before going any further, it will be useful to introduce some technical terms. We will call the items within a logic model, theory of change, etc., "variables" . Also, we will use terms based on those used by Pearl (2000, p. 13) to refer to variables according to their place within a network of variables, such as the network shown in Figure 1 and the other diagrams in this article:

- any variable which has an arrow pointing to another variable can be called a "parent" of the latter;
- any variable which has an arrow pointing to it from another variable can be called a "child" of the latter;
- any variable in a Theory which has no "parents" (i.e., there are no arrows pointing to it) can be called a "no-parent" variable;
- any variable which has no "children" (i.e., it has no arrows pointing away from it) can be called a "no-child" variable;
- any variable in a Theory which is neither a "no-parent" variable nor a "no-child" variable can be called an "intermediate" variable.

In a traditional logic model, the "no-parent" variables are the ones upon which we intervene and the "no-child" variables are the ones we value. Often these models are presented graphically in a bottom-up sense, in which case the no-parent variables are all those at the bottom, usually given names like "Inputs" or "Activities", and all the variables which we value (the ones we care about, the ones we want to change) are at or near the top – usually given a name like "Goal".

Value is a central part of evaluation, perhaps its very core, as Scriven (2012) has argued persuasively. When looking at a Theory of Change, the question of what we value is just: "Which are the things we want to influence, which motivate us, and without which we wouldn't bother with the whole project?" When we look at the graphical representations of project theories, we find that the underlying, implicit notion of value represented is a simple hierarchical one (even if the causal theory represented is not hierarchical). Some authors such as Dhillon and Vaca (2018, p. 8) do provide symbols for "result" or "outcome" for which a use is not necessarily dictated by position within the causal network. More generally, however, we do not have the differentiated tools we need to represent such an important concept as value within our theories. The method nearly always presented for constructing a theory of change is "backwards mapping" (Anderson, 2005, p. 12): to start with the variable(s) which we want to influence and work backwards from them. This procedure tends to pre-empt the question of the value of the intervening steps. Intermediate variables are characterised merely as "preconditions", getting their value, if they are valued at all, automatically, because they serve as means to an end.

Rigid planning formats, popular as they are, have been widely criticised (Chambers & Pettit, 2004; Davies, 2004, p. 104; Earle, 2002, p. 2) as too restrictive to be useful for accurately modelling a wide variety of project theories and the broad range of different factors which may influence them. Specific criticisms are levelled at the use of a fixed number of "layers" (Inputs, Outputs, Outcomes etc); at the insistence that one variable may have only one "child" (Davies, 2004, p. 111); at the exclusion of factors beyond the control of the project (Mayne, 2015, p. 224); amongst other issues. A variety of more flexible templates have been introduced which address some of these issues, some of them called "theories of change". For example, see Anderson (2005).

This gives rise to a problem I call The *Flexibility Problem*. If we are to use a more flexible format for theories of change than for traditional logic models, one in which we can no longer assume that only the no-child variables are valuable, nor that we intervene on all the no-parent variables, how then can we show which variables we value, and upon which we intervene? There is no systematic, accepted way to do either of these things, although individual project models sometimes employ various symbols, colors or other strategies for these purposes. Dhillon and Vaca (Dhillon and Vaca, 2018, p. 80), in a recent paper on the graphical presentation of theories of change, suggest the use of a special symbol for "activities", although they use it not only for "no-parent" variables but also for intermediate variables.

More generally, if we provide a way to explicitly mark which variables are valued, rather than assuming without discussion that they get their value automatically due to their position within a project diagram, can this help facilitate practical and theoretical discussion about value within

theories of change and within evaluation more broadly? For example, Davidson (2015, p. iii), following Scriven, gives a good summary of the argument that it is not enough in an evaluation to simply report scores on variables, or even differences made by interventions to scores on variables, but to evaluate those scores: simply put, is this score good enough? Following this line of argumentation, is it possible to represent within a theory of change not only which variables are valued but which levels of achievement on those variables count as, for example, inadequate, acceptable or outstanding?

The Definition Problem: from Theory to Theory of Change

The second problem addressed by this article, the *Definition Problem*, is as follows: What is the difference between a theory showing the causal influences within and around a project and, more specifically, a theory of change for the project? Can we provide definitions of "Theory of Change" and "Theory" which show how the two are related?

Theories about how projects or programs are supposed to work (Chen, 1990; Weiss and Others, 1995) have often been hailed as a central, or even the central, concept within evaluation, and for good reason. But there is no universal or even dominant definition of what, in evaluation, constitutes a theory, or of the difference between a theory and a theory of change – if, indeed, there is one. These two terms are far from enjoying the kind of consensus (Vogel, 2012, p. 3) which is the case with, say, the OECD-DAC definitions of "effectiveness", for example.

Weiss (1995) defined a theory of change as "a theory of how and why an initiative works", which is not so different from that given by Scriven (1981): "a hypothesis about the way that a program brings about its effects." These are a good start. Can we go on to relate "theory of change", "logic model", and "theory" to one another in a more satisfying and useful way?

Some authors treat the traditional logic model as just one part of a theory of change, namely the part which concerns only the intervention variables and their direct causal consequences, ignoring other influences. For an overview of this issue, see Blamey & Mackenzie, (2007, p. 445). Mayne (2015, p. 3), also referring to Patton (2008, p. 336) says that what turns a logic model into a theory of change is just precisely the addition of "no-parent" variables beyond the control of the project, which he calls "causal assumptions". Dhillon and Vaca (2018) on the other hand suggest that including such assumptions is one factor which turns an ordinary theory of change into a "strong" theory of change.

In the present article I intended to identify, as far as possible, general distinctions and definitions based on structural characteristics, in the belief that these should prove more robust than those driven by specific evaluation contexts, issues or agendas. For example, Mayne (2015, p. 122) explores the specific layers which many real-life theories of change will usually include – "behavioural changes", "direct benefits", etc. These distinctions are useful in many contexts but may not be applicable in every case.

Limitation

In this article I will focus only on the formal, structural aspects of what constitutes a Theory of Change and not on the many important political and practical questions such as who constructs theories of change, for whom , why, and using what processes. Having said that, providing ways to make value visible within Theories of Change would certainly facilitate political, practical and ethical discussion about who values what – within and between stakeholder groups. I will also ignore issues of measurement and how one could, or should, measure the variables within a theory of change, or their value. Finally, my driving interest is in the concepts underlying theories of change and not primarily in the also important issue of how exactly we should or could visualise them graphically, as for example discussed by Dhillon and Vaca (2018) and Vaca and Vidueira (2016). The "heart" and "action" symbols introduced here are basic "Unicode" symbols (available to copy and paste as text on different computing platforms) which are generic and easy to draw by hand. In real-life applications, more attractive realisations of these ideas could be used. The actual symbols are unimportant for the focus of this article, which is to make value explicit within Theories of Change.

Definition of "Theory"

The definition of "Theory of Change" which I present in the next section below is, quite literally, theory-based: a Theory of Change is a special type of Theory. So first it will be necessary to provide a definition of "Theory". This definition is based loosely on the work of Judea Pearl (Powell, 2018; Pearl, 2000; Pearl & Mackenzie, 2018), who has provided a formal and mathematically rigorous (yet

non-parametric) treatment of the kind of causal networks which may underlie theories of change, as well as a set of related tools to reason about causality.

I introduce the term "Theory" (with a capital "T") as a new concept which is intended to be close enough to most existing usages of the familiar word "theory", as follows:

> A "Theory", in some particular context, is a model (e.g. a belief, claim or hypothesis; a description, map or picture) presenting how two or more variables (causally) influence one another: what leads to what.

Figure 1 shows a simplified Theory about the influence of a training course for some hypothetical new teaching method, first upon teachers' creativity, and then upon student achievement.

Definition of "Theory of Change": a Theory plus value (❤) and intervention (▶)

Now at last, we will (loosely) define a "Theory of Change":

> A Theory of Change is somebody's Theory which shows how their intervention on one or more variables (marked with "▶") causally influences variables which they value (marked with "❤"). In other words, it shows how they can get what they want.

This definition highlights what makes an ordinary Theory into a Theory of Change. I use capital letters to clarify that "Theory of Change" is another new concept: I am not claiming that this is the definition which everyone meant or should have meant all along when they talked about "theory of change". But this new concept is close enough to previous concepts of "theory of change" to be a useful possible replacement.

Figure 2 shows that we intervene on one variable (Amount of training which teachers receive on the new teaching method, marked with a "▶"); our intervention sets this variable to "complete package delivered" rather than "nothing" and it shows what we value (Student academic achievement, marked with a "❤"); and it shows how these two are connected.

This definition is formulated in terms of a specific "someone" – an agent who believes, values things, and intervenes. So "a Theory of Change" is, at least from a linguistic point of view, just like "a plan" – a plan has to be somebody's plan, and if they don't believe it, it isn't their plan . This leaves open the possibility of separate or even overlapping theories which include different agents who may differ in what they believe, control or value. For the moment, to keep things simple, we will just assume that it is "we" who do the believing, intervening and valuing. Nevertheless, the important point is that constructing or reconstructing a Theory of Change involves two difficult modelling tasks: modelling not only (somebody's) causal theory but also (somebody's) "valuation theory" – at this point,

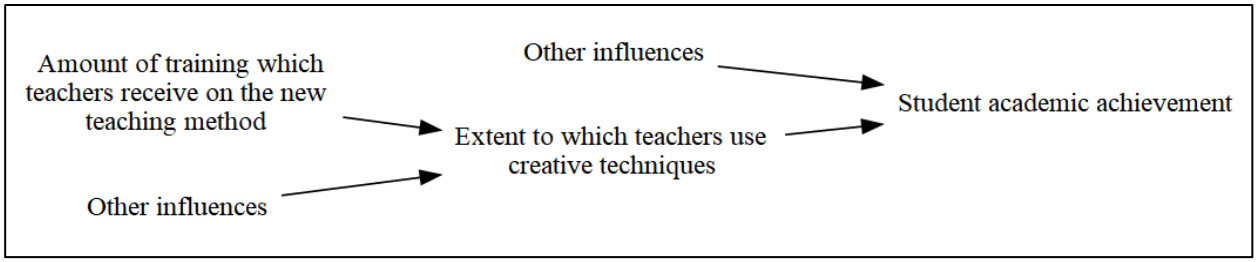

Figure 1. A Simplified Theory About the Influence of a Training Course

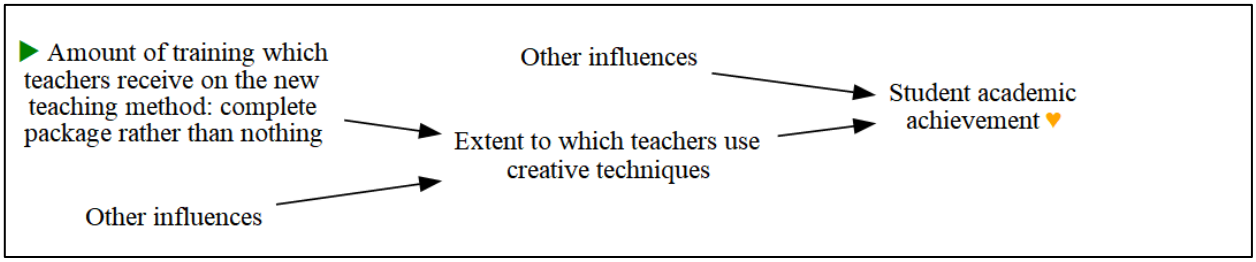

Figure 2. A Simplified Theory of Change Showing "Heart" and "Intervention" Symbols

"the valuation theory" is simply the set of variables marked as being valuable, but I generalise this idea below.

As defined here, if a four-year-old child plans (whether implicitly or explicitly) to whine and whine until her father buys her ice cream, she has a Theory of Change . The conception of "Theory of Change" presented here is completely agnostic about the kinds of Theory involved and about the kinds of steps which are likely to be involved in a particular case. As such this is very different from that taken by Mayne (2015, p. 122). My aim is to find some general principles for modelling how people plan and implement behavior so there is no need to restrict our understanding of "Theory of Change" to, for example, only projects which are intended to benefit society, improve human rights, etc. I have also made no use of the typical semi-technical terms ("output", "intermediate result", etc.) in the present article as these terms are often useful in specific circumstances but can be very difficult to use generically.

For many organisations, a "Theory of Change" is often primarily a broad and inspirational overview. If we can understand such an overview as a map of what influences what and how we can intervene to get what we want, I would still understand it as a Theory of Change. If it is nothing but an inspirational picture, I would not.

Value, intervention and Theory of Change mutually define one another: what someone values is what they try to maximise, given the opportunities they have to intervene and the theory they have about what leads to what. If one knows the Theory of Change which a particular agent has, one has a good chance of working out what they value simply by observing where and how they intervene in different circumstances.

Valued variables should be of the form "more is better". I suggest only marking a variable as valued, with the "❤" symbol if it is of the form "more is better": it should be ordered (i.e., we should be able to distinguish more of it from less of it), and more of it should always be better in terms of the way we value it. In a specific context we might mark "number of required vaccinations actually received by the child" with a "❤" because more of it is better, whereas we would not do this with "body-weight of the child"; although the latter is an important variable in many contexts, more of it is not always better.

Possible extensions to this rule are discussed below.

Predicting the difference made by an intervention. When we intervene on a binary variable like "the grant is awarded, false/true", an intervention simply sets it to "true" (the state of the variable with the intervention, aka "factual" or "intervention" state) rather than "false" (the state of the variable without the intervention, aka "base" or perhaps "counterfactual" state). Similarly, we can think of an intervention on a continuous variable as setting it to one level rather than another – but as the variable has more than two levels, it will be necessary to highlight which level is the "intervention" level and which is the contrasting, or "base", level (for example, holding 20 training workshops rather than zero workshops). Generally, a single intervention will set the levels of several "no-parent" variables in this way, and it might use feedback from downstream variables to adjust the release of resources in response to need; and these variables might stretch or repeat over time, and so forth.

Real-life theories of change nearly always show not only, trivially, the difference made by the implementation on the intervention variables themselves but also the likely consequences on all the variables downstream of them which can be deduced using causal inference. Figure 3 does this, though in a rather vague and limited way.

In Figure 3, the variable names include very minimal information about the differences made under implementation ("receive", "increase", "improve" ...) – so minimal that it would be quite hard to falsify this Theory (how small does an

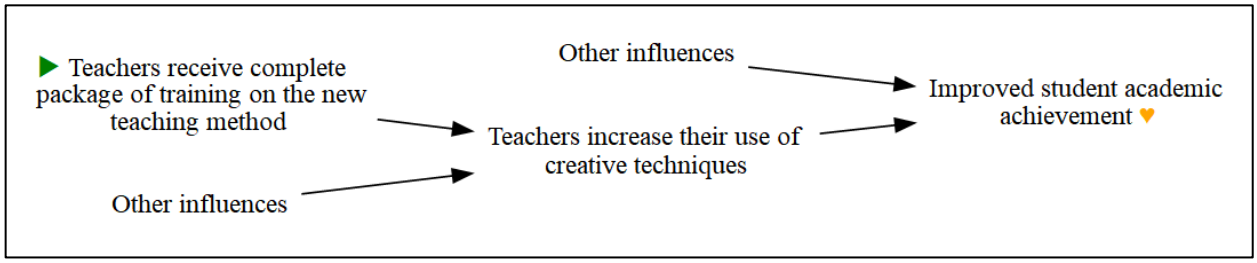

Figure 3. A Simplified Theory of Change in Which the Labels of the Variables are Expressed in Terms of the Differences Made by the Intervention

improvement have to be in order to still count as an improvement?). In general, it would be desirable to estimate more precisely the difference made to the downstream variable(s) by the intervention, given what we know about the relevant causal influence.

The concept of making a difference was central to the work of the philosopher Daniel Lewis, who revolutionised ideas about causation: "We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it" (Lewis, 1973, p. 557). Naively, we can think of this "difference made" as a change over time, as the difference between a baseline and an endline score on a variable. In practice of course, baseline scores often change anyway, even without our intervention: the real "difference made" which should interest evaluators is the difference between the actual or "factual" score and the "base" score which would have been observed if the intervention (had) never happened. "Theory of Change" is actually a misnomer – we should be talking about "Theory of Difference-Making" – but it is probably too late to do anything about this now.

In practice, Theories of Change rarely explicitly mention the strength or nature of the causal connections or the evidence for them. Instead, the process is usually reversed: we simply guess, with or without evidence, when constructing the Theory, that a certain difference made to one variable (e.g., giving the complete package of training rather than nothing) should be enough to cause some given difference to another (say, 15% more creative techniques), resulting in diagrams like Figure 4. Working backwards from such a diagram, our beliefs about the power and nature of the causal connections can be deduced from the variable labels, which include phrases like "X rather than Y" or "Z% improvement" – in terms of differences which we hope will be made to them.

In some cases, we can treat this "difference" as a literal subtraction of numbers. If the percentage of children vaccinated is 90% following some intervention, and we would expect only 70% without it, we can say the intervention made a difference of 20%. In this case we might see a variable labelled like this: "Percentage of children vaccinated improved by 20%". In other cases, the difference made cannot be literally expressed as a subtraction and we might prefer to specifically mention both the intervention state and the contrasting base state, perhaps using the phrase "rather than". For example, "teachers frequently use a whole range of creative techniques (rather than hardly ever using any)". Most real-life theories of change gloss over these niceties and simply combine the name of the variable and the difference made into something as vague as "Improved creativity", leaving any details for the more boring parts of the planning documentation.

To summarise the ideas in this section: interventions consist in setting the levels of one or more variables, marked with a "▶" symbol (or other suitable symbol), to one level ("the intervention level") rather than another ("the base level", which they would have had without the intervention). Often, intervention variables are in any case expressed in binary false/true form, in which case the intervention simply sets them to true rather than false ("do this" rather than "do not do this"). Usually, when Theories of Change are presented, the consequences of this intervention on all the downstream variables (according to what we know about the nature, strength etc. of the causal links) is actually noted as part of the variable label in more detail (e.g., "20% increase in vaccinations" or even "12,000 rather than 10,000 vaccinations") or less detail ("increased vaccinations"). In particular we are interested in differences made to the variables we value, which are marked with a "♥" symbol (or some other suitable symbol).

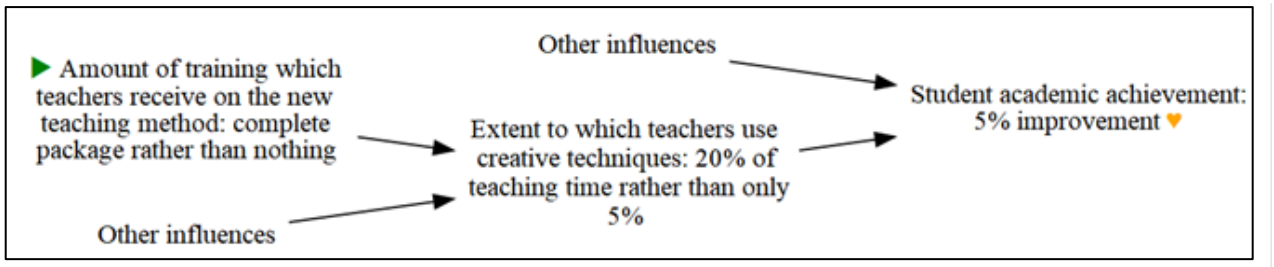

Figure 4. A Simplified Theory of Change in which the Labels of the Variables are Expressed in Terms of the Differences made by the Intervention, with More Detail Given About the Differences

Where do Valued Variables and Intervention Variables Appear in a Theory of Change?

Marking some "no-child" variables as intervention variables. Within logical frameworks or logic models, all the "no-parent" variables are also intervention variables (the points where we can intervene). Theories of change are distinguished from such models in part because they explicitly include "no-parent" variables which are not intervention variables - at least not by "us".

The suggestion made here is to formalise this distinction by marking intervention variables with the "▶" symbol and leaving other "no-parent" variables unmarked, whether we think of these as completely external influences such as the weather, or as those parts of an explicit "causal package" (Mayne, 2015, p. 124) upon which we ourselves do not intervene.

We will only use the "▶" symbol on "no-parent" variables, and not on intermediate variables. This is because there is something incoherent about saying "We decided to do X; but also, our decision was caused by factors A, B and C".

Intermediate variables can be valued. As mentioned above, it is usually assumed without discussion that only no-child variables can be ultimately valuable. If intermediate variables are treated as valuable at all, it is only in virtue of their being links in a chain, and in proportion to their distance from the end of it. Project management may be criticised for focusing "merely" on intermediate variables. There may be good reasons for this kind of criticism, such as when someone has been trying to imply that project success consists in, say, attendance at endless workshops. It is often right to ask about an intermediate variable: "What is the real use of that?" But it is wrong to express this point by banning anyone from recording intermediate variables as valued. There really are times when we need to be able to say "Alongside improving student outcomes, we also want the teachers to use creative techniques (an intermediate variable); this is simply something we value in its own right", as shown in Figure 5. Perhaps we have a conception of teacher professional development in which this is a key factor; perhaps we just love all things creative. In other words, we might value it because it is the kind of thing which contributes to things like the valued variable causally downstream, or perhaps because it might causally contribute to some other valued variable which has not been mentioned, but (most interestingly) for no reason at all: we simply value it.

It is surprising how seeing a variable as a link in a causal chain which leads to something valued seems to discourage us from considering it, at the same time, as valuable in its own right. But there is nothing logically incoherent about someone valuing an outcome which also causally reinforces another outcome which they also value. Of course, a sceptic can say of any variable, not just an intermediate one, "aha, but why is that valuable, why is student academic achievement valuable, what does that lead to, what is it good for?" And indeed, this might provoke a substantive discussion about work-life balance, the problems of stress in young people, etc. But we must know where to stop and accept that some things just are valued in their own right. Otherwise we will fall into an "infinite regress" in which the sceptic keeps asking "but why is that valuable?", and the discussion never ends. As Wittgenstein wrote, (1978, p. 3) "Explanations come to an end somewhere".

For any of these reasons, if in fact we (also) value an intermediate variable we can mark that with a "♥" symbol, as is the case with the variable concerning creative techniques in Figure 5.

There is nothing intrinsic to any given variable (neither its duration, nor its purported sustainability, nor its position in a diagram, nor what it involves – e.g., human behavior or technical achievements) which says whether or not it is suitable for being valued by us. A variable is only

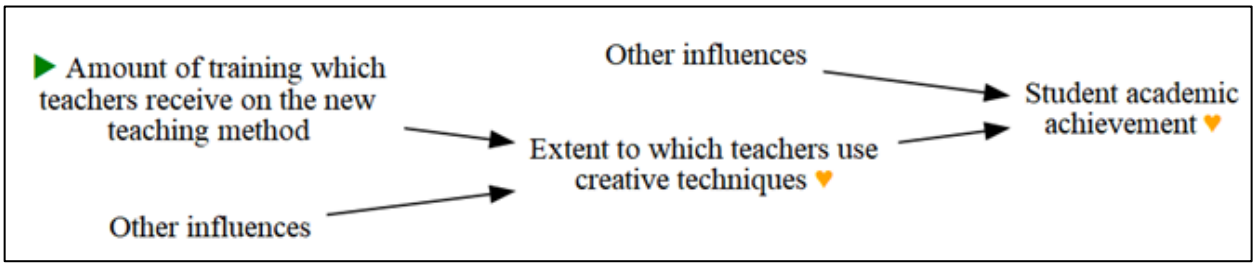

Figure 5. Simplified Theory of Change in which an Intermediate Variable is Valued as well as a "No-child" Variable.

valued if someone happens to value it, as the thought experiment in Figures 6 and 7 remind us. Each of these two stories is plausible in its own context. In the first, the students are told that winning games is ultimately only important because it promotes work ethic and team spirit; in the second, they are told that team success is what really counts and that is the real reason for promoting work ethic and team spirit. Both stories are perfectly coherent.

Some philosophers have suggested (see Donnelly (2013) for an overview) that there is in fact a right and proper place for this kind of "why is that valuable?" questioning to end. Such a place might be for example, "human rights" Some approaches to project planning do implement this idea in practice, as in "Rights-Based Approaches". Adherents of this theory would claim that variables are only ultimately worth influencing if they contribute to improving or protecting human rights. Other authors suggest concepts such as "well-being" as typical or generic final carriers of value for most theories of change (Mayne, 2015, p. 123).

But limiting our theories of change in this way means limiting their generality. For example, if we say that only Human Rights are ultimately valuable, there could be no theory of change of, say, a biodiversity intervention unless one really wanted to perform the acrobatics of arguing that all biodiversity interventions are actually valuable because, and only because, they improve well-being or human rights. In the present article I am taking the stance that evaluators need a tool-kit which, having helped them to model the project theory, also allows them to model the way an arbitrary stakeholder or client happens to value things, without telling them what they ought to be valuing. (This stance does not imply that just because someone says they value some arbitrary thing, they do in fact value it. Firstly, Claiming that something is valuable is not a fact-free choice separate from what others value. Secondly, it only makes sense if the person's actual behaviour is generally in alignment with the claim. We touch on this point again below).

We allow intermediate variables to be valued. This might seem like a relatively trivial step, but it can have significant ramifications for the way we think about, monitor and manage projects. This step takes us beyond any "Results-Based" approach (Kusek & Rist, 2004) in the sense that it is not restricted to valuing only the ends of causal chains. In this way it can also encompass approaches like Outcome Mapping (Earl, Carden, & Smutylo, 2001), which specifically values "Progress Markers" – changes in the attitudes and behaviour of immediate partners (Guijt, 2008, p. 2) – acknowledging that these changes may well, hopefully, go on to cause other desirable changes, but essentially shifting the focus from such "ends" to the progress markers themselves. More broadly, this approach encourages us to re-visit the importance of how we do things (such as acting out of principle) as well as acting in order to reach a particular goal. "Acting out of principle" is a value which is in a constant tension with the goals it might help reach further down a causal chain; it is meaningless without them but cannot be reduced to them.

This step also encourages us to free ourselves from other restrictions imposed upon project planning by hierarchical templates such as those involving timing and duration. If we no longer automatically assume that the value of a variable comes purely and automatically from the number of links separating it from the end of a causal chain, we also no longer need to assume that the most

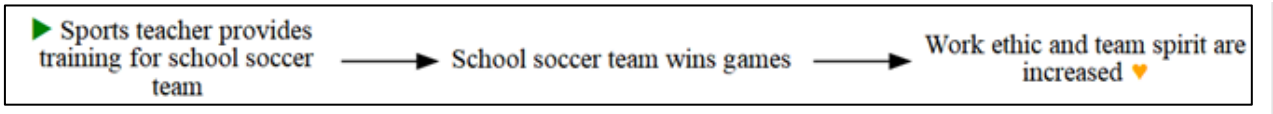

Figure 6. Simplified Theory of Change in which success of school soccer team is explained as being important only because it promotes work ethic and team spirit. To that end, the teacher intervenes to provide training designed primarily to improve chances of winning games.

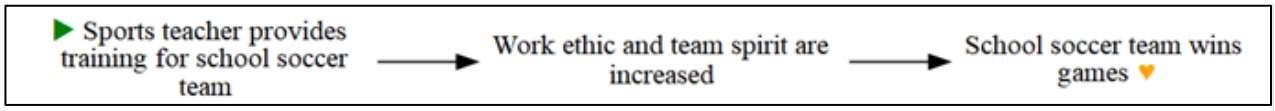

Figure 7. Simplified Theory of Change in which work ethic and team spirit in school soccer team are explained as being important because they promote team success. To that end, the teacher intervenes to provide training designed primarily to improve work ethic and team spirit.

valuable variables are necessarily just exactly those which take longest to achieve or which will sustain the longest.

"No-child" variables need not be valued. There is no reason why a Theory of Change should not include "no-child" variables which are not valued. For example, we might want to note the possibility of some interesting downstream consequences which we do not directly care about ourselves – beyond our intervention, beyond even the variables which we value. For example, we might note that our project to open a new cultural center might have some implications for traffic flows.

Intervention variables can be valued. Having acknowledged the use of the "❤" symbol to mark not only "no-child" variables but also intermediate variables as valuable, we now realise there is nothing to stop us valuing intervention variables themselves. For example, perhaps I value helping my neighbor for its own sake. I have been brought up to be a helpful person and I just love that role. I also, of course, value the consequences, but that is not my only motivation.

A related example is what is sometimes called the "humanitarian imperative" (International Federation of the Red Cross and Red Crescent Societies, 1994, p. 3): I might claim that it is a good thing, or an essential thing, in and of itself, to lend a helping hand to people in need – for example after a disaster. I might also get involved in important and complicated arguments about the possible unintended negative consequences versus the positive consequences of such help; but this may have nothing to do with my valuing the act of helping in its own right . Again, the temptation to judge the act of helping as "merely" a means to an end – presented, as it is, as part of a causal chain, and therefore not one which can be "really" valued – is a strong one which should be resisted.

Extensions

This basic idea of explicitly marking the variables which we value within a Theory of Change can be extended in various ways.

Highlighting Relative Priority by Using More Than One "❤" Symbol

It can be useful for stakeholders during project implementation, as well as for ex-ante evaluators, to know the relative importance of different valued variables, for example to help decide where resources should go, or should have gone. While there might sometimes be good reasons for not having or publicising such a prioritisation, here is a suggestion for how to do it when required: If there is more than one valued variable, we can use varying numbers of "❤" symbols to show their relative priority. For example, we might put two symbols next to "Student academic achievement ❤❤" and just one by "Extent to which teachers use creative techniques ❤", to show the greater priority of the former. This can be done in a purely illustrative manner, or in conjunction with a more precise weighting (see final section). This kind of weighting could be expressed in a different form, for example like this: "❤ = 2.5", "❤ = 4", etc.

Highlighting Negative Value and Cost

What about variables like "morbidity" or "stress" or "environmental damage" which have a negative valence? These break the "more is better" rule given above. We could use the "❤" symbol for them but this might be misleading. Instead I suggest using a different symbol which suggests a negative valence such as "☹" or some other suitable symbol, reserving the heart symbol for positive valence.

Similarly, although variables which represent *costs* could also be marked with a "☹" symbol, we could also consider a special way to mark them, perhaps with one or more "$" symbols. For example, stakeholders could be reminded of which "no-parent" variables were particularly expensive by marking them with a greater number of "$" symbols.

Highlighting Valued Variables Which Are Not Ordered

There are many other ways in which a variable might be intrinsically important to us, but we cannot say of it that "more is better".

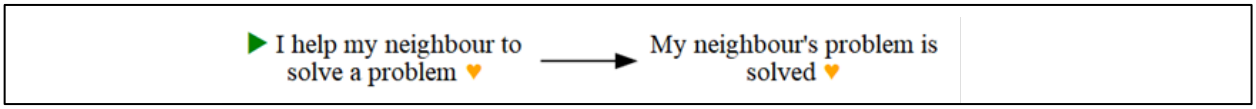

Figure 8. Simple Theory of Change in Which a "no-parent" Variable is Valued as Well as a "no-child" Variable.

For example, in Ainsworth's theory of children's emotional and developmental attachment, the "attachment style" of a child to a main caregiver is seen as a very important factor going forward; some attachment styles ("anxious-ambivalent"; "anxious-avoidant"; "disorganized-disoriented") are seen as being less beneficial for the child, whereas "secure" is more beneficial (Ainsworth, Waters, & Walls, 1978). But I would not recommend characterising the variable "attachment style" as valuable without further comment because it is, as such, not even an ordered variable: we cannot sort the different styles into a single dimension of better versus worse. Their clinical relevance cannot be reduced to points on a good/bad continuum. We could perhaps use a "❤?" symbol to show that something here is valuable but not in a "more-is-better" way. We can contrast this with an alternative, ordered variable which represents simply how secure a child's attachment is (and which loses some of the other aspects of the original variable): this could certainly be marked with the "❤" symbol.

Dealing with Multiple Stakeholders

Up to now, we have only seen value ❤ and intervention ▶ symbols used globally to show what "we", the makers of the Theory, value and control. Using the same ideas, I will zoom out to take a perspective which includes multiple stakeholders. This has similarities with an evaluation approach introduced by Hansen and Vedung (2010) which keeps the project theories held by different stakeholders separate rather than trying to always establish a consensus theory, which they argue is the approach used in an overwhelming majority of cases.

When different stakeholders or "agents" interact, they usually control different variables. This can be done by writing the name of the agent before the ▶ symbol, as in Figure 9. We can use ❤ symbols, preceded by the name of the agent who values this particular variable, in the same way. Alternatively, it might be more convenient to color the ❤ and ▶ symbols differently for each stakeholder or stakeholder group, along with an appropriate legend for reference.

This kind of presentation can be useful for underlining that different agents might do different things or put a different emphasis on the same things as they have different motivations. It suggests a way of modelling stakeholder motivation which is very different from the conventional, more behaviourist perspectives familiar from public health research such as COM-B (Michie, van Stralen, & West, 2011). It might help understand the behaviour of stakeholders and stakeholder groups in terms of their own (cognitive, if not explicitly formulated) theories of change – the way they view the world and how they can get what they want.

In Figure 9, all the agents value the main outcome, improved teaching, whereas the increased pay to which the teacher is entitled after completing the training is a positive motivation for the teacher but a negative one for the Ministry. In addition, part of the trainer's motivation is simply the act of providing pro-bono training itself.

In this case, the agents differ (but overlap) on two of the aspects of the Theory of Change (the variables on which they can intervene and the variables which they value) but share the third aspect, namely the Theory itself, the variables and causal connections shown here. What if they differ on the Theory itself?

- If their different Theories mostly overlap but one or two stakeholder groups take account of one or two variables which are ignored by others, it is possible just to mark

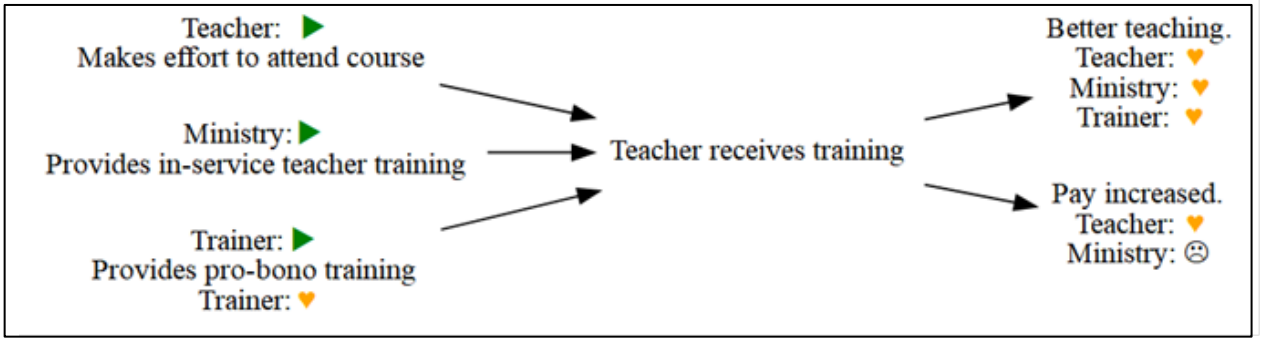

Figure 9. Simplified Theory of Change in Which Three Stakeholders Differ in What They Value and Where They Intervene.

that fact on the relevant variables using the "agent: …" notation used already for value and intervention in Figure 5 (or by using different colours).

- It might also be possible to mark different ideas about the causal links by marking the arrows in the same way, i.e. by using the "agent: …" notation or by using different colours.
- But when there are large differences between the Theories, this approach becomes very unwieldy and it will be necessary to present a series of separate Theories, using common elements wherever possible.

It is interesting to note that one of Chen and Rossi's (1980) original papers on Theory-Based Evaluation focuses also on multiple goals and the possibility of re-assessing originally posited goals.

Modelling Valuation Theories

So far this article have presented, firstly, two symbols for turning a project Theory into a Theory of Change by highlighting which variables are to be intervened upon, which are valued, and how the former influence the latter; and secondly, more details on how this idea could be implemented and extended in various ways.

The final part of the present article takes the idea of simply marking which variables are valued and generalises it to modelling everything about the way a stakeholder calculates and compares value within and between projects: their complete valuation theory. "Valuation theory" is related to the concept of "normative theory" as discussed by Chen: "guidance on what goals and outcomes should be pursued or examined" and is even closer to "normative theory" as discussed by Hansen and Vedung (2010, p. 300), with reference to Chen: "Notions concerning why the various aspects of the situation that are supposed to be affected by the intervention are preferable or not preferable to the situation without the intervention or with another intervention." However, Hansen and Vedung do not give many further details of this concept.

In the simplest case, an organisation might well care about, say, both environmental benefit and number of children vaccinated due to a project without ever wanting to combine the two into one measure. There are often good reasons for keeping variables separate even when they seem to measure the same thing – for example, a rescue mission might have saved five lives but sadly led to the death of one of the rescuers. It might be quite abhorrent, and would probably be of little use, to say that the net outcome was "plus four lives". In a Theory of Change with valued variables, if we do not need to ask or answer questions about how much we value specific changes on each valued variable, or about their combined value, as long as they are "more is better" variables, we can simply mark them as valued and leave it at that. Our evaluation report could mention the valuation of both without trying to combine or synthesise them.

But what if an evaluation task does involve modelling more sophisticated valuation theories in which value is defined in terms of transformations and combinations of other variables? We have not yet covered this possibility, although it was already

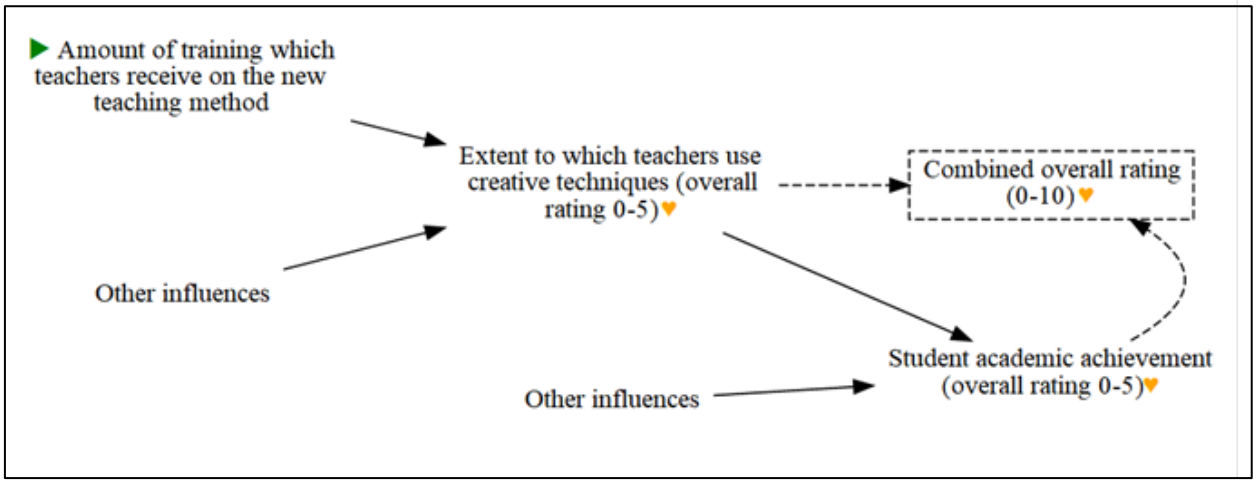

Figure 10. A simplified Theory of Change Showing a Valued Variable Which is Defined in Terms of Other Variables.

implied when we added the idea, above, of marking variable priority and marking variables which were associated with cost.

For example, a client or stakeholder may want to define some overall score or assessment by combining the scores on many related valued variables such as a cumulative number of events over time, or the scores of a cohort of students. These can be combined by taking, for example, a numerical total or average by taking note of the number of students who exceed a minimum threshold, or by some other method.

We can actually show these combinations directly within the Theory of Change by introducing defined variables which are also valued variables, as in Figure 10. Suppose our client tells us that student academic achievement and teachers' use of creative techniques are both equally valuable. We agree to rate both variables on a 0-5 scale and construct an overall rating which is simply the sum of the two. In Figure 10, we model the latter with a defined variable, marked by a dashed border to show that it is not necessary or even logically possible to collect additional data for it; the "raw" data from the two variables which define it is all we need. In addition, a dashed arrow is also used to show a definitional, rather than causal, relationship (Powell, 2017). This network of dashed definitions is the "valuation theory", built on top of the causal theory .

Another example: suppose a local non-governmental organisation in a flood-prone area is working on flood resilience. They want to ensure that every village has both an early warning siren and an early warning plan. They are not interested in villages only installing a siren or in villages only adopting a plan. They want to see villages with both. Again, we can model this situation using dashed lines, as in Figure 11.

In Figure 11, the arrowheads are joined to show some kind of interaction, and the label "AND" specifies that both parent variables have to be true for the child variable to be true. But where does the "❤" symbol go? On the defined variable B or the variables S and P which are part of its definition? In this case at least, S and P are only valuable in combination so perhaps we should just put the symbol on B.

There are myriad other ways in which our valuation of a project might involve defined variables which transform and combine other variables. For example, we will need to be able to deal with variables which are stretched across time – and potentially into the future – as well as with variables which are associated with some particular time-point, such as an end-line. For example, we often need to define a variable which collects the cumulative value of another variable. Given a Theory of Change which includes a variable representing the number of visitors to a youth center over time, we can define a variable which collects the total number of visits over a given period, accompanied perhaps by another valued variable which collects information about the percentage of female visitors.

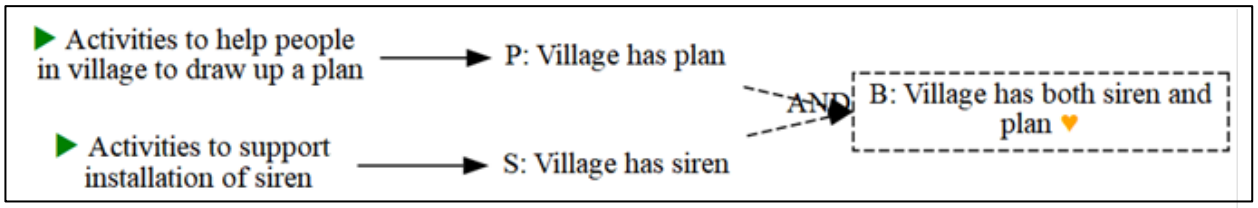

Figure 11. Simplified Theory of Change in Which the Single Valued Variable B is Defined as S "AND" P (i.e., it is only true when both S and P are true).

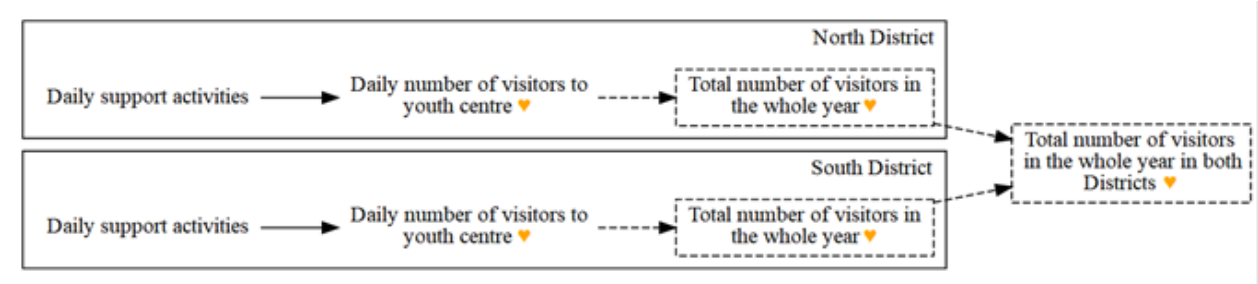

Figure 12. Fragment of a Theory of Change and Valuation Theory for a Youth Project (the right-hand variable is defined in terms of variables which are themselves defined in terms of others).

We can also define valued variables on the basis of other variables which are themselves defined, as in Figure 12.

It is not always necessary to even show all the intermediate variables in a Theory of Change if only the final, "no-child" variable is valued, especially if they are themselves only defined variables, as in Figure 12. By all means, we could live without these intermediate variables and note the details of the corresponding calculations separately. In the valuation Theory just as in the causal Theory, the level of detail which should be shown depends on the use.

Often it is not enough to simply report that there was a very high score on, or very substantial difference made to, a valued variable. We need to know is that good enough? Did it meet a certain target or standard? To allow the asking and answering of this kind of question, it will be necessary to allow the interpretation of specific levels of specific variables (i.e., not just "more / less valuable", but, for example, "good enough / not good enough" and more specifically to address questions like "Was the change made quickly enough?"). Essentially, this establishes variables which are not only valued but have a specific form in which the lowest values are interpreted as something like "very poor, very unsatisfactory" and the highest values are interpreted as something like "very good, very satisfactory". Scriven (1967) calls this kind of scale a "goal scale". We can use an anchor symbol ("⚓") instead of a heart in this case. The anchor symbol says: " We have cut to the chase: specific scores on this scale are not just arbitrary, but mean things like accept, reject, outstanding, not good enough, etc."

The use of evaluative rubrics (King, McKegg, Oakden, & Wehipeihana, 2013) is a very promising way to facilitate this process and provide the necessary interpretations of the levels of the variables in question .

We can use the same ideas to extend a simple valuation theory into the logical skeleton of an evaluation. We can define new variables to provide evaluative comparisons, for example comparing benefits with costs. Some classic evaluation questions such as "cost-effectiveness" can be understood in this way as comparisons between variables which are usually already present in a Theory of Change. In this way, we can see an evaluation as, amongst other things, constructing a valuation theory on top of a Theory of Change: a valuation theory in which the values are constructed by the evaluator in interaction with the Terms of Reference and with stakeholders. This (e)valuation theory often differs in some ways from that intrinsic to the project, perhaps including additional (defined) criteria which were not part of the original project plan.

Using defined variables in our (e)valuation theories also allows us to model how our valuation of some (numerically measurable) variables does not seem to be linear. For example, in the case of the body weight of a child, we might value medium weights the most and extreme weights less. In this way we can provide a new valued variable, called something like "healthy weight score" which is defined in terms of the physical weight using a non-linear, inverted-U-shaped, function, giving a low score for extreme weights and a high score for moderate weights.

Real-life theories of change very often include "goal-level" variables which are not actually caused by the variables pointing to them but which are a summary of them. Using dotted lines can help to avoid mixing up causal with definitional links (Powell, 2017).

As a final point, it will also be essential to show how this task of modelling a valuation theory – like all other evaluation tasks – can work with non-numerical, as well as numerical, variables. In Figure 10, if the scores were expressed simply as "outstanding" in both cases, then even without a numerical score we should be able to summarise these two results at least as "outstanding" and certainly not as "poor": a non-numerical reasoning

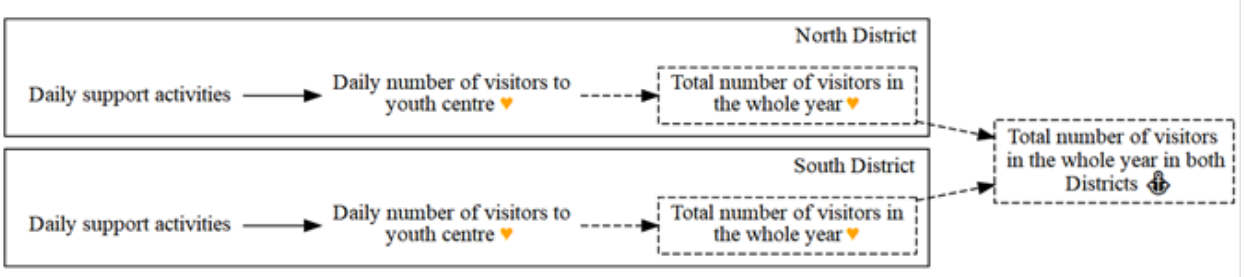

Figure 13. As Figure 12 (but the right-hand variable is an "anchor" variable: for example, specific low scores or ranges of low scores mean something like "inadequate", and high scores mean something like "adequate").

process – see Scriven (2012) – which we could call "soft arithmetic".

The evaluation task which I have here described as "modelling the valuation theory" corresponds very well with the list of "evaluation-specific methodologies" given by Davidson (2015, p. 6) and can be seen as a central one in evaluation.

This section has discussed some essential extensions to the basic definition of a Theory of Change given above.

Conclusion and Challenges

In the Introduction I presented the Flexibility Problem: (If we are to use a more flexible format for theories of change, how can we show which variables we value, which we intervene on) and the Definition Problem (What is the difference between a Theory showing the causal influences within and around a project and, more specifically, a Theory of Change for the project?). I argued that the "❤" and "▶" symbols are a direct answer to both problems. The body of the present article has suggested a definition of "Theory of Change" as a special kind of Theory which involves these two concepts, in the course of which it was first necessary to provide a loose definition of "Theory". I have paid particular attention to the "❤" symbol and the task of modelling what we value within a Theory. Showing explicitly, within a Theory, which variables are valued can help bring the issue of value within Theories of Change to the forefront.

The tools presented only allow us to do quite rudimentary modelling of value. For one thing, the underlying definition of Theory presented here needs to be enriched, for example, to include Theories which include emergent, changing or hard-to-predict variables and links, feedback loops, etc. But, in particular, the way we model value needs to be extended to address at least some of the following issues:

- valuing the difference made on valued variables as a consequence of some intervention. Above, I suggested that evaluations need to be able to assess the difference which an intervention makes to valued variables. But does this mean that we are dealing with some kind of subtraction – if an intervention improves some valued outcome score from, say, 5 to 7, can we say that the intervention has produced "two units of value"?
- distinguishing between valuation which is implicit and valuation which is explicit. Does it make sense to say to a client "In your plan you've forgotten to say that you also really value X, … you do, don't you? … shall we add it?" Does it perhaps even make sense to say "this agent actually values X, but won't admit it"?
- capturing the way an evaluator might feel ethically or professionally compelled to critique a client's own valuation, perhaps from the perspective of a wider, shared system of values, effectively saying: "These things that you value, they aren't really so valuable", or "You said you value X, but doesn't that conflict with your commitment to human rights"? What happens if the evaluator actually manages to shift the client's views?
- capturing how clients may value not only specific variables but a whole process, for example when a donor wants to know about any detrimental impacts of any part of a project on the environment.
- modelling what happens when a valuation theory is changed, for example during a project, perhaps in accordance with a fixed set of "higher" values or as part of a process of adaptation of values.

In conclusion, we can note that the approaches introduced here have had some interesting and important side-effects. Three of the most interesting are as follows:

- They suggest that it is the task of evaluators to model how stakeholders value aspects of a project just as much as it is their task to model the causal chains within a project.
- These approaches can help model the fact that stakeholders may value variables which are not at the end of a causal chain. This throws a new light on the debate between results-based and principles-based programming.
- They open up a way to understand the behaviour of stakeholders and stakeholder groups in terms of their own theories of change – the way they view the world and how they can get what they want – rather than from the kind of behaviourist perspective more familiar to most evaluators.

Sources

Ainsworth, M. D. S., Waters, E., & Walls, S. (1978). *Patterns of attachment: A phsychological study of the strange situation*. Mahwa, NJ: Lawrence Erlbaum Associates.

Anderson, A. A. (2005). *The community builder's approach to theory of change*. Washington, DC: The Aspen Institute.

Blamey, A., & Mackenzie, M. (2007). Theories of change and realistic evaluation. *Evaluation, 13*(4), 439–455. https://doi.org/10.1177/1356389007082129

Chambers, R., & Pettit, J. (2004). *Logframe: A critique*. Retrieved from https://www.betterevaluation.org/sites/default/files/simonhearn_en_Logframe_A_Critique_199-1.pdf

Chen, H. T. (1990). *Theory-driven evaluations*. Thousand Oaks, CA: Sage.

Chen, H.-T., & Rossi, P. H. (1980). The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science. *Social Forces, 59*(1), 106–122.

Coleman, G. (1987). Logical framework approach to the monitoring and evaluation of agricultural and rural development projects. *Project Appraisal, 2*(4), 251–259. https://doi.org/10.1080/02688867.1987.9726638

Davidson, E. J. (2015). Question-Driven Methods or Method-Driven Questions? How We Limit What We Learn by Limiting What We Ask. *Journal of MultiDisciplinary Evaluation, 11*(24), i–x. Retrieved from http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/414/394

Davies, R. (2004). Scale, complexity and the representation of theories of change. *Evaluation, 10*(1), 101–121. https://doi.org/10.1177/1356389004043124

Department for International Development. (2011). *Guidance on using the revised Logical Framework (DFID practice paper)*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/253881/using-revised-logical-framework-external.pdf

Dhillon, L., & Vaca, S. (2018). Refining theories of change. *Journal of MultiDisciplinary Evaluation, 14*(30), 64–87.

Donnelly, J. (2013). *Universal human rights in theory and practice*. Ithaca, NY: Cornell University Press.

Earl, S., Carden, F., & Smutylo, T. (2001). *Outcome mapping: Building learning and reflection into development programs*. Ottawa: International Development Research Centre. Retrieved from https://www.idrc.ca/en/book/outcome-mapping-building-learning-and-reflection-development-programs

Earle, L. (2002). *Lost in the matrix: The Logframe and the local picture*. In Paper for INTRAC's 5th Evaluation Conference. The Netherlands. Retrieved from http://pdf2.hegoa.efaber.net/entry/content/907/Lost_in_the_Matrix.pdf

Guijt, I. (2008). *Critical readings on assessing and learning for social change: A review*. Brighton, UK. Retrieved from www.ids.ac.uk/ids/bookshop

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

Holland, J. H. (2014). *Complexity: A very short introduction*. Oxford: Oxford University Press. https://doi.org/10.1093/actrade/9780199662548.001.0001

International Federation of the Red Cross and Red Crescent Societies. (1994). *Code of conduct*. Retrieved February 12, 2019, from https://media.ifrc.org/ifrc/who-we-are/the-movement/code-of-conduct/

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

King, J., McKegg, K., Oakden, J., & Wehipeihana, N. (2013). Rubrics: A method for surfacing values and improving the credibility of evaluation. *Journal of MultiDisciplinary Evaluation, 9*(21), 11–20. Retrieved from http://journals.sfu.ca/jmde/index.php/jmde_1/article/viewFile/374/373

Kusek, J. Z., & Rist, R. C. (2004). *Ten steps to a results-based monitoring and evaluation system: a handbook for development practitioners*. World Bank Publications.

Lewis, D. (1973). Causation. *The Journal of Philosophy*, *70*(17), 556–567. https://doi.org/10.2307/2025310

Lewis, D. K. (1986). *On the plurality of worlds*. Malden, MA: Blackwell.

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

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), 1–11. https://doi.org/10.1186/1748-5908-6-42

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Los Angeles: Sage.

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge: Cambridge University Press.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York: Hachette Book Group.

Powell, S. (2018). The book of why: The science of cause and effect. *Journal of MultiDisciplinary Evaluation, 14*(31), 47–54. Retrieved from http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/507/453

Powell, S. (2017, December). Visualizing theories of change: How not to confuse causes and definitions. *EVal*. Retrieved from https://www.elval.org/en/vizualizaciji-teorije-promjene-kako-razlikovati-uzroke-i-definicije/

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*, 39-83. Chicago, IL: Rand McNally.

Scriven, M. (1981). *Evaluation thesaurus* (3rd ed.). Inverness, CA: Edgepress.

Scriven, M. (2012). The logic of valuing. *New Directions for Evaluation*, (133), 17–28. https://doi.org/10.1002/ev.20003

Shaw, I., & Crompton, A. (2003). Theory, like mist on spectacles, obscures vision. *Evaluation, 9*(2), 192–204. https://doi.org/10.1177/1356389003009002005

Vaca, S., & Vidueira, P. (2016). Using graphical perception principles to improve the systems thinking tools' data visualization: Revisiting the systems dynamics model. *Journal of MultiDisciplinary Evaluation, 12*(26), 18–24.

Vogel, I. (2012). *Review of the use of "Theory of Change" in international development: Review report*. Department for International Development. Retrieved from www.isabelvogel.co.uk

Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. P. Connell (Ed.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts* (pp. 65–92). New York: Aspen Institute. https://doi.org/10.1177/1356389003094007

Wittgenstein, L. (1978). *Philosophical investigations*. Oxford: Blackwell.