# Does this policy work here?

Using evidence for better policy design, prediction and evaluation

Nancy Cartwright, Durham University
Eileen Munro, London School of Economics
Brendan Kelters, Durham University

# Does this policy work here?
## Using evidence for better policy design, prediction and evaluation

Nancy Carwright, Durham University
Eileen Munro, London School of Economics
Brendan Kelters, Durham University

Professor Nancy Cartwright
Centre for Humanities Engaging Science and Society (CHESS),
Department of Philosophy, Durham University, 50 Old Elvet, Durham, DH1 3HN, UK
nancy.cartwright@durham.ac.uk

Professor Eileen Munro
Department of Social Policy, London School of Economics, 2 nd Floor, Old Building,
London School of Economics and Political Science, Houghton Street, London, WC2A 2AE
e.munro@lse.ac.uk

Dr Brendan Kelters
Centre for Humanities Engaging Science and Society (CHESS),
Department of Philosophy, Durham University, 50 Old Elvet, Durham, DH1 3HN, UK
brendan.t.kelters@durham.ac.uk

**A guide on using evidence for**

- Better policy design and implementation
- More reliable prediction about whether a policy will work where it is implemented
- More accurate post hoc evaluation of whether a policy worked as intended

**To achieve these aims this guide will show you how to**

- Develop an *information-rich theory of change* showing the process by which a policy is supposed to achieve its targeted outcomes
- Classify and use *evidence* about whether a policy is likely to work if implemented as planned
- Develop an overall assessment of how much warrant you have that the process envisaged by a theory-of-change will go through start to finish; that is, *how likely your policy is to succeed*
- *Leverage* such assessments before implementation to predict how likely you are to be able to implement a policy successfully and after implementation to evaluate how well a policy achieved its goals
- Use the methodology to assess how well a policy is working
- Use such assessments, supported by well evidenced information-rich theories of change, to *enhance policy design and implementation*
- And to underwrite a post hoc evaluation of whether the policy in fact succeeded

# What we do here

Suppose you are considering adopting a policy in your local setting and there is evidence that it has worked elsewhere. Or, alternatively, you are designing a home-grown policy. How do you select, design and implement the policy to maximise your chances of success? What kinds of evidence can help you evaluate whether it will work if implemented in this way? And how should you use that evidence to arrive at an overall evaluation of your likelihood of success? If you have already implemented a policy, how do you evaluate how well it has worked? We aim to help answer these questions.

Many readers of this guide will be familiar with the 'Evidence-Based Practice' approach where the focus of implementation and evaluation is on the policy or practice *itself* and what it might accomplish and where Randomised Controlled Trials (RCTs) are taken to be the best source of evidence about this. In contrast, we focus on *the context in which the policy or practice is used*. This allows us to tackle a well-known pitfall in RCT-based policy evidence: what works in studied populations often fails to work elsewhere. In the social world there are few actions that have reliably similar effects on different occasions and in different contexts. As is widely acknowledged, the same policy that has worked well in one setting can fail miserably in another.

So, what can you do to estimate whether a proposed policy will work in your setting, given how you expect to implement it? Or, for purposes of evaluation, how do you check afterwards if your policy did what you hoped for? Of course, if there have been several studies of the policy in a range of settings, their findings will be helpful, especially if those settings are *similar* to yours in the right ways. But then, how do you know what makes for a *relevant* similarity, since usually the reason for doing an RCT in the first place is ignorance about the confounding factors that can affect whether, and to what extent, the policy produces the targeted outcome?

For a sound estimate of whether a policy will work in your setting and of how best to implement it, we urge that there is no alternative but to look to your setting itself and try to understand how the policy would work *there*. You need to ask: 'How would the policy lead to the intended outcome here?' i.e. 'What is the causal pathway?' and 'What does it take for such a pathway to operate start-to-finish in this setting?'

But what should you focus on? What *kinds* of features of the process, setting and the way the policy will be implemented matter for describing causal pathways and establishing whether they will actually work? These are the questions we address in the first part of this pamphlet. The second part is to help you organise the evidence you collect *about* these features in a way that makes clear the role each piece of evidence plays and where more evidence is needed. Together, this rich step-by-step account of what is needed for a policy to work and of the evidence for these requirements being met, should put you in a stronger position from which to plan, tailor and implement policy.

# Part 1: Constructing a rich theory of change

A step-by-step account of how a policy works is called the policy's **theory of change (ToC)**. It details the **intermediate steps** by which the policy is supposed to bring about the targeted outcomes.

You are standardly advised to start with a basic '**boxes-and-arrows**' theory depicting how the process is supposed to unfold step-by-step. We start there too. But then we explain how to *enrich* that theory to include a variety of different kinds of facts that matter if each step is to lead to the next. This enriched ToC can then be used as a guide for policy design and implementation. It can also be used, in ways we will describe, to organise and assess the evidence that your policy will do the job that you want it to.

## *Constructing a basic boxes-and-arrows ToC*

Basic boxes-and-arrows theories of change are very familiar now and are widely recommended. Here is an example depicting a way in which the introduction of a tax on sugar-sweetened beverages (SSBs) is expected to decrease their sales.
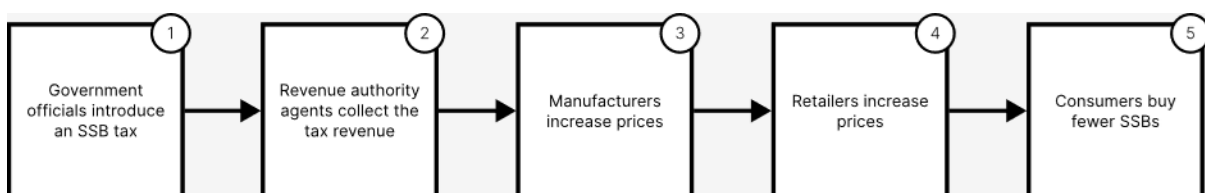


*Figure 1: Example of a basic Boxes-and-Arrows Diagram[1]*

This diagram shows a single **causal pathway** by which the introduction of an SSB tax is supposed to decrease the purchase of SSBs. Sometimes though there are multiple pathways by which the policy might contribute to a targeted outcome. This requires a branching diagram of different pathways of boxes and arrows.

---

[1] Reproduced from Cartwright et al (forthcoming), based on work by Miriam Alvarado.

To **construct a basic theory of change for a policy**, set out the policy interventions you are considering making at the beginning and the outcomes that you're hoping for at the end. Next, diagram the significant intermediary events that you theorise need to occur in turn if the policy is to lead to the outcome(s) you've specified, connecting these intermediaries together according to your theory of what causes what on the path towards the outcome(s). Each connection in a diagram like this, detailing a particular intermediary or intervention causing a particular other intermediary or outcome, is a **causal step**.

While you may use short labels in your boxes, it is important to provide as full a description as possible of what the cause and effect at each step are supposed to be like (normally in a key) as an aid in figuring out what kinds of evidence to look for in checking whether steps occur as theorised. Important facts to mention often include both the expected time lapse between the cause and the effect and their relative sizes. Sometimes, say, something that looks like the expected effect occurs, but it is too big to have been produced by that cause, or it occurs too soon or too late.

You will note that these boxes-and-arrows diagrams don't contain any loops but go in one direction only (they're **acyclic**). This is because these diagrams depict relations between causes and effects occurring at specific times and nothing can cause anything earlier than itself. In describing a system, the ongoing interactions of causes and effects are often depicted in feedback loops but these ignore this time factor, trading temporal specificity for simplicity. In our approach it is important to 'unwind' causal processes which work in this way into their time sequence so that future versions of the same kind of cause and effect appear later in the ToC.

## *Making a boxes-and-arrows ToC more useful*

To make the basic theory more useful for planning, prediction and evaluation, consider for each step just how that step is supposed to occur. What must happen for the cause at that step to produce the effect at the step? What must be in place for that to happen and what might prevent it from happening? What might help or hinder the step's occurrence? We suggest you think in terms of these categories for each step:

- **Support factors** for the cause at the step to produce the effect at the next. *Things that help*, either because they're needed for the cause to work.

- **Derailers** that stop the cause from producing the effect, so *things that hinder*.
- **Detractors** that hinder the cause producing the effect. *Things that hinder to some degree.*
- **Safeguards** that stop derailers and detractors from disturbing the causal interaction. *Things that help by stopping hinderance.*
- The **activities** by which the cause produces its effects.
- The **tendency principles** under which these activities occur.
- Characteristics of the social/economic/legal/cultural /geographical/physical/etc **underlying systems** that afford these activities.

**Support factors.** The specific factor that you focus on and call 'cause' in a given step is seldom enough on its own to bring about the effect. Almost always other factors not mentioned must be in place as well. We call these **'support factors'** (represented by what are often called **moderator variables** since they moderate whether – or how much – the highlighted cause will produce its effect). For example, consider the step from Box 4 to Box 5 in Figure 1. Retailers raising prices of SSBs will not by itself cause consumers to buy fewer of them. What more must be in place as well for that to happen? Suppose that this is to happen by what economists call a 'price effect'. Then at least these four further facts must hold:

S4.1. Consumers aim to maximise utility
S4.2. Consumers notice the price rise
S4.3. Consumers take the price rise as a disutility
S4.4. The consumers' disutility due to price > utility to them of consuming the SSBs

If any one of these fails and there is no adequate substitute for it in place, then a rise in prices will not produce less consumption by a price effect.

**Derailers** and **detractors** are, respectively, conditions that can prevent a cause from bringing about its effect or reduce the contribution it makes (derailers may thus also be considered absent support factors and vice versa). For example in the case study that Figure 1 was based on, the step from Box 2 to Box 3 ('revenue authority agents collect the tax revenue' to 'manufacturers increase prices') was detracted from by manufacturers figuring out how to produce SSBs more cheaply, so they did not need to increase prices so much.

**Safeguards** are conditions which thwart derailers and detractors. When you are designing your policy and deciding how to implement it, you should think about

possible derailers for each step and build in whatever safeguards are practicable. Of course, even when you have envisaged a derailer, it's not always possible to guard against it. For instance, what could have been done in advance to safeguard against manufacturers discovering cheaper ways to produce SSBs?

**Activities** are the interactions by which causes produce effects. When a cause produces its effect, it doesn't just sit there and then a bit later the effect pops into existence. The cause *does something* to produce the effect: it engages in some **activity.** It is important to think through what the activity is supposed to be at each step since this provides a good clue as to what support factors will be needed and what might derail, detract from or safeguard the process.

Consider again the step from Box 4 to Box 5 in Figure 1. We suggested that this might take place via a price effect. In this case what the price rise does is to *decrease the utility of consuming SSBs*. But Box 4 could produce Box 5 in another way, via a *signalling effect*. The tax was introduced
in Barbados accompanied by messaging that it was being introduced as a public health measure, thereby warning the public 'SSBs are very bad for your health'. In this case, whether-or-not the price rise appreciably reduces the utility of consuming SSBs, it can also *remind* consumers of these health warnings. But for this reminder to lead to a reduction in consumption, different support factors need to be in place, like:

> S4.1. Consumers care about their health
> S4.2. Consumers perceive the price rise
> S4.3. Consumers associate the price rise with the SSB tax
> S4.4. Consumers recall the health warnings

Because of the important role of activities in identifying support factors, detractors, derailers and safeguards, it is helpful to use vivid and detailed descriptions of them in your enriched ToC. Very generic descriptions, such as 'causing' or 'bringing about', will be of little help.

**Tendency principles.** We assume that it is not arbitrary which activities a cause can initiate and which outcomes can be produced from these – at least it is not arbitrary in cases where you can hope to be able to predict or explain. There is some 'systematicity' to it; these things happen in accord with principles that we can learn and that we can learn how to use. The easiest to use are of course principles that are supposed to be universal and exceptionless, like Newton's second law of motion, since these can be expected to hold everywhere and under all circumstances. But there is also a vast store of knowledge that can be put

to use concerning looser, often qualitative, 'ceteris paribus'-style principles telling what a cause can contribute in cases where the principle obtains.

These are often expressed as generics, without any explicit scope or range indications. Many are familiar everyday principles that we all regularly appeal to in explaining and predicting what happens, like 'People avoid actions they expect to get punished for', 'Parents care about the welfare of their children', and 'People act to maximise their expected utility'. They are also often the result of social science research. For example, here are some that apparently economists tend to agree on: 'Universal health insurance coverage will increase economic welfare in the United States', 'Addressing biases in individuals and institutions can improve both equity and efficiency', 'Climate change poses a major risk to the US economy', 'Tariffs and import quotas usually reduce general economic welfare', 'Flexible and floating exchange rates offer an effective international monetary arrangement'.[2]

Even when you know what activity is supposed to occur, knowing the tendency principle under which it occurs can be of further use in identifying support factors and derailers. For instance, consider again the step from Box 4 to Box 5 in Figure 1 in the case where the activity is taken to be 'reminding consumers of the health warnings'. This activity can reduce consumption in two different ways: under the principle 'People tend to avoid things they think will hurt them' or 'People tend to pursue things they think will be good for them'. Which of these is to come into play matters for identifying further support factors. For example, if the first is intended, the government health warnings should stress the dire consequences –diabetes, heart trouble, etc – of too much sugar. But if the second, the warnings should instead stress the benefits of eating healthily: feeling better, being able to do more, living longer, etc.

You may ask, 'Why *tendency* principles?' This is a term that JS Mill used to describe the principles of political economy. 'Tendency' marks out two things. First, though widespread, at least across certain ranges of contexts, tendency principles do not hold everywhere and when they do, they may need support factors etc. to be in place in order to operate and sometimes need triggering as well. Second, they do not usually tell you what will actually happen but only what the cause tends to do, in the sense of what it contributes to the overall effect. That's because other causes might also contribute to that effect at the

[2] Geide-Stevenson, D., & La Parra-Pérez, Á. Consensus among economists 2020—A sharpening of the picture. *The Journal of Economic Education*, 55(4), 461–478, (2024).
https://doi.org/10.1080/00220485.2024.2386328

same time. Note that what we call 'tendency principles' are sometimes called 'mechanisms', especially in realist evaluation.

**Underlying systems** are the relatively stable social, economic, legal, cultural, geographical, physical, and so on arrangements that obtain in the setting. For example, the settings in which social services to families operate typically provide a legal framework, buildings, equipment, salaries, and oversight mechanisms as well as a host of cultural and social norms and expectations. It is important to think about these since they set *what the causal possibilities are* in that setting – what tendency principles can obtain there and what it takes for these principles to be brought into play. Norms of politeness provide a simple example. As we all know, a gesture that might signal good manners in one culture can cause offense in a culture where the norms are different.

There's no clear-cut line between what is underlying and what is part of the causal process you are focusing on. Since it is facts about the underlying system in a setting (and often complex arrangements of these facts) that makes it possible for a tendency principle to obtain there, these *can* be thought of as support factors for the cause cited to bring about its effect in accord to with the tendency principle. In practice, it seems best to distinguish what is underlying and what is part of your process based on *plausible changeability*. If a support factor, or the obtaining of a tendency principle in your setting, might plausibly change so as to help your policy work as intended, or frustrate it working, then you should highlight it by counting it as part of the causal process you're examining. If the factor or principle is very unlikely to change, you can consider it underlying.

As long as the setting affords the step going ahead, it does not need scrutiny. If you find that the setting doesn't afford the step going ahead due to some plausible changeable feature, then it is better to classify that feature as a derailer. If the setting doesn't afford the step going ahead due to an invariable feature, on the other hand, then the step can never be expected to work under any plausible conditions; a useful finding.

As you reflect, discuss and explore the components of your theory of change - the support factors, derailers/detractors, safeguards, activities and tendency principles - details about them should be added to each step in your theory of change to produce an **enriched theory of change**.

For example, consider Figure 2 which shows a single causal step, prepared as part of a study into the implementation of a new policy in child protection (in-

volving changing the audit system from a file-based review by a manager to a conversation between the practitioner and the manager).[3]
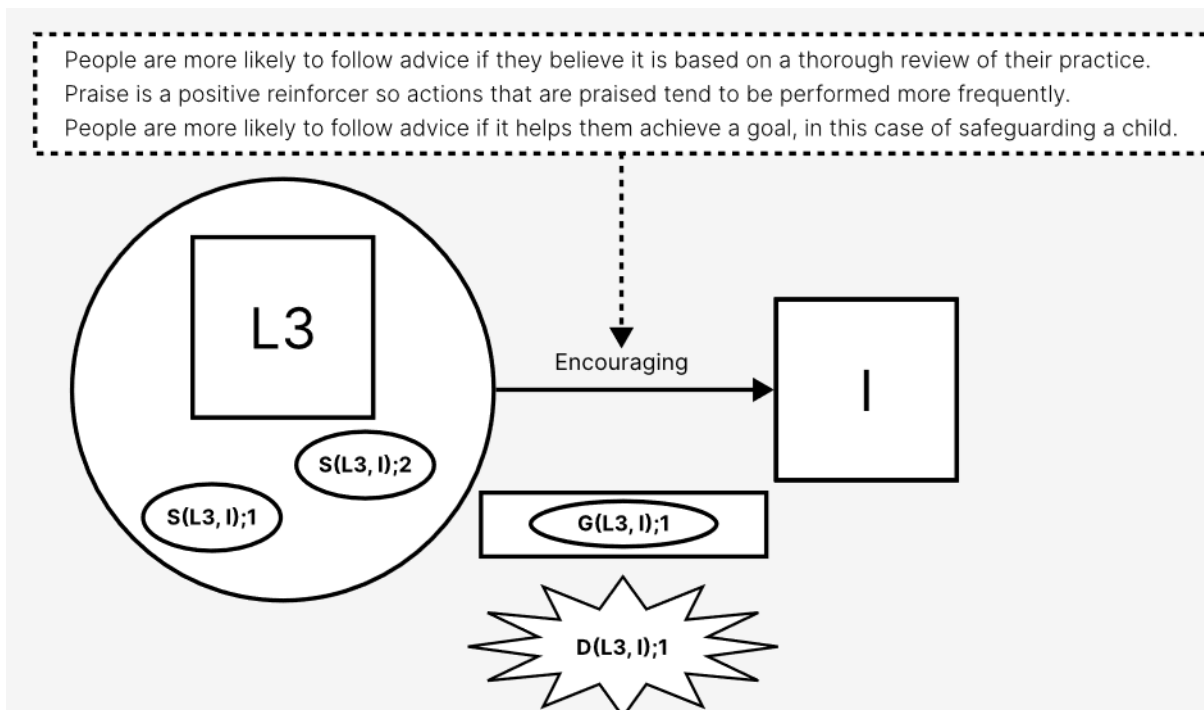


*Figure 2: A Single Causal Step from a Box-and-Arrow Diagram Representing a Detailed Theory of Change*

- L3 is the theorised **cause** at this step: 'managers praising and criticising practitioner behaviours under the new policy'
- I refers to the theorised **effect**: 'practitioners working in ways consistent with the new policy'
- S(L3, I);1 and S(L3, I);2 name the **support factors** for L3 to contribute to I: 'practitioners find new audits more thorough' and 'practitioners learn right lessons from new audits' respectively.
- L3, S(L3, I);1 and S(L3, I);2 are enclosed within a circle that marks out the **overall combination of conditions** needed for L3 to contribute to I.
- 'Encouraging' is the **activity** by which factors in the circle acting together are to contribute to L3 and the principles linked to it are the tendency principles under which it is to operate.
- D(L3, I);1 is a **detractor**: 'practitioners' lack of time/resources needed to implement new approach' theorised to potentially lessen the effect of this activity.
- G(L3, I);1 is a **safeguard** against this detractor: 'managers find time for audits, encouraging practitioners to find time/resources to practice new approach'.

---

[3] From Cartwright et al (forthcoming), Cambridge University Press, based on work by Eileen Munro.

Such diagrams for each step can also be pieced together to depict whole causal pathways from original policy interventions to intended outcomes. Such diagrams, and the enriched theories of change they represent, can become large and complex, but the extra detail they include tells us a lot about what is needed for a policy to work in a context. You can compare whether the identified support factors, safeguards, tendency principles and underlying systems are in place in your context, while identifying and locating any distinct derailers or detractors, and potential safeguards against them, guided by the rich description given to each causal step.

See for example Figure 3, which diagrams a pathway in which Figure 2 appears as only the final step. (For a key to the new symbols not explained in Figure 2 see Cartwright et al, forthcoming, Cambridge University Press.)
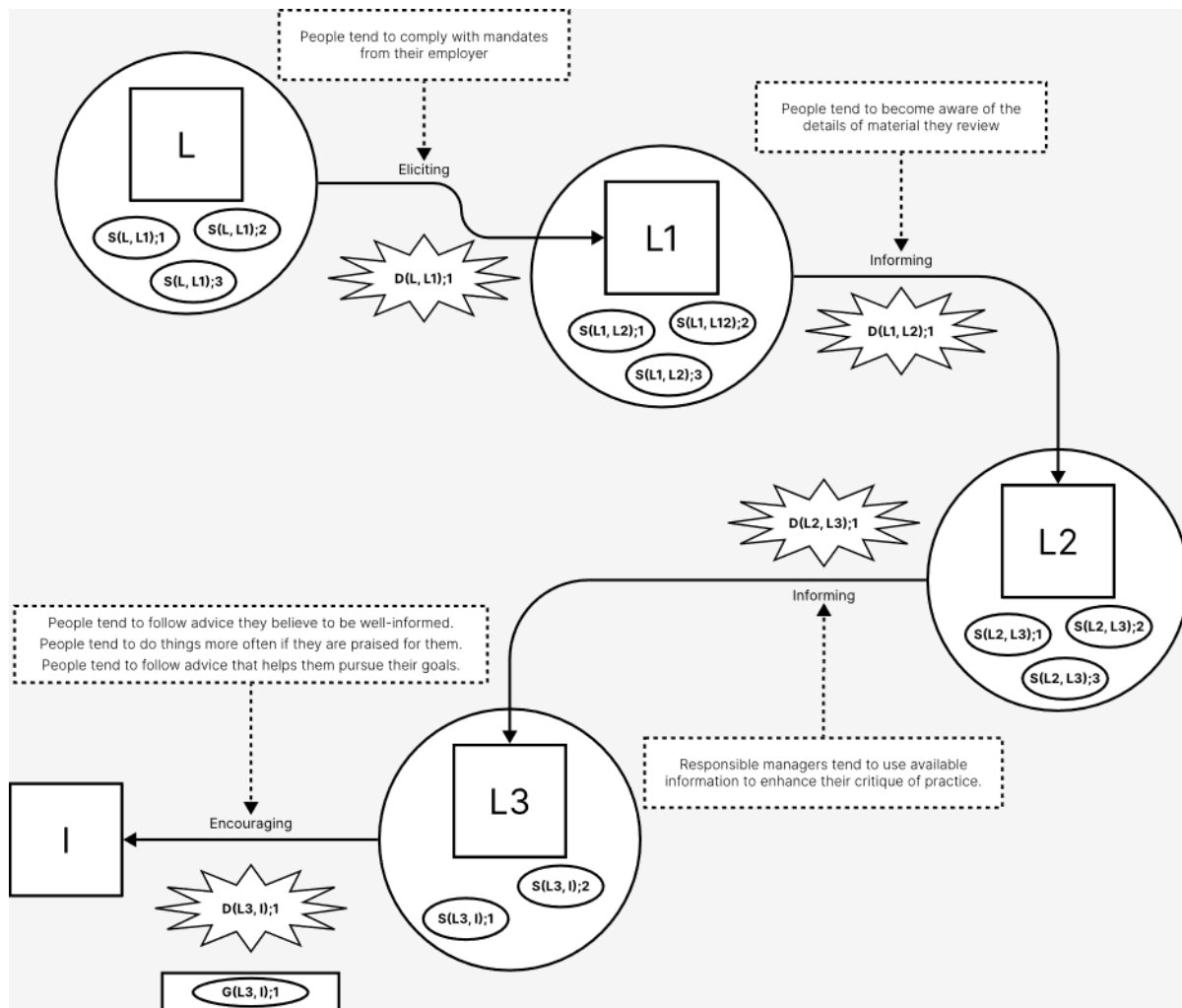


*Figure 3: A Detailed Theory of Change for one policy in the implementation of the child protection programme Signs of safety in a specific UK jurisdiction[4]*

---

[4] From Cartwright et al (forthcoming), based on work by Eileen Munro.

Any theory of change developed at this initial stage will only be tentative and preliminary, subject to revision in light of evidence in pursuit of a better understanding of the local causal dynamics that determine the effectiveness of your policy. Given this, the business of preparing a theory of change and testing that theory occur in constant negotiation, the intermediate steps in the theory being re-envisaged in response to evidence about how things can actually work in your context.

# Part 2: What counts as evidence?

Your enriched theory of change diagrams the process by which your proposed policy is supposed to lead to your intended outcomes. But will it do so in your setting as you plan to implement it there? Here we use the enriched theory of change to catalogue different categories of evidence that can help answer that and suggest how to use them.

The warrant that the whole process you hope for goes through depends in large part on the warrant that each of its steps take place: certainly, strong warrant that one step will not take place is strong warrant against the process succeeding, and strong warrant that each step will take place is strong warrant in favour of success. Of course, how much warrant you need depends on what decisions you are making and how costly wrong decisions are versus how costly it is to invest more effort in gathering more evidence.

Since the success of the whole process depends on the success of the steps in it, we propose that you focus on gathering and evaluating evidence about the facts that matter to each of those steps. Your warrant that a step occurs – or not – will naturally depend both on how certain your evidence about the various facts is and what role these facts play in ensuring that the step goes through.

To keep things simple, in what follows we will suppose your evidence is fairly certain at each stage. When, as is often the case, it is less than certain, the warrant it provides (either for or against) will naturally be weaker. There will of course usually be much more evidence possible after the fact for post hoc evaluation of whether a policy genuinely contributed to the effect, than one might have for ex ante prediction that it will.

You will notice that we deal in purely qualitative assessments, talking for example of 'weak' or 'moderately weak' or 'strong' warrant. That's because in a host of practical cases the kind of background information necessary to properly employ formal systems of inference is not available. For instance, you will very often not have any good grounds for making the basic quantitative probability assessments necessary to use Bayesian methods. In cases where you do, that's all to the good. In those cases, the quantitative results about causal processes should mesh easily with our qualitative ones since the latter are based on facts discussed in Part 1 about the nature of causal relations and how causes operate.

# *What facts matter?*

The facts that matter to a step's success include:

## 1. The proposed cause occurs

If the cause proposed by a causal step doesn't occur, the step doesn't occur, so warrant that a step's cause is absent yields equivalent warrant that the step fails. Warrant that it occurs provides some warrant in favour of the step but this warrant is weak unless supplemented by further information.

## 2. The proposed effect occurs

Like the cause, if the effect proposed at a causal step doesn't occur, the step doesn't occur, so any warrant that a step's effect is absent yields equivalent warrant that the step is absent. Warrant that a step's effect does occur provides warrant in favour of the step occurring, but only weak warrant. The warrant is generally stronger if you also have good evidence that the cause occurs (will occur, did occur). But it would be even better to have further evidence that the envisaged cause actually contributes to the effect, for instance evidence that the expected activity will occur or that all the support factors are there that are necessary for the cause to produce the effect.

## 3. Other causes sufficient to produce the effect are absent

If there is evidence that other causes make sufficient contributions to account for the effect, this yields equivalent warrant against the step. The exception is in cases where there is a threshold: then different causes may all genuinely operate without producing any change in the effect.

Conversely, if you have evidence that no other causes make sufficient contributions to account for the effect *and* that the effect occurs, you have at least as much warrant that the step succeeds. This can be, however, hard to prove, since you must uncover and discount all plausible alternate causes.

## 4. The timing and size of the effect are correct

What counts as the correct timing and size for a step's effect depends on the character of the cause and the principle under which the activity occurs. Some activities produce effects quickly, others slowly. Some produce large contributions to the effect, some small. Evidence that the effect occurs with the size and timing expected yields fairly strong warrant that the step occurs. Evidence that the timing or size of the effect isn't as expected provides moderately strong warrant that the step does not occur, warrant which becomes much stronger when coupled with evidence that other causes operate that could account for the timing or size of the effect.

## 5. The required support factors occur

Warrant for the occurrence of all the support factors required at a step, coupled with evidence that the cause occurs, is strong warrant for the step (unless there is warrant that unguarded derailers occur). Evidence that not all the required support factors occur yields warrant against the step. Evidence you don't really know which support factors are required limits how much warrant one can get here, either way.

## 6. Detractors and derailers are absent (or guarded against)

Evidence that detractors and derailers do not occur, or are all guarded against, yields little warrant that a step occurs without warrant that the cause occurs or that all support factors occur. Warrant for all three together provides equivalent warrant that the step occurs. Conversely, warrant that sufficient detractors and derailers exist yields equivalent warrant that the step doesn't occur.

## 7. The required activity obtains, start-to-finish, and the related tendency principle operates

Evidence that the activity specified for a step occurs yields very strong warrant for that step, and evidence that it doesn't yields very strong warrant against the step. Of course, it will be hard to find evidence that an activity will occur ahead of time that you haven't already taken into consideration (for instance, evidence about whether the cause or its support factors or derailers will obtain). But you may have evidence that the related tendency principle operates in your setting and that the conditions needed for an activity of the right sort (described by this

principle) to make the right sort of impact, contingent on the cause occurring, will be present. Coupled with evidence that the cause occurs, this provides fairly strong warrant that the activity will occur, especially given evidence that all support factors will obtain and detractors will be absent or guarded against. This in turn provides good warrant that the step occurs.

### 8. The underlying system being right to afford the tendency principle obtaining and operating

Evidence from the nature of the underlying system that the requisite principle does not obtain or cannot be brought into play yields strong warrant against the step occurring. Warrant that the underlying system does afford the principle and its operation yields very weak warrant that the step occurs, since background conditions being right for a causal step to occur doesn't say much about its actual occurrence.

## *Marshalling and evaluating the evidence*

We suggest that, for each step in your theory of change, you have a separate **page** with **8 columns** where you collect what evidence you have about that step together, one column for each of the 8 categories above. This will facilitate evaluating the warrant for each of those kinds of facts. Then your understanding of the role each of those facts plays in determining whether the step occurs will help you evaluate the overall warrant you have for – or against – a step occurring. This work can be conducted post hoc or ex ante, though, as we noted, working ex ante limits what evidence is available.

Reviewing the strength of warrant for each step then helps you make a sensible judgement about the success of the whole process. But remember that lack of evidence is not evidence against! For example, if, for a step, you have strong evidence that the cause occurs and that the theorised tendency principle often operates in your setting, this together provides medium warrant of the step occurring as theorised if you have no evidence to the contrary. If, at the same time, though, you have evidence that a needed support factor is missing, this weakens the case substantially, suggesting that things can't work as you theorise.

In this kind of case when you are doing ex ante prediction, and aiming to maximise your policy's chances of success, it may well be worth the effort to try to get that support factor (or a surrogate for it) into place.

If you are doing post hoc evaluation and you also have evidence that the effect occurred, this suggests revising your theory of how the effect was in fact achieved. Perhaps the supposedly needed support factor wasn't in fact needed, or perhaps the effect was caused by something other than the step's cause. For a proper evaluation, you need to probe such possibilities and refine your theory of change. By this process a theory of change can be developed which is well-grounded in the available evidence and which can be used to assess whether the policy did what it was supposed to.

We have so far been talking (unless otherwise specified) as if your evidence is certain – which it seldom in fact will be. First you may not be certain that the factual claims you are using as evidence are in fact true. For instance, look at the first support factor S.4.1 for the step 4 to 5 in the SSB tax example: 'Consumers care about their health'. You might feel entitled to take this support-factor claim for granted, but perhaps you'd like to find some evidence to warrant it? You might, for instance, cite high gym enrolment or reduced cigarette consumption as evidence for it. However, firstly, you might not be altogether certain that the enrolment figures or sales figures you have are accurate. Secondly, you may not be entirely certain that the evidence claim really bears on the claim it supports – what its relevance is. For instance, how much do high gym enrolments, even if entirely accurate, support the claim that consumers care about their health?

Given the two different sources of uncertainty about evidence – uncertainty in whether the facts employed as evidence are true and uncertainty about how relevant they are and in what way – we suggest of thinking of evidence claims not as single claims but as pairs of (1) **factual claims**, often (but not necessarily) expressing pieces of empirical evidence, coupled with (2) **relevance claims** which explain how these factual claims bear on the conclusion to be established.

For example, suppose you are claiming that a strong wind caused tiles to be blown off a roof and suppose this account is captured in a simple single-step theory of change. In assessing whether the putative cause – the strong wind – occurred, you might cite this matched pair as evidence: the factual claim 'a nearby anemometer detected a strong wind' and the relevance claim 'if a nearby anemometer detected a strong wind then it was very likely windy'. These together, if warranted, warrant the claim that the wind occurred, and this warrant, in

turn, may be employed in your judgement of whether the step in question occurred, as described above.

This process of evidencing a causal claim can quickly become long and complex. One way of handling this is to draw up **evidence-role maps** showing how claims offered as evidence for steps in a theory of change relate to the things they're evidencing. This can be accomplished longhand by assigning claims codes or names that capture their relations to one another, however it can be easier to lay things out in a diagram. For a sample diagrammatic representation of this method, see Figure 4.
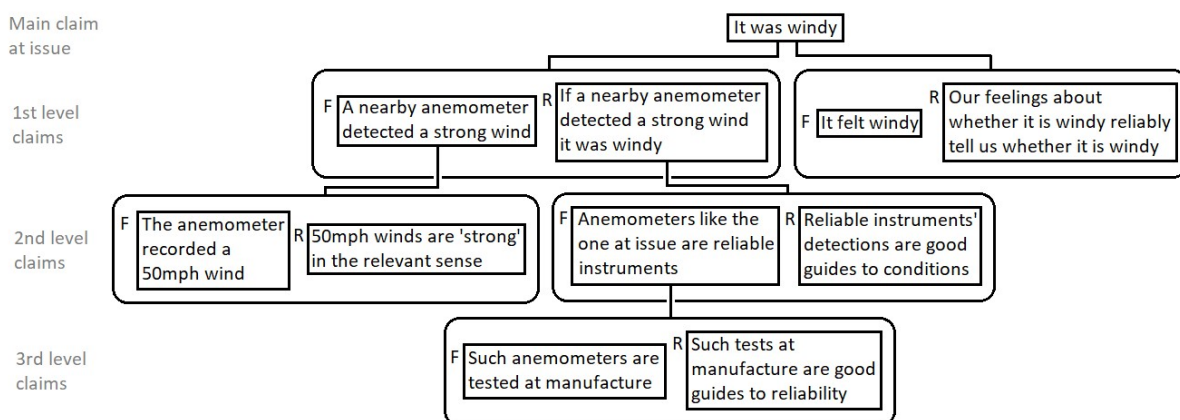


*Figure 4: A simple evidence-role map*

In figure 4, factual and relevance claims (annotated F and R respectively) are organised into groups which together warrant lower-level claims of fact or relevance, including the 'main claim' that is the fact pertinent to the truth of the singular causal claim at issue. An advantage of organising your evidence in this way is that it allows you to quickly see how a piece of evidence contributes to the overall claim or how weaknesses in a piece of evidence tell against the claim.

For example, suppose you discovered that the anemometer in question, in the case depicted in Figure 4, was badly rusted. This might suggest that the testing of the pristine instrument at manufacture cannot be relied upon to indicate current reliability, rendering suspect the 3rd level relevance claim offered here ('such tests at manufacture are good guides to reliability'). This, in turn, would weaken the case for the factual claim 'anemometers like the one at issue are reliable instruments', the relevance claim 'if a nearby anemometer detected a strong wind it was windy', and the case for the overall main claim (that the cause proposed by the relevant step of the theory of change here occurred). Another part of the case for this main claim, though, coming from feelings about windiness and their implications and shown in Figure 4, wouldn't be impacted. Any

independent case for anemometers like the one at issue being reliable also wouldn't be impacted.

This should all be obvious in a simple case like this, but in more complex cases with more complex evidence involved being clear in this or some similar way about the structure of the argument for the obtainment of the facts pertinent to the causal claims you're making can be very helpful in testing and providing credible evidence to back up your claims.

Once you have warrants for the causal steps in a theory of change, depending on your results, you can then refine your theory of change (by revising poorly or negatively warranted steps as appropriate, and retesting against your evidence) or, if you conclude the case for your theory is strong enough, use it to aid in your policy evaluation. In applying your resulting theory in this way, remember that a theory of change can be no stronger than its weakest link; it can only provide as much warrant for the overall effect it posits being brought about by the cause it posits (for desired outcomes being caused by policy changes) as there is for the least well-warranted step in the causal pathway you've theorised connecting the two.

# Part 3: Using the evidence more effectively

What can you learn from working through the justifications for the causal claims involved in policy in this way?

Most clearly, you can gain insight into, and better express, how well-justified the claims are, given your evidence. A well-constructed theory of change, step-by-step validated by well-mapped and high-quality evidence, allows you to make a strong case for or against a causal claim. Equally, you might discover a lack of such a case which might otherwise be missed, no less important a finding in working towards well-evidenced and effective policy. Further, you may do all this by drawing on evidence sensitive to the context in which the policy is deployed, allowing local variations in how things work and what can work to reveal themselves during the step-by-step validation and refinement of your theory of change. This gives a basis for interrogating the helpfulness of imported policies, post hoc and ex ante.

A further benefit is that, by organising the evidence for your theory of change in the way we suggest, you should gain insight into where gaps in this evidence exist, or where more proof is needed. This will be wherever evidence-role maps show a reliance on claims that seem in need of further justification, or where what seemed to be a sound theory of a causal step turns out to rely on unsafe assumptions about any of our eight kinds of fact pertinent to the step in question. In such cases you can return to evidence collection armed with greater clarity on what needs checked, and thereby more efficiently develop a case for or against the policy you are evaluating.

The evidence-role maps that we recommend may look daunting, but it should be remembered that quantitative methods involve conducting statistical tests and these are equally long and complicated when written out in full. Such detail is needed in both approaches for making judgments about what is going on in the social and physical world.

The effort put in will, we believe, help you to produce high quality policy evaluation in a way that others will find credible.