

Can I make this policy work here?

Using evidence for better policy design, prediction and evaluation

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

CHESS Working Paper No. 2025-01 Durham University

Original: March 2025

Revised: September 2025



CENTRE FOR HUMANITIES
CHESS
ENGAGING SCIENCE AND SOCIETY

Can I make this policy work here?

*Using evidence for better policy design,
prediction and evaluation*

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

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

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

Dr John Pemberton
Centre for Humanities Engaging Science and Society (CHESS),
Department of Philosophy, Durham University, 50 Old Elvet, Durham, DH1 3HN, UK
john.m.pemberton@durham.ac.uk

This booklet is an overview of the method proposed in Cartwright, N., Munro, E. C Pemberton, J. (forthcoming), *Causal Processes and their Warrant: A Practical Guide*. Cambridge: Cambridge University Press. This work was supported by the Arts and Humanities Research Council (Grant Ref: AH/X006727/1). It was also supported by British Academy South Africa grant *Integrating Evidence for Contextualised Public Health Policy: Lessons from South Africa* (Grant Number: EPG\100488). We are grateful for the help of Likhwa Ncube, postdoctoral researcher on the BA grant.

What we do here

If you are deciding about whether to use a policy, you will want to know: Will it work here? What do I need to check before I start? What should I do to improve chances of success? How do I get a good estimate of those chances? And how will I know if it worked afterwards? We aim to help answer these questions.

Many readers of this guide will be familiar with the ‘Evidence-Based Practice’ (EBP) approach where the focus of implementation and evaluation is on the policy or practice *itself* and what it might accomplish and where. The EBP approach takes Randomised Controlled Trials (RCTs) 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 real life, the same policy often works in one place but fails in another even when implemented with high fidelity. Context makes all the difference.

So, what can you do to estimate whether a proposed policy will work in your setting, given how you expect to implement it there? And how do you figure out how best 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 ‘the right ways’, what is a *relevant* similarity? Often 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 contend that there is no alternative but to *look to your setting itself* and to try to understand how the policy would work there when implemented in different ways. 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 here in our 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 carry through? 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 that 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.

In summary, this is a guide on using evidence for

- Better policy design and implementation
- More reliable prediction about whether a policy will work in your setting
- More accurate post hoc evaluation of whether a policy has worked there 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 – *how* the proposed interventions are to lead step-by-step to the hoped-for outcomes
- Classify and use *evidence* about whether a policy is likely to work if implemented as planned
- Develop a rough overall assessment of how much justification you have for believing that the process envisaged by your 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 to allow you to fix things that are not proceeding properly or abort the policy before too much is invested if it is failing
- Use such assessments, supported by well-evidenced information-rich theories of change, to *enhance policy design and implementation*.

Part 1 describes how to construct an information-rich theory of change that depicts what is necessary for your policy interventions to lead to your hoped-for outcomes. Part 2 describes how to use your information-rich theory of change to catalogue the evidence you collect and understand what role it plays in supporting – or undermining – a claim of policy success. We conclude in Part 3 with a brief summary.

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 **successive steps** by which the policy is supposed to lead to the targeted outcomes.

The standard advice is to develop a “**boxes-and-arrows**” **diagram** that shows the main sequence of steps. We start there too but then explain how to enrich this to give more detail of *how* you get from the very first policy interventions to the final targeted outcomes, what helps or hinders each step and how these steps depend on the local context. This enriched ToC will be a practical tool for **design, implementation and evaluation** of the policy.

Constructing a basic boxes-and-arrows ToC

Basic boxes-and-arrows theories of change are very familiar now and are widely recommended. See Figure 1 for an example, depicting a way in which the introduction of a tax on sugar-sweetened beverages (SSBs) is expected to decrease their sales.

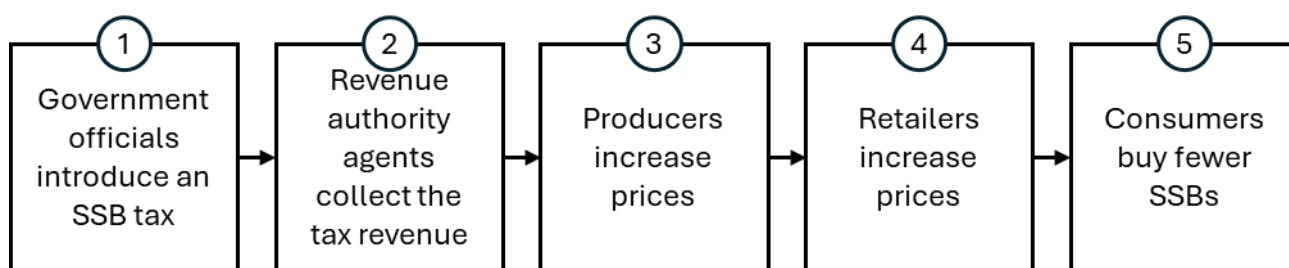


Figure 1: Example of a basic Boxes-and-Arrows Diagram¹

This example shows a single direct causal pathway. In reality there may be **multiple pathways** by which the tax reduces consumption, which we shall illustrate later with price and signalling effect pathways for the SSB tax. A complete ToC may need to show branching routes to the outcome, but for simplicity we'll follow this single main pathway.

Figure 1 is an example of a ToC in the usual form with the proposed **policy interventions** at the start, the **desired outcomes** at the end and in between a series of **intermediate cause-effect steps** by which the interventions are expected to eventuate in the outcomes. We will now show how to enhance the boxes-and-arrows ToCs by filling in *what is supposed to happen at each step to ensure that it will lead to the next*. By doing this for each step in the ToC you develop a richer ToC that can be the basis of further exploration. This exploration can help you develop and analyse evidence that is useful for discovering, before implementation, a policy's chances of success and what can be done to maximise them, as well as, after implementation, to what extent the policy achieved its goals.

¹ Adapted from Cartwright et al (forthcoming). After Alvarado, M., Penney, T. L., Unwin, N., Murphy, M. M., and Adams, J. (2021). 'Evidence of a health risk 'signalling effect' following the introduction of a sugar-sweetened beverage tax'. Food Policy, 102(1).

Making a boxes-and-arrows ToC more useful

The boxes in a standard boxes-and-arrows ToC generally have very short labels. The first thing we urge you to do is to **describe each cause and each effect in the ToC more fully** in a key or notes. The more you understand about exactly what the features at each step have to be like for the process to go through, the better you will be able to see if they have been (or can be) achieved, and how. Include the expected **timing** (how long should elapse between cause and effect?) and the expected **size of effect** (is it small or large?). These details can help you check whether what happens in practice matches what was planned. They also help you figure out what must happen to ensure that the features you depict in the ToC are in place.

Next, 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 next, and **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: *features that help* because they're needed for the cause to work.
- **Derailers and detractors.** Derailers stop the cause from producing the effect: *features that block it*. Detractors decrease the strength of the effect: *features that hinder to some degree*.
- **Safeguards** that lessen the effects of derailers and detractors: *features that help by eliminating obstacles and hinderances or by diminishing their effects*.
- The **activities** by which the cause produces its effects: *actions that terminate in the targeted effect*.
- The **tendency principles** under which these activities occur: *principles established to hold frequently in settings like this*.
- Characteristics of the social/economic/legal/cultural/geographical/physical/etc. **underlying systems** that allow these principles to obtain and the related activities to happen: *features and arrangements of the setting that ensure that the necessary principles and activities can hold there*.

To detail these a bit:

- **Support factors**

A single cause rarely works alone. For example, in the step from Box 4 ('retailers increase prices') to Box 5 ('consumers buy fewer SSBs') in Figure, price increases will not automatically reduce consumption. Retailers raising prices of SSBs will not by itself cause consumers to buy fewer of them. What more must be in place 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. Consumers' disutility due to price is greater than the 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 via a price effect.

- **Derailers and detractors**

These are, respectively, conditions that can prevent a cause from bringing about its effect or reduce the contribution it makes. For example in the Barbados study that Figure 1 was based on, the size of the effect in Box 3 ('producers increase prices') from the cause in Box 2 ('revenue authority agents collect the tax revenue') was detracted from by producers figuring out how to produce SSBs more cheaply so they did not need to increase prices so much.

- **Safeguards**

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 or detractor, it's not always easy to guard against it! For instance just what could have been done in advance to stop producers developing cheaper products in the Barbados SSB tax case?

- **Activities**

These are the actions by which causes produce their 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 clues as to what support factors will be needed and what might derail, detract from or safeguard the process.' buy fewer SSBs') in Figure 1. We suggested that this might take place via a price effect. In this case what the price rise does is *decrease the utility per unit of cost 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 these:

- S'4.1. Consumers care about their health
- S'4.2. Consumers perceive the price rise
- S'4.3. Consumers associate the price rise with the SSB tax
- S'4.4. Consumers recall the health warnings.

So: knowing which activity is intended helps identify the right support factors.

- **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; the activities happen in accord with principles that we can learn and that we can learn how to use. These principles describe the general patterns that the activity exemplifies. Many are familiar everyday principles that we 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’.

These are called ‘tendency’ principles because they do not describe what always follows when the causes acts but rather what the cause ‘tends’ to do. Often what a cause ‘tends’ to do is not the same as what actually happens because other causes operate as well and what actually happens depends on all of them together.

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 (‘retailers increase prices’) to Box 5 (‘consumers buy fewer SSBs’) 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 under the principle ‘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. If the first is intended, the government health warnings should stress the dire consequences – diabetes, heart trouble, etc – of too much sugar. If the second, the warnings should instead stress the benefits of eating healthily: feeling better, being able to do more, living longer, etc.

Our ‘tendency principles’ are sometimes called ‘mechanisms’, especially in realist evaluation.

- **Underlying systems**

‘Underlying system’ is the label we give to the relatively stable social, economic, legal, cultural, geographical, physical and so on arrangements that obtain in the setting. For example, the settings in which health-orientated taxes are introduced typically provide a legal framework, collection infrastructure, facts of human physiology, resource availability and a particular economic and political backdrop, as well as various cultural and social norms and expectations. It is important to think about these since they set *what the causal possibilities are* in the setting – what tendency principles can obtain there and what it takes for these principles to be brought into play.

For example, consider the step from Box 3 (‘producers increase prices’) to Box 4 (‘retailers increase prices’) in Figure 1. Whether this step is causally possible in a setting depends on the underlying conditions of the setting. If, say, retail sales of the product are all in state-run not-for-profit stores, this familiar principle may not operate.

Generally, if the underlying system supports some tendency principle by which a step's cause tends to bring about its effect, then the step can work. If the system doesn't support any such principle the step may fail in that setting under any realistic conditions.

There is no clear-cut line between what is 'underlying' and what is part of the causal process you are focusing on. A useful strategy for deciding whether you should count a factor as part of the underlying system or as part of your intended process is to consider *ease of changeability*. If the current system is currently missing factors necessary for a step to go through, can they reasonably be changed or would that require many factors in the setting all to change together in an integrated way? Since systemic change is so much harder, the sooner you recognise that it might be needed the sooner you can get started on it or decide it is too hard and begin considering alternatives.

Enriching a boxes-and-arrows ToC

Now you can draw a far richer, far more informative ToC. 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 can be added to each step in your ToC diagram to produce an enriched theory of change.

For example, consider Figure 2, which diagrams the step from Box 1 ('government officials introduce an SSB tax') to Box 2 ('revenue authority agents collect the tax revenue') in Figure 1 discussed above, including illustrative examples of the kinds of extra details we suggest.

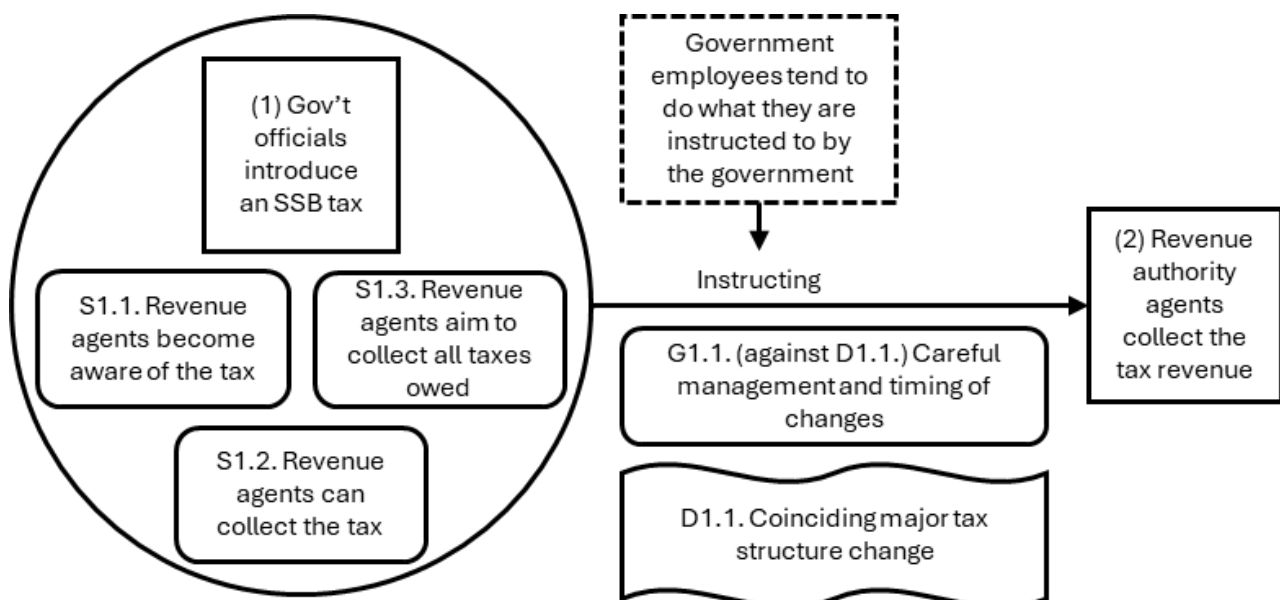


Figure 2: A Causal Step from a Diagram for a Detailed Theory of Change

In this diagram the cause and effect at the step are numbered (1) and (2) and support factors are numbered S1.1, S.1.2 and S1.3. The combination of (1) with all three support factors, represented by the circle enclosing them, acts to produce the step's effect (2) through the activity 'instructing'. This activity works in accord with the tendency principle 'government employees tend to do what they are instructed to by the government'. This step could be derailed or detracted from, say, if the introduction of the sugar tax coincides with the introduction of a major change to the basic tax system so that there is a delay in making the necessary administrative changes. One safeguard against this would be careful management and timing by the government of new tax initiatives and changes to tax structure.

In this illustration we've written our various details into the diagram in an abbreviated way for simplicity. In practice and especially with larger diagrams showing many steps, it's often better to use numeric codes and a key where factors are described in greater detail.

Figure 3 shows an example of such a larger diagram that enriches the causal pathway represented in Figure 1 by including the additional details we recommend. The step represented in Figure 2, notice, appears as only the first step in this larger diagram. Such diagrams and the enriched theories of change they represent can become large and complex. But the extra detail they include tells you a lot about what is needed for the policy to work in your context. You can compare whether the identified support factors, safeguards, tendency principles and underlying system arrangements 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.

Note that any theory of change developed will generally, at least initially, only be tentative and preliminary, subject to revision in light of the evidence you assemble about the local causal dynamics that determine the effectiveness of the policy in your setting. Given this, the business of preparing a theory of change and testing the theory against your context will generally occur iteratively, the intermediate steps and all that goes with them being reworked in response to evidence about how things work in your context.

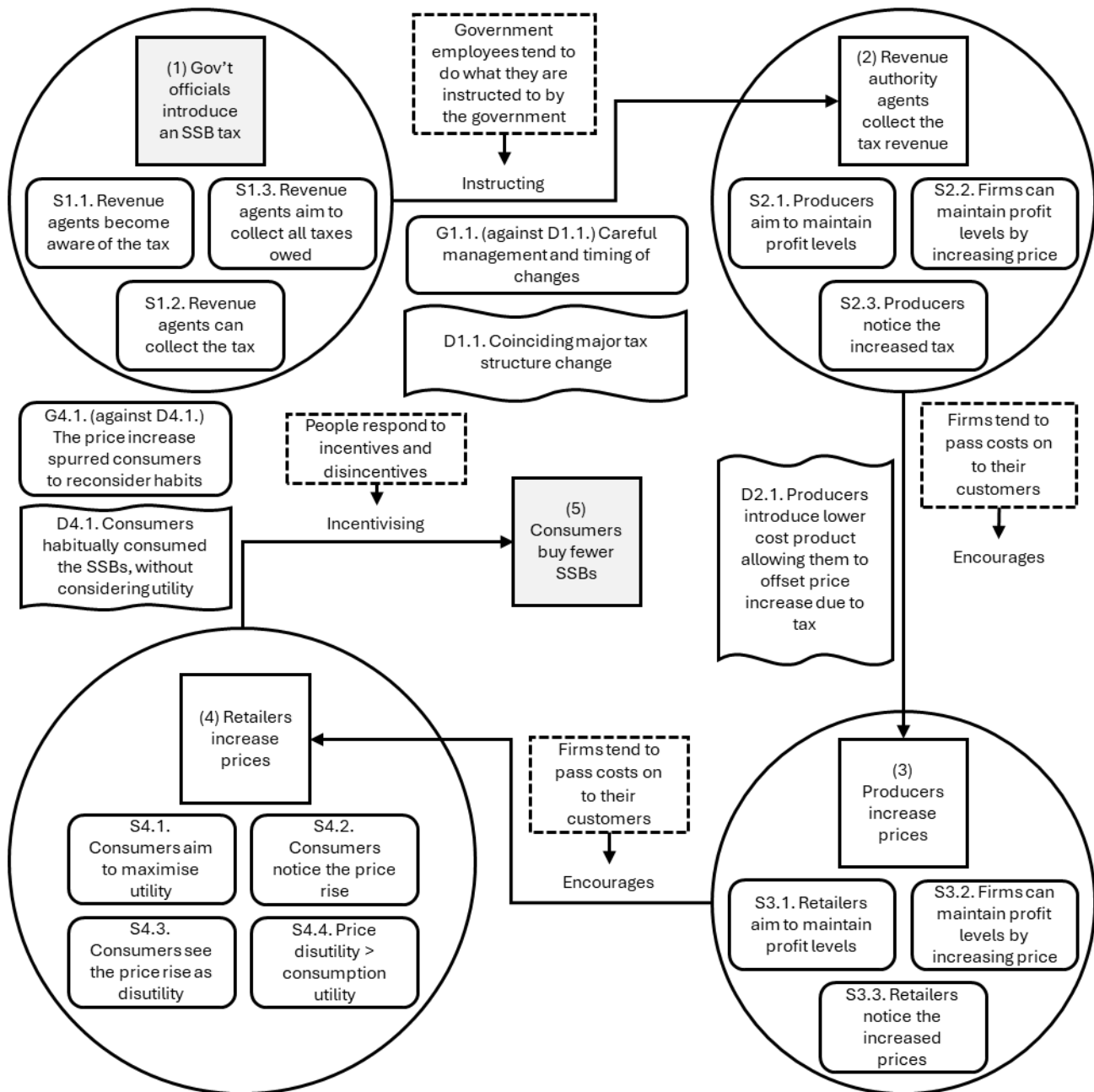


Figure 3: A Detailed Theory of Change showing a price effect pathway by which ‘Government officials introduce an SSB tax’ (Box 1) may cause ‘Consumers buy fewer SSBs’ (Box 5)

Part 2: Understanding the evidence

Your enriched theory of change illustrates 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.

Your enriched ToC shows how the policy is **supposed** to work in your setting, step by step. But did it (for post hoc study) or will it (for ante hoc study) do so? To answer this, you can test each step of the ToC using **evidence**. The strength of the whole process depends on the strength of each step. If there is strong evidence a step will not happen, the overall policy is unlikely to succeed (at least in any way requiring the problematic step). If there is strong evidence each step can or did occur, the policy is more likely to succeed.

To keep things simple, in what follows we will begin by supposing your evidence claims themselves are fairly certain. When, as is often the case, they are less than certain, the support they provide (either for or against policy success) 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 contributed to the outcome than you can have for ex ante prediction that it will contribute.

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 evaluation 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 which Bayesian analyses should also be consistent with.

For simplicity of illustration we will generally suppose that you are gathering evidence for a post-hoc evaluation of whether the policy contributed as expected. For ex ante policy design and prediction the same kinds of evidence are relevant though naturally much less will be available than after the fact.

We suggest that you catalogue the evidence step-by-step. You will then want to consider for each step all the things that must happen there if that step is to be successful and then try to gather evidence about whether they occurred.

What facts matter at each step?

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

1. The proposed cause occurs

If the cause does not occur, the step fails. For example, in the step from Box 4 ('retailers increase prices') to Box 5 ('consumers buy fewer SSBs') in Figure 3, if retailers do not increase SSB prices the step to consumers buying fewer SSBs by the overall 'price effect' pathway depicted in Figure 3 cannot happen. Evidence that the cause at a step occurs provides some support for the success of that step, but on its own, it is weak unless combined with evidence that the effect occurred or, even better, that the expected activity occurred and was completed.

2. The proposed outcome occurs

If the expected effect does not happen, the step fails. For example, at the last step depicted in Figure 3, if consumers did not buy fewer SSBs (the effect in Box 5 doesn't occur) then it follows that the step, and the overall 'price effect' pathway depicted by Figure 3, didn't occur. In practice this may be because the cause wasn't accompanied by some needed support factors, or some unguarded derailer occurred or the underlying system wasn't right. Evidence that the effect occurred provides some support for a step occurring, but stronger support comes when this is combined with evidence that the cause contributed to it. It is also important to secure evidence about the size and timing of the effect. Did the effect have the size expected and did it occur at the time expected if the cause contributed as expected?

3. Other causes sufficient to produce an effect of the size that occurred at the time it occurred are absent

If other causes are present that are sufficient to produce an effect of the right size that occurs at the right time, the cause at that step may not have contributed. In such circumstances you may even see the final outcome you hoped for but not as a result of your intervention (at least in the way envisaged by your ToC). For example, considering the last step of Figure 3 again, if there was an unrelated public health campaign sufficient to cause a drop in consumer purchase of SSBs of the scale observed (the effect in Box 5), this would count against the proposed cause in Box 4 having contributed its expected bit and overall against the 'price effect' pathway having produced the outcome. If you can rule out all such other causes, your case that your cause worked is stronger. But finding reasonable evidence to decide can be difficult, especially with multiple factors influencing outcomes.

4. The required support factors occur

Recall, support factors are conditions needed for the step to succeed. For example, in the last step of Figure 3, for the cause in Box 4 ('retailers increase prices') to have contributed to the effect in Box 5 ('consumers buy fewer SSBs') at least all of these needed to be the case:

- S4.1. Consumers aim to maximise utility
- S4.2. Consumers notice the price rise
- S4.3. Consumers see the price rise as disutility
- S4.4. Price disutility is greater than consumption utility.

If any needed support factors were missing then the step would have failed and any decrease in SSB consumption (even one of the expected size) must have been brought about in some different way – or else your hypothesis that that factor was necessary was mistaken.

5. Detractors and derailers are absent (or guarded against)

If derailers or detractors exist without safeguards, the step will fail or not be as big as hoped for (which may matter to whether the next step can succeed). For example, when a SSB tax was introduced in Barbados, its effect on consumption was diminished by what Figure 3 calls detractor D2.1 – ‘producers introduce lower-cost product allowing them to offset price increase due to tax’ – without any safeguards in place to limit its impact on the causal pathway and policy outcome.² In a case where you are doing an ex ante study, anticipating and checking for such detractors and derailers can help anticipate troubles.

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

The cause only works as intended if the requisite activity happens. For example, in Figure 3 where a price effect is expected, the activity in the last step is ‘incentivising’. If this incentivising doesn’t occur – for example, if prices rise but that does not reduce consumers’ utility, maybe because the consumers do not perceive the rise as significant – the step fails.

7. The underlying system is right to allow the tendency principle to obtain and operate

As we explained in Part 1, the underlying system in your setting must allow the tendency principle to operate for a step to occur as envisaged by your ToC. For example, returning to the last step of Figure 3, the tendency principle invoked is ‘people respond to incentives and disincentives’. Plausibly this obtained in the Barbados context that inspired this example, but it needn’t obtain in every context. In a cultural context where people feel duty-bound to act in specific ways even in minor matters regardless of incentives, it needn’t be the case that ‘people respond to incentives and disincentives’ enough for the kind of ‘incentivising’ activity employed by the last step of Figure 3 to deliver an outcome like Box 5 (‘consumers buy fewer SSBs’).

² Taken from Cartwright et al (forthcoming). After Alvarado, M., Penney, T. L., Unwin, N., Murphy, M. M., and Adams, J. (2021). ‘Evidence of a health risk ‘signalling effect’ following the introduction of a sugar-sweetened beverage tax’. Food Policy, 102(1).

Marshalling and evaluating evidence using evidence-role maps

We have so far been talking as if your evidence is certain – which it seldom in fact will be. Coping with this unfortunately makes matters more complicated, especially since – as we'll explain – uncertainty here can come from at least two different sources.

Given that each of the 7 kinds of fact we've highlighted can impact your support that a step occurs, we propose that for each step in your ToC you have a **separate page** with **7 columns** where you collect together what evidence you have about that step, one column for each of these 7 facts. This will make it easier to evaluate the amount of support for each of these facts. Then your understanding of the role each of these facts plays in determining whether the step occurs will help you evaluate the overall support you have for – or against – each step occurring. This work can be conducted before or after the fact. As we noted, working beforehand limits available evidence, but it is certainly a great help in designing policy and deciding whether – and how – to implement it.

Given that you want to determine support for each of the 7 facts we've highlighted for each of the steps in your ToC, how do you estimate evenly roughly the support you have for each of these facts?

Support must be derived from evidence, so let's take it that you have a fixed bundle of evidence to work with (though in actual research you will often work with an evolving bundle of evidence). How can this evidence warrant any one of our seven facts?

Most obviously, evidence can simply imply a fact. For example, if you have as a piece of evidence 'remembering seeing a cause occur', this might be taken to *imply* 'the cause occurred'. However implications like this are rare when dealing with evidence of the kinds needed for ToCs.

There are, as we noted at the start of this section, two sources of weakness for pieces of evidence: uncertainty about the **facts offered as evidence** and limitations on or uncertainty about **the relevance of these facts**.

For instance, look at the cause in Box 4 ('retailers increase prices') for the last step in Figure 3. What evidence could warrant that this occurred? You might, as indeed the post hoc Barbados SSB-tax evaluation did, cite the records of a chain of stores as evidence for it. However, first you might not be altogether certain that the records you have about the chain's prices are accurate. To whatever extent this is likely, this reduces the support for retailers increasing prices supplied by those records. Second you may not be certain that the evidence you cite decisively bears on the claim you suppose it supports. *How much*, say, can the records of a single chain support the claim that in general retailers increased prices? Perhaps *a bit*, but also plausibly not enough to prove it beyond doubt even if the chain's records are beyond doubt.

Given these two different sources of uncertainty about evidence we suggest thinking of evidence claims not as one single claim but as a matched pair of claims: as **factual claims**, often (but not necessarily) expressing pieces of empirical evidence, coupled to **relevance claims** which explain how these factual claims bear on the conclusion to be established.

Going even further, this same approach can also be applied to any claims offered as evidence themselves and so on until you have fully articulated the argument for your claim.

This process of evidencing a 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 are evidencing. These are, in effect, diagrammed arguments for these claims.

Articulating an evidence-role map can be accomplished longhand by assigning codes or names to claims that express their relations to one another, however it can be easier to lay things out in a diagram. For illustration see Figure 4, which expands on the example of retailers increasing prices, adding additional evidence claims.

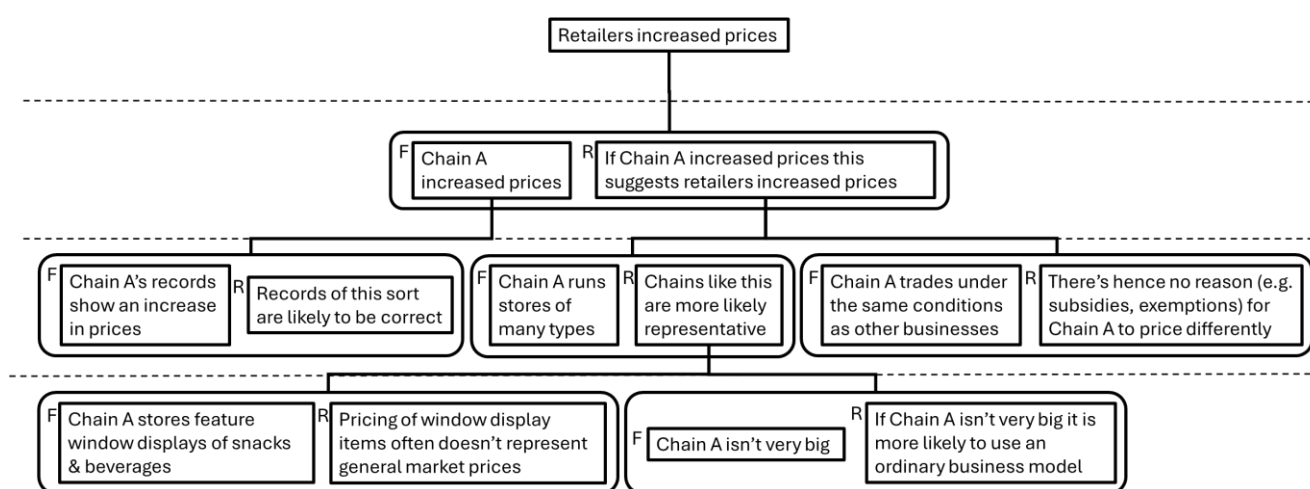


Figure 4: A simple evidence-role map

In Figure 4, factual and relevance claims (annotated *F* and *R* respectively) are organised into groups which together support (or oppose) claims) claims of fact or relevance diagrammed above them, including the ‘main claim’. Note that the evidence diagrammed can provide support either for or against the claim it bears on. For example, see the bottom line in Figure 4. If Chain A’s stores aim to attract customers by window displays of snacks and beverages, they are less likely to be representative in their pricing of these. It can then be very helpful to devise a way of marking evidence claims that provide positive support differently from those providing negative support, perhaps with different colours, and also to indicate how strong you estimate that support to be. Note though that we don’t expect you will generally be able to do this with any precision.

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 it. You need only track the connections up to see which claims are affected and hence check the impact on the overall support for or against the top claim. This may seem obvious in a simple case like this, but in more complex cases involving more complex bodies of evidence, being clear about just how your evidence bears on your claims can be very helpful.

Assessing the chances of success

Recall that we do not expect that in general you will be able to do this with any quantitative precision. But understanding your evidence and just what it is doing for you should make your qualitative assessments far better grounded and accurate. Throughout this process it is of course important to remember that *lack* of evidence is not evidence *against*.

We suggest you begin by reviewing the strength of support for each step since this helps you make a sensible judgement about the success of the whole process. In what follows we will generally assume you are doing a post hoc evaluation. But the same kinds of assessments apply for ex ante planning and prediction.

For example, if, for a step, you have strong evidence that the cause occurred and that the tendency principle often operates in your setting, this generally provides what we would label ‘medium support’ that the step occurred when you have no evidence to the contrary. If at the same time, though, you have good evidence that a needed support factor was missing, this weakens the case substantially, suggesting that things didn’t work exactly as you wanted. In this kind of case when you are doing ex ante prediction and aiming to maximise your policy’s chances of success, it will be worth some effort to try to get that support factor into place.

If you are doing post hoc evaluation and you have evidence that a necessary support factor was missing but you also have evidence that the effect occurred, this suggests revising your theory of how the effect was achieved. Perhaps that factor wasn’t in fact needed, or perhaps the effect was caused by something other than the step’s cause. In the latter case this provides good reason to suppose your interventions did not actually contribute even if the expected outcome occurred – unless the process you theorised was not after all the one by which your interventions operated. 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 that is well-grounded in the available evidence and that can be used to assess whether the policy did what it was supposed to in the way expected.

Evidence-role maps highlight exactly where your evidence is weak. We hope that their usefulness for better understanding what your evidence shows and hence for **better planning, prediction and post hoc evaluation** is clear.

You can use such information to plan **targeted evidence** collection to increase your confidence in each step, instead of gathering data indiscriminately. And you can do that both before and after the fact. For instance, consider how achieving intended results via the pathway shown in Figure 3 relies on retailers increasing prices. Evidencing that this occurred in the way shown in Figure 4 involves evidencing that a sample chain whose prices you reviewed is representative of the market. If you were to do this without knowing that your sample chain operated many types of stores and you believe this to be relevant, then checking this would be a priority in building your overall case, as shown by the relevant evidence-role map.

A similar strategy helps when doing **pre-implementation planning**. By mapping out your evidence, you can organise an overall case that the steps in your policy ToC will go through as desired, or not. In doing so you can detect places where you do not have enough support to

provide sufficient confidence that factors necessary for a step will be present to bet on its success. This alerts you to where you need to make changes in the setting so that there is more possibility of the step occurring and hence of overall policy success.

Below are some **rules of thumb** about how evidence for the different kinds of facts required at each step in your ToC bear on whether the step worked (or for ex ante planning, will work) and thereby whether your policy interventions did (or will) contribute as hoped to the desired outcomes. Note that you can only make these judgments relative to the assumption that your ToC is correct so you should bear in mind that your ToC itself may be faulty. In what follows we mean by 'good' evidence evidence-pairs that supply strong reason for belief, as could be shown by relevant evidence-role maps.

- Good evidence that the cause (fact 1), effect (fact 2) and activity (fact 6) specified for a step all occur (or will occur) as described (including at appropriate times with appropriate sizes) provides strong positive support that the step did (or will) work. Good evidence that any of these did not provides strong support against.
- Good evidence that the effect (fact 2) occurred (or will occur) and that nothing besides the cause described by your ToC was (or will be) sufficient to produce an effect of that size (fact 3) provides strong positive support that the step did (or will) work. This is often called 'Holmesian inference'. (Note that these not provide strong support if you have good reason to think the cause did not occur, in which case you need to review your understanding of what the causal possibilities are and how good your evidence is.)
- Good evidence that the effect (fact 2) occurred (or will occur) and that other causes operated that were sufficient to bring it about can be strong support that the cause did not contribute. This then is strong support that the step failed and in turn that the intervention did not contribute to the final outcome, at least not by the expected pathway. This can be the case even if the expected final outcome occurs. (The exception is when there are 'threshold' effects: your cause contributed but that is not visible in the size of the effect because its contribution couldn't make the effect any bigger. In this case your cause did what it was supposed to – and your intervention may also have done what it was supposed to – but they were redundant.
- Good evidence that the cause (fact 1) and support factors (fact 4) needed for a step occurred (or will occur) and that all detractors and derailers are (or will be) absent or guarded against (fact 5) provides strong positive support that the step did (or will) work.
- Good evidence that any one of the seven facts relevant to a step working did or will not occur is by itself a case against the step working as specified by the ToC.

Part 3: Better ToCs, better appraisals

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 assemble strong support for or against a causal claim. Equally, you might discover a lack of such support that might otherwise be missed, which is no less important in working towards well-evidenced and effective policy. Further, you are doing 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 likely effectiveness both of imported policies and of home-grown ones, both 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 can gain insight into where gaps in this evidence exist. 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 seven kinds of fact pertinent to that step. In such cases you can return to evidence collection armed with greater clarity on what needs to be checked and thereby more efficiently develop support for or against the policy you are evaluating. Of course before the fact, evidence about which of the necessary factors are already in place and which not and whether your local setting can reasonably easily allow you to get those that are missing will be invaluable for designing how to implement your policy and for predicting the chances of success.

The enriched ToCs and evidence-role maps that we recommend may look daunting. But it should be remembered that reliable quantitative methods also involve long and complicated procedures when written out in full – just think about how many steps go into conducting statistical tests and how much substantive background knowledge must be fed into them. 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, ex ante, to plan wisely when deciding whether an intervention can work in your context and, post hoc, to produce high quality policy evaluation in a way that others will find credible.