Journal of MultiDisciplinary Evaluation
Volume 16, Issue 37, 2020

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

Program Logic Foundations: Putting the Logic Back into Program Logic

Andrew J. Hawkins
ARTD Consultants
Charles Darwin University, Australia

Background: Program logic is one of the most used tools by the public policy evaluator. There is, however, little explanation in the evaluation literature about the logical foundations of program logic or discussion of how it may be determined if a program is logical. This paper was born on a long journey that started with program logic and ended with the logic of evaluation. Consistent throughout was the idea that the discipline of program evaluation is a pragmatic one, concerned with *applied* social science and effective action in complex, adaptive systems. It gradually became the central claim of this paper that evidence-based policy requires sound reasoning more urgently than further development and testing of scientific theory. This was difficult to reconcile with the observation that much evaluation was conducted within a scientific paradigm, concerned with the development and testing of various types of theory.

Purpose: This paper demonstrates the benefits of considering the core essence of a program to be a proposition about the value of a course of action. This contrasts with a research-based paradigm in which programs are considered to be a type of theory, and in which experimental and theory-driven evaluations are conducted. Experimental approaches focus on internal validity of knowledge claims about programs and on discovering stable cause and effect relationships—or, colloquially, 'what works?'. Theory-driven approaches tend to focus on external validity and in the case of the realist approach, the search for transfactual causal mechanisms—extending the 'what works' mantra to include 'for whom and in what circumstances'. On both approaches, evaluation aspires to be a scientific pursuit for obtaining knowledge of general laws of phenomena, or in the case of realists, replicable context-mechanism-outcome configurations. This paper presents and seeks to justify an approach rooted in logic, and that supports anyone to engage in a reasonable and democratic deliberation about the value of a course of action. It is consistent with systems thinking, complexity and the associated limits to certainty for determining the value of a proposed, or actual, course of action in the social world. It suggests that evaluation should learn from the past and have an eye toward the future, but that it would be most beneficial if concerned with evaluating in the present, in addressing the question 'is this a good idea here and now?'

Setting: Not applicable.

Intervention: Not applicable

Research design: Not applicable.

Findings: In seeking foundations of program logic, this paper exposes roots that extend far deeper than the post-enlightenment, positivist and post-positivist social science search for stable cause and effect relationships. These roots lie in the 4th century BCE with Aristotle's 'enthymeme'. The exploration leads to conclusions about the need for a greater focus on logic and reasoning in the design and evaluation of programs and interventions for the public good. Science and research are shown to play a crucial role in providing reasons or warrants to support a claim about the value of a course of action; however, one subordinate to the alpha-discipline of logical evaluation and decision making that must consider what is feasible given the context, capability and capacity available, not to mention values and ethics. Program Design Logic (PDL) is presented as an accessible and incremental innovation that may be used to determine if a program makes sense 'on paper' in the design stage as well as 'in reality' during delivery. It is based on a configurationalist theory of causality and the concepts of 'necessary' and 'sufficient' conditions. It is intended to guide deliberation and decision making across the life cycle of any intervention intended for the public good.

Keywords: *Program logic; program theory; theory of change; program design logic; logic of evaluation; theory of causality; INUS condition*

Introduction

The classic *Foundations of Program Evaluation* (Shadish et al., 1991) describes the five attributes of any theory of evaluation: Theory of Social Programming, Theory of Knowledge, Theory of Valuing, Theory of Use, Theory of Practice. More recently, Donaldson and Lipsey (2006) discuss the role of theory in evaluation and set out the importance of Program Theory, substantive Social Science Theory, and Evaluation Theory. Both of these important works provide a useful description of the universe of 'theories' that evaluators should be aware of when approaching evaluation—they encourage us to think, what are programs, what can we/ should we want to know about them, and how can we know it? Substantive social science theory that describe the world and evaluation theory that prescribe how to do evaluation (Alkin & Patton 2020) are essential considerations for evaluation. This paper simply argues that programs themselves should not be considered as a type of theory and that doing so can lead to negative consequences for evaluation.

This paper is about a fundamental re-orientation towards programs. It proposes a move from treating programs as theories subject to scientific analysis, to courses of action subject to logical analysis. In concrete terms, this paper is concerned with a distinction lying at the heart of much program evaluation—the distinction between program logic and theory of change. It is the central purpose of this paper to demonstrate that once we view a program as a plan, or a proposition about the value of a course of action, rather than a theory, it is much easier to understand and evaluate logically. This also provides for more cost-effective evaluation in line with Scriven's principle that an evaluation should not cost more than the value of the information it provides (Scriven, 1976).[1]

Science is about "what kinds of things there are, as well as how the things there are behave" (Bhaskar, 2008, p. 20). Evaluating a theory is about whether it provides an accurate, useful or 'valuable' description of reality. Scientists should, and do, spend much of their time evaluating their theories, a process that involves much deliberation in the pursuit of collective knowledge. It may seem reasonable then to classify our interventions into the world as a type of thing whose behaviour may be understood using a scientific approach: in treating programs as theories. It is the central claim of this paper that in so doing we confuse the domain of science with the domain of engineering. Taking this approach ignores the useful distinction between knowledge about the world, and *applied* knowledge for making change in the world. This slight slip in our application of the word 'theory' has far reaching consequences for the practice of evaluation.

Treating programs as a type of theory tends towards a paradigm of scientific research and evaluation that is concerned with testing theory. Arguments for the important but supportive role for theory in program evaluation are further developed throughout this paper. However, the primary claim is that programs are not, on close examination, best understood as theories at all. It does not help to talk about lower case 't' theory or 'theory incarnate' (Pawson, 2013) as these terms obscure the fundamental nature of programs. The core essence of a program is considered to be a proposition about the value of a particular course of action: a plan or argument drawing on a set of reasons to suggest it will be effective. What little discussion there is on logic and the theory of argumentation in the literature on program evaluation tends to be focused on the forming of evaluative judgements (Schwandt, 2015).

This paper seeks to describe a primary role for logic in program design and evaluation. In

[1] Scriven looms large over this paper, not just for the injunction about cost effective evaluation, but in his suggestion that evaluation must be considered alongside logic as a candidate for 'master transdiscipline' (Scriven 2008). This paper treats logic and evaluation as two sides of the same coin and logical evaluation as the alpha discipline for considering the value of a proposed course of action for the public good. Program Design Logic described later in this paper provides a tool to conduct logical evaluation at any stage of the design or delivery of a program or intervention.

so doing, it seeks to pivot evaluation from a discipline concerned with testing theory to one focused on evaluating arguments. The results of a failure to more seriously engage with logic in program evaluation are revealed in 'program logic' diagrams that appear to provide a wish list of intended outputs and outcomes, rather than a serious attempt at explicating the 'logic' of any particular intervention. The focus on program theory tends toward evaluation focused on a single underlying 'theory of change'. Program logic is relegated to 'a tool that describes the theory of change underlying an intervention, product or policy' (Frechtling, 2007) despite the fact that innumerable theories must be leveraged for a program to be effective. There is considerable irony in the fact that very few, if any, theories of change have a corresponding 'theory of causality' or description of how change itself comes about. This has major implications for our discussion of logic in program design if we are to avoid this same shortcoming and provide a description of how programs bring about, generate or lead to change.

The paper concludes by describing a tool to support evaluation in the form of an incremental innovation called Program Design Logic (PDL). PDL provides a simple and coherent system for developing evidence-based policy and programs. It offers a diagrammatic model similar to an outcomes hierarchy, using the logic of 'necessary' and 'sufficient' conditions in place of 'outputs' and 'outcomes'. An evaluation using PDL will first determine if a course of action makes sense 'on paper', before we attempt to determine if it makes sense 'in reality'. It is the hope that PDL is used ex-ante by politicians, public servants, journalists, and citizens as much as by evaluators. In this way, PDL can be considered to be a tool for 'prospective evaluation', as set out by the US GAO, that 'focuses on a systematic method for providing the best possible information on, among other things, the likely outcomes of proposed programs (Datta, 1990). It may also be used to answers questions about the 'coherence' of a prospective policy or program, as set out in the updated DAC guidelines for evaluation (OECD/DAC Network on Development Evaluation, 2019).

The Ancient Origins of Logical Deliberation About the Value of a Course of Action

Aristotle is credited with developing the first systematic means of critical thinking and logic. In the *Art of Rhetoric* (2012) Aristotle concerned himself not with formal logical mathematical proofs but with means for considering propositions, including deliberation about the best form of action in a given circumstance. His concept, the 'enthymeme' is a form of logic based on syllogism. It has a series of premises that, if true, lead to a conclusion.

An enthymeme is a logical syllogism of the same form used in deductive logic but with two distinct features. First, some assumptions are unstated and, second, there are premises in the argument—as well as the conclusion—that are probable rather than certain. The movement from premise to conclusion is like an informal argument (Toulmin, 2003a) based on warrants, or reasons, that allow us to make inferences from the facts associated with one premise to the next or from the premises to the conclusion. This paper suggests that programs as propositions can be considered as an 'enthymeme' and a special case of deliberation when the subject is the value of a proposed course of action.

To compare an enthymeme with deductive logic, let us consider the most famous example in deductive logic, 'Socrates is a man, all men are mortal, therefore Socrates is mortal'. This is a valid argument using deductive logic as the conclusion is already inherent in the premises and no new information is created by the argument. If the premises can be established, then the form of the argument is the guarantee of its truth. Science may be used to classify Socrates as a man, it may also be used to gather inductive evidence that 'all men are mortal'. The work of establishing the premises, that is, the reasons or warrants for believing the premises, is provided by science; however, the conclusion that Socrates is, himself, mortal belongs to logic, in this case, deductive logic. Of course, while we say science established the premises, they will have themselves been established as the result of a conclusion formed using scientific methods (something along the lines of death rates of 100% of humans over time, leading to the logical conclusion that all men are mortal) and, as such, are also the result of logic. The point here is not only that science proceeds under the supervision of logic but that, once established, premises don't need to be re-tested. Instead, they can be marshalled to develop an argument about the value of a course of action. For example, once it is established that immunisation for polio protects against polio, our argument about the value of an immunisation program just needs to make sure we implement it faithfully. While uncertainty always exists, and adverse consequences are possible, this means that the evaluation of a program of immunisation can focus on issues for implementation, rather than on measuring outcomes, i.e. in terms of polio rates.

Consider an argument for rescuing Socrates in the form of a deliberative enthymeme. We might propose, 'Socrates has been condemned to death. If we rescue Socrates, he will survive. If he survives, the youth of the future will be better educated; therefore, we should attempt to rescue Socrates.' The first point is to note the missing assumptions—these might include 'Socrates is a man' and 'all men are mortal' because we can assume the audience will already agree that Socrates can be harmed by death. The next is to note the lack of certainty. There is clearly a lot of room for deliberation about the validity or well-groundedness of the argument for attempting to rescue Socrates. This specific argument contains a large degree of uncertainty and also questionable inferences. There is an assumption that if we rescue him, he will survive; but maybe Socrates has already been poisoned? There is an inference that if we attempt to rescue him, we will succeed; but maybe all past attempts at rescuing condemned prisoners from this location have failed? Knowing the latter would provide evidence that might prevent us from inferring that we should attempt to rescue Socrates. We may even question whether it can be inferred that the youth of the future will benefit if Socrates is rescued, given that Plato and others had already written down much of what he had to teach. We may not have sufficient warrants or good reason to infer a number of steps in the argument, up to and including that we should attempt to rescue Socrates. We might conclude that this argument or enthymeme does not appear to be sound. Unlike deductive logic, this does not follow with certainty from the validity of the premises and the form of the argument but requires a judgment to be made, it is therefore informal.

Let us take a further step and consider an argument about a particular method for rescuing Socrates. Perhaps it could be argued that we should bribe the prison guard, so that we can switch the hemlock with a harmless substance (Socrates has decided if the authorities say he should die then being the responsible citizen that he is, he will probably drink the hemlock and might alert the guards that he is being rescued). We then argue that, as Socrates is being transferred from his confines to the ritual place of death, we send in a group of people from Thebes to abduct him. We then support him in exile, to the greater glory of their city and the benefit of their youth. Of course, all sorts of reasons will have to be given for how to replace the hemlock, with what substance, how much to bribe the guard and which Thebans can be trusted, etc. There is, of course, much to be discussed about the logic of this plan of action, which is not reducible to the form of the argument but, instead, includes the artful selection and presentation of compelling evidence. Incidentally, the importance of art, as much as science, to an argument is how rhetoric received its bad name. It does not follow, however, that just because persuasive techniques exist, that there is no inherent logic that can be interrogated in relation to the soundness and well-groundedness of an informal argument.

Now, let us consider the value of a program of rescuing political prisoners more generally. We are now moving from one-off actions to something that more closely resembles a program. We will be concerned with whether the program is likely to be and is, in fact, generally effective, rather than the merits of a specific plan. In developing this program, we might develop a set of criteria for who should be rescued, which reflects our 'values'. We might describe how much effort we should exert to rescue these people and set out a range of options that might be more or less effective in different situations relative to cost.

These options might be established based on explanations using theories of human behaviour, data about how and why they were or were not effective in the past, and how much they cost. Once options coalesce into a program, we would expect compelling reasons to be provided as to why this program will work. We might expect specific evidence to support a claim about the benefits of a particular method in a particular situation. We would expect these reasons to be compelling enough to encourage us to contribute funds to the program and we would want to know if these reasons turned out to be good ones. Once funded, we may wish to hold this program to account and ask how many attempts were successful and, further, where and when and under what circumstances they were successful. We may even refine our program as a result. This sequence of reasoning might describe a typical program development or policy cycle (Althaus, Bridgman, & Davis 2017), with lots of reasons and arguments provided along the way. The role of theory here is as a special class of warrant or reason to infer some element of the program will be effective. It should be clear by this stage why this paper argues for the primacy of logic and argumentation over theory when it comes to program design and its evaluation. Programs are built on theories but they are not, themselves, a theory; they are a specific argument and should be evaluated as such.

Despite this long tradition in argumentation to draw upon, this is not how program logic has developed. Program logic has received substantial attention in the evaluation literature and by public servants commissioning and conducting evaluation of public policies and programs (Frechtling, 2007; Funnel & Rogers, 2011; McLaughlin & Jordan, 2015). While there is much discussion about the form of logic models and their use in communication, including the need for 'decreased mental effort' and 'perceived message credibility' (Jones et al., 2019), there is relatively little explanation of the logical foundations of program logic.

Program Logic in Practice

Program logic, in practice, may be defined as a diagram or model that sets out the important components of an intervention and how they are combined to deliver a specified outcome. This is usually in the form of a series of boxes that contain propositions linked by arrows. There are numerous different forms this can take—the two most common are the outcomes hierarchy and the log frame. The former being a vertical series of outcomes that are expected to occur if the intervention is effective. The latter referring to the intended inputs, outputs, outcomes and impacts of an intervention. In both cases, if there is any description of the logic behind the sequence of boxes it is referred to as 'if-then' logic—an implicitly successionist theory of causality to which we will return later. There are a number of logical problems with both these models that will be briefly discussed.

Logic Chains

The Kellogg (W.K. Kellogg Foundation, 2004) and Wisconsin models (Taylor-Powell & Henert, 2008) set out a 'chain of outcomes' between inputs, outputs and outcomes. It has benefits in clearly communicating key information about a program. It typically displays information such as the resources and activities as well as the intended outputs and outcomes. It is primarily designed to satisfy a need 'to enhance program performance through outcome accountability' (Hernandez, 2000). Partly for this reason, these models tend to look more like a wish list than a logical proposition. They provide a series of claims about outputs and outcomes but do not articulate any logic that may be subjected to interrogation. It may focus on intended outcomes that may become the subject of an evaluation and assumptions—'the beliefs we have about the program, the people involved and how we think the program will operate' (Taylor-Powell & Henert, 2008)—but provide no clear means of identifying flaws in each step or the overall logic of the design at an early stage. The approach of Renger and Titcomb (2002) places more emphases on considering 'antecedent conditions' as well as 'target activities' and measurement issues;

however, all these models focus on evaluation in terms of accountability for outcomes. These types of models are sometimes considered to provide a 'road map', which may be very useful for communication and plans for measurement, but less so for interrogating the validity of a propositions about the value of this course of action. Although they take the form of a causal model or an argument structure—a series of boxes and arrows—they are neither good arguments nor causal models, lacking warrants in the case of the former and probability coefficients in the case of the latter.

Outcomes Hierarchy

This approach sets out a series of outcomes to be obtained in sequence. It identifies the conditions that are to be achieved and places them in an 'outcomes hierarchy' or sequence (Rogers & Funnell, 2011). Typically, each condition statement is broken down into its component parts and defined in a 'data matrix' in terms of the attributes of that condition statement, and then into outcomes that may be observed and/or measured in existing or new data sets for monitoring and evaluation. Together, the outcomes hierarchy and data matrix provide a roughly sequential and sufficiently detailed checklist for the evaluator to follow and to determine the extent to which each intended condition was, in fact, brought about. This approach also identifies assumptions that the program is making and external factors that will also affect results relevant to the success of any program. This is the version of program logic that I was taught by my mentor, Chris Milne. I have used it valuably numerous times and it is the version that I have adapted in designing Program Design Logic. I owe a great debt of gratitude to Chris for the many discussions we have had about program logic, as I do to Gill Westhorp who helped me understand the power and value of theory in evaluation even if this paper has a different focus.

The main drawback of the outcomes hierarchy for program design is the absence of a theory of causality that would allow someone to critique a program design in the design phase. There often appears to be an implicit assumption that one outcome causes the next. Most often, what is described, however, is that one condition lower down the hierarchy is a precondition for another higher order outcome. This is far from being a cause and it means the outcomes hierarchy is far from being a causal model. Assumptions are listed but they are usually disembodied from each step in the hierarchy—often only appearing as general assumptions about the operating context. Sometimes even core social science theories or 'theories of change' are described as assumptions. There is nothing wrong with an outcomes hierarchy for designing a post hoc evaluation—it is very useful for identifying key issues for an evaluation to examine. What it is less good for is surfacing the logical flaws in an intervention design at the outset. The tacit acceptance of 'if-then' logic is rooted in a linear view of causality that tends to overlook the idea that assumptions are equally as important as the conditions brought about by the intervention itself. This kind of program logic can lead to the development of programs with unrealistic expectations, money wasted trying to measure outcomes that could not logically occur, and, ultimately, frustration, cynicism, lack of progress and a high risk of program failure.

Common Problems with Logic Models

Regardless of the form of program logic, it is a common critique that it provides an artificially linear 'chain of outcomes'. It is a critique often rebutted with 'it's just a model'. However, there is a deeper point to be considered. Standard program logic diagrams suffer from an implicit assumption about how change occurs (i.e. theory of causality), which is often reflected in program logic being seen as having a 'causal chain'. As such, pipeline models and outcomes hierarchies are sometimes described as displaying 'if-then' thinking. The Kellogg guide to program logic actually uses the phrase 'if this happens, then we hope this next thing will happen'. Unfortunately, this approach does not provide any evidence to put a brake on overly optimistic ideas about a program. On inspection, pipeline models are not structured causal models in the form meant by scientists interested in analysis; they do not include mediators, confounders, probability statements or any of the core elements used in causal diagrams (see Pearl, 2019). The outcomes hierarchy addresses this problem somewhat better than the pipeline model, but it is still deficient. What we find in an outcomes hierarchy is an articulation of 'if not x, then not y'. For example, if someone is aware of a program, it does not cause them to enrol in that program; however, it may be a necessary precondition. That is, if they were not aware of the program, then they would not enrol.

The realisation that neither the pipeline model or the outcomes hierarchy provided the 'chain of cause and effect' that is so often referred to when discussing program logic was hard to understand. This realisation was the first step towards the idea that what we are setting out in an outcomes hierarchy is most often a claim that 'if not this early condition, then not this later condition'. The second step was much more difficult. It emerged from trying to understand what exactly is meant when we say that a program causes some outcome. I spent many years trying to work with a realist 'generative' theory of causality as applied to an entire program. I couldn't make it work until I realised that this was not the only theory of causality, and that maybe different theories of causality are more or less useful for different problems.

Theory of Causality

In this paper, we will deal with only a very small number of candidate theories of causality that are useful for evaluation. We focus on those theories most associated with 'manipulable' causes as we are concerned with the effects of programs, rather than, more generally, with how things in the world come to be. A theory of causality is very different to a specific theory of change. A theory of causality describes how we think programs, or interventions in general, lead to, generate, result in, or cause change. Pawson (2008) has outlined three candidate theories of causality.

- The presence of something is invariably followed by the presence of something else (*successionist*).
- The configuration of certain somethings immediately brings about a new something (*configurationalist*).
- The presence of something with certain latent powers interacting with the latent powers of something else in a certain context creates a new something (*generative*).

The successionist approach derives from the empirical and positivist tradition, whereby 'causation' cannot be observed directly but is a construct of the mind built up from experience. This approach follows the position of Hume and John Stuart Mill that a causal relationship exists if, 1) the cause preceded the effect, 2) the cause was related to the effect, and 3) we can find no other plausible explanation for the effect other than the cause (Shadish et al., 2002). Programs are conceptualised as a complicated series of cause and effect relationships. The theory of causality is based on what can be observed and echo's Hume's search for a 'constant conjunction of events.' The goal is the 19th century positivist dream of Comte to find laws that govern human behaviour, and then perhaps unconsciously but perniciously, to shape a world that seeks to control the individual through knowledge of these laws (see Comte & Lenzer (ed), 1975). In program logic, there is often an attempt to render these forces into a 'causal chain' using 'if-then' thinking as discussed above. This theory of causality is so often implicit, that the history of modern evaluation has in many ways been a debate about the virtues and drawbacks of randomised controlled trials or RCTs as a means of measuring the size of an observed effect following a purported cause. As an example, a discussion about causation by two lions of the field, Cook and Scriven, is in effect a series of arguments for and against RCTs (Cook et al., 2010).

This successionist approach works well when causes can be investigated with a counterfactual and an experimental design. The RCT allows for causal inference when the assumptions that underpin it are met; a stable intervention that can be sufficiently specified, ample sample size, detectable effect sizes and most controversially, when knowledge is to be gained by controlling for rather than incorporating context into the equation (Hawkins, 2016). Applied to more complicated and complex entities, such as programs delivered by people, it is difficult to be precise about the part of an intervention each intended beneficiary is experiencing; therefore, it is difficult to be precise about what parts of the intervention caused anything. This is crucial for any causal claim with external validity that is for knowledge derived from the past but can be useful for replicating past effects in future contexts. Lest we be seen as overly critical about successionist causality, there is substantial scope for development of a 'new science of cause and effect' (Pearl, 2019) in program evaluation as much as in fundamental science. Using Bayesian statistics that bring context into the equation, this approach uses casual diagrams that articulate hypothesised causes, effects, mediators and confounders. This is the era of 'personalised medicine', where modelling is used to find out what works for you, rather than whether an intervention works per se. These new methods can deal with non-linear dynamics—a core source of critique by realists of the positivist position (i.e., the view that context matters and that different people react in different ways to the same stimulus). These new methods of causal diagramming could provide a new structure for program logic that is able to describe the causality inherent within a program; however, this has not yet been achieved by program logic commonly in use.

A configurationalist theory of causality recognises that a number of factors need to come together. Rather than a 'causal chain', there is a 'causal package' that brings about a change. This is the approach Nancy Cartwright (2012) uses to describe evidence-based policy. It is also the theory that underpins Charles Ragin's work and the method of Qualitative Comparative analysis (Ragin, 2009). Using this conception, causal analysis is like baking a cake; for success, you need the right combination of ingredients, mixed in the right way, and placed in the right context (i.e., an oven at the right temperature). The strict order is not always important (i.e. wet before dry ingredients or the other way around) and it doesn't make sense to say how much of each ingredient causes the cake—you need all of them as well as the right environment for the cake to bake. You may not need to get very detailed about things such as how the 'rising agent' in the cake works, unless the rising agent turns out to not be working. This is the theory of causality considered most useful for articulating the initial logic of a program. Using this conception, a program is a sufficient condition. We will turn to this component of necessary and sufficient conditions after we deal with one more theory of causality.

A generative theory of causality explains the underlying mechanism of change and the features of context that are important for change to occur. It refers to abstract, latent or dormant mechanism that make up a 'real world' and have 'real effects' (Bhasker, 2008). A program or intervention works through actions that leverage these mechanisms (often the reasoning and resources of individuals) in certain contexts to generate outcomes[2]. To continue with the cake metaphor, the cause of the cake rising is the release of carbon dioxide, which is due to bringing the leavening agent, baking soda or yeast, into a context of moisture and heat. As with the configurationalist approach, it still doesn't make sense to say how much the leavening agent did it and how much the heat of the oven did it—they were both necessary. But further to the configurationalist approach, it specifies the mechanism—the 'rising' occasioned by the release of carbon dioxide. The realist generative conception is perhaps the most useful way for understanding how change occurs for individuals and for developing theories about what works for whom, under what circumstances and how. Realist Mechanism, Context, Outcome configurations may provide the most ambitious articulation of the casual power inherent in an intervention.[3] In this author's experience,

[2] We don't cause gravity; we arrange the conditions between a ball and an edge to leverage and allow gravity to cause the ball to fall. There is an infinite regress here, in looking for 'initial causes', because each condition is itself a result of actions that have conditions and causes; for example, the causes of me deciding to push the ball towards the slope. It will not be useful for our discussion, however, to descend into this abyss. In program logic, we are changing conditions that allow for the unleashing, firing or strengthening of causal mechanisms that already exist in the world.

[3] CMOs refer to the context before the purported causal mechanisms. This can be confusing because the features of context are only relevant in relation to the causal mechanisms. It may be less confusing to refer to a MCO, ordering the causal recipe in terms of mechanisms then

however, these are too detailed and granular as a starting point for thinking about a program and its evaluation. A realist approach requires substantial time and expertise and attention to the nuances of program design that are not often apparent in the everyday world where policies and programs are developed. A realist approach may be useful when building a program, when trying to understand why an intervention that was thought to be logical is not effective, or when trying to work out how to modify or target an intervention for maximum impact; though it may require too much depth of analysis for everyday program evaluation.

The conclusion of this brief survey of theories as applied to program logic is pragmatic. It builds on an understanding that the goal of evaluation is to reduce uncertainty about the value of interventions in the world. In evaluation, where time and money are major concerns, we are restricted to the 'fair price of causal information'. We are not concerned with research with the much more difficult goal of determining some enduring truth about the nature of the world. The theory of causality we invoke will depend on the questions we can afford to answer. It will also be built on an understanding of nature and science as 'stratified'. Nature into the domains of the real, actual and observable (Bhaskar, 1979) and science into the different disciplines for interrogation of nature, including physics, chemistry and biology (Toulmin, 2003b). We use physics to understand the behaviour of sub-atomic particles, chemistry to understand the behaviour of molecules, and biology to understand the behaviour of animals; however, we do not talk about animal behaviour in terms of valance shells, neutrinos or dark matter because it is an unhelpful level of granularity and goes beyond the strata of science with which a science is concerned. None of these stratifications are 'right' in a philosophical sense but are more or less useful depending on the degree of granularity required to answer a specific question. Instead of finding the 'right' theory of causality, the pursuit should be concerned with the most useful one in the circumstances. While evaluation is distinct from science, the analogy is that the understanding of the causal powers within a program may require us to draw on theories of causality with different 'depths', depending on how deep we can afford to go. In the same way as almost every major evaluation theorist says, "questions come first and method choice comes second" (Cook et al., 2010, p. 113), we may say that what you are trying to explain comes first and your theory of causality comes second.

The conclusion as applied to program logic is that it may be best when designing an intervention to put aside a successionist or generative theory of causality. There was no succession in the logic diagrams and the generative trees were obscuring the forest. At the design stage, and when considering the program as a whole, it may be most appropriate to invoke a configurationalist theory of causality and view a program as a 'causal package'. This was the third step in the development of PDL. This position follows the work of Nancy Cartwright (2012) in developing evidence-based policy. It developed into PDL with two further developments, a consideration of necessary and sufficient conditions, and the recognition that the essential nature of a program is not a type of theory but an argument about the value of a proposed course of action. We will discuss these developments in the next two sections.

INUS Causality

To appreciate the foundations of necessary and sufficient conditions, it is instructive to consider Hume's 1739 definition of cause, which claims, 'We may define a cause to be an object followed by another and, where all the objects, similar to the first, are followed by objects similar to the second. Or in other words, where if the first object has not been, the second never had existed.' As Judea Pearl (2019) points out, these two sentences are not equivalent. Pearl takes issue with the first sentence for confusing correlation with causation, while he views the second as an attempt to introduce counterfactual thinking. It can, however, also be seen that the first

the context in which it is effective; this may make the outcome easier to follow, even if the logical 'form' of the MCO or CMO is equivalent.

sentence provides a definition of cause as *sufficient*, because observing the first object is sufficient for observing the second. The second sentence adds that the first object must be *necessary* for the second. Thus, in combination, we are describing two different elements of causality—necessary and sufficient conditions.

This is the approach of seeing causes as INUS conditions (Mackie, 1974). On this account, a cause is an Insufficient but Non-redundant (i.e., it is needed) part of an Unnecessary (i.e., there are other ways) but Sufficient condition. The famous example of this is the faulty electrical circuit, which is considered as the cause of a house burning down. While the short circuit was necessary for the house to catch fire in the way that it did, this was not the only way it could catch fire, and was not, on its own, enough as the fire also required the presence of oxygen and the house to be built of combustible materials.

Applied to program design, while an intervention is rarely *necessary* or the only way to achieve something (e.g., there is more than one way to increase social cohesion), it must be *sufficient* for its objectives (e.g., to bring together diverse members of society in a social sporting exercise). The program will have components that we think are necessary and, when all achieved, must be sufficient to achieve an outcome. The program, however, might not be sufficient to cause a very ambitious change or 'ultimate intended outcome' to which it only contributes. On this account, it is questionable whether it is ever useful to measure how much an intervention contributed to this type of outcome as the program was never designed to be sufficient for achieving it (i.e., whether a homelessness program is intended to solve homelessness, rather than to merely address one aspect of it).

We will return to INUS causality as it has a central place in PDL and represents the fourth step in the development of PDL. For now, it is important to note that in avoiding use of the term 'theory of change', PDL is engaging explicitly with theories of causality, but is not concerned with testing 'program theory' that is so often the focus of modern evaluation practice.

Conflation of Program Logic with Program Theory or Theory of Change

The centrality of 'theory of change' and 'program theory' in modern theory-driven evaluation practice can be explained by an examination of evaluation history. The origins lie in the early experimental and positivist approaches to evaluation (Shadish, Cook, & Campbell, 2008) and dissatisfaction with the usefulness of 'black box' evaluations of programs. These invovled careful observations of outcomes but provided little explanation of how the outcomes were achieved or what precisely would be required to replicate the outcomes elsewhere. One early legend of the field, and original proponent of experimental design Carol Weiss who struggled with its practical application and ability to identify program effects opened an edited book with 'This book is dedicated to overcome the kind of naivete that I began with' (Weiss, 1972). Weiss (1995) often discussed the importance of theories (note the plural) of change that underpin a program. This concern with reasoning also led to Rossi championing a 'theory-driven' approach to evaluation (Shadish, Cook, & Leviton, 1991)[4]. Pawson took this position to its logical conclusion when he referred to programs as 'theories incarnate' (Pawson, 2013). Described in the realist evaluation reporting standards 'In a realist evaluation, the assumption is that programmes are 'theories incarnate'. That is, whenever a program is designed and implemented, it is underpinned by one or more theories about what ""might cause change", even though that theory or theories may not be explicit" (Wong et al., 2016, p. 1). This idea is that programs are composed of actions and these actions may or may not fire certain 'mechanisms' that generate change, depending on the circumstances and context in which the action is applied. The

[4] While all three of these theorists were. Like Weiss, early proponents of experimental evaluation, they all modified their views over time as they confronted the problems of the internal and external validity of knowledge claims that derive from experimental evaluation. 'Internal validity' being concerned with claims that the program (or some part of the program) caused an outcome, 'external validity' being concerned with the claim that similar outcomes from rerunning the program can be expected in future times and places.

identification of mechanisms and outcome configurations is the basis of a realist approach to evaluation. Pawson is concerned with replicable, or portable CMO configurations. In order to be scientific, Pawson and other realists have had to dispose of the idea of a program or action having a certain causal force (and therefore being replicable) and restrict themselves to dealing with the psychological and sociological 'mechanisms' that have causal powers and, that if leveraged by program or action, may be expected in some, but not other, circumstances and contexts to generate change.[5] While not everyone is a realist, many evaluators place central importance on the idea of theories for program design and on theory-driven evaluation. These evaluators may or may not make use of a form of logic model but tend to focus deeply on the reasons why a change should be expected. These tend to have been written as 'theories of change' rather than as simple reasons.

A theory of change may describe *how* or *why* the move from the problem to outputs and outcomes in a logic model is expected to occur. A theory of change (ToC) may be written about a whole program (i.e. program theory of change) OR about a component within a program. Extending an example from Funnel and Rogers (2011), a theory of change that is also a program theory may be written as:

> IF we provide apples to people who have insufficient vitamin C AND people eat those apples THEN we expect their levels of vitamin C will clinically rise significantly enough that they will avoid scurvy BECAUSE apples contain enough vitamin C to overcome a deficiency.

While a theory of change about a component within a program may be written as:

> IF we shine the apples that are delivered to people who like shiny apples but leave them un-waxed for those who have more organic tendencies AND we place the apples by a person's desk rather than in a common bowl THEN they will eat the apples BECAUSE attractiveness and proximity support the decision to eat apples.

In the examples above, these statements are offered as theories of change. They are, in fact, claims supported by reasons, which makes them, in other words, parts of an argument. Theories are about what things there are in the world and how they behave. A program is just one way amongst many potential means for bringing about change. A good program, like any good piece of technology, should be built on sound theory. A theory plays an important role in providing reasons why a program or component is thought to be effective BUT a program, itself, is not a theory. A program will often draw on a main substantive social science theory (e.g. 'the theory of scurvy') but the theory is only important in so far as it provides a warrant or reason to think the program activities will be effective if implemented as intended (often by following a 'theory of action'). For example, in a parenting program, there may be a theory about why improving certain mothers' parenting skills in a certain way will lead to better parenting, but there may also be a theory as to why distinct approaches to marketing the program will work differently for mothers depending on their circumstances, motivations, media consumption, etc. A 'theory of action' sounds a lot more like a plan than any kind of theory – we must be cautious of scientism. This paper argues that it is unhelpful to consider interventions primarily from the perspective of 'theory' and that is more appropriate and useful to conceive of a program as a logical argument about a course of action supported by evidence or reasons to think that it will be effective.

I was recently asked at a conference why I no longer believe that programs are theories incarnate. I responded, 'I don't think it's wrong, I just don't think it's the most useful way to see a program. Is catching a plane to

[5] In so far as realists seek knowledge of CMOs, rather than seeking to warrant programs as effective, it would appear sensible for realists to avoid the term 'program theory' at all. The realist evaluation discourse is about the adequacy of CMO configurations for explaining observed patterns or 'demi-regularities'. Inevitably 'jobbing' realists (to use a Pawson term) must have conversations about a program or initiative as a whole. In these situations, this paper suggests the logic of the proposition could be explicated in realist terms by the extent to which program activities are founded on an understanding of the underlying mechanisms generating an actual problem and are appropriately designed to harness CMO configurations to make a meaningful difference to that problem.

get home from a conference, theory incarnate? Is flying a plane, theory incarnate? Is flight, theory incarnate?' And this is the nub of it, successful programs leverage many theories—but to treat the overall program as a type of theory leads to evaluation where the program is the unit of analysis. This is despite the realist unit of analysis being context-mechanism-outcome configurations. It seems in the everyday world of public policy the program often remains the unit of analysis—this paper suggests dealing with this reality and focusing program evaluation by ensuring programs are logical—including but not limited to, that they are leveraging appropriate theories. There is a grave risk that too much discussion of theory provides a fig leaf for programs that sound good 'in theory' but may not be adequately specified to have much chance of being useful.

Catching a flight home is 'a plan' and one that can be evaluated without much useful recourse to theory, just some rational decision making. Flying a plane might be theory incarnate but there are many, many theories about the world that are being incorporated (the Bernoulli effect is a key one but so are theories about gravity, wind dynamics, air pressure, metallics, etc.) and, when we launch the plane, we are not testing these theories but trying to leverage or make use of them. Flight itself may well be theory incarnate but 'flight' in the abstract is a long way from a practical means of getting home from a conference.

Those delivering social programs are more like social engineers than social scientists and methods of evaluation should be more akin to those that are useful in applied engineering than in pure science. A program should be evaluated like technology is evaluated; not how scientific theories are evaluated. For example, if we are evaluating a battery, we will see whether it works and how long it lasts, we will rarely need to get into the theory of chemical reactions and theories of electron valance shells; if we do, it will not be to test these theories, so much as to make sure we are properly leveraging them. If the battery is not working as intended, despite all logical connections, then we may wish to seek to determine if we are properly leveraging an established theory. In rare cases, we may develop a new theory but experimenting on people in society to identify new theories of behaviour should not be a primary goal for public policy and its evaluation.

The founder of modern argumentation, Stephen Toulmin, in *Return to Reason*, argued that the success of science in some domains and the quest for certainty and specialisation of thought into disciplines has led us to valorise 'abstract theories of the world that apply in general and, therefore, nowhere in particular' (Toulmin, 2003b). Once it is realised that a program is not a theory but a proposition, it becomes obvious that applying methods that are scientific in some domains of research will not make an evaluation scientific—this is well known 'scientism' that must be called out. This does not mean that there is no place for RCTs—there certainly is; however, the decision-making criteria should not be the hierarchy of methods but the method that will provide the best answer to a specific question. Once we realise the evaluand for what it is—a proposition about a course of action rather than a scientific theory—we can be a lot more accommodating of the different methods that may generate evidence for better decision making and reduce uncertainty about the value of that course of action.

Conclusion of the Discussion of Theory

Developing and testing theories is the focus of scientific research. Evidence-based policy is about putting theories to work. Program logic is sometimes referred to as a model of a program's theory of change. This assumes the program is a theory about how to achieve ultimate outcomes rather than a proposed course of action for a specific time and place. The pernicious result of thinking about programs as theories of change is the logical fallacy that it is useful to measure how much a program contributed to an ultimate outcome—an outcome for which it was never designed to be sufficient. This limits much 'impact evaluation' using 'scientific methods' such as RCTs to a means of providing accountability by measuring outcomes that can be attributed to the program. When applied to complex social programs in complex social systems the results of RCTs are unlikely to generate scientific evidence to better inform future decisions as we navigate the complex and adaptive systems into which we intervene (Hawkins 2016). The realists worked out the problems of the experimental design, but the systems thinkers have most fully leveraged the insights of complexity to propose effective means of navigating uncertainty (Kurtz & Snowden 2003, Renger 2015).

Program Design Logic

To recap, this article suggests that a program or intervention is, at its core, a proposition that a certain course of action will lead to a certain set of outcomes. An *evidence-based* program provides good reasons to expect actions will lead to outcomes in the future. Program Design Logic (PDL) provides a framework to determine if a proposition makes sense 'on paper' and outlines how it may be determined if it makes sense 'in reality'.

What do we mean by 'makes sense'? Thomas Schwandt (2015) has engaged heavily with the concept of warrants in evaluation. He has done so primarily in the context of evaluators forming valid judgments about the value of a program or intervention. He is concerned with how a fact or outcome provides evidence about the value of an intervention. PDL simply steps back one step further and considers the program itself to be a form of argument that must be sound and well grounded. That is, that there are reasons to think or warrant that the program is likely to be effective, before the first participant is even enrolled.

A PDL is simply a visual depiction of the necessary and sufficient conditions—with space to identify assumptions and external factors. Assumptions are crucial to the functioning of the program and will often need to be evaluated. External factors will moderate the extent to which the program or intervention outcomes contribute to ultimate intended outcomes, sometimes called contingent conditions. As we discussed, an intervention is rarely the only way to achieve something (e.g. there is more than one way to increase social cohesion) but it must be sufficient for its objectives. It will have components that we think are necessary and when all achieved are sufficient for achieving the objective. In the philosophy of causation, each component (i.e., output) is considered to be an INUS condition. That is an insufficient but non-redundant (i.e., it is needed[6]) part of an unnecessary (i.e., there are other ways), but sufficient condition (i.e., the program). It is important to be realistic about what the intervention is sufficient for and what it will only contribute towards. Typically, the overall intended outcome is more ambitious than what the program can reasonably be expected tdeliver on its own.

PDL uses an outcomes hierarchy approach with two major differences. First, it has an explicit configurationist theory of causality. Second, it uses the language of 'necessary' and 'sufficient' conditions in place of 'outputs' and 'outcomes'[7]. While outputs and outcomes are important—the distinction between the two is the source of much confusion and a matter of perspective rather than an absolute attribute of some or condition. One person's output is another person's outcome and vice versa, whereas a condition statement simply is. For example, an increase in a child's self-esteem may be an outcome to a parent; however, it may simply be an output of some training program whose ultimate outcome is increased school attendance. In government, there is much angst spent over the distinction between outputs and outcomes. If we instead focus on the condition statements that a PDL suggests are necessary to our intervention and sufficient for bringing about a better state of affairs, we can focus on the conditions that are most doubtful, most variable, or most affected by individual differences and features of context etc., rather than arguing over whether something is an output or an outcome.

The key features of PDL shared with an outcomes hierarchy are the boxes with condition statements. These are propositions or conditions in the form of 'who or what is in what state'. These can be arranged in a

[6] At least it is thought to be needed. For cost effective programs, the focus should initially be on what is truly necessary or the 'minimal viable product'. Over time, there may be opportunities to test the proposition that some aspect of a program is actually redundant or superfluous.

[7] The term 'outcome' is used interchangeably with 'condition' because this is familiar language; although it would be more correct to say condition. Outcome may imply that the thing immediately prior caused it—an inherently successionist way of thinking. Established language, however, is hard to shift.

hierarchy as some steps are logically prior to others (you can't put a cake in the oven until you have mixed the ingredients), but we are not claiming that one thing leads to the next. We are claiming that one or more intended objectives or ultimate outcomes are achieved by progressively achieving subordinate conditions. It is actually pretty simple to set out the intended conditions that some program activities are meant to bring about. The tricky bit is being realistic about the extent of possible achievement and in surfacing assumptions.

Typically, an overall intended outcome is more ambitious than what the program is resourced to deliver on its own. The options are to change the program design to make it more ambitious—essentially doing more things to increase the likelihood that the program is sufficient for higher order outcomes. The other option is to accept the ambitions for which the program is logically sufficient and focus on achieving these. Measuring contribution towards higher order outcomes may be useful for accountability and to compare the outcomes of different interventions, but it is not a sensible or cost-effective approach for scientific or formative evaluation if the program was never hypothesised or designed to be sufficient for some goal.

In a PDL (see Figure 1)—as in Aristotle's enthymeme—there are usually propositions that are not explicitly stated. A program will also almost always require implicit or explicit assumptions about the context for the necessary conditions to provide a sufficient 'causal package'. These assumptions are things the program is relying on to be the case for it to be effective, but about which it is not doing anything. For example, if we are building schools to educate children, we are assuming that there will be teachers available to be hired to teach in the schools. Major assumptions should be checked because the conditions brought about by assumptions are just as important as the conditions brought about by an intervention in determining whether an intervention is sufficient for some new state of affairs.

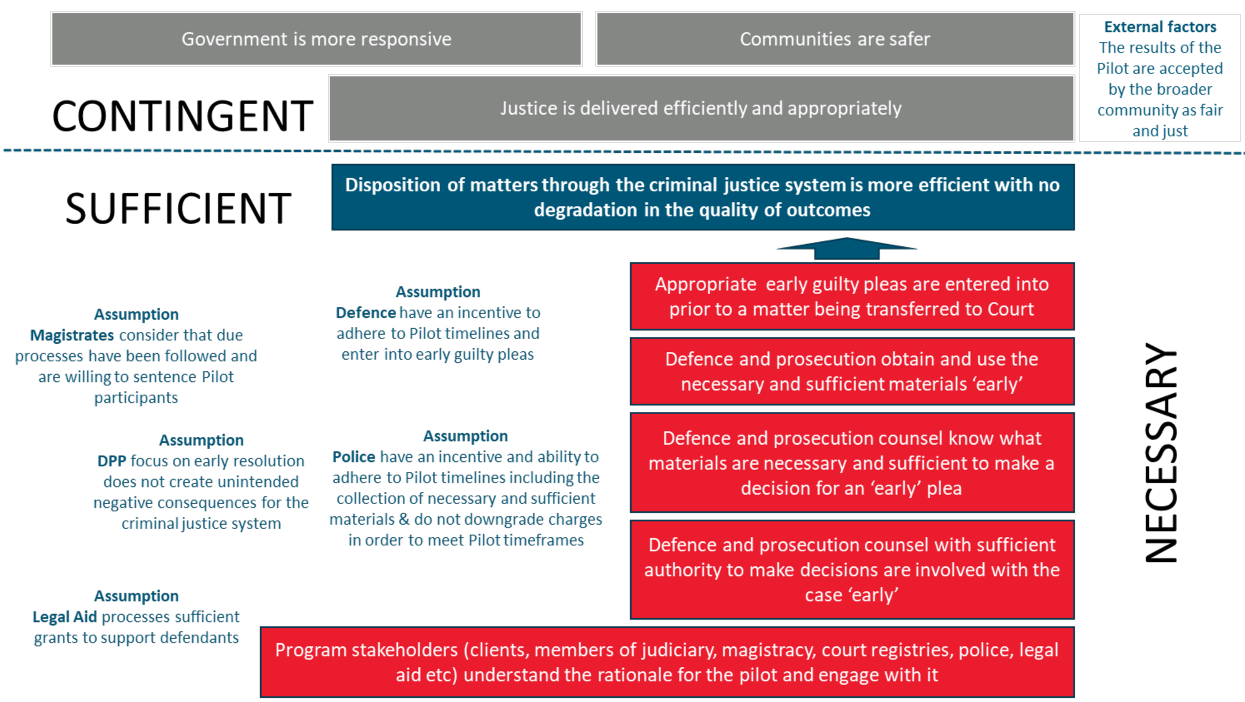

Figure 1. *Example of a PDL.*

Warrants, Assumptions, and Theories of Change

A sound argument or proposition about the value of a course of action needs warrants or reasons to support the claim. A warrant is what allow us to infer or move from the observation of some fact towards claiming that this fact is 'evidence' for something. This is where the theories of change are used in PDL.

In addition to a one-page diagram, PDL will sometimes include an articulation of the reasons why we expect it will be effective in the form of 'if we do xx, we will achieve yy because of zz'. This may be considered by some to be a theory of change that seeks to hypothesise or claim how and when the program is working, or likely to work, to deliver its intended objectives. These reasons will include a mix of common sense reasons as well as what others have referred to as capital 'T' and lower case 't' theory. In its most mature form, we would expect a PDL to be accompanied by realist theories that have been developed and tested and include mechanism, context and outcome configurations. Not all PDL, however, will need to be or can afford to be so well developed as to include realist theories of the mechanism and contexts by which each premise is achieved. Initially, it just needs to 'pass the pub test', assuming this test were conducted over half an hour and involved drawing on napkins rather than a loud and glib single slogan.

Evaluation of Program Design Logic 'On Paper'—Is it Sound?

A valid PDL must provide a sound argument. It must follow that if the premises of the argument, which are the necessary conditions, are brought about and the assumptions hold, then the conclusion, which is the sufficient conditions, will follow with near certainty. This does not mean that we expect the program activities to always, for everyone, lead to the necessary conditions (a task for the next step in evaluation), but we do expect that if they were achieved, then the outcomes or sufficient conditions would follow.

> Necessary conditions + assumptions = sufficient conditions or intended outcomes

While this article stresses that the focus should be on evaluating what a program is logically sufficient for achieving, at times, it will be useful to evaluate the contribution that the program made to longer term outcomes.

> Sufficient conditions + external factors = longer term or contingent outcomes

Evaluation of a Program Design Logic 'In Reality'

Once a PDL is evaluated to be sound, the next step is to evaluate it 'in reality'; that is, to implement methods to determine the extent to which the premises in the argument are well grounded. Here, the steps will be very similar to any evaluation informed by an outcomes hierarchy. The goal may be to test the weakest links in the argument—either in relation to necessary conditions that are expected to be brought about by actions or assumptions that may be 'heroic'. The goal may be to understand for whom and in what circumstances the program activities fire certain mechanisms that generate change for some but not all people, communities, organisations, etc.

It will also be useful when considering the efficiency of some intervention to test whether the necessary conditions were actually necessary. They may be superfluous or redundant. It may be that the same conditions are being achieved in locations where some 'necessary' conditions do not hold. For example, in the PDL in Figure 1, it may not actually be necessary that both defence and prosecution counsel have sufficient information to negotiate an early resolution.

The inferences in the argument of a PDL are of the form, 'if not this, then not that', or if not P, then not Q. This form of reasoning is referred to as modus tollens. Once all conditions considered necessary are laid out in the proposition, we infer that these will be sufficient for some new condition or outcome. This form of inference is using modus ponens, 'if all these conditions, then this condition', or if all these Ps, then Q. For a discussion of modus tollens and modus ponens, see Dagli, 1998.

The warrant that supports us to say if not P, then not Q, is that one is 'logically antecedent' or that one is a necessary precondition of the other—'if a person is not aware of the program, then the person cannot enrol in the program' (which is not the same thing as 'no one *is* enrolled in the program'. Someone may be enrolled without their knowledge, but the person cannot enrol in the program of their own volition, if they are unaware of the program).

The warrant that supports us to say if 'all these Ps, then Q' is the theory of configurational causality. That is, if we bring all these things together in the right way, we will achieve some outcomes.

In public policy, the warrants that permit the inferences or the reasons we should accept the interferences are drawn from the wide variety of evidence that is available for the development and evaluation of public policy. It may include results from past research or evaluation, current theories about how the world works, or common sense.

Some Potential Criticisms of PDL

There are a number of criticisms that could be levelled at PDL—that this is nothing new and that we cannot expect a logical model to be helpful because real life is too messy to place in a model, and our actions will rarely, if ever, pan out as we hoped. And even if it does pan out as we hoped, different people will respond to our program differently, leading to hopelessly heterogeneous results. Here some of these potential criticisms are addressed.

PDL is nothing new. Outcomes hierarchies have been around for long time. What has not been around for a long time is an approach to evaluate the logic of a design. In many respects, PDL is using an outcomes hierarchy approach with an explicitly configurationalist theory of causality. It is used to surface implicit assumptions and inform discussion about the value of proposed course of action. It seeks to move the discussion away from abstract theories of how the world works, to remove the fig leaf of 'theory' and provide a method for identifying problems with the logic of how a proposed course of action is intended to work. This approach does not require knowledge of any particular theory. Using PDL, logical flaws may be spotted and debated by any reasonable person. PDL aims to provide a tool for this rationale discourse between politicians and the public and between the commissioners and providers of public programs and interventions. PDL aims to be useful for evaluation, as other forms of program logic can be, but its higher aim is to support rationale discourse about the design of public policy—once certain values and objectives have been agreed upon. In this sense, PDL provides a tool for Prospective Evaluation Synthesis (PES), as developed by the GAO. PES focuses on proposals and on the conceptual, operational and empirical questions of 'Logically, should the proposal work?', 'Practically, could the proposal work?' and 'Historically, have activities conceptually and operationally similar to the proposal worked in the past?" PDL fuses these questions into a single question of "Logically, should it work here?' The reasons or warrants that will be marshalled will be variously conceptual, operational and empirical. PDL and the associated data matrix and specification of methods to answer key evaluation questions also guide empirical data collection about the proposed course of action, if and when it is implemented, to answer the question 'Does it work here?'

PDL is over simplistic. The real world is much messier than the model describes. A PDL may appear aloof from the vicissitudes and complexities of life. It may give an appearance that social policy is simple and linear. A PDL is not attempting to describe a complex reality. The purpose of a PDL is not to ignore the messy reality, where complex interdependencies determine the extent to which any intervention actually delivers change. PDL is about having a logical plan. Even if 'no plan survives contact with the enemy', it is necessary for deliberative and democratic public policy that people can debate the value of a course of action. It is attempting to bring to the surface the implicit logic to a proposed course of action. A PDL may be more or less detailed, depending on which parts of the proposition require the greatest justification—the more contested, the greater the need for accompanying text to provide the reasons and warrants as to why a certain set of conditions are expected to be sufficient for generating a desired state of affairs.

A PDL focuses on what a program may be sufficient for achieving. A program may have some fairly lofty ambitions; however, a sound PDL will be clear about what program activities should be sufficient for achieving (given certain assumptions), and what they may only contribute towards (in addition to external factors). Many programs are not sufficiently designed or resourced to achieve higher order outcomes. As per Rossi, (1985) few interventions will be strong enough, or sufficient, for actually addressing the big problems of poverty, crime and drug addiction that plague our societies. If a PDL is not ambitious enough, either more actions are required or there must be greater reliance on assumptions, which may turn out to be 'heroic'. Program designers can increase the scope of action to bring more things under the control of the program and leave less to assumption. They may also debate the extent to which assumptions are reasonable or not, but PDL simply points out that these are the assumptions on which a design is reasoned to be a good or effective course of action.

PDL is not scientific. This is correct—PDL focuses on the same logic that is necessary in science, but it does not explain how and why change happens. People are different and will react differently to the same program; this is the central insight of realist approaches to evaluation. PDL is not setting out to test theories but to leverage them. PDL is an argument structure that says when certain premises are true, then some conclusion is true. It must be valid and well grounded. The validity comes from an assumption that the conclusion (the outcome for which the program is sufficient, i.e., that participants learn parenting skills) will follow if all the premises are true. For example, 'participants engaged with the program'. The well groundedness of the PDL comes from evidence as to whether, and to what extent, each premise was supported in reality. We know that how different people engage with each activity will vary; we also know there is a rich world of theory that will seek to explain which people, in what circumstance, will or will not engage with the activity.[8]

PDL starts as an argument that is valid if the premises are true and the assumptions hold. This is just an initial step. Further evaluation of the extent to which each premise holds for different people in different circumstances (e.g. what lead different people to engage), and knowledge of the extent to which assumptions about the content into which the intervention was placed, actually hold, should lead to refinement of the PDL—maybe new actions or conditions are considered necessary, maybe old ones are no longer considered necessary. This experience should lead to a stronger argument that uses evidence of the extent to which each premise is well-grounded and how it may be further achieved. In the theory-driven evaluation literature, this is similar to how theories of action are designed to activate theories of change—in PDL, theories are special cases of reasons to think it will work. The evidence obtained from an evaluation is often about what worked; it cannot claim to be about what works unless it is based on a sound approach to the development and testing of social science theories. This is, however, outside the scope of PDL and often too ambitious a goal for the effective use of public money in developing cost-effective ways of addressing aspects of social problems that we are, honestly and in good faith, actually setting out to address.

Despite the goal of a scientific evaluation—either on the positivist or realist account—most evaluators, especially later in their careers, experience considerable humility about the ability of elevation to identify stable cause and effect relationships. Pawson and Tilley (1997) have shown why this is the case—because program evaluators ignored the mechanisms that, in context, were responsible for any outcomes. It was never the actions themselves. When combined with Rossi's (1985) Iron Law, this may give pause to anyone with the audacity to design a program. It should cause one to question the utility of seeking to evaluate a program by measuring its outcomes in terms of an average effect size.

[8] Realist evaluation will often move from an explanation of what happened to suggesting that this is evidence of what will happen in future similar contexts. A central problem for realist evaluation, as with all approaches, moving from evidence of what happened in the past to what will hold in the future.

Some complexity science theorists go further and identify situations where any attempt at identifying stable cause and effect relationships is untenable (Kurtz & Snowden, 2003). Approaches to evaluation, informed by the science of systems, also tend to focus less on measuring attributable outcomes and place a greater focus on monitoring system conditions (Renger, 2015).

PDL is too cold and rational. PDL provides a tool to evaluate whether a proposed course of action is logical 'on paper' as well as to guide data collection activities to determine whether it makes sense 'in reality'. The focus is on logic and what we think we know, rather than on participatory methods, appreciating values, directing innovation or creative problem solving. This is true. PDL is not an approach to work out *the* rational approach to solving a problem and is silent on this process. Similarly, PDL does not seek to place logic over ethics or adjudicate whose values matter. PDL starts when a decision is made about a particular approach to solving a problem; it is concerned with whether that approach is sound. There may well be a great need for public policy to look more seriously at the root causes of problems and to engage in more participatory and creative co-design of solutions, but this is a task that precedes PDL.

A word of warning. In informal propositions and argumentation, there is no form of argument that allows us to draw conclusions from the validity of the form alone, as is the case in deductive reasoning. The strength of an informal argument is up to the reasoning ability of the proponents of a course of action and the type of evidence that a particular audience will find credible and be persuaded in their decision. This deliberative from of argumentation is the basis of ancient and modern democracies alike. It is exactly what is required in the modern business case, prospective evaluation, new policy proposal or political platform. And here it is important to issue a warning—given the existence of logical fallacies and the influence of style or rhetoric on an audience, it is imperative to be aware of these 'tricks of the trade' in evaluating any argument. Evaluation should address the strength of reasons given as to why a course of action is reasonably likely to be effective—it should stick to facts and logic, rather than rely on the style of an orator. PDL provides a tool to set out the inherent logic of any proposed course of action, aside from the method of delivery or rhetorical devices commonly used to persuade.

Conclusion

The central goal of this paper is to position logic as a core requirement in all stages of the policy and program design cycle—particularly in the design phase before an intervention is released into the world. Throughout the paper, we have valorised reasonableness and humility over truth and certainty—emphasising logic, rational debate and decision making under time and resource constraints. This paper has been critical of 'program theory' and 'theory of change' as concepts that risk over-extending theory and hiding under-developed logic, leading to programs that with some rational interrogation could be seen to have a high risk of failure before the first participant is enrolled.

The central claim of this paper is that a public policy program is, in essence, not any type of theory, but a proposition about the value of a course of action. But this paper should not be read as a retreat from theory. Far from it. PDL is based on a configurational theory of causality that is built with necessary and sufficient conditions. Treating programs as a form of theory has led to much confusion and waste in the expenditure of evaluation budgets on research projects and, somewhat ironically, a general failure to engage with any actual theory of causality. Too much focus on theory can cast a veil over a program and exclude reasonable people from participating in a discussion of why a program may or may not be effective. It can lead to a sense of inevitably about a particular policy or program, when any social problem can be addressed leveraging many different theories—a left-wing or a right-wing person will have different ideas about any given problem, its root causes, means by which it may be addressed, or even whether a problem exists at all.

It follows from the conception of a program as a plan or proposition that evaluation should be concerned with the validity and well-groundedness of that proposition. The validity of the policy or program proposition is whether the premises, if all true, would lead to the conclusion or intended outcomes. To translate into common policy language, if the outputs were all achieved, whether it is likely that the short-term outcomes would follow. The well-groundedness of the argument is the extent to which the premises, i.e. outputs, are likely to be, or were actually achieved. This requires warrants, reasons and, in many cases, theories of the way the world works and empirical data about what actually happened. Theory, however, is subordinate to logic.

This paper has sought to demonstrate how PDL can provide a tool for those responsible for making decisions about the design and implementation of a policy or program. It is concerned with evidence-based policy and prospective evaluation. PDL may deliver benefits by identifying and subjecting to examination the reasons to think some proposed course of action is likely or is, in fact, generating intended outcomes. PDL like most other forms of program logic is also designed to identify evidence that may demonstrate which parts did work. The intercalation of 'design' between program and logic signifies a focus on ensuring that whatever is designed, by whatever means, provides sound reasons to think it and its component parts, will work with a specific group of people in a specific time and place. PDL is silent about the adequacy of a program in addressing root causes and not intended for adjudicating between different social science theories or ideologies. It is not a tool for research but for evaluation based on a configurationalist theory of causality and the concepts of 'necessary' and 'sufficient' conditions.

References

Alkin, M. C., & Patton, M. Q. (2020). The Birth and Adaptation of Evaluation Theories. *Journal of MultiDisciplinary Evaluation, 16*(35), 1–13.

Althaus, C., Bridgman P. & Davis G. (2017). *The Australian Policy Handbook 6th edition*, Allen & Unwin.

Aristotle. (1992). *The Art of Rhetoric.* Penguin Classics.

Bhaskar, R. A. (2008). *A realist theory of science.* London: Verso.

Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better.* Oxford University Press.

Comte, A. & Lenzer, G. (ed). (1975). *Auguste Comte and Positivism: The Essential Writings.* Transaction Publishers.

Cook, T. D., Scriven, M., Coryn, C. L. S. & Evergreen, S. D. H. (2010). Contemporary Thinking About Causation in Evaluation: A Dialogue with Tom Cook and Michael Scriven. *American Journal of Evaluation, 31*(1), 105–117.

Dagli, M. M, (1998). Modus Ponens, Modus Tollens, and Likeness. *The Paideia Archive: Twentieth World Congress of Philosophy*, Volume 8, pp 45-52.

Datta, L, E. (1990). Prospective evaluation methods: The prospective evaluation synthesis. GAO/PEMD-10.1.10. *United States General Accounting Office*, 119.

Donaldson, S. & Lipsey, M. (2006). Roles for theory in contemporary evaluation practice: developing practical knowledge. In Shaw, I. F., Greene, J. C., & Mark, M. M. *The SAGE Handbook of Evaluation,* 57-75. SAGE.

Fournier, D. (1995). Establishing evaluative conclusions: A distinction between general and working logic. *New Directions for Evaluation*, 68.

Frechtling, J. (2007). *Logic modelling methods in program evaluation.* San Francisco, CA: Jossey-Bass.

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

Hawkins, A. J. (2016). Realist evaluation and randomised controlled trials for testing program theory in complex social systems. *Evaluation, 22*(3), 270–285.

Hernandez, M. (2000). Using logic models and program theory to build outcome accountability. *Education & Treatment of Children, 23*(1), 24-41.

Jones, N. D., Azzam, T., Wanzer, D. L., Skousen, D., Knight, C. & Sabarre, N. (2019). Enhancing the effectiveness of logic models. *American Journal of Evaluation.* 1098214018824417.

Kurtz, C. F. & Snowden, D. J. (2003). The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM Systems Journal*, *42*(3), 462–483.

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford: Clarendon Press.

McLaughlin, J. A. & Jordan, G. B. (2015). Using logic models. *Handbook of Practical Program Evaluation: Fourth Edition,* 62–87.

OECD/DAC Network on Development Evaluation (2019) *Better Criteria for Better Evaluation Revised Evaluation Criteria Definitions and Principles for Use.*

Pawson, R. & Tilley, N. (1997) *Realistic Evaluation*. London, UK: SAGE Publications.

Pawson, R. (2013). *The science of evaluation: A realist manifesto*. SAGE Publications.

Pawson, R. (2008). *Causality for beginners.* NCRM Research Methods Festival 2008.

Pearl, J. (2017) *The Book of Why: The New Science of Cause and Effect*. Illustrated edition. New York: Ingram Publisher Services Us.

Ragin, C. (2009). Qualitative comparative analysis using fuzzy sets (fsqca). In Rihoux, B. & Ragin, C. C. *Applied Social Research Methods: Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques, 51*, pp. 87-122. Thousand Oaks, CA: SAGE Publications, Inc.

Renger and Titcomb. (2002). A three step approach to teaching logic models. *The American Journal of Evaluation*, *23*, 493-503.

Renger, R. (2015). System evaluation theory (SET): A practical framework for evaluators to meet the challenges of system evaluation. *Evaluation Journal of Australasia*, *15*(4), 16–28.

Rogers, P. (2000) *Causal models in program theory evaluation*. In Rogers, P., Hacsi, T. Petrosino, A. & Huebner, T. (eds), Program theory in evaluation: Challenges and opportunities, *New Directions in Program Evaluation*, 87, 47-55. San Francisco, CA: Jossey-Bass Publishers.

Rossi, P. (1985) *The iron law of evaluation and other metallic rules*: Paper presented at State University of New York, Albany, Rockefeller College.

Schwandt, T. (2015). *Evaluation foundations revisited: Cultivating a life of the mind for practice*. Stanford University Press.

Scriven, M. (2008). The Concept of a transdiscipline: And of evaluation as a transdiscipline. *Journal of MultiDisciplinary Evaluation, 5*(10), 65–66.

Shadish, Cook & Campbell. (2002). *Experimental and quasi-experimental designs for generalised causal inference.*

Shadish, Cook and Leviton. (1991). *Foundations of program evaluation, theories of practice*. SAGE publications.

Taylor-Powell, E., & Henert, E. (2008). *Developing a logic model: Teaching and training guide.* Madison, WI: University of Wisconsin Extension, Cooperative Extension, Program Development and Evaluation. http://www.uwex.edu/ces/pdande

Toulmin, S. (2003a). *The Uses of Argument (2nd edition).* Cambridge: Cambridge University Press.

Toulmin, S. (2003b). *Return to reason*. Harvard University Press.

W.K. Kellogg Foundation. (2004). *Using Logic models to bring together planning, evaluation, and action: Logic model development guide*. Battle Creek, MI.

Weiss, Carol (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. *New Approaches to Evaluating Community Initiatives*. Aspen Institute.

Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J. & Greenhalgh, T. (2016). RAMESES II reporting standards for realist evaluations. *BMC medicine*, *14*(1), 96.