How We Model Matters: A Manifesto for the Next Generation of Program Theorizing

Sebastian Lemire
University of California, Los Angeles

Jane Whynot
University of Ottawa

Steve Montague
Performance Management Network

Abstract: *In this concluding article, grounded on the exemplary contributions contained in the preceding pages, the guest editors scale the proverbial soapbox and present a manifesto to guide the pursuit and advancement of the next generation of program theorizing. Formulating ten declarations for program theory development and examination, the modest hope of the authors is to motivate and inspire reflective evaluation practitioners to broaden their views, approaches, and techniques for future program theorizing.*

Keywords: contribution analysis, *manifesto, program theorizing, program theory, realist evaluation, reflective practice, theory-based evaluation*

Résume : *Pour conclure, en s'inspirant des contributions exceptionnelles des articles précédents, les éditeurs de ce numéro spécial se lancent et soumettent un manifeste pour orienter la poursuite et l'avancement de la prochaine génération de théories d'intervention. En formulant 10 propositions sur l'élaboration et l'analyse des théories d'intervention, les auteurs entretiennent l'espoir de motiver et d'inspirer les évaluatrices et évaluateurs pour qu'ils élargissent leurs horizons, leurs approches et leurs techniques pour la théorisation future des interventions.*

Mots clés : *analyse de contribution, manifeste, théorie d'intervention, élaboration de théories d'intervention, évaluation réaliste, évaluation fondée sur la théorie*

REFLECTIVE PRACTICE AND PROGRAM THEORIZING IN EVALUATION

As stated in the Introduction, the overarching aim of this special issue is to promote reflective practice in program theorizing: to expand and strengthen both the conceptual and technical foundations of program theories in evaluation. This

Corresponding Author: Sebastian Lemire, Social Research Methodology Division, University of California, Los Angeles, Moore Hall 2005, Los Angeles, CA 90095-1521, USA; slemire@ucla.edu

emphasis on reflective practice emerges in part from a broader push in evaluation circles toward evaluative thinking (Vo & Archibald, 2018). More specifically, the notion of reflective practice is rooted in the recent work by Thomas Schwandt (2015), who describes the reflective practitioner as someone who develops the knowledge and skills to design and implement methodologies in a variety of un-familiar situations, has the capacity to comprehend local contexts, knows how to adapt methodological principles and practices accordingly, and has the ability to determine courses of action and move forward in a way that balances methodo-logical, practical, and ethical considerations.

As scholar-practitioners, we have over the past years collectively developed hundreds of program theories across a broad range of settings, geographical loca-tions, and contexts, deploying a wide range of methods and approaches for devel-oping and testing our program theories, experiencing both successes and failures. And as is true of most practitioners, we have learned through trial and error about what has worked (and what has not), allowing us over time to advance and refine our understanding and appreciation of real-world program theorizing. Our first-hand experience of how program theories unfold (and sometimes fold) in real-world settings has in fundamental ways shaped our scholarly and practical work.

As aspiring reflective practitioners, we have also come to realize that the methodological quality (however we choose to define it) of our theory-based evaluations was never the product or property of a specific design, data-collection method, analytic approach, or even specific features of our program theories, although we falsely assumed this to be the case on more than one occasion. In-stead, we have come to realize that the *rigor* more commonly resided or emerged from the analytical reasoning embedded in our program theorizing—the active interplay and integration of analytical competencies, technical know-how of methods and techniques, contextual and ethical awareness, and, perhaps of most importance, our integration of these in developing and verifying the program theories. As Van Melle, Gruppen, Holmboe, Flynn, Oandasan, and Frank (2017, p. 752) declare, "it is all about rigor in thinking."

With these observations as our backdrop, the position we hold is that the notion of reflective practice provides a worthwhile framework for the proposed pursuit of next-generation program theorizing. And with the notion of reflective practice as our guide, we now turn to our program theory manifesto.

A SOAPBOX MANIFESTO FOR THE NEXT GENERATION OF PROGRAM THEORIZING

From our viewpoint, striking changes are called for in program theorizing for it to rise to its potential. For far too long, program theorizing in evaluation has been characterized and depressed by linear, overly simplistic program depic-tions (Coryn, Noakes, Westine, & Schröter, 2011; Weiss, 1997). The all-too-common mechanistic and perhaps ritualistic development of less than inspired program theories, followed by limited, if any, attention in subsequent program

implementation, has resulted in piles of half-implemented or even unused program theories. Admittedly, we too have added pebbles to this pile. As such, and as theory-driven soapbox orators, we seek to preach our practice (and practice our preach) toward widening the range of strategies for improving the development and potential use of program theories.

Toward these ends, the following 10 declarations collectively comprise what we take to be a "Program Theory Manifesto":

1. Promote inclusion and representation
2. Strengthen the linkage between evaluation and decision making
3. Develop situated, contextualized, systems-oriented models
4. Focus on program archetypes
5. Explore different types of causal explanations
6. Flip, stack, and layer program theories
7. Put "theory" back into program theory
8. Strengthen the testability of program theories
9. Pursue theory-based synthesis and accumulation of knowledge
10. Do more with less, more quickly

These declarations represent what we take to be useful avenues to pursue in future program theorizing. They are grounded in, and inspired by, the important work of other scholars and practitioners, informed by our experiences as practicing evaluators and shaped by our ongoing exchanges on the topic of theory-based evaluation. The declarations range in scope and focus; some promote specific methods or techniques; some advocate for more fundamental perspectives on what program theories are, or even what program theories could or even should be; yet others simply call attention to ideas or practices in the work of other theory-based practitioners and scholars that we find inspiring. Underlying this diversity is a shared commitment to Schwandt's (2015, p.33) idea of evaluation practice as "actions informed by situated judgements." As Schwandt suggests,

> [p]ractice decisions can be enlightened by conceptual or theoretical knowledge that serves as an aid in thinking through options in a situation that a practitioner faces. It is helpful to think of this kind of knowledge as a repertoire of principles, concepts, insights, and explanations that professional practitioners can use as heuristic tools "to think with."

We wholeheartedly agree. And in line with the above, it deserves to be said that the declarations are just that: aids in thinking about future program theorizing. Accordingly, these declarations are not to be viewed as off-the-shelf solutions or recipes to the many pesky pitfalls of program theorizing, nor do the declarations collectively represent an exhaustive list of worthwhile practices to pursue. Rather, the declarations—individually and collectively—represent what we take to be improvements and advancements that are both grounded, yet innovative in

the context of current practice, as well as feasible and desirable avenues for future program theorizing.

DECLARATION 1: PROMOTE INCLUSION AND REPRESENTATION

We envision program theory as something that is only as valuable as when it is shared. This requires both the broad and targeted engagement of a diverse set of stakeholders at various junctures and levels throughout the development, implementation, reporting, and use of program theory. Indeed, calls have repeatedly been made for participatory approaches to theory-based evaluation (see Hansen & Vedung, 2010, on theory-based stakeholder evaluation; and Koleros & Mayne, this issue, on actor-based theories of change). This engagement requires the adoption of a pluralistic framework that is sufficiently flexible to include conceptual, technical, and practical contributions from various stakeholders, including program beneficiaries as experts on their own experience, sources of expert input including academics from other sectors/disciplines, and the decision makers responsible for allocations of funding. Such engagement cuts to the core of reflective practice. While we have learned a significant amount from those who have gone before us, the capacity for learning is unlimited; there is a potential contribution to be made by each stakeholder. Moving from stage to stage within a detailed theory of change from the level of inputs through to the ultimate outcomes without engaging others in the development, implementation, reporting, and use dimensions seems foolish.

Opening both evaluators and the evaluation function to learning allows for the engagement of various diverse groups of stakeholders from a multitude of disciplines to collaboratively contribute to the articulation of how interventions are anticipated to work, and perhaps more importantly offer insight into why interventions do not work. This is not to say that all stakeholders need to be involved at all stages or throughout all levels. Homework is required to determine which stakeholders can contribute most effectively at the various program-theory junctures. At various points in the process we should be asking the following questions: What would happen if a broader set of stakeholders were involved in the various stages and levels of program theory? Would we uncover key assumptions faster? Would we challenge our own assumptions more effectively? Would we waste fewer already scarce resources on implementation if the theory of implementation was flawed and program beneficiaries could tell us this from day one? The recent coverage of Zimbardo's debunked Stanford Prison Experiment (Reicher, Haslam, & Van Bavel, 2018), which was widely acclaimed in psychology circles, demonstrates that the study's impacts could have been much different if only study researchers' and participants' views had been heard, recognized, and valued differently, much earlier. And the Stanford Prison Experiment is not unique in this regard.

While it is useful to bring diverse viewpoints to bear on our theories, we should remind ourselves that the participation of diverse groups of stakeholders

does not automatically lead to the achievement of equitable outcomes. Equity concerns and efforts need to be reflected in program theory in much more systematic ways to ensure transparency regarding inclusion/exclusion decisions, particularly in identifying relevant assumptions underlying and potential adverse consequences emerging from our evaluations. This involves providing space to ensure that multiple voices are represented. Rogers (2016, p. 202) draws attention to the theory/equity deficit and outlines that "while theories of change are increasingly used in evaluation, they rarely address equity issues adequately." With respect to Bledsoe's (2005) contribution on using program theory in underserved communities, and more recent efforts (Bledsoe & Donaldson, 2015; Donaldson & Picciotto, 2016) that actively promote the role of program theory in equity discussions, encouragement is offered on the need for evaluators to step forward to explore what culturally responsive evaluation means in all facets and types of evaluation, ranging from program to policy. Evaluators are further encouraged to think about how all evaluation approaches are imbued with culture (Bledsoe & Donaldson, p. 23). We can be more aware.

While specific attention is paid to the engagement of stakeholders in developing impact theory, the implementation dimension of program theory is just as important as the alignment with systemic barriers to access and participation, as noted by Greene (2016). The relationship among program theory as an approach, participation, and equitable outcomes and valuation needs to be explored in greater detail. The intentional inclusion of elements affecting equitable outcomes, including gender (Whynot et al., this issue), needs to be explicitly incorporated into program theory. Incorporating reach into program theory is an early step in the right direction (Mayne, 2015; Montague & Porteous 2013). Conceptually, equity may be embedded in program theories as key components and expressed through the articulation of assumptions, but its representation is muted when it requires inferences to be made by readers. More can be done to situate equity issues explicitly in program theory.

Promoting advancement in related areas requires building the capacity to learn more effectively from many in a program-theory-specific context; this also requires a mindset that supports hearing from others who have not perhaps been part of these discussions to date. Certain epistemological orientations and subsequent assignations of evaluator roles align themselves with greater facility. Regardless of paradigm, we advocate for a recognition of the dynamics related to power and influence in theorizing. These variables sway the value assigned to who is speaking and who is being heard. Program theory has jointly been held historically in two different sets of hands: those of the intervention and those of the evaluators. Enacting change to incorporate the collaboration of a broader set of stakeholders is not an easy undertaking, particularly working in silo bureaucracies that are heavily dependent upon processes and protocols to engage with other functions. The message regarding the value of engaging a diverse set of stakeholders needs to be clearly communicated by organizational leadership, and for that to happen there need to be related accountability mechanisms in place,

but furthermore, for sustainability, the value associated with broader engagement needs to be demonstrated.

DECLARATION 2: STRENGTHEN THE LINKAGE BETWEEN EVALUATION AND DECISION MAKING

We aim to strengthen the linkage between evaluation and decision making through program theorizing. The evaluation function in policy environments has historically failed to maximize its potential in contributing to decision making (Leeuw, 1991). This is not a new challenge, but we are positing that program theory could be a potential solution. This challenge emerges in Weiss's (1998) critical work, where she notes that evaluation is one of many inputs in the governance of public funds, summarizing that evaluation and research are used in a manner that could be characterized as indirect at best. Dobell and Zussman's (1981) articulation of specific challenges encountered by the evaluation function include its misunderstanding by senior managers and program staff, tacit resistance by program managers, and the failure to consider the true nature and information needs of the user. These challenges, in addition to existing evaluation policy requirements, the program level focus of evaluations, and the public nature of evaluation reports (Bourgeois & Whynot, 2018), were additionally identified as findings that impeded strategic evaluation utilization by decision makers. The many disconnects between evaluation and decision making are relevant for discussion, as we suggest that program theory as envisioned in this issue could potentially provide one means by which to bridge the divide to evaluative thinking.

How we identify our evaluand and how it interacts within its existing systems is a lost opportunity. Internationally, work on the Sustainable Development Goals (SDGs) reminds us of this. For example, gender equality not only is a standalone goal but is also woven amongst multiple other SDGs. These goals serve to remind us that policy makers are not dealing with simple problems but rather complex ones, and with what has been increasingly referred to as the "wicked policy problems" (Head, 2008; McGrail, 2014). With cross-cutting issues facing decision makers worldwide in areas of concern such as climate change, these policy problems will not simplify themselves, nor will they disappear of their own volition. Complexity and wickedness necessarily involve a broader view of the landscape to address non-linearity and its potential paths, the inputs and influences of multiple actors, their associated technical and political considerations, and most importantly the "small shifts that may produce large differences in the outcomes of the systemic dynamics" (Peters, 2017, p. 386). With more sophisticated taxonomies evolving related to complexity and wickedness (Alford & Head, 2017), we are suggesting that more sophisticated understandings of program theory can contribute to fostering evaluative thinking amongst decision makers. Ultimately, it is people who will reap the benefits or bear the burden of policy and decision makers getting theory "right."

DECLARATION 3: DEVELOP SITUATED, CONTEXTUALIZED, SYSTEMS-ORIENTED MODELS

Experience suggests that logic models that set out as stark box-and-wire diagrams with limited attention to narratives, explanations, and contexts can do more harm than good (Freer & Lemire, this issue). Without elaboration, these depictions become no more than desires or slogans in boxes subject to political "force fitting."

Schwandt (2018, p. 131) importantly reminds us that as evaluators we can never "investigate something like a policy, program, practice or strategy in its totality—that is, in terms of all of its interconnections and relationships, and from all different perspectives and viewpoints." Rather, what we can do, and what we should always strive for, is to be transparent about how we select which aspects of our evaluand are highlighted and who is engaged in making these decisions. Building on this proposition, we would emphasize three elements as key to model improvement:

1. theory situating;
2. contextualizing; and
3. systems orientation.

Depictions of theory at the program, policy, or broader initiative level must be situated in terms of the level and focus for analysis and juxtaposed with other levels of theory. Theories can, and should, be contemplated at micro-, meso-, and meta-levels. Much of the discussion in this volume has been at a meso-level, although actor-based theories as well as theories that situate and position girls' and women's experiences in STEM may touch on micro-theories to some extent. Lemire and Christie in this issue directly illustrate an approach to meta-modeling. The point is that different levels of theory can work to support or detract from each other. For example, evaluators need to be cognizant of whether, and to what extent, a micro-theory about how to reduce individual drug-addiction harm is congruent with broader program theories of deterrence and remediation for drug addiction and, in turn, how these are consistent with meta-theories (e.g., can harm-reduction strategies and theories work within a broader strategy that amounts to a "war" on drugs?).

Following the situating of theories, contextualizing is crucial. The notion of singular best practice should be eliminated. All practices need to be judged in context. The realist mantra of "what works (to what extent) for whom under what conditions and why?" can be a powerful means to recognize the inherent complexity in most human-based systems (Pawson, 2006). It also ensures the consideration of context as part of the performance story of any initiative. It should also encourage a fundamental focus on learning and improvement in addition to accountability and a normative assessment of value. Whether identified as "factors," "assumptions," "pre-conditions," or "risks," good program theory should include contextual considerations.

Ultimately, a systems orientation needs to be fostered while avoiding hyper-complexity in description. Several articles in this special issue have, in addition to advocating the adoption of a systems view, noted the value in recognizing key system elements such as the change theory and the action or implementation theory and strategy (e.g., Montague, this issue). The match or mismatch of these elements can explain a lot of observed performance. Another key element in systems is that they have actors. Koleros and Mayne (this issue) note the benefits of actor-based theories in the conduct of contribution analysis in complex settings. One could argue that such actor-based modeling can also help explain another key element of systems—the relationship and feedback loops that create virtuous and vicious circles, performance distortions, as well as emergent outcomes and "bends" in implementation (Pawson, 2006).

Finally, it deserves mention that another useful strategy may involve leaving room for ambiguity in certain aspects of the program theory (Dahler-Larsen, 2018). As Dahler-Larsen succinctly argues, there are concrete steps that evaluators can take to include and use specific types of ambiguity in their development and examination of program theories, including the use of what Dahler-Larsen refers to as "Janus Variables"—variables that work in two ways (p. 6).

The implications or impact of not getting program theory right may not be felt in the short term at the individual program level, but it will undoubtedly raise its implication head in years to come at broader policy, country, or global levels. The achievement of results is a shared responsibility that necessitates moving beyond the stakeholders that have been historically consulted in the course of evaluation activities.

DECLARATION 4: FOCUS ON PROGRAM ARCHETYPES

Evaluation has for the most part been fixated on programs. How to think about and best portray programs is no trivial task. The problem is that program evaluation as commonly practiced does not encourage a systematic approach to accumulating knowledge on programs. Evaluations are all too often narrow and "local" in scope. They often miss obvious similarities in their impact pathways with other programs simply because the subject matter and clientele may be different. Moreover, program theories and logic models have often been developed from "scratch." Realist evaluators like Pawson have noted that

- all programs are associated with some kind of theory or theories (they are theories incarnate); and
- there are as many unique programs as there are programs; however, there are only a limited number of program theories; therefore,
- one can accumulate a good deal of diversified knowledge about the circumstances and success of program theories which can be of value to programs.

Program theories in turn can be classified by the nature of the program as policy instruments. As an example, take the Bemelmans-Videc, Rist, and Vedung (1998) definitions of carrots, sticks, and sermons representing incentive, deterrent, and information programs. Each of these instruments has distinct results and delivery logic, characteristics, and recognized circumstances associated with success and failure. Funnell & Rogers (2011) propose that archetypal models can be built for these policy instruments, whereby refined categories can be established. For example, in the deterrence (stick) archetype, there might be subcategories such as summary conviction notices of variance-type actions, with little or no penalty. The relative success of such actions has been found to depend heavily on other contextual factors, such as the potential for shame ("reputational loss" is often a term used) with public exposure. It has also been associated with the understanding that more significant penalties can and will be used—thereby forming a deterrence period (Ayres & Braithwaite, 1992). Other types of deterrents may more easily stand on their own. Over time, one can begin to glean from study, work, and experience the factors that recur in association with certain types of outcomes for a given mechanism. One can also consider the implementation and design characteristics that best fit a given policy instrument and its success characteristics.

The notion of archetypal program theories applies to both policy instruments and to action/implementation, including governance and delivery dimensions. Is the governance and/or delivery through a multi-government and multi-agency agreement? Unilateral delivery? Delivery with privately contracted elements? Delivery with voluntary or semi-voluntary third-party elements? Is the delivery mode hierarchical, matrix, or network-based? These are but a few key questions and factors that are important for understanding the complex effects that action theories and change theories can have on each other and as such demand our attention. While the distinction of archetypes and correlational success factors may experience a few false starts, this orientation may just produce some profound insights.

DECLARATION 5: EXPLORE DIFFERENT TYPES OF CAUSAL EXPLANATIONS

A central aim of theory-based evaluation is to explain how (or why) programs work (or fail to work). Despite this central role of causal explanation, there is limited attention to the different kinds of causal explanation potentially pursued in the context of theory-based evaluation. For instance, an important distinction could and should be made between *how* or *why* a program brings about a desired outcome. Explaining *how* a program works entails determining and describing the "active ingredients" that either individually or collectively elicit the desired outcome. These active ingredients are typically in the form of program activities or outputs that in different configurations elicit a specific outcome. By identifying and describing these critical ingredients, or (for lack of a better term) causal

This ahead of print version may differ slightly from the final published version.

recipes, theory-based evaluation offers important insights into how the program works. This is a configurational explanation for the outcomes.

In marked distinction, explaining *why* a program works entails determining and describing the program mechanism(s), that is, the underlying processes generating a specific outcome. This entails making explicit the individual or social psychological processes triggered by the program. By specifying these underlying mechanisms, theory-based evaluation offers important insights into why the program works, why the program makes a difference for the participants. This is a mechanism-based explanation.

The position we hold is that these are fundamentally different types of explanations, requiring different types of data and analysis, resulting in different types of information, and potentially supporting different kinds of decision making. Accordingly, the type of explanation pursued should emerge from the type of informational need motivating the evaluation. As just one illustration, if the aim of the theory-based evaluation is to support future program design, a configurational explanation focusing on how individual program components (or configurations of these) lead to a desired outcome might be advantageous. This type of information will support the identification of specific program components (critical ingredients) that all things considered should be included in future programs. Conversely, if the aim of the theory-based evaluation is to understand the underlying reason(s) why program participants change their behavior (or fail to change) in response to the program, a mechanism-based explanation is perhaps more appropriate. Of course, we may also pursue understanding both how and why the program works (see Lemire & Christie, this issue).

There are of course several other types of explanations to be pursued in theory-based evaluation. To be sure, many situations might call for a combination of explanations, especially in layered program theories (see Koleros & Mayne, this issue). The point to be made here is that evaluators should be reflective of the types of causal explanations pursued and the type of information provided, and the extent to which these match the information needs motivating the evaluation. What makes something an explanation in theory-based evaluation? What makes one explanation better than another? What is the explanatory logic of these different types of explanation? And wherein lies the explanatory strength? These are but some of the fundamental questions that must be pursued if we are to better understand and further advance the explanatory potential of theory-based evaluation.

DECLARATION 6: FLIP, STACK, AND LAYER PROGRAM THEORIES

How we model matters. Simply consider the marked difference between a program modeled according to a simple logic model, an embedded program theory, or a Context-Mechanism-Outcome (CMO) configuration. Each of these types of models depicts a markedly different representation and, in effect, understanding of the program being evaluated. Underlying all the creative labels for program

Table 1: Common types of contributory relationships in program theories

Model	Type	Explanation	Description
Simple causal model	successionist	A causal model depicting a contributory relationship between a program activity and one or more program outcomes	A leads to C
Simple causal package model	configurational	A causal model depicting a contributory relationship between two or more program activities and a program outcome	A plus B leads to C
Simple causal mechanism model	generative	A causal model depicting a contributory relationship between a program activity and one or more outcomes, specifying the underlying mechanisms of the relationship	A leads to C because of D
Complex causal model	configurational/ generative	A causal model depicting a contributory relationship two or more program activities and one or more outcomes, specifying the underlying mechanisms of the relationship	A plus B leads to C because of D
Embedded causal package model	contextual configurational	A causal model depicting a contributory relationship between two or more program activities and a program outcome, specifying the contextual conditions under which the relationship holds	A plus B leads to C, under condition E
Embedded causal mechanism model	contextual generative	A causal model depicting a contributory relationship between a program activity and one or more outcomes, specifying the underlying mechanisms of the relationship as well as the contextual conditions under which the relationship holds	A leads to C because of D, under condition E
Embedded complex causal model	contextual configurational/ generative	A causal model depicting a contributory relationship between two or more program activities and one or more outcomes, specifying the underlying mechanisms of the relationship as well as the contextual conditions under which the relationship holds	A plus B leads to C because of D, under condition E.

theories, the position we hold is that practical applications of program theories typically involve seven fundamental (arche-)types of models, as described in Table 1.

The point to be made here is not that one type of model is inherently better than another—nor would we encourage any evaluation practitioner to try to memorize them. What the table illustrates is that there is a broad array of types of relationships to be modeled as part of program theorizing—many different

opportunities for modeling and understanding how and why programs work (or fail to work). Moreover, most program theories might advantageously use a variation of the above model types, using one type of model for some parts of the program theory and other types of models for other parts of the program theory, depending on the degree of specification called for (this type of differentiation is illustrated by Koleros & Mayne, this issue).

In addition to these types of models, there is also a wide range of modelling techniques available to evaluators, including simple box-and-arrow diagrams, causal loop diagrams (Williams & Hummelbrunner, 2011), an interconnected system of stock-and-flow diagrams (Williams & Hummelbrunner), or some fanciful combination of these. Each of these modelling techniques would depict the underlying program logic in markedly different ways, promoting different program understandings and, in effect, result in dramatically different insights and conclusions about how and in what way the program works (or fails to work). The different models offer different lenses through which we see the program.

Accordingly, we need reflected conceptual and practical pluralism in how we model programs, whereby the use of different modelling techniques emerges from active reflection and decision making in regard to the specific purpose and role of our program theories. To illustrate, we may, as Koleros and Mayne propose in this issue, use multiple layers in our models, using different types of modelling for each layer. One theory is then nested (Mayne, 2015) in another theory of change within the same model. We may also use hybrid models by combining modelling techniques within the same theory of change (without layers). Yet other innovative modelling techniques include "stacked" and even "3D" logic models (Grim, Castillo, & O'Quinn, 2018).

The complexity in and surrounding social programs necessarily demands a broader range of modelling techniques. By broadening the range of models we employ in our program theorizing, we actively more away from the all-too-common linear, one-way, overly simplified box-and-wire diagrams. We also move closer to the real-world complexity of social programs that are often characterized by feedback loops (balancing and reinforcing loops), reverse causality, causal interactions, asymmetrical causal relationships, contextual contingencies, and so on. The simple aim here is to reinvigorate our toolbox and practice by considering a broader range of modelling techniques when developing program theories, recognizing the relative strengths and weaknesses of each technique. We call for reflective practice in our modeling.

DECLARATION 7: PUT "THEORY" BACK INTO PROGRAM THEORY

We need to put the "theory" back into program theory. The potential role and value of social science theory is a persistent topic in theory-based evaluation discussions. Yet, despite sustained interest, both the extent to which and how evaluators use social science theory in their program theorizing is limited. The

selection of *which* social science theories also matters, as seen in the case of equity considerations (see Whynot et al., this issue). The position we hold is that social science theory, especially behavioral change theory, potentially serves well to inform and enhance our program theorizing. Inspired by Riemer and Bickman (2011), we strongly recommend a stronger push for "theory knitting," that is, using "one or more social science theories as the conceptual grounding for both the design of the program and the program theory" (Lemire, Christie, & Nielsen, 2019).

The idea of theory knitting is nothing new in the social sciences and can be traced back to work on theory development in psychology by Kalmar and Sternberg (1988). In the context of evaluation, Leeuw and Donaldson (2015) have further promoted this linking of domain theories, theories on policy-making processes, and behavioral theories, offering a compelling example of theory knitting in relation to producing public goods. In a forthcoming publication, Lemire et al. (2019) review and showcase published examples of different types of theory knitting in theory-based evaluations.

Indeed, the case for theory knitting is easily made. As noted by Mark, Donaldson, and Campbell (2011), the active and explicit use of social science theory promotes the specificity of key program-theory components, including purported mechanisms whereby the testability of the program theories is enhanced. On a similar note, Vaessen and Leeuw (2009) compellingly argue that using specific behavioral theories to substantiate and explain causal mechanisms can serve to enhance the internal validity of causal claims. Equally important, and in relation to external validity, the use of social science theory also supports the generalizability of these conclusions. As argued by Donaldson and Lipsey, the integration of social science theory and program theory "constitutes a (if not the) major way that evaluation contributes to social betterment by way of knowledge development" (cited in Donaldson & Crano, 2011, p. 145). We agree.

DECLARATION 8: STRENGTHEN THE TESTABILITY OF PROGRAM THEORIES

A central aim of theory-based evaluation is to establish what Mayne (2001) terms "plausible association" between a specific set of program components and outcomes. As Mayne goes on to argue, this involves evidentiary confirmation of key elements of the program theory, examination of influencing factors, and disproof of alternative explanations (2001, p. 7). These are necessary steps in constructing credible contribution stories, regardless of the type of theory-based evaluation that is pursued.

There are many potential ways to address this concern. One central stepping stone is to enhance program theory specification by developing more fine-grained descriptions of the most salient aspects of the program theory to be empirically examined (Vaessen, 2016). As indicated in Table 1, different levels of model specification (granularity) can be pursued, ranging from simple models (depicting A leads to B) to more complex models (depicting A plus B leads to C because of D,

under condition E). The methodological point here is that specifying these aspects of a program theory with more precision enhances the evaluator's ability to make predictions about confirmatory/disconfirmatory patterns in the data, which in turn allows for more robust empirical verifications of the program theory. To aid in this empirical verification, Mayne (2017) has developed and promoted theory-robustness criteria that emphasize, first, overall understanding, structural logic and clarity, and alignment between activities and anticipated results, followed by specific result and assumption criteria.

In logical extension of enhancing the specificity of program theories, another way to enhance the testability of program theories is by using structured analytical approaches and techniques. Brousselle and Champagne (2011) have proposed logic analysis as an approach for using existing social science theory to examine the plausibility and validity of a program theory. Logic analysis critically examines "the program's strengths and weaknesses, elucidates the links between the program's design and the production of desired outcomes, and identifies contextual influences" (Brousselle & Buregeya, 2018, p. 156).

Another strategy is to consider the strength of evidence in favor of one or more key elements of the program theory. For instance, by gauging the breadth of evidence in favor of the most salient components of the program theory, including the span of evidence across different data sources, methods, and implementation sites. There is a rich and broad literature on how to conduct quality assessment of qualitative, quantitative, or mixed methods evidence, including operational frameworks to structure and guide such appraisals. Lemire, Nielsen, and Dybdal (2012) have proposed the Relevant Explanation Finder, an operational framework promoting a more systematic and transparent assessment of rival explanations and influencing factors. The framework has been applied and further developed in a wide range of settings (e.g., Biggs, Farrell, Lawrence, & Johnson, 2014). The use of the REF or other similar frameworks may serve to establish a stronger chain of evidence in support of plausible association and lend transparency and structure to the identification and examination of rival explanations and influencing factors, ultimately giving credibility to the causal conclusions drawn.

Yet a third strategy is to pursue analytical approaches such as process tracing, whereby more fine-grained causal mechanisms can be examined and verified (Schmitt & Beach, 2015). A key strength of process tracing is that it relies on clearly specified tests, whereby the presence or absence of hypothesized mechanisms can be determined. In this way, the method makes explicit the weight of evidence in favor or against the specific mechanisms, which in turn enhances the credibility of the conclusions drawn. In a similar vein, Qualitative Comparative Analysis, a case-based analytical approach, can also be partnered with program theorizing in identifying the causal conditions (e.g., specific program components) that either individually or collectively promote a specific outcome (Befani, Ledermann, & Sager, 2007; Lemire & Christie, this issue).

There are of course many more structured analytical strategies and techniques for both developing and empirically verifying program theories. By using

these structured analytical strategies, we ensure a more systematic and transparent approach to program theorizing, whereby the conclusiveness and credibility of our causal conclusions are enhanced. We strongly encourage more applied work on this topic.

DECLARATION 9: PURSUE THEORY-BASED SYNTHESIS AND ACCUMULATION OF KNOWLEDGE

Taking stock of what we know is important for learning. The findings included in this special issue suggest that there is an immense opportunity for cumulative learning at this time. The calls for archetypal program theories and theory knit-ting are at least in part motivated by the opportunity to cumulatively learn from theory-based evaluations in ways not as of yet systematically or widely pursued in evaluation circles. As compared to the Cochrane Collaboration, for example, which might focus on addressing specific research questions using data from controlled trials and rigorously pre-screened statistical studies, the accumulation of evidence under this kind of effort would accept a wider range of studies but in turn would encourage transparency, discussion, and dialogue about both data and interpretations of theory.

Over the past few years, there has also been a growing interest in mixed methods synthesis approaches, and, in logical extension, theory-based synthesis approaches. The most common approach, at least in evaluation circles, is that of Pawson's realist synthesis approach (Pawson, 2006). In the present special is-sue, Lemire and Christie illustrate and promote meta-modeling, an alternative approach to developing program theories across existing evaluations (see also Lemire, 2017).

Pursuing theory-based syntheses come with a number of benefits. For one, systematically combining the findings from a comprehensive pool of diverse studies provides robust information on the extent to which, and how, programs work across different settings, populations, and times. Moreover, synthesizing a broader range of evidence—qualitative and quantitative—allows for answers to a broader range of questions, including "how" and "why" programs work (or fail to work). These types of questions are difficult to answer in traditional systematic reviews relying on meta-analysis of quantitative findings. For another, theory-based synthesis approaches also hold potential to move beyond "a thin description of the evidence to produce higher order syntheses resulting in the production of new knowledge and/or theory" (Pope, Mays, & Popay, 2007, p. 6). The pursuit of "thicker" understanding of how and why programs work across different settings, contexts, and times is in our view a worthwhile effort.

DECLARATION 10: DO MORE WITH LESS, MORE QUICKLY

There is an iterativeness that has been associated with program theory's de-velopment, implementation, reporting, and use that has significant resource

implications. These resources include capacity, time, and finances that are typically in short supply in most organizations, including various orders of government. These resource implications are highlighted in an era in which the evaluation function is pressured "to do more with less, and—do it more quickly, please" (we are Canadian after all). In the Canadian federal government, resources allocated to the evaluation function, including contracting dollars, and full-time equivalents (FTEs) have been steadily declining over the last decade. Coupled with the former 2009 *Policy on Evaluation*'s requirement to address all direct program spending (DPS) with no new funding, and the recently implemented 2016 Policy on Results overhaul of the government's results structure, this tension is very real and unlikely to change in the foreseeable future.

Practically speaking, we suggest that there are several factors that will contribute to making program theory more efficient over the duration, including the following: (a) building program-theory capacity and competencies, including the resulting ability to scope appropriately, (b) the adoption of a longer-term perspective on behalf of decision makers, which features the accumulation of learning about what works, and (c) the demonstration of its value. Akin to the development of any other competency, we suggest that the more program theory is used, the better we will get at it. This applies at first at the individual level, and then within institutional settings based on principles forwarded in double- and triple-loop learning. Building competency, cumulatively learning from ourselves and others, and scoping where and how program theory may be employed most effectively should all result in a reduced resource burden by better targeting the expensive parts of evaluation—the data collection and analysis. By having strong theory-based knowledge about the key actors and critical influence points in our interventions, we can more efficiently target the people and resources required to understand how things are working.

Additionally, the value of getting program theory and, accordingly, investing in its development, will ultimately serve as a more efficient use of government resources. It is our experience that organizations sometimes spend millions of dollars doing the wrong thing *by design*. Investing resources up front to develop an understanding of key pre-conditions and factors that would allow a given design in a given context to work seems far superior to spending on patchwork fixes and corrective actions downstream.

CONCLUDING THOUGHTS

A central aim of this issue is to promote reflective practice in program theorizing: to expand and strengthen both the conceptual and technical foundations of program theories in evaluation. Grounded on the idea of reflective practice, we have proposed 10 declarations that in our perspective will serve to advance reflective practice around the development, refinement, examination, and use of program theories. Ultimately, our manifesto and this issue as a whole both elaborate on Carol Weiss's (1995) famous adage that when it comes to modern evaluation practice, there is truly nothing as practical as good theory.

REFERENCES

Alford, J., & Head, B. W. (2017). Wicked and less wicked problems: A typology and a contingency framework. *Policy and Society*, *36*(3), 397–413. https://doi.org/10.1080/14494035.2017.1361634

Ayres, I., & Braithwaite, J. (1992). *Responsive regulation: Transcending the deregulation debate*. New York, NY: Oxford University Press.

Befani, B., Ledermann, S., & Sager, F. (2007). Realistic Evaluation and QCA: Conceptual parallels and an empirical application. *Evaluation*, *13*(2), 171–192. https://doi.org/10.1177/1356389007075222

Bemelmans-Videc, M.-L., Rist, R. C., & Vedung, E. (Eds.) (1998). *Carrots, sticks & sermons: Policy instruments and their evaluation*. New Brunswick, NJ: Transaction Publishers.

Biggs, J. S., Farrell, L., Lawrence, G., & Johnson, J. K. (2014). A practical example of Contribution Analysis to a public health intervention. *Evaluation*, *20*(2), 214–229. https://doi.org/10.1177/1356389014527527

Bledsoe, K. (2005). Using theory-driven evaluation with underserved communities: Promoting program development and program sustainability. In S. Hood, R. H. Hopson, & H. T. Frierson (Eds.), *The role of culture and cultural context: A mandate for inclusion, the discovery of truth and understanding in evaluation theory and practice* (pp. 175-196). Charlotte, NC: Information Age Publishing.

Bledsoe, K., & Donaldson, S. (2015). Culturally responsive theory-driven evaluation. In S. Hood, R. H. Hopson, & H. T. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 3-28). Charlotte, NC: Information Age Publishing.

Bourgeois, I., & Whynot, J. (2018). Strategic evaluation utilization in the Canadian federal government. *Canadian Journal of Program Evaluation*, *32*(3), 327–346. https://doi.org/10.3138/cjpe.43179

Brousselle, A., & Champagne, F. (2011). Program theory evaluation: Logic analysis. *Evaluation and Program Planning*, *34*(1), 69–78. https://doi.org/10.1016/j.evalprogplan.2010.04.001

Brousselle, A., & Buregeya, J.-M. (2018). Theory-based evaluations: Framing the existence of the new theory in evaluation and the rise of the 5th generation. *Evaluation*, *24*(2), 153–168. https://doi.org/10.1177/1356389018765487

Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation, 32*(2), 199–226.

Dahler-Larsen, P. (2018). Theory-based evaluation meets ambiguity: The role of Janus variables. *American Journal of Evaluation*, *39*(1), 6–23. https://doi.org/10.1177/1098214017716325

Dobell, R., & Zussman, D. (1981). An evaluation system for government: If politics is theatre, then evaluation is (mostly) art. *Canadian Public Administration*, *24*(3), 404–427. https://doi.org/10.1111/j.1754-7121.1981.tb00341.x

Donaldson S. I., & Crano, W. D. (2011). Theory-driven evaluation science and applied social psychology: Exploring the intersection. In M. M. Mark, S. I. Donaldson, &

B. Campbell (Eds.), *Social psychology and evaluation* (pp. 141-160). New York, NY: Guilford Press.

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

Greene, J. C. (2016). Advancing equity: Cultivating an evaluation habit. In S. I. Donaldson & R. Picciotto (Eds.), *Evaluation for an equitable society* (pp. 49-66). Charlotte, NC: Information Age Publishing.

Grim, E., Castillo, I., & O'Quinn, E. P. (2018). *The logic model repair shop: An introduction to 3D logic models*. A Presentation at EERS 2018. Retrieved from https://www.slideshare.net/IsaacCastillo6/the-logic-model-repair-shop-an-introduction-to-3d-logic-models

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

Head, B. W. (2008). Wicked problems in public policy. *Public Policy*, *3*(2), 101–118. Retrieved at https://search.informit.com.au/documentSummary;dn=662880306504754;res=IELHSS

Kalmar, D. A., & Sternberg, R. J. (1988). Theory knitting: An integrative approach to theory development. *Philosophical Psychology*, *1*(2), 153–170. https://doi.org/10.1080/09515088808572934

Leeuw, F. L. (1991). Policy theories, knowledge utilization, and evaluation. *Knowledge and Policy*, *4*(3), 73–91. https://doi.org/10.1007/BF02693089

Leeuw, F. L., & Donaldson, S. I. (2015). Theory in evaluation: Reducing confusion and encouraging debate. *Evaluation*, *21*(4), 467–480. https://doi.org/10.1177/1356389015607712

Lemire, S. (2017). *Meta-modeling social programs: Methodological reflections on a practical example* (Doctoral dissertation). University of California, Los Angeles.

Lemire, S., Christie, C. A., & Nielsen, S. B. (2019). Mending the theory gap in evaluation—moving towards theory knitting. In M. Palenberg and A. Paulson (Eds.), *Evaluation and the pursuit of impact*. New York, NY: Taylor & Francis.

Lemire, S., Nielsen, S. B., & Dybdal, L. (2012). Making contribution analysis work: A practical framework for handling influencing factors and alternative explanations. *Evaluation*, *18*(3), 294–309. https://doi.org/10.1177/1356389012450654

Mark, M. M., Donaldson, S. I., & Campbell, B. (2011). The past, the present, and possible futures of social psychology and evaluation. In M. M. Mark, S. I. Donaldson, and B. Campbell (Eds.), *Social psychology and evaluation* (pp. 4-27). New York, NY: Guilford Press.

Mayne, J. (2001). Addressing attribution through contribution analysis: Using performance measures sensibly. *Canadian Journal of Program Evaluation*, *16*(1), 1–24.

Mayne, J. (2015). Useful theory of change models. *Canadian Journal of Program Evaluation*, *30*(2), 119–142. https://evaluationcanada.ca/system/files/cjpe-entries/30-2-119_0.pdf

Mayne, J. (2017). Theory of change analysis: Building robust theories of change. *Canadian Journal of Program Evaluation*, *32*(2), 155–173. https://doi.org/10.3138/cjpe.31122

McGrail, S. (2014). Rethinking the roles of evaluation in learning how to solve "wicked" policy problems: The case of anticipatory techniques used to support climate change mitigation and adaptation. *Evaluation Journal of Australasia*, *14*(2), 4–16. https://doi.org/10.1177/1035719X1401400202

Montague, S., & Porteous, N. L. (2013). The case for including reach as a key element of program theory. *Evaluation and Program Planning*, *36*(1), 177–183. https://doi.org/10.1016/j.evalprogplan.2012.03.005

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. London, England: Sage.

Peters, B. (2017). What is so wicked about wicked problems? A conceptual analysis and a research program. *Policy and Society*, *36*(3), 385–396. https://doi.org/10.1080/14494035.2017.1361633

Pope, C., Mays, N., & Popay, J. (2007). *Synthesizing qualitative and quantitative evidence: A guide to methods*. Milton Keynes, England: Open University Press.

Reicher, S., Haslam, A., & Van Bavel, J. (2018). Time to change the story. *The Psychologist*. Retrieved from http://thepsychologist.bps.org.uk/time-change-story

Riemer, M., & Bickman, L. (2011). Using program theory to link social psychology and program evaluation. In M. M. Mark, S. I. Donaldson, and B. Campbell (Eds.), *Social psychology and evaluation* (pp. 104-138). New York, NY: Guilford Press.

Rogers, P. (2016). Understanding and supporting equity: Implications of methodological and procedural choices in equity-focused evaluations. In S. I. Donaldson and R. Picciotto (Eds.), *Evaluation for an equitable society* (pp. 199-216). Charlotte, NC: Information Age Publishing.

Schmitt, J., & Beach, D. (2015). The contribution of process tracing to theory-based evaluations of complex aid instruments. *Evaluation*, *21*(4), 429–447. https://doi.org/10.1177/1356389015607739

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

Schwandt, T. (2018). Evaluative thinking as a collaborative social practice: The case of boundary judgment making. In A. T. Vo & T. Archibald (Eds.). *New Directions in Evaluation*, *158*, 125–138.

Vaessen, J. (2016). What is (good) program theory in international development? (Blog post). Retrieved from http://ieg.worldbankgroup.org/blog/what-good-program-theory-international-development

Vaessen, J., & Leeuw, F. L. (2009). Interventions as theories: Closing the gap between evaluation and the disciplines. In J. Vaessen and F. L. Leeuw (Eds.), *Mind the gap: Perspectives on policy evaluation and the social sciences* (141-170). New Brunswick, NJ: Transaction Publishers.

Van Melle, E., Gruppen, L., Holmboe, E. S., Flynn, L., Oandasan, I., & Frank, J. R. (2017). Using contribution analysis to evaluate competency-based medical education programs: It's all about rigor in thinking. *Academic Medicine*, *92*(6), 752–758. https://doi.org/10.1097/ACM.0000000000001479

Vo, A. T., & Archibald, T. (2018). Evaluative thinking. *New Directions in Evaluation*, *158*, 139–147. https://doi.org/10.1002/ev.20317

Weiss, C. H. (1995). *Nothing as practical as good theory*. Washington, DC: Aspen Institute.

Weiss, C. H. (1997). How can theory-based evaluation make greater headway? *Evaluation Review*, *21*(4), 501–524. https://doi.org/10.1177/0193841X9702100405

Weiss, C. H. (1998). If program decisions hinged only on information: A response to Patton. *Evaluation Practice*, *9*(3), 15–28. https://doi.org/10.1177/109821408800900302

Williams, B., & Hummelbrunner, R. (2011). *Systems concepts in action: A practitioner's toolkit*. Stanford, CA: Stanford University Press.

AUTHOR INFORMATION

Sebastian Lemire is a postdoctoral scholar in the Social Research Methodology Division in the Graduate School of Education and Information Studies, University of California, Los Angeles.

Jane Whynot is the past president of the National Capital Chapter of the Canadian Evaluation Society, a contract instructor at Carleton University, and a Ph.D. candidate at the University of Ottawa.

Steve Montague is a partner in Performance Management Network Inc. and an adjunct professor at Carleton University