

Does Your Implementation Fit Your Theory of Change?

Steve Montague
Carleton University

Abstract: *A brief review of evaluation findings in almost any given domain typically reveals that most and sometimes all major findings deal with the implementation of initiatives—also known as action theory. Moreover, the findings regarding implementation frequently allude to mismatches between the type or level of implementation occurring and the fundamental nature of the initiative. Case examples will illustrate that while all permutations and combinations of change and action theories cannot be summarily assessed, one can use case analysis to draw some lessons to suggest that some combinations are essentially toxic, while others provide at least a reasonable chance of success. The implication is that further systematic coding and analysis of change theories, action theories, and in particular their combinations in programs could produce useful insights for both evaluation and public-policy decision making.*

Keywords: *action theory, implementation, program theory, public policy, theory of change*

Résumé : *Une revue rapide des résultats d'évaluation dans presque n'importe quel domaine révélerait que la majorité des études abordent l'implantation des interventions (théorie de l'action). De plus, les résultats d'analyse d'implantation indiquent fréquemment un manque de correspondance entre le type ou le niveau de mise en œuvre et la nature fondamentale de l'intervention. L'analyse d'exemples indique que même s'il est impossible d'évaluer toutes les permutations et les combinaisons de théories de l'action et de changement, il est possible d'utiliser ces cas pour tirer certaines leçons qui suggèrent que certaines combinaisons sont essentiellement, toxiques, alors que d'autres indiquent une chance raisonnable de réussir. Une codification et une analyse systématiques plus poussées des théories du changement, des théories de l'action, et de leur combinaison peut conduire à des observations utiles autant pour l'évaluation que pour la prise de décision en matière de politiques publiques.*

Mots clés : *théorie de l'action, implantation, théorie d'intervention, politiques publiques, théorie du changement*

Corresponding author: Steve Montague, Carleton University, 1945 Lauder Drive, Ottawa, ON, K2A 1B2, Canada; steve.montague@pmn.net

Theories of change and program theories have been much discussed (as they are in this special edition); however, what has not been focused on lately are the key differences between what has been called the theory of change and the action or implementation theory. According to [Rogers \(2014\)](#), a “theory of change” essentially explains how activities are understood to produce a series of results that contribute to intended impacts. [Chen, Pan, Morosanu, and Turner \(2018\)](#) have recently gone on to distinguish the theory of change and the action theory quite definitively. They note that the change model describes the causal process generated by the program and distinguish the action model from the change model as a “systematic plan for arranging staff, resources, settings and support organizations, to reach a target group and deliver intervention services” (p. 54). The [Treasury Board of Canada Secretariat \(TBS\) \(2010\)](#) has noted that such theories should include assumptions, risks, and external factors that describe how and why a program is intended to work. The TBS goes on to say that theory connects “the program’s” activities with its goals. It is inherent in the program design and is often based on knowledge and experience of the program, research, evaluations, best practices, and lessons learned. I contend that the problem with the TBS’s statement is that it appears to conflate “theory of change” (i.e., an explanation of how and why a certain type of intervention will make a difference) with the action or implementation theory. As noted above, action (some have called it implementation) theory has been distinguished from change theory by [Chen \(2005, p. 23\)](#), who states that action theory

specifies the major activities a program needs to carry out, ensuring an environment for the program that is supportive or at least not hostile, recruiting and enrolling appropriate target group members to receive an intervention, hiring and training program staff, structuring modes of service delivery, designing an organization to coordinate efforts, and so on.

As an extension of Chen’s comments, see also [Sager and Andereggen \(2012\)](#), [Mayne and Stern \(2013\)](#), [Montague and Porteous \(2013\)](#), [Renger, Bartel, and Foltysova \(2013\)](#), and the six distinguishing elements of an action model ([Chen et al., 2018](#)). Indeed, Chen’s program-theory models typically show reasonably elaborate action theories.

The problem is that many modern logic models, and certainly those found in Canadian studies, suffer from the conflation of these ideas, as shown in the TBS definition above. They either do not show the “action theory” at all or they “blend” in the key elements of Chen’s action theory.¹ What results from this conflation is a framing for analysis that may not anticipate design, delivery, or implementation weaknesses that relate to *how a program is implemented*—as compared to weaknesses that are a result of problems or gaps in the change theory. Note that, in Chen’s model, more conceptual space is dedicated to the Action Model than to the Change Model and yet in most logic models, it is the other way around.

AN ILLUSTRATIVE EXAMPLE OF THE ISSUE

In order to illustrate this point, an example is in order. Consider an education or training program. The theory of change follows [Kirkpatrick and Kirkpatrick's \(1994\)](#) learning theory, which states that there are four levels of change connecting what [Funnell and Rogers \(2011\)](#) categorize as an individual-based theory of reasoned action:

1. *Reaction*: what participants thought and felt about the training;
2. *Learning*: the resulting increase in knowledge and/or skills, and change in attitudes;
3. *Behaviour*: transfer of knowledge, skills, and/or attitudes from classroom to the job (change in job behaviour due to training program); and
4. *Results* (some evaluators might call these ultimate outcomes or impacts): the final results that occurred because of attendance and participation in a training program and due to the behavioural changes that ensued after the training (these are benefits and can be monetary, large-scale performance-based, etc.; they typically connect to mission goals).

(For an example of such a theory with assumptions, please see Koleros and Mayne in this special issue.) A typical logic model or theory of change description might simply show inputs and activities and outputs leading to the above-noted expected change. The problem here is that even if a broad range of assumptions and contextual factors are considered, when mapped simply to the change-theory logic they may only peripherally relate to the soundness of the design and delivery of the training.

Many times in evaluations of training, the design and delivery of training are seen as critical to its success. The mode, medium, content, timing, and physical conditions surrounding the training are critical. How often have we seen that training offered to people has suffered from poorly designed materials, learning environments, timing, or tailoring (relevance and suitable format), or from delivery problems with those teaching, relationships between teachers and students, or linkages to supportive infrastructure, institutions, or individuals? How often have we seen a failure to reach the appropriate students in the first place? Such elements and assumptions are rarely included in theory of change depictions and logic models, yet experienced educational evaluators (and empirical evidence) suggest that delivery and design components can make a huge difference to the success of educational investments. The quality of teaching is particularly important; see, for example, [Chetty, Friedman, and Rockoff \(2014\)](#), whose statistical study suggests that teacher quality is associated with huge differences in student outcomes.

Lesson learned #1: For educational programs (theories of change), pay attention to the pedagogy and the quality of the teaching.

THE PREPONDERANCE OF THE PROBLEM

A quick perusal of recent evaluation findings shows that observations of design and delivery (action or implementation theory) elements are more common than observations regarding the theory of change. A quick review using Google of available online evaluation reports notes that items such as delivery timeliness, collaborative support, data and information sharing, the delivery of findings, and the clarity of roles and responsibilities were found to be prevalent. These components are all related to the manner in which a program is delivered, rather than focusing on its anticipated results. In summary, some of the learning in evaluation reports relates to how programs/initiatives are delivered, yet most logic models and frameworks are either silent or give short shrift to the implementation (action) theory that serves the theory of change. Therefore, evaluation learning in perhaps its area of highest potential is unsystematic and almost accidental.

Other articles in this issue explore “useful” models for complex settings, nested models, actor-based models, and methods of getting away from “mechanistic” approaches to depicting theories. In this article I look at simplified case examples to illustrate the value in recasting activities, outputs, and outcomes by area in order to distinguish and recognize the theory of action or implementation (these words will be considered synonymous) as separate but related to the application of a theory of change (ToC) itself. These cases will broadly illustrate the application of this thinking at different levels and contexts, but they will also note “lessons learned” in order to show some immediate practical results from adopting such an approach.

SUMMARY CASES FOR CONSIDERATION

Case 1: “Cash for Clunkers” vs. infrastructure support as economic stimulus

In the fall of 2008, policy analysts and economists debated what to do about that year’s recession. They accepted the need for large government stimulus packages, but their debates centred on where to put the money. Two types of stimulus packages that were often discussed in North America were major public works (infrastructure) and tax “break” programs. Logically, the theories worked as follows:

1. Public infrastructure:
 - invest in (needed) infrastructure;
 - construction and “supply community” employment will be directly and indirectly created;
 - the economy will be stimulated / “saved” / maintained;
 - needed public infrastructure will be put in place; also enabling
 - supporting goods and services markets to be maintained.

2. Tax break:

- provide a tax cut to taxpayers;
- those people will spend more money in the economy; and
- the economy will be stimulated / “saved” / maintained.

While not all economists agreed, option (1) tended to be favoured over option (2), essentially because the idea was that public infrastructure investments left a legacy to help further growth (think “Marshall Plan” for war-ravaged Europe or Asia), while tax breaks could serve to more strongly exacerbate deficits: one-time infrastructure spending creates a “temporary” deficit and builds assets, compared to tax breaks, which could cause more permanent structural deficits, especially since tax cuts are politically hard to reverse. Economists also noted issues with distributive effects (tax cuts tend to favour those who pay more taxes, namely the wealthy) and problems of absorption. People might tend to save money since times were uncertain, so the stimulus would not work its way into markets as “fully” as other incentives. This latter concern appeared to come true, as American savings rates jumped to their highest levels in the decade right around this time, and the wealthiest Americans—the primary beneficiaries of the tax cuts of the early 2000s—are shown to have saved disproportionately more than lower-income earners ([Board of Governors of the U.S. Federal Reserve System, 2016](#)). In the end, governments went mostly for infrastructure investments, with some “sweetening” of social programs; for example, in Canada the employment insurance requirements were adjusted.

So what happened? Tim [Kiladze \(2010\)](#) of the *Globe and Mail* noted that investors were far too optimistic about the extent to which public stimulus funding to infrastructure would boost the fortunes and bottom lines of the construction industry (and the economy). Kiladze quoted a market analyst as follows:

Very few meaningful stimulus related infrastructure projects were launched in 2009. In fact, somewhat ironically, some infrastructure spending was actually delayed ... as [provincial, state, and local] governments awaited funding from the federal stimulus coffers.

By contrast with the above, and assuming that some stimulus was needed to boost the economy, one of the few Canadian federal government programs that arguably seems to have worked well—perhaps almost too well—in terms of spending money quickly was euphemistically called “Cash for Clunkers” ([Elliot, 2009](#)). The Canadian equivalent, “Retire Your Ride,” was essentially a rebate program to support people trading in their gas-guzzling, polluting older vehicles for more fuel-efficient, newer vehicles. The national program was in a way like a tax cut, but one-time, temporary, and based on a specific activity. It took off, meeting its annual targets of retiring 50,000 vehicles each year once the full program had been launched. Its evaluation noted that the program would have met its vehicle-retirement target of 200,000 in 48 months if it had not been delayed in its implementation by 21 months ([Environment Canada, 2011](#)). A large number of

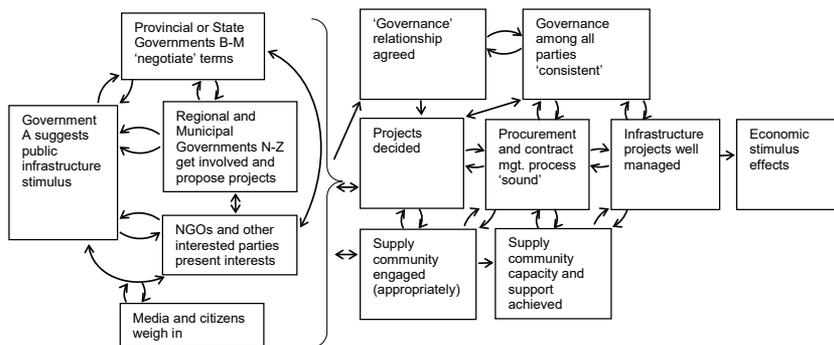


Figure 1: A simplified logic flow for infrastructure spending of a central government in a federal system.

those “retirements” were achieved in 2009, which, according to reports, provided a significant boost to automobile sales, possibly acting as a key push to reinvigorate the automobile sector (Elliot). At a minimum, sales were accelerated during a time when spending stimulus was needed. Why was this kind of program able to work while the best-laid “grand design” programs for infrastructure fail? The answer lies most likely not in the (mostly macro-)economic theory of one stimulus type versus another once in place; instead, it arguably lies in the implementation network or pathway of reach and relationships that these different initiatives require. Let us reconsider the infrastructure results pathway, including the reach and relationships “implementation” logic (see Figure 1).

Figure 1 shows that the reach, roles, and relationships for the delivery of typical infrastructure investments are myriad, complex, and politically dynamic. This means that, for one thing, they tend to take time. The number of major capital infrastructure projects that have been completed on time and on budget (in North America at least) can practically be counted on one hand. The fact is that the structure of the roles, relationships, and authorities required to “action” the stimulus assistance, even when streamlined, militates against speed in delivery. As noted by Kiladze (2010), the Government of Canada infrastructure program announced in 2008, but delivered late, likely had the perverse effect of “de-stimulating” the economy by delaying otherwise shovel-ready investments. By contrast, the key reach, relationships, and results for “Cash for Clunkers”-like programs are much more direct (see Figure 2).

Lesson learned #2: infrastructure support programs in multi-jurisdictional settings tend to make for slower than expected stimulus funding.

Clearly, other levels of government, members of civil society, mass media, and citizens can still affect the relationship between the government provider and the user, but this relationship is still much more direct and “authoritative” than a program to create public infrastructure. In other words, government “A”—in this case the U.S. or Canadian government—has an unambiguous authority to offer

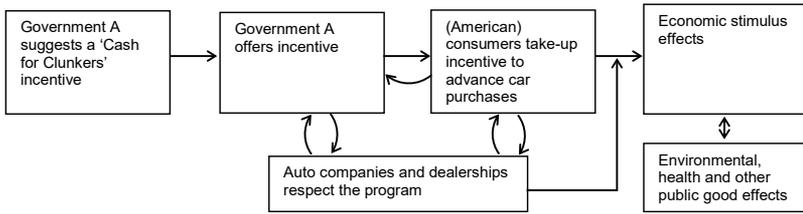


Figure 2: Simplified logic flow for “Cash for Clunkers” (“Retire Your Ride”)

citizens an incentive (a “carrot” or “bribe”) to trade in their vehicles for new ones. In most infrastructure investments, there are several jurisdictions and dozens of competing interests in play.

Lesson learned #3: stimulus delivery speed = F(# +complexity of authorities, the number of key actors, transaction amount, relationship (strength, trust), technical complexity, and other factors).

The main point is that public policy makers need to examine the reach, relationships, and roles implied by particular schemes when deciding on a given course of action. This is especially true when the initiative is required within a tight time frame for stimulus spending. Infrastructure investments may be a “good” investment in terms of economic theory, but as quick stimuli they are often structurally handicapped. The implementation reality does not typically fit the theory of change in this case.

Case 2: Repayable contributions for high-technology innovation

Another way to look at implementation and change theory “fit” is to describe them along with their requisite assumptions and then review evidence. [Table 1](#) summarizes a brief review of repayable contributions for innovation programs. The reviews used here involve more than a dozen studies, either conducted or reviewed by the author, of different types of innovation programs run over the past four decades by the Government of Canada. Some programs included repayability clauses and consortium or partnered delivery, while some did not. All related to some level of innovation commercialization as at least one intermediate outcome. Three theories are identified in [Table 1](#): theory of need or theory of the problem, theory of change, and theory of implementation/action. For more on problem identification and analysis discussion, see the [United Nations Development Programme \(2009, pp. 39–43\)](#).

Lesson learned #4: Proceed with extreme caution regarding the inclusion of repayability clauses in innovation funding agreements if you are dealing with projects that are a fair way from commercializing, but **also** if repayment triggers have widely interpretable definitions of success and consortium recipients have access to highly competent lawyers and accountants.

Table 1: Summary of findings from past innovation contribution program reviews

Theory	Reality observed
<p>Problem theory (<i>cause–effect</i>): There is insufficient innovation in sector X caused by a lack of domestic investment to bring innovations to market.</p>	<p>Investment in sector X becoming increasingly multinational ... not clear that domestic investment gap hinders innovation but rather gap in international investor confidence plus other policy, standards and marketplace barriers.</p>
<p>Change theory (<i>how a contribution mechanism should work</i>) A contribution will provide needed cash to companies to reduce their burden and costs so as to encourage them (or allow them) to move innovations from discovery to commercialization.</p>	<p>Cash does appear to enable many proponents to move forward faster in many cases than they would otherwise; however, if negotiations and payments go too slowly, then the lack of certainty creates risk aversion and may discourage highly innovative and less predictable investments in favour of stable and “safe” investments proposed by proponents. (See repayability below.)</p>
<p>Implementation (Action) Theory: Repayability in the contribution will address WTO concerns about unfair support. It will also ensure greater discipline in the innovations to focus on getting product to market.</p>	<p>“Repayability” clauses were found to work against the need for nimble investment, since consortia are often involved and repayment liability is handled by lawyers and accountants. (Intellectual property disputes among consortium members and the Government of Canada were also raised as barriers.) This “delay” and increased uncertainty create a negative feedback loop, which leads to a slow, hard process, which in turn leads to negative reactions, which lead to fewer good applicants and projects. In any case, when repayment has been contingent on commercial success, most funded innovation programs in Canada have had a very modest repayment rate. It seems that the payment triggers can be quite easily avoided—especially by large and more sophisticated contribution recipients.</p>
<p>Department will partner with the sector X industry association in order to select and deliver projects – based on the premise that the sector industry association will know and represent the needs of the sector for innovation.</p>	<p>Associations were not always found to represent all important sector interests—causing internal political disputes and lack of trust in the process, reducing reach and slowing processes, reinforcing a negative feedback loop, leading to fewer good applicants, which led to less success in areas of newer development (i.e., less innovative innovation projects may get funded because of a bias toward the more developed interests of larger players who carry greater weight in associations and partnered program management initiatives and who create “a success to the successful”² feedback loop.)</p>

Case 3: Administrative Monetary Penalties (AMPs)

We can also review the use of theory regarding needs or problem areas, change theory, and implementation/action theory in the area of regulatory control mechanisms. The idea of an Administrative Monetary Penalty (AMP) has a long history and connects to a deterrence pyramid “theory” stemming from Australian coal mining over many decades. An enforcement pyramid subjects regulated firms to escalating forms of regulatory intervention. These typically escalate from persuasion, a warning letter, civil penalty, criminal penalty, licence suspension, and then licence revocation (Ayres & Braithwaite, 1992). Table 2 summarizes the theory of the need or problem, the theory of change and the theory of action / implementation.

Table 2: Summary of findings from past AMPs review

Theory	Reality observed
<p>Problem Theory</p> <p>Regulators need an ability to moderate an otherwise harsh response (i.e., too big a jump between warning and criminal penalty).</p> <p>Regulators have inefficient existing means available (i.e., need to improve efficiency in regulation).</p>	<p>Not all regulators need a civil sanction, and not all civil penalties fit easily into the “scaled” pyramid.</p>
<p>Change Theory (how a monetary penalty sanction will work)</p> <p>A civil commercial penalty (the Administrative Monetary Penalty—AMP) is needed to fill out the deterrence pyramid and to allow for scaled deterrence.</p>	<p>AMPs often do not work as planned due to both theory being misapplied to context and implementation factors. The AMPs mechanism (as a theory of change) works where there exist</p> <ul style="list-style-type: none"> • a high level of regulated party commitment to the basic intent of the Act (low level of willful non-compliance); • controlled inspection conditions; • low complexity in terms of regulatory clauses and transactions; • a significant proportion of commercial transaction “value” represented by AMP—the economics of the marketplace; • a belief on the part of the regulated party that enforcement actions will be upheld; and • a complementary naming and shaming mechanism is in place (i.e., charged names are published)—good for established companies with a potential for reputational loss.

(Continued)

Table 2: (Continued)

Theory	Reality observed
<p>Implementation (Action) Theory</p> <p>Agencies will readily apply AMPs to gain efficiencies in sanctioning violators and will improve the cost-effectiveness of the whole system.</p>	<p>There can be significant cultural differences between and among implementing parties which significantly impact AMPs' efficiency and effectiveness.</p> <p>The key stakeholders include:</p> <ul style="list-style-type: none"> • associations (supplier and consumer) • policy makers and program proponents • inspectors • enforcement officers • legal counsel • "review" institutions <p>Other key implementation factors include:</p> <ul style="list-style-type: none"> • clarity of language defining violations • knowledge by inspectors and investigators of what constitutes a violation • "commitment" to the promotion of regulatory compliance by inspectors and regulators • level of engagement with regulated parties—and their representatives • consistency of the interpretation of legal responsibilities and authorities of all concerned parties and burden of proof / sufficiency of evidence

The main findings summarized here are from an evaluation study conducted for the [Canadian Food Inspection Agency \(CFIA\) in 2011](#). This study undertook an in-depth review of theory available in the literature before examining the use of AMPs at CFIA.

Lesson learned #5: AMPs fit much better for some types of applications (i.e., straightforward, high volume, inadvertent breach, strong supporting community) than for others. In addition, the implementation/action characteristics required for AMPs to work require a real commitment to deterrence by inspectors, information sharing practices, and clear/easy to interpret authorities.

WHAT THE EVALUATION TEAM TRIED

In cases 2 and 3 above, a systematic approach was used in real evaluation studies to come up with the findings noted in [Tables 1](#) and [2](#). In summary, the study team

- i) extracted the implementation theory from a conventional logic model;
- ii) lined it up with the change theory;

- iii) drew from research, experience, and analysis key assumptions and enabling factors to examine what factors were important and how they connected; and
- iv) tested the “lined-up” theories with real case evidence.

The study team, led by the author, constructed models to test key stages of significant cases for evidence of how key factors influenced performance and to test possible alternative explanations for results. The teams found that case-by-case assessment allowed study teams to pinpoint evidence to attempt to validate contribution claims and then help explain key factors for success. This altered depiction enabled the teams to identify issues regarding funding structure, governance, activities flow, and sector engagement/participation, which could be shown to directly influence the nature of the change theory, behavioural results, and impacts. More importantly, the dialogue allowed all concerned to separate issues, observations, and “learning” related to program governance and implementation from concerns related to a broader theory of change. In other words, the approach likely helped avoid a rush to judgement regarding the merits of an overall approach to regulation or innovation by noting that the characteristics of the implementation design and delivery had a profound effect on how the change theory worked. Therefore, if one reads between the lines in the 2011 AMPs evaluation, the study is not saying that AMPs don’t work; it is saying that this instrument works better in some applications than others, due to some key enabling conditions.

The laying out of a results chain showing how implementation (action) theory conditions change theory helps to establish the true (complicated and/or complex) reality of a program. It can also help all stakeholders to understand that change occurs in and among different groups. Such an insight in turn suggests that there is merit in planning initiatives with different expected results chains for different groups (see the article by Koleros and Mayne in this special edition for one approach that illustrates this concept).

PROPOSED APPROACH

So analysts like [Chen \(2005\)](#) and realist evaluators like [Pawson and Tilley \(1997\)](#), [Funnel and Rogers \(2011\)](#), and most recently [Brousselle and Buregeya \(2018\)](#) have distinguished action/implementation from change theory. Brousselle and Buregeya suggest that logical analysis, contribution analysis, and realist evaluation have a grounding in critical realism and that we may be observing a fifth generation of evaluation. If so, it will require models that are intuitively accessible to a wide range of users; that is, the models can’t get too complicated. It is hoped that the cases presented above show the practical value (lessons learned) from thinking about how each component works, and most importantly how each component works *with* (or against) the other in terms of how implementation (design and delivery) arrangements match up with theories of change.

As noted, traditional logic models do not extensively describe theory of any kind, let alone both theory of change and theory of action/implementation. Attempts to include both can be “unwieldy,” to say the least. So what does an evaluator do? Need one rely on narratives and side descriptions to adjust conventional logic models? Can a comprehensible approach be undertaken? Over the past five years, along with other evaluation colleagues, I have developed an approach to systematically and sequentially consider the action/implementation theory and the change theory. This approach is summarized in [Figure 3](#).

The first step is to recognize the action/implementation logic or theory that is involved. The process evaluation questions related to “how” a program is delivered are relevant, along with governance questions, attributes, and so on that may be drawn for program-authorizing documentation and reviews such as internal audits. The next step is to establish the theory of change that accompanies the program. There may be more than one. For example, there may be an educational component, suggesting a theory such as [Kirkpatrick and Kirkpatrick’s \(1994\)](#) learning model, as well as a financial contribution designed to reduce cost burden and/or encourage particular behaviour (a “carrot” program). There may be some kind of potential penalty or sanction for those not ultimately complying (a “stick” program). Each of these theories should be “modelled” as applicable (see [Bemelmans-Videc, Rist, & Vedung, 2011](#)).

Research can next be done on the theories to determine which factors have been important in shaping success. In addition, it will be important to look for

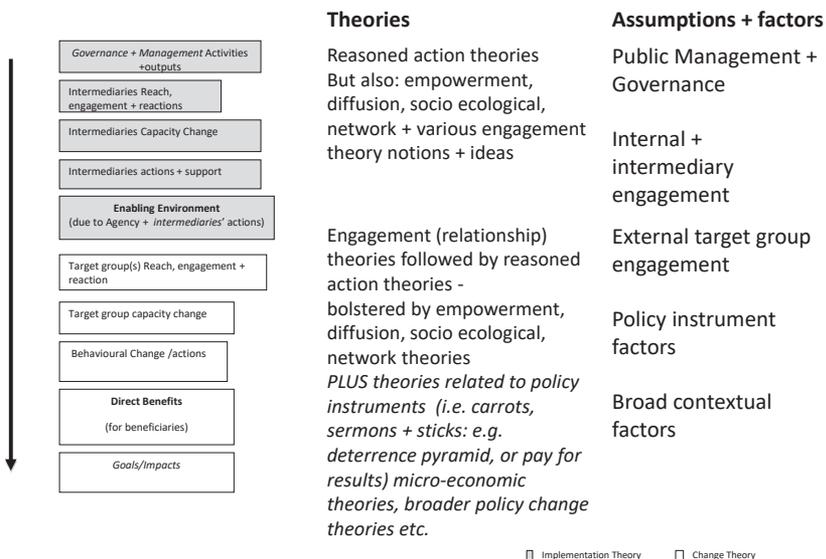


Figure 3: Theories of implementation (action) and change, key assumptions and factors. Montague, S. First used in 2015. *Using Realistic Contribution Analysis for Process and Impact Evaluations*, CES Annual Learning Event, February 25, 2015

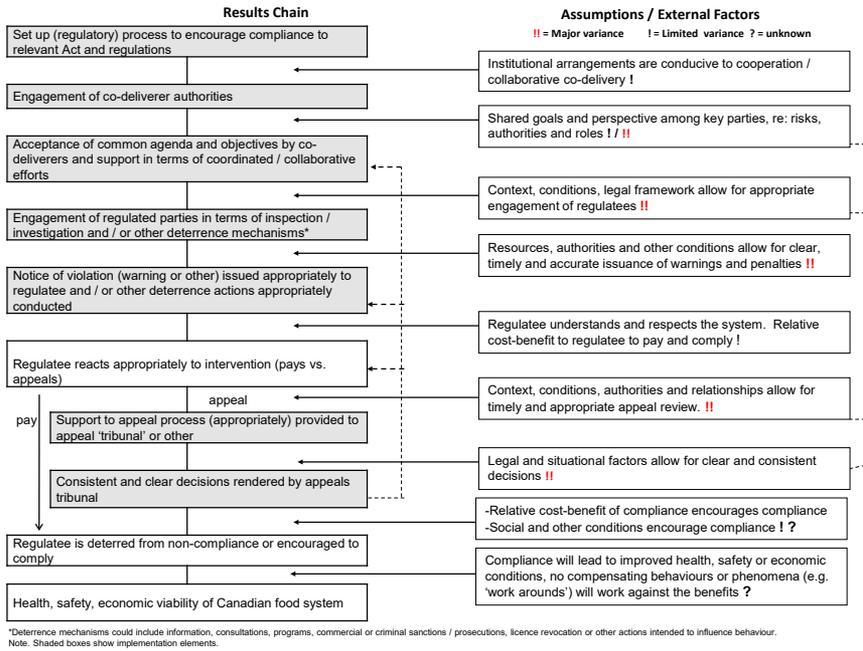


Figure 4: Regulatory initiative results chain for contribution analysis: Summary of AMP observations. CFIA. (2012). *Evaluation of Administrative Monetary Penalties (AMPs)*. <http://inspection.gc.ca/about-the-cfia/accountability/other-activities/audits-reviews-and-evaluations/evaluation-of-amps/amps/eng/1337024520304/1337025417391>

studies and analyses of how the proposed implementation strategy has worked in combination with the stated change theory. Figures 4 and 5 show the results of such an analysis: one regulatory application and one innovation support program.

The importance of this approach is that it can provide insights with only minimal original data collection and can guide data collection so as to invest gradually, in a targeted fashion and as needed, to support evaluation questions. This can potentially save thousands of dollars in data collection. If, for example, early parts of the results logic show that the implementation strategy is not reaching the target users, then one need not proceed much further with an analysis of impact on that target group.

In each of the examples shown, the insights had different emphases, but they also had some common elements. In the implementation of financial stimulus via infrastructure investments (as opposed to direct cash rebates for trading in one’s vehicle), the broader enabling environment would seem to play the fundamental role in determining success. Very little can be done about the federal system that exists in the United States and Canada, and these systems are in fact structured, respectively, by design and by evolution in order to temper the will of one government over others. This structural condition of the political economy—including

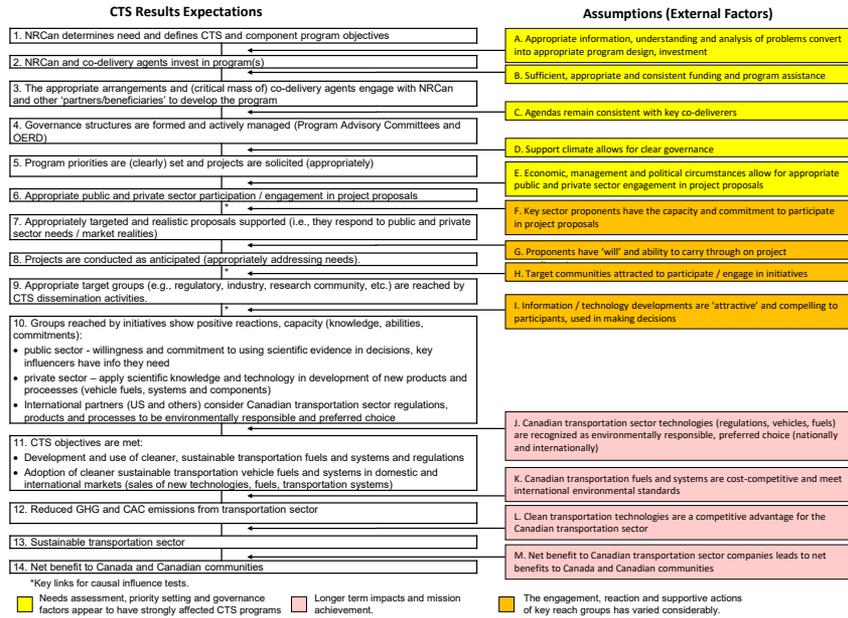


Figure 5: CTS results chain factors and assumptions. Natural Resources Canada. (2013). *Evaluation of Clean Transportation Systems Portfolio*. https://www.nrcan.gc.ca/evaluation/reports/2013/14844#3_2

the vast geography of both countries—would seem to create what amounts to a fundamental barrier to conducting large-scale centrally directed infrastructure projects in a short time frame, assuming that these liberal democracies stay in place as they are. In Canada, infrastructure in the form of a national railway system was at once a source of the nation’s emergence and the cause of its first major scandal. The interesting thing about the building of the Canadian transcontinental railway is that, according to Pierre Berton (1970, 1971), the construction element of the railway was done quickly and, despite some deplorable treatment of immigrant workers, efficiently. The main issues were political—with regional politics dominating from the outset. In this case the “implementation” theory writ large relates to the characteristics of the political system in place, its distribution of authorities, and therefore its complicated authorizing environment. (See also Mark Moore’s (1995) Strategic Triangle, in which he describes the “value proposition”—akin to the theory of change, “operational and resourcing considerations,” and “authorizing environment.”)

In the second case of repayable contributions for innovation, the use of a theoretically appropriate condition (i.e., repayability), appealing to both those concerned with politics and those interested in economics, does not work in practical application because of both the behavioural effects such a clause triggers and the uncertainty it brings. In the case of one of the innovation programs I

reviewed, agreements and payments were taking many months, causing concern and indeed hardship to all but the most patient and well-capitalized investors. The conclusions of this evaluation include the following:

In an era of Federal Government deficit reduction, TPC has been asked to cover a myriad of industrial policy objectives. The program was given a wide reaching mandate to create jobs and foster innovation in three very different sectors. At the same time, TPC's role in the innovation assistance process was limited by the focus on requirements for investment payback, as well as being subjected to intense international scrutiny. These constraints have led to several logical inconsistencies in its set-up, as compared to previous and other existing contribution programs:

- The program focuses on repayment of investments, yet it must fund (increasingly since the WTO decision of 1998) high risk technology innovation—often in emerging areas of technology.
- TPC must operate in a consistent, transparent and “fiscally responsible” manner, leading to a lengthy multi-step assistance process, yet it is by definition mandated to fund projects of high technology risk and market uncertainty—areas which require speed and flexibility.
- The program was essentially developed in many respects as a “son of DIPP.” The Defence Industries Productivity Program (DIPP) was designed to serve the needs of the mature, cold-war aerospace and defence industry of the 1970s and 1980s. The approach is likely not consistent with a program being asked to assist sectors (even that of the modern aerospace industry) facing a completely different market situation.

In conclusion, TPC has, in theory and rhetoric, been established to serve several publicly stated industrial innovation goals and sectors, while subjecting itself to the constraints of economic development assistance in the modern era. In reality it would appear that the “one-size fits all” approach that has been taken up to now may not be up to the task. (*Industry Canada, 2003*, pp. 5–6)

Regrettably, while innovation programming has evolved away from a “plodding” risk-averse delivery culture, it has not, at this time, moved all that far. This perhaps speaks as much to the Canadian government’s weakness in systematically accumulating evaluative evidence and knowledge as it does to a failure to distinguish the implementation factors important to program success. (The theme of knowledge accumulation through better and more consistent theory-based evaluation is a theme expressed elsewhere in this special issue.)

In the third case, a regulatory instrument—AMPs—is first situated in its theoretical place in terms of regulatory instruments (i.e., on a continuum between warnings and criminal prosecutions) and then examined for its fit with respect to the logical expectations for the theory. In essence, the theory fails in certain application areas (i.e., low-value cargo, difficult inspection conditions, multiple groups involved in processes, the lack of basic “will” to comply in the target and surrounding communities, and technical testing difficulties) and has certain implementation considerations and constraints (i.e., motivations of inspecting staff,

vaguely worded guidance in terms of legal or regulatory clauses, limited penalty levels versus cargo values, and the presence or absence of complementary deterrents or incentives, including a “sympathetic” quasi-judicial and judicial appeals system). When these and possibly some other conditions are negative, AMPs are doomed to fail. This was found for applications under the *Health of Animals Act* and regulations of the CFIA, but several of these conditions were also found to be true in other cases outside of food safety. The [CFIA’s \(2011\) Evaluation of Administrative Monetary Penalties](#) summarized the factors guiding use in terms of evidence from the research and evidence from interviews as follows:

Scholars have noted that AMPs are appropriate when the following elements are present, and their observations are supported by the interviews conducted for this evaluation:

- a large volume of cases is likely to be processed annually (that is, many transactions are being inspected);
- the regulator had stronger sanctions but the monetary penalties could be used to moderate a harsher response;
- speedy adjudication to the enforcement scheme is important;
- specialized knowledge (for example, technical expertise) and agency expertise in the resolution of disputed issues was needed;
- issues of law are rare;
- consistency of outcome was important; and
- there is a likelihood that an agency or group of agencies will establish an impartial forum in which cases can be efficiently and fairly decided.

Interviewees made the following suggestions of key factors affecting AMPs success:

- The regulated party has a high level of commitment to the basic intent of the legislation, so there is a low level of willful non-compliance. For example, associations and companies are willing to work with the CFIA to ensure compliance;
- The regulations contain clear language defining violations;
- There are controlled inspection conditions. For example, the inspection takes place in a regulated public facility vs. a remote privately run location;
- Transactions are not complex;
- Inspectors and investigators share an understanding of what constitutes a violation;
- Inspectors and all other regulator[y] program staff share a commitment to the promotion of regulatory compliance;
- There is a consistent interpretation of the legal responsibilities of all concerned parties, and of the burden of proof and of evidence;
- Significant proportion of commercial transaction “value” represented by AMP (i.e., the cost of the AMP versus the value of the shipment); and,
- Regulated parties believe that enforcement actions will be upheld.

In this case, sorting these observations after the fact into implementation and theory of change factors can be useful, because in general terms,

implementation elements (penalty levels, regulation clarity, inspection and appeals process changes, communications and internal education investments) urge one to consider how to *do things right*. On the other hand, the broader factors related to the target areas and the basic fit of the deterrence theory suggest how to *do the right thing* (e.g., if we are dealing with chronic non-compliers who see the AMP as a cost of doing business and have a business model that essentially relies on “borderline” practices, then it will likely take more than AMPs to be successful in bringing them into compliance). In the case of the latter—complementing AMPs with the publishing of the names of AMP recipients (naming and shaming) subsequently worked for at least one chronic offender, according to direct correspondence between a CFIA official and the author (see [Pawson, 2006](#), for an elaboration of factors allowing the use of naming and shaming to work).

So while each of the cases briefly examined here looks at different levels of policy, programming, and instrument use, has different levels of observable evidence, and appears to relate to slightly different definitions of implementing or action theory—if not both action/implementation and change theory—these cases also have something in common. All three cases suggest that sorting or synthesizing a policy, program, or initiative by its change theory and its implementation/action theory can be beneficial in understanding why and how phenomena occur. Ultimately, the practice should be helpful in accumulating knowledge and evaluative evidence.

CONCLUSIONS AND IMPLICATIONS

Canadian evaluators appear to have led a worldwide trend to develop what [Brouselle and Buregeya \(2018\)](#) have called a fifth generation of theory-based evaluation. This generation combines theory-based approaches to determine logical consistency and likely impact (logical analysis, contribution analysis) with explanatory features surrounding program mechanisms (realist evaluation). They see a coalescing of approaches using critical realism as a foundation to focus on “the explanatory power of contextual characteristics, implementation processes, and causal pathways to show, by identifying expected effects and impacts, how an intervention’s activities and outputs lead to outcomes” (p. 64). This paper suggests that in order to continue to foster this trend, which has evolved somewhat unconsciously up to now, it is useful to more clearly codify implementation/action theories and change theories, and furthermore to consider them together whenever one looks at a policy, program, or project intervention. Brouselle and Buregeya also suggest that we may be witnessing the evolution of a composite approach that follows the logic of a given intervention:

As the intervention unfolds, several implicit causal mechanisms result in the cumulative success or failure of the entire intervention or some of its components. Theory based approaches to evaluation are used to shed light on these mechanisms that operate in open systems and are embedded in multiple social systems. (p. 164)

This paper briefly models just such an approach and attempts to integrate theory-based approaches in order not only to make an attempt to show logical connections and contributions but also to help explain how and why certain results occur through specified contextual considerations applied at different parts of the results chain or impact pathway. Thus the whole may be greater than the sum of its parts in terms of explanation.

There would seem to be a major opportunity in the future to continue to consistently codify both implementation/action theories and designs and change theories. This accumulated knowledge and learning could focus on what works, to what extent, for whom, and why, under specific application conditions—including the implementation characteristics, the change theory characteristics, and the combination of the two. I have suggested a few practical lessons learned from such efforts. Imagine an open-access learning system that might collectively put forward such learning for review, challenge, embellishment, and refinement. Such a practice might finally embed evaluative thinking into public management—perhaps in a kind of global Socratic forum where “learnings” such as those sprinkled throughout this paper can be discussed.

At the very least, we as evaluators might systematically address Patton’s (2018) ninth principle of evaluative thinking: “Evaluative thinking looks at the connections between processes and outcomes, and that means distinguishing them and measuring both” (p. 23). I have proposed a small addendum to this principle—namely, that evaluative thinking should also systematically look at the fundamental “fit” of implementation processes with desired outcomes.

NOTES

1. This practice may have been a result of a lack of familiarity with the Chen model, but also the somewhat complicated depictions and language used in its communications such that many practitioners do not see the difference between the action or implementation theory and the essential change theory.
2. “Success to the successful” is a famous archetype in systems thinking. See, for example, Senge (1994); Kim (2018).

REFERENCES

- Ayres, I., & Braithwaite, J. (1992). *Responsive regulation transcending the deregulation debate*. New York, NY: Oxford University Press.
- Bemelmans-Videc, M., Rist, R. C., & Vedung, E. (2011). *Carrots, sticks, and sermons: Policy instruments and their evaluation*. New Brunswick, NJ: Transaction.
- Berton, P. (1970). *The national dream: Building the impossible railway*. Toronto, ON: McClelland and Stewart.
- Berton, P. (1971). *The last spike: The great railway 1881–1885*. Toronto, ON: McClelland and Stewart.
- Board of Governors of the U.S. Federal Reserve System. (2016). *Report on the economic well-being of U.S. households in 2015*. Retrieved from <https://www.federalreserve.gov/2015-report-economic-well-being-us-households-201605.pdf>

- Brousselle, A., & Buregeya, J.-M. (2018). Theory-based evaluations: Framing the existence of a new theory in evaluation and the rise of the 5th generation. *Evaluation*, 24(2), 153–168. <https://doi.org/10.1177/1356389018765487>
- Canadian Food Inspection Agency [CFIA]. (2011). *Evaluation of administrative monetary penalties*. Retrieved from <http://inspection.gc.ca/about-the-cfia/accountability/other-activities/audits-reviews-and-evaluations/evaluation-of-amps/amps/eng/1337024520304/1337025417391>
- Chen, H.-T. (2005). *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective*. Thousand Oaks, CA: Sage.
- Chen, H.-T., Pan, H.-L. W., Morosanu, L., & Turner, N. (2018). Using logic models and the action model/change model schema in planning the learning community program: A comparative case study. *Canadian Journal of Program Evaluation*, 33(1), 53–58. <https://doi.org/10.3138/cjpe.42116>
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679. <https://doi.org/10.3386/w19424>
- Elliot, H. (2009). Did Cash for Clunkers work? *Forbes*. Retrieved from <https://www.forbes.com/2009/09/02/cash-for-clunkers-toyota-ford-lifestyle-autos-auto-sales.html#43bc9c723f62>
- Environment Canada. (2011). *Evaluation of National Vehicle Scrappage Program*. Retrieved from <http://www.ec.gc.ca/doc/ae-ve/2011-2012/1447/ec-com1447-en-s4.htm#s4-3>
- Funnell, S., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco, CA: Jossey-Bass.
- Industry Canada. (2003). *Evaluation of the Technology Partnerships Canada Program*. Retrieved from <https://ito.ic.gc.ca/eic/site/ito-oti.nsf/eng/00191.html>
- Kiladze, T. (2010, November 23). The great infrastructure boom that wasn't. *Globe and Mail*. Retrieved from <https://www.theglobeandmail.com/globe-investor/investment-ideas/the-great-infrastructure-boom-that-wasnt/article1315521/>
- Kim, D. (2018, May). *Success to the successful: Self-fulfilling prophecies*. Retrieved from <https://thesystemsthinker.com/success-to-the-successful-self-fulfilling-prophecies/>
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (1994). *Evaluating training programs: The four levels*. San Francisco, CA: Berrett-Koehler.
- Mayne, J., & Stern, E. (2013). *Impact evaluation of natural resource management research programs: A broader view*. Retrieved from <http://aci.gov.au/publication/ias084>
- Montague, S., & Porteous, N. L. (2012). 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>
- Moore, M. (1995). *Creating public value: Strategic management in government*. Cambridge, MA: Harvard University Press.
- Patton, M. Q. (2018). A historical perspective on the evolution of evaluative thinking. In A. T. Vo & T. Archibald (Eds.), *Evaluative thinking*. *New Directions for Evaluation*, 158, 11–28. <https://doi.org/10.1002/ev.20325>
- Pawson, R. (2006). *Evidence-based policy: A realist perspective*. London, England: Sage.

- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Thousand Oaks, CA: Sage.
- Renger, R., Bartel, G., & Foltysova, J. (2013). The reciprocal relationship between implementation theory and program theory in assisting program design and decision-making. *Canadian Journal of Program Evaluation*, 28(1), 27–41.
- Rogers, P. J. (2014). *Theory of change*. Methodological Briefs: Impact Evaluation 2. Florence, Italy: UNICEF Office of Research. Retrieved from <https://www.unicef-irc.org/publications/747-theory-of-change-methodological-briefs-impact-evaluation-no-2.html>
- Sager, F., & Andereggen, C. (2012). Dealing with complex causality in realist synthesis: The promise of qualitative comparative analysis. *American Journal of Evaluation*, 33(1), 60–78. <https://doi.org/10.1177/1098214011411574>
- Senge, P. (1994). *The fifth discipline: The art and practice of the learning organization*. New York, NY: Random House.
- Treasury Board of Canada Secretariat [TBS]. (2010). *Supporting effective evaluations: A guide to developing performance measurement strategies*. Retrieved from <https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/centre-excellence-evaluation/guide-developing-performance-measurement-strategies.html>
- United Nations Development Programme. (2009). *UNDP handbook on planning, monitoring and evaluating for development results*. Retrieved from <http://web.undp.org/evaluation/handbook/documents/english/pme-handbook.pdf>

AUTHOR INFORMATION

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