

Why are There Unintended Consequences, and What are the Implications for Doing Evaluation?¹

By
Jonathan A. Morell, Ph.D.
734 302-4668

jonny.morell@altarum.org, jamorell@jamorell.com

Available for downloading at: http://www.jamorell.com/Jonny/web_new/vita.htm#consequences

Some months ago I became intrigued with the question of why programs have unintended consequences, and the implications of their seeming inevitability for doing evaluation. To spur my thinking about the topic, I posted a query to Evaltalk, the listserv of the American Evaluation Association (<http://www.eval.org>). Not to my surprise, and most certainly to my delight, a large number of insightful messages appeared. This document summarizes my current beliefs about evaluation and unintended consequences, as informed by my reading and the discussion spurred by my Evaltalk post. Of course I make no claim to being right about what follows, and I know that better answers must be out there. The best I can say is that as of this writing, the following explanation works for me. Comment and critique is needed and invited.

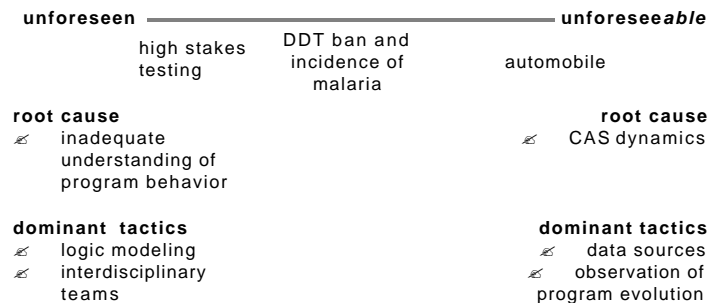
Unforeseen and Unforeseeable Consequences

A critical distinction is between consequences that are unforeseen, and consequences that are unforeseeable. However, these are actually two points on a continuum, and depending on where one is on that continuum, there are differences in why uncertainty exists, and what tactics evaluators should use.

In the real world, even at the extremes, there are no pure cases. Still, at the extreme “unforeseen” point, a reasonable amount of the surprise could plausibly be identified by means of rigorous, interdisciplinary inquiry, while at the extreme “unforeseeable” point, adaptive and non-linear

phenomena make prognostication impossible. As one moves along the continuum, the tactics needed to do better evaluation change. To anticipate later conclusions: Tactics in the “unforeseen” region tend toward methods and procedures for logic modeling. Tactics in the “unforeseeable” region deal with measurement (e.g. use of existing data) and

Figure 1: Consequences, Tactics, and Causes



¹ ? 2003 Jonathan A. Morell

observational methods to detect unanticipated change. A pictorial overview of the situation, along with examples arrayed at various points, is shown in Figure 1.

Unforeseen Consequences – Root Causes and Evaluation Tactics

There are many situations wherein rigorous logic modeling, encompassing an interdisciplinary team of people with deep expertise, would reveal many consequences that were not articulated when a program or a policy was born. Essentially, the problem is a failure to look, not a failure to reveal the unknowable.

The high stakes testing example falls into this category. While the specifics may have been unpredictable, many contours of the consequences were eminently knowable. One source of insight comes from systems theory. A second comes from the field of education. The high stakes testing movement is, in many ways, an effort to apply a single, overriding, outcome measure on the multidimensional educational system. A systems perspective has much to say about the consequences of such efforts. It tells us that systems will divert resources, and change services, to meet criteria set for them. It tells us that sometimes pursuit of a single objective may be necessary, but that such action is only useful over a short period of time, or under extraordinary circumstances, as for instance, when an organism is fighting for survival. It tells us that long run survival and growth almost always requires the joint optimization of a number of conflicting objectives. And from the field of education, we know that when particular objectives are set, or tests mandated, classroom teachers will mobilize to “teach to the test”, regardless of other stated objectives of the education system.

None of this is to argue that high stakes testing is a good or a bad idea, either as an objective or as a way to force organizational change. Nor is it to say that had the consequences been thought out in more detail, that people would have, or should have, acted differently. It is only to say that given our state of knowledge at the time, many unforeseen (and hence unexamined) consequences could have been revealed, and presumably, could have been entered into the policy debate. (Whether high stakes testing is a good idea, and whether it was implemented properly, are questions I leave to others.) From the point of view of evaluation, the essential point is that there are cases when “unforeseen” means “easily observable but not looked for”.

Unforeseeable Consequences – Root Causes and Evaluation Tactics

The explanation that works best for me comes from evolutionary biology. I do not know if what follows is best seen only as metaphor, or if it actually describes reality. I suspect the latter. I reason as follows.

- ✍ Any single program (aka organism) that we may wish to evaluate is but one inhabitant of a complex ecology made up of many other programs and policies.
- ✍ The organism of interest is itself not unitary, but is made up of many sub-units, each of which can be seen as an organism in its own right.
- ✍ The organism is surrounded by many other organisms, also of varying levels of size and complexity.

- ✍ To say that we are evaluating a program is to say that we are focusing our evaluation knowledge and wisdom on only one of many organisms in a complex ecology.
- ✍ To say we look at a logic model for a program is to say that we are developing a logic model for only a single organism within this complex ecology.
- ✍ Our organism of interest is, like all its constituents and neighbors, competing for resources and the opportunity to grow and thrive.
- ✍ As change in a single organism takes place, the evolutionary opportunities and challenges of the other organisms in the environment are affected. They will begin to evolve to adjust to their changing environment.
- ✍ All this evolution and co-evolution takes place on a fitness landscape whose shape may be changing.
- ✍ To say a program has unintended consequences is to say that the behavior of this complex ecology cannot be determined in advance. This is a reasonable view, because as the organism changes, each inhabitant of the environment has many choices as to how to respond, which in turn affect the system as a whole.
- ✍ Because of these dynamics, unlike programs at the “unforeseen” end of the continuum, here, no matter how carefully we look, the course of events cannot be discerned.

To continue this line of reasoning, why does it seem that so many unintended consequences are also seen as undesirable? (I have not counted, but I believe this is the case.) It is because:

- ✍ Any evaluation takes place from the point of view of one organism in the complex environment.
- ✍ The resource base in the environment, at least in the short run, is relatively stable. Thus much change is zero sum, or close to it.
- ✍ Therefore if we define “good” change from the point of view of only one organism climbing its fitness landscape, most responses from surrounding organisms will not be seen as “good”.

To illustrate, consider the case of the automobile. When the automobile first came on the scene, it held great promise to improve the quality of life in urban areas. (Just consider the amount of manure generated in places like New York City in the late 19th and first decades of the 20th centuries.) As the automobile developed, what changes took place in the environment?

The events unfolded at a time when scientific knowledge about consequences of CO₂ emissions were not only unknown, but unknowable. Only recently did we develop the combination of computing power and remote sensing needed to do the appropriate analyses.

The feedback loops were decades long, thus making it difficult to relate action to consequence.

In contrast to the high stakes testing example, the scenario here shows consequences of action that are unknowable because the course of events can only be explained in terms

of the continual interaction of many organisms, with the ecosystem characterized by a complicated array of cross linkages, symbiotic relationships, competitive pressures, and changing fitness landscapes. Earlier I said with respect to high stakes testing: “While the specifics may have been unpredictable, many contours of the consequences were eminently knowable”. In the case of the automobile, the opposite is true. No matter how hard we could have looked, no matter how interdisciplinary the searchers, consequences could not have been known.

How should evaluators do their work if they suspect that the program under scrutiny is susceptible to these kinds of dynamics? First, there must be an emphasis on routinely collected, versus specially collected, data. By routine I mean data whose collection is embedded in a program’s standard operating procedures, or which is collected outside the program, but none the less useful. Why this recommendation? Because special data, be it from tests, questionnaires, interviews, or observation, is usually constructed with a specific set of program outcomes in mind. If those outcomes were to change, the evaluator would be left with data that do not bear on the new realities that must be researched. While there is no guarantee that existing data will be useful, there is at least, the possibility of reconfiguring analysis strategies to take the new realities into account. The second approach is for the evaluator to keep the program under close scrutiny in order to determine whether important shifts, either in the program or its environment, are taking place.

Unforeseen and Unforeseeable – Insight from an Intermediate Case

The worldwide ban on DDT made a lot of sense because data were emerging which clearly showed environmental risk from a substance that had a very long half-life. However, as a result of the ban, the single best means of fighting malaria disappeared, resulting in much illness and death in many parts of the world. What happened?

- ✍ The importance of DDT as a malaria prevention measure did not outweigh other environmental concerns.
- ✍ Assumptions (implicit ones, at least) must have been made about the timely and economically feasible development of other preventative and curative measures.
- ✍ Intermediate solutions (e.g. judicious use of DDT in combination with mosquito nets) were not seriously considered.
- ✍ Drug companies, who might benefit from malaria drugs, operate in a different ecosystem from the one inhabited by those concerned with the effect of DDT on the environment. In the pharmaceutical industry ecosystem, there is no evolutionary advantage to developing malaria drugs. In fact, quite the opposite is true. Adaptive behavior for pharmaceutical companies competing with each other is to put resources into drugs for the industrialized world.

I see this as an intermediate case between those of high stakes testing and the automobile because it has similarities to both. On one hand, revealing the true consequences of action required long feedback loops. It took time to realize that non-DDT based solutions would not be realized. On the other hand, the disconnect between drug companies and malaria drug development was by no means certain. At the time, I think it would have been quite

plausible to expect the developed world to provide incentives to drug companies, thus putting the DDT ban, malaria in the developing world, and the pharmaceutical industry, on the same, mutually reinforcing, fitness landscape.

Placing Settings on the Continuum – the Problem of Knowing the Difference

The discussion so far has had the great benefit of hindsight. We know a lot about high stakes testing, the history of DDT, and the automobile. The problem for evaluators is that we deal with current programs and current problems. Faced with real challenges, it is not so easy to discern whether consequences that are neither seen nor understood are in some theoretical sense, unknown or unknowable. The simplistic advice is “do it all” – interdisciplinary logic modeling, close observation of program evolution, inclusion of as much routinely collected data as possible, identification of powerful competing activities. However, there is a collective lesson that goes beyond particular evaluation tactics, and extends to how we interact with clients. First, it is important to bring to discussions a sense of how much we already know, and do not know, about the program at hand. For instance, there is an extensive literature on the evaluation of HIV/AIDS prevention programs, while there is considerably less information on programs which use computer and networking technologies to help patients manage chronic diseases. In my experience one cannot assume that program developers know the state of knowledge, or understand the implications of having a deep, or thin, knowledge base. The issue here is not whether this knowledge is brought to bear on the planning of a program, as evaluators are seldom get involved in the planning business. But the knowledge is still relevant to evaluation, and it is our obligation to give our clients a sense of realistic expectations for what empirical data will reveal.

A second implication is that the diversity of input into logic modeling and design may be more important than the precise disciplines that are represented. (Assuming of course, some reasonable connection between expertise and program content.) This is so because it is impossible in advance to know all the relevant dimensions of the topic. Moreover, even if one did know, the total number would probably prove to be impractical. I have a strong sense that what matters is the wisdom and insight that comes from synthesizing multiple perspectives, than the specific perspectives themselves. The question is how robust the evaluation design is. Robustness comes from multiple perspectives, and not from *particular* multiple perspectives.

Third is the importance of evaluation within time frames. Many of the uncertainties in the DDT and automobile cases came from the length of the feedback loops involved, and as a consequence, the opportunity for change to work its way through an ecosystem, and back to the activity of interest. The longer the follow-up of an evaluation, the greater the likelihood that good evaluation will require tactics for dealing with the *unforeseeable*.