The Logic in Logic Models Part 2: Messages in Connections Among Model Elements

Jonny Morell
jamorell@jamorell.com

Abstract

This is the second of two blog posts on the logic in logic models. This one recapitulates and extends the content of the first. (Go here for Part One.) The title of this post uses the term "logic model" but I transition to "model" to reinforce the idea that what we call "logic models" are actually models to drive inquiry, and models in that sense have a considerable epistemological literature. I discuss three levels of model specificity. The first is the siloed model that is specific only at a high level of abstraction (e.g., outputs → outcomes). This model form lists, but does not specify relationships among elements within each high-level category. The second form is the "box and arrow" layout that is so common in evaluation. The third adds to the "box and arrow" form with additional information on relationships, e.g., designations of how strong or likely relationships are likely to be. We know less about how programs work and their consequences than we think we do. Therefor while more detail and information in a model is desirable, extreme caution is needed when moving from higher to lower levels of detail. As long as we use models exclusively for evaluation purposes, there is no reason to reach consensus on a single model. Beliefs about program operations and impacts can differ, so there can be every reason to test multiple models. Models with different levels of specificity in different regions can be practical and necessary. Practical because we may know more about one region than another. Necessary because if complex system behavior is involved, it may be impossible to drive the model to a low level of specificity. Blog posts I plan (but do not promise) to write in the future are: 1) mixing specificity and ambiguity in a single model, 2) models that exhibit complex system behavior, and 3) implications of using models for different purposes.

Table of Contents

This blog post is the second in an occasional series I am writing for the purpose of expanding evaluators' considerations of what logic models are, what they are good for, and what they might look like. A theme running through this post is that qualifiers can be added to representations of relationships among elements in a model and that so doing will strengthen our methodology, data interpretation, and conversations with customers and stakeholders.

I am *by no means* advocating that all the relationships discussed here should always be used. But there are times when some should be used for some parts of models. My intention (and hope) is to expand evaluators' recognition of what the possible relationships are, when they should be applied, and where.

A word about terminology. The title of this piece uses the familiar term "logic ,model" to draw the attention of the evaluation community. From now on, however. I will only use the term "model" because what we call "logic models" are visual descriptions of relationships that guide empirical inquiry. Across the sciences these depictions are called "models" and are subject to a great deal of epistemological thought (Frigg & Hartmann, 2018) . We need to pay attention to this work if we are to employ models to their greatest advantage when doing evaluation. (Evaluators use models for reasons other than guiding inquiry. These uses will be the subject of another blog post that will be written sometime in the undetermined future.)

Ambiguity in Models is a Feature not a Bug

My favorite model is the familiar Kellog model. Figure 1 is an example based on work I did in transportation safety. I like this model form because it is modest and honest. It acknowledges our ignorance. The story it tells is:

Operations	Activities	Outputs	Outcomes	Impact
Legislation	Rulemaking	Rules	Reduced defects	Reduced fatalities
Funding	Inspection	Reports	Reduced failures	Less environmental harm
Industry	Enforcement	Penalties		Less property loss
Industry standards	Investigation	Information		Reliable delivery
State programs	State grants			

Figure 1: Model 1 - Silo Model - Transportation Safety

- Operations → activities → outputs → outcomes → impact.
- We know the components of each category.
- Within categories we are ignorant with respect to relationships (if any) among components.
- We are ignorant with respect to how activity in one category affects activity in the next category.
-

The best we can say is that if the right things happen in one category, desired changes will take place in the subsequent category.

I am not familiar with evaluations that subject the elements of this kind of model to empirical investigation. One would think that results of an evaluation would decrease ignorance as to what was connected to what within silos. Also, I'd think that at least some connections across adjacent stages could be identified. (I am sure that models like this are out there. If anyone has examples please pass them along to me.)

The title of this post is: *Messages in Connections Among Elements*. I am fond of this model form because it acknowledges that we know almost nothing about connections among elements. At the very highest level, we know that some unknown combination of operations affect some unknown combination of activities, and so on through impact. In general, I think we know a lot less than we think we do about how programs work and what they accomplish. Using this model is a public announcement of how little we know.

The Equality Assumption in "Box and Arrow" Models

Frequently people who construct models think that they understand programs with more specificity than the column format discussed above. Such beliefs lead to constructing models using the "box and arrows" format. These models are elaborations, or variations on a theme, of simple causal models, as in Figure 2. (Models can mix specificity with ambiguity. I am ignoring such models in this post. My plan is to deal with them in the future.)

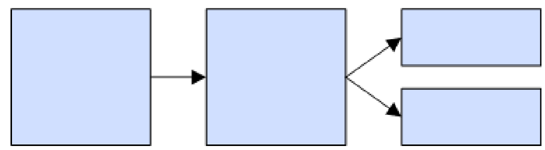

Figure 2: Many Evaluation Models are Elaborations of Simple Causal Models

Frequently people who construct models think that they understand programs with more specificity than the column format discussed above. Such beliefs lead to constructing models using the "box and arrows" format. These models are elaborations, or variations on a theme, of simple causal models, as in Error! Reference source not found.. (Models can mix specificity with ambiguity. I am ignoring such models in this post. My plan is to deal with them in the future.)

As is common in evaluation models, all the connectors in Figure 2. have the same format. I interpret this similarity as expressing the implicit assumption that all the relationships depicted in the model are equally consequential. And because models drive methodology and data interpretation, all our work is founded on the equality assumption. We all do this, including me. But I think we should do better. I think that we should add information to our models. (I have never seen a box and arrow model that did not make the equality assumption, but there must be some. If anyone has examples, please send them to me.)

Example: The Confidential Close Call Reporting System

To explain what I mean by "adding information" I'll present an example that is (very) loosely based on a large-scale evaluation that I led of the Federal Railroad Administration's Confidential Close Call Reporting System [(C^{3}RS) (Federal Railroad Administration, 2019)]. I adapted the model to let me discuss the consequences of expressing unequal connection relationships.

C3RS Program Theories.C^{3}RS operates on two program theories.

Near misses indicate potential for accidents

This theory posits that near misses differ from accidents in random (and often small) ways. According to this theory, any near miss could have been an accident and that, therefore, understanding why near misses occur, and decreasing their incidence, will decrease the incidence of accidents. This is an intuitively appealing theory that does have considerable validity. On the other hand, there are reasons not to accept it completely (Winkler, 2019, Wright 2004). The back and forth about near miss theory notwithstanding, that theory was the prime driver of C^{3}RS.

Diverse Input is Important to Determine the Reasons for Near Miss Incidents

The program theory here is drawn from a considerable literature that claims that the factors that cause near misses and accidents span multiple boundaries of expertise and system behavior (Salmon et.al. 2011). Consequently members of C3RS's near miss investigation committees were drawn from the different labor unions involved in train driving, railroad management, and the Federal Railroad Administration. The added advantage of this membership is that any recommended action called for approval by each of these stakeholder groups.

History and Context: Railroads in the United States

By lowering the cost of transporting goods by orders of magnitude, railroads had a truly transformative impact on life in the United States. While the Interstate Highway system (which facilitates shipping by truck) provides competition, it is still true that railroads are by far the most cost-effective way of transporting goods (Tolliver 2013). In their early years railroads were dangerous places to work. In response to all the injuries that occurred, in 1908 the United States passed the Federal Employees Liability Act (FELA). A critical aspect of FELA is that a railroad employee can sue his employer. You can imagine the arguments that FELA induces. It matters whether an employer is 45% or 50% responsible for an accident. As you can also imagine, FELA is not conducive to free and open conversations about the root and contributing causes of accidents. And yet, without such discussions, it is difficult to eliminate conditions that are conducive to accidents. Enter C^{3}RS, which was implemented and evaluated in two large freight railroads and two large passenger railroads.

C³RS Program Model

The model developed is shown in Figure 3.. (This is a loose adaptation of the model we actually used. The figure was modified to suit the needs of this blog post.)

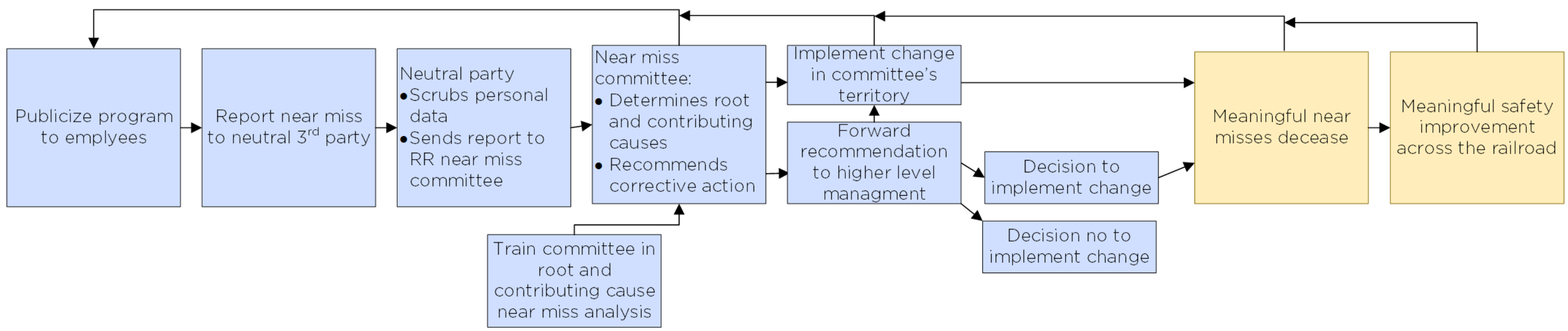

Figure 3: Model Two-- C3RS

1) C³RS is publicized to the workforce.
2) The publicity leads to reporting near misses to a neutral third party. (At the beginning the third party was the Bureau of Transportation Statistics (BTS). It later transitioned to NASA, which was already running the Aviation Safety Reporting System.)
3) The neutral third party scrubs the report of identifying information and forewords it to a labor/management committee. (Representatives of the FRA are also sometimes involved.)
4) At some time prior to the flow of near miss reports, the committee is trained in root and contributing cause near miss analysis.
5) The committee implements a near miss fix if the problem can be solved at their level or forwards a recommendation to higher level management.
6) As a result of implementing change the incidence of near misses decreases.
7) Because near misses decrease, safety improves.
8) Throughout, feedback on C3RS activity is publicized, thus increasing workers' willingness to report near misses.

Note that the connectors in the model have the same format and are unlabeled. This is the equality assumption that is the norm in evaluation. There is nothing wrong with this model. It served us well. As are all models, this one is a simplification of reality that worked to drive our methodology and data interpretation. But it is still the case that the equality assumption does not reflect the very real possibility that not all connections have the same valence. What might this model look like if we dropped this assumption? There are too many possibilities to enumerate, but Figure 4 on the next page shows one
.

> For anyone interested in railroad safety, it is important to bear in mind that the model leaves out a great deal of essential activity. Unions and management have to agree to participate in a process that is alien to the way they do business. The FRA has to grant rule waivers to some federal reporting requirements. A neutral third party has to be recruited. Funding needs to be secured.

A Model that Specifies the Nature of Relationships – Continuing the C^{3}RS Example

Model Three does not assume that all relationships in the model are the same. Although relationship specifications differ, the architectures of Models Two and Three are the same. There is nothing exotic or difficult in specifying the additional detail in Model Three because precision is not needed. All that is needed are some reasonable ballpark estimates. But the story told is much more enlightening than what we can glean from Model Two. How so?

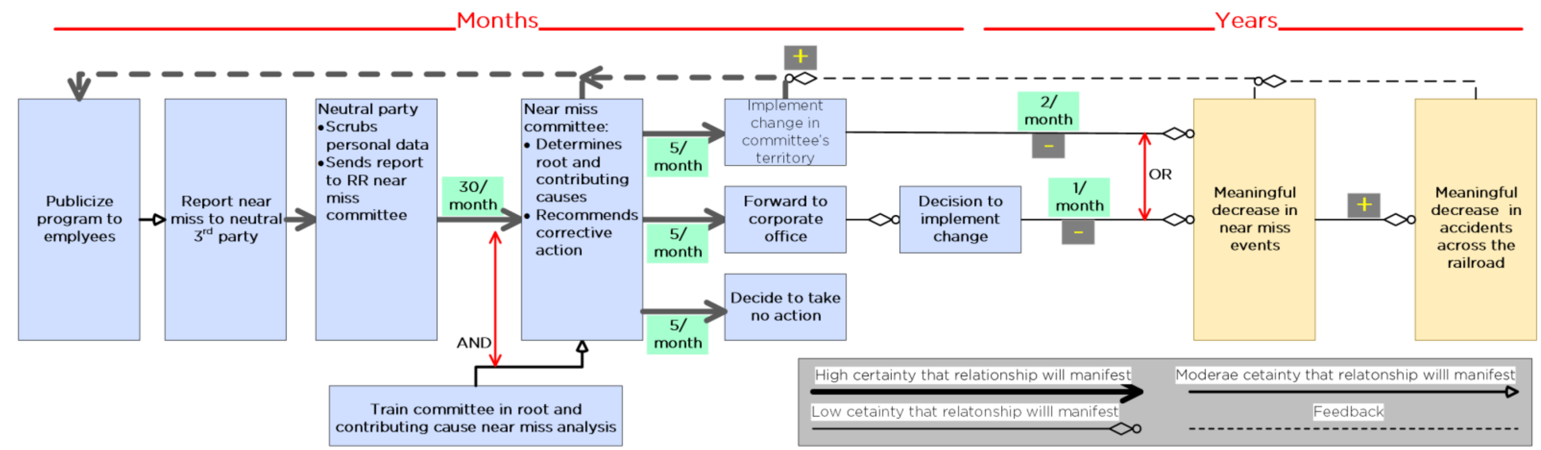

Figure 4: Figure 4: Model 3 - C3RS Model Two with Connection Specificity

Certainty of relationship:

Line thickness and arrow shapes signify an estimate of how certain it is that a relationship will manifest. For example, at the start of the program there might be only moderate confidence that workers would report near misses. After all, doing so is a radical departure from long-standing practice. But we can have confidence that BTS and NASA have the skills and motivation to assess near misses and to communicate their findings

.

Throughput:

How much activity takes place along the connectors between model elements? This question is answered by the green rectangles that accompany some of the connectors. The model shows that thirty reports come in to the committee each week, but only fifteen are processed – five for immediate correction, five passed on, and five flagged for no action. (Keep in mind that I made these numbers up for the sake of illustration. They are not reflective of the data in the real C^3RS evaluation.)

Conditionals:

This model specifies two conditionals that Model One does not. One is the AND designation for the two inputs to into "near miss committee problem analysis". The model reflects a program theory that asserts that successful problem solving requires that both of two conditions obtain – near miss reporting and problem-solving training. Either one alone would not suffice.

The second is the OR designation for the two paths to a meaningful decline in near miss events. The program theory here asserts that either path alone would lead to meaningful change. Also note that Model Two conflates these two paths. It was not until I puzzled through the ways that change could occur that the need to separate them occurred to me. Therein also lies a story about model building worth remembering.

Directionality:

In the outcome section of the model, the direction of change is specified. Implementing change has a negative effect on the occurrence of near miss events. Decreasing near miss events decreases accidents. Communication with workers increases their reporting of near-miss conditions. Specifying the direction of change is not the most consequential change from Model Two (Figure 3) to Model Three (Figure 4). After all, the directionalities are obvious without the plus and minus signs. But I find it useful to cast my eyes over a model and get a sense of what goes up and down as a result of program activity.

Feedback:

Feedback activity is depicted with dotted lines to differentiate them from forward activity. Visual differentiation to indicate directionality of change is not all that consequential. Still, it is often hard to scan models and quickly discern what is flowing forward and what is flowing backward. Using different forms for causal directions helps.

Timing:

Model Three is not drawn to scale, but it does make it obvious that the time scale from program inception to near miss analysis and some corrective action is in the order of months, while the time scale for meaningful change in safety is years.

Incomplete specification:

Note that the extra detail in Model Three is spotty. For instance, throughput is specified among some elements, but not others. I did this to indicate that it is not necessary to do a thorough job of specification. All that matters is to add the detail that may affect the evaluation or that represents an important aspect of program theory. More is better, but only some is OK.

Choice of indicators:
The different formats of connectors in Model Three represent degrees of confidence that an activity will occur. But there is no requirement to make this choice. One might pick an effect size measure, i.e. how much influence does one activity have on another? Or perhaps a signifier for the difficulty of detecting change? Or whatever makes sense for the evaluation. Also, there is no requirement to apply only one designation, or the same ones over all parts of the model.

Implications for Evaluation

It is clear that Model Three (Figure 4) provides much more guidance for designing an evaluation than does Model Two (Figure 3). The possibilities are too numerous to mention, but here are a few that strike me as particularly consequential.

Realistic Expectations Regarding Program Activity and Change Timelines
Model Three makes it clear that many close call reports will not get processed, that only a few of those that are processed will be acted upon, and that improvements in safety are far removed in time from when necessary corrective actions are identified. This is not necessarily a bad thing. It may just be an accurate representation of what is involved in making changes that decrease accidents. With such a representation, all involved can proceed with realistic expectations about process and outcome.

Realistic Expectations go Beyond the Practical and Extend to Program Theory
Realistic expectations are more than practical guidance about what should be measured when, or how stakeholders could justify a program. They are also an expression of program theory, i.e. an expression of how an accident reduction program operates in its environment.

Throughput
Model Three expresses a hypothesis about how many close call reports will be submitted, and the rates at which they will flow through the system. By stating this hypothesis, attention is drawn to the need to measure it over time. The need to know becomes a high-profile variable in a way that it does not in Model Two.

Methodology that Links Cases to Overall Change
The program theory claims that low-volume improvements over time can result in a meaningful improvement in safety. This implies a methodology that links specific cases of corrective action to specific changes in process or hardware, and which then aggregates the consequences of those changes over time.

Testing Conditionals
The Model specifies two hypotheses. One is that in order for the committee to function effectively, its members *must* be trained in some manner of systematic root cause analysis. This is not an unreasonable hypothesis, considering the long history of success with such programs (Latino & Latino, 2019; Percarpio et al., 2008). On the

other hand, an alternate hypothesis is not unreasonable, i.e., that special training may be a nice to have, but is not essential. After all, those committees are populated by members of labor and management who have long experience with railroading, and who have deep knowledge of the equipment they use and the processes that affect their behavior. This alternate hypothesis implies the need for an evaluation methodology that uses some kind of no-treatment control or delayed implementation of training design.

The second conditional is exceedingly consequential. It claims that meaningful system-wide safety improvement can result *either* from an accumulation of local change, *or* from system-wide change, but that both are not required. This hypothesis is important because it has implications for future designs of C^3RS or C^3RS-like programs, and also for organizational theories of safety improvement. As a practical matter it would be very difficult to test this hypothesis because once the C^3RS machinery is in operation, it will generate all the near miss improvements that it possibly can, both local and system wide. I suppose one could identify all the local and system-wide change, look at each group separately, and construct a counter-factual. I am not enamored with this design, but I can't think of another way to do it.

Choosing a Model Form

In this article I moved along a dimension of specificity in models and argued that different levels of specificity can be appropriate in different conditions. My bias is to favor less specificity because I tend to believe that we know less about how programs work and what they will accomplish than we think we do.

But I also believe that there are times when we can feel confident enough to specify relationships between elements in a model. The C^3RS example is a case in point. Its format (Model 2 Figure 3) is similar to many models found in the evaluation literature. Models like that often meet evaluation needs. They show what the elements in a program are, how those elements are connected, and what consequences are expected from program activity. The model points to what needs to be measured and what relationships should be investigated. It is also a clear articulation of program theory. As a guide for designing an evaluation, it is certainly superior to the siloed model (Model One Figure 1). IF that is, we are confident enough to build models like Model Two, we should.

The leap from Model Two to Model Three is both easy and instructive. I have never been involved in an evaluation where informed judgments could not be made about the kind of information that was added to Model Three, at least over some parts of the model. Nor have I ever been involved in an evaluation that would not have benefitted from adding that information. My advice is to do as I say, not to do what I do. Construct models like Model Three. Push back on your stakeholders' and team members' beliefs that the Model Two form will suffice. Argue that it is not.

When we construct models we operate on the assumption that we must reach consensus – one model to satisfy all. Why? It seems perfectly reasonable that people will differ on beliefs about how a program will operate, what outcomes it will produce, and how those outcomes will affect each other. Is there any reason not to test multiple models? Actually, there are good reasons to stick with one model, but I think we can work around them. One reason is time and cost. For instance, different models might have different needs for data or research design. Another reason is that what we think as an evaluation model is actually a model that serves two purposes – evaluation and planning. Planning cannot proceed by following multiple models. We can often get around the first difficulty with clever design. We can escape the second difficulty by building separate models, one for planning and one for evaluation. (This is another topic I hope to write a blog post about.)

Finally, I have been discussing different forms of evaluation models as if they were mutually exclusive. They are not. There is every reason to construct models that mix forms – specificity when possible, ambiguity when not. In fact, mixing degrees of specificity is not only practical, but it is often required because programs often exhibit complex system behavior. (Morell, 2021; Morell, 2024). I plan another blog post to address complex behavior in models.

References

Federal Railroad Administration. (2019). *Confidential Close Call Reporting System (C3RS) Lessons Learned Evaluation – Final Report*. https://railroads.dot.gov/elibrary/confidential-close-call-reporting-system-c3rs-lessons-learned-evaluation-final-report

Frigg, R., & Hartmann, S. (2018). Models in Science. *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2018/entries/models-science/

Latino, M. A., Latino, R.J., , & Latino, K. C. (2019). *Root Cause Analysis: Improving Performance for Bottom-Line Results, 5th ed.* CRC Press. https://doi.org/https://doi.org/10.1201/9780429446573

Morell, J. A. (2021). A Complexity-based Metatheory of Action for Transformation to a Green Energy Future. In R. D. Van den Berg, C. Maro, & M.-H. Adrien (Eds.), *Transformational Evaluation for the Global Crises of our Times* . (pp. 345 - 370). IDEAS. https://ideas-global.org/ideas-book-transformational-evaluation-for-the-global-crises-of-our-times/

Morell, J. A. (2024). A Complexity-based Plan for Evaluating Transformation. *Journal of Multidisciplinary Evaluation*. https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/867

Percarpio, K. B., Watts, B. V., & Weeks, W. B. (2008, 2008/07/01/). The Effectiveness of Root Cause Analysis: What Does the Literature Tell Us? *The Joint Commission Journal on Quality and Patient Safety, 34*(7), 391-398. https://doi.org/https://doi.org/10.1016/S1553-7250(08)34049-5

Tolliver, D., Lu, P., & Benson, D. (2013). Comparing rail fuel efficiency with truck and waterway. Transportation Research Part D: Transport and Environment, 24, 69-75. https://doi.org/https://doi.org/10.1016/j.trd.2013.06.002

Winkler, M., Perlman, Y., & Westreich, S. (2019). Reporting near-miss safety events: Impacts and decision-making analysis. Safety Science, 117, 365-374. https://doi.org/https://doi.org/10.1016/j.ssci.2019.04.029

Wright, L., & van der Schaaf, T. (2004). Accident versus near miss causation: a critical review of the literature, an empirical test in the UK railway domain, and their implications for other sectors. Journal of Hazardous Materials, 111(1), 105-110. https://doi.org/https://doi.org/10.1016/j.jhazmat.2004.02.049