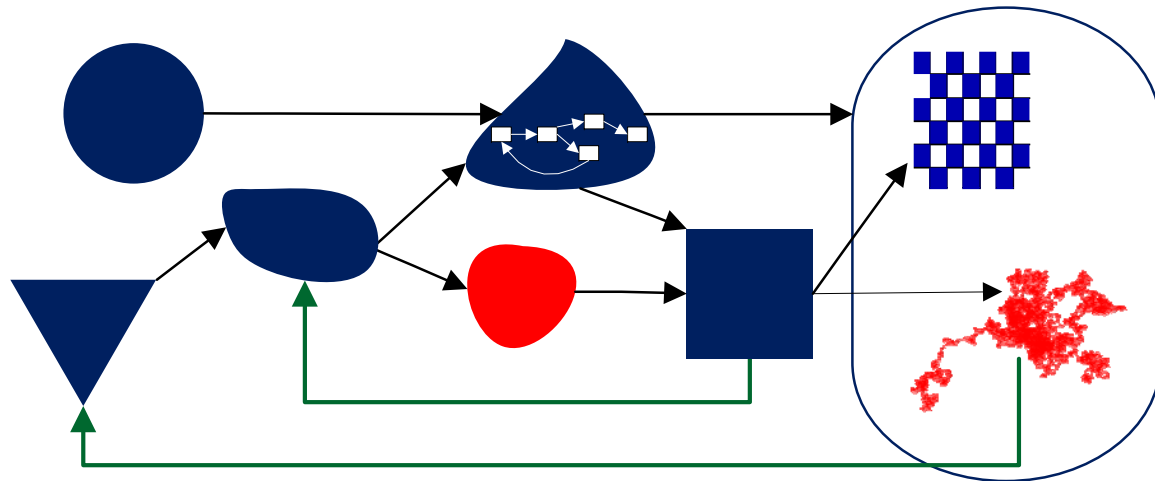


Program Logic, Program Theory, and Unintended Consequences: Understanding Relationships. Implementing Action

Professional Development
Workshop
Canadian Evaluation Society
Ottawa, June 15th – 18th, 2014



Jonathan A. Morell, Ph.D.
Director of Evaluation, Fulcrum Corporation
Editor, Evaluation and Program Planning
jmorell@fulcrum-corp.com
734 302-4668

Website: www.jamorell.com
Blog: www.evaluationuncertainty.com

Game plan for workshop

Part 1:

- Community building and expectation setting

Part 2: Core of workshop

- 2.1 Models for depicting and understanding program theory: power and limitations
- 2.2 Implications of model form for unintended consequences
- 2.3 Designs for dealing with unintended consequences

Part 3: Optional topics if time and interest

- Joint optimization of information density and readability
- Program theory and the practicalities of data and methodology
- Working with stakeholders and interested parties to develop program theory.

Community Building and Expectation Setting

- Name
- Affiliation
- What are some of the things you want to know when the day is over?

Do you know of a program that had an unexpected outcome?

- What happened?
- Why?
- What did the evaluation do about it?
- How might the outcome have been anticipated?
- Was it impossible to anticipate the outcome?

Vocabulary

I tend to use terms interchangeably

- Logic model and program theory
- Outcome and impact

If you need me to be precise, let me know, and tell me why the distinction is important to you.

What do I want you to get out of this?

Response to surprise

- Crisis response → advance planning

Adding “surprise” to evaluation planning

- Funding
- Deadlines
- Logic models
- Measurement
- Program theory
- Research design
- Information use plans
- Defining role of evaluator
- Logistics of implementation
- **Planning to anticipate and respond to surprise**

Class exercise

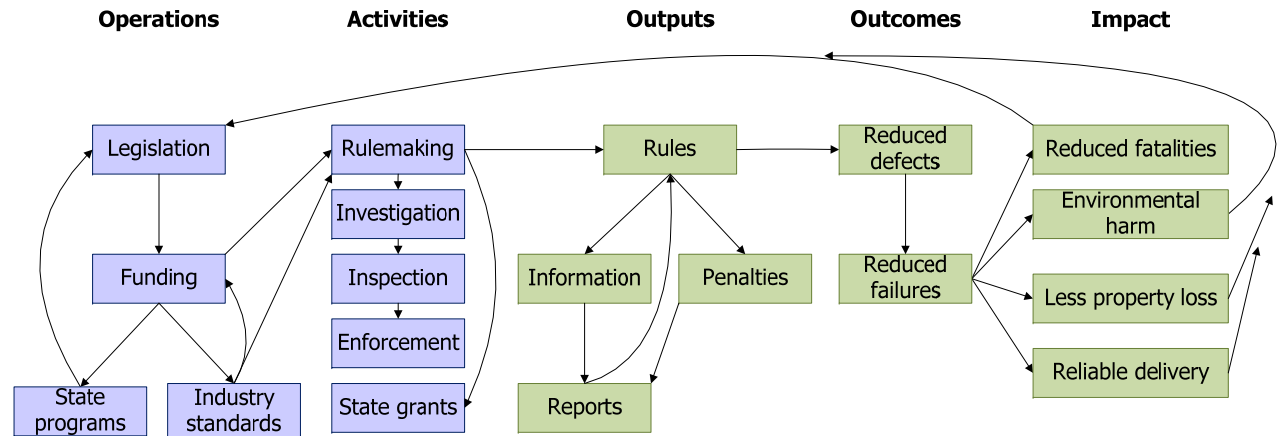
- Break into groups
- Someone volunteer a logic model
- Report back
 - What was being evaluated?
 - How did the model get developed?
 - What was the model used for at different stages in the evaluation life cycle?
 - When all is said and done, did you really need the model?

Part 2: Core of workshop

2.1 Models for depicting and understanding program theory: power and limitations

What can LMs be used for?

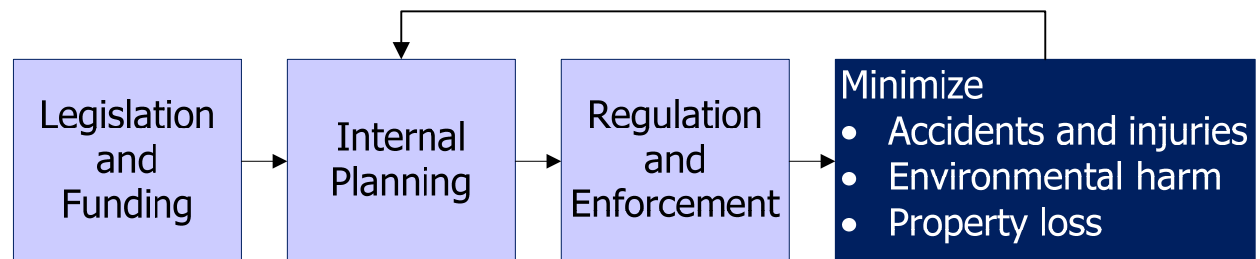
Explanation



Evaluation

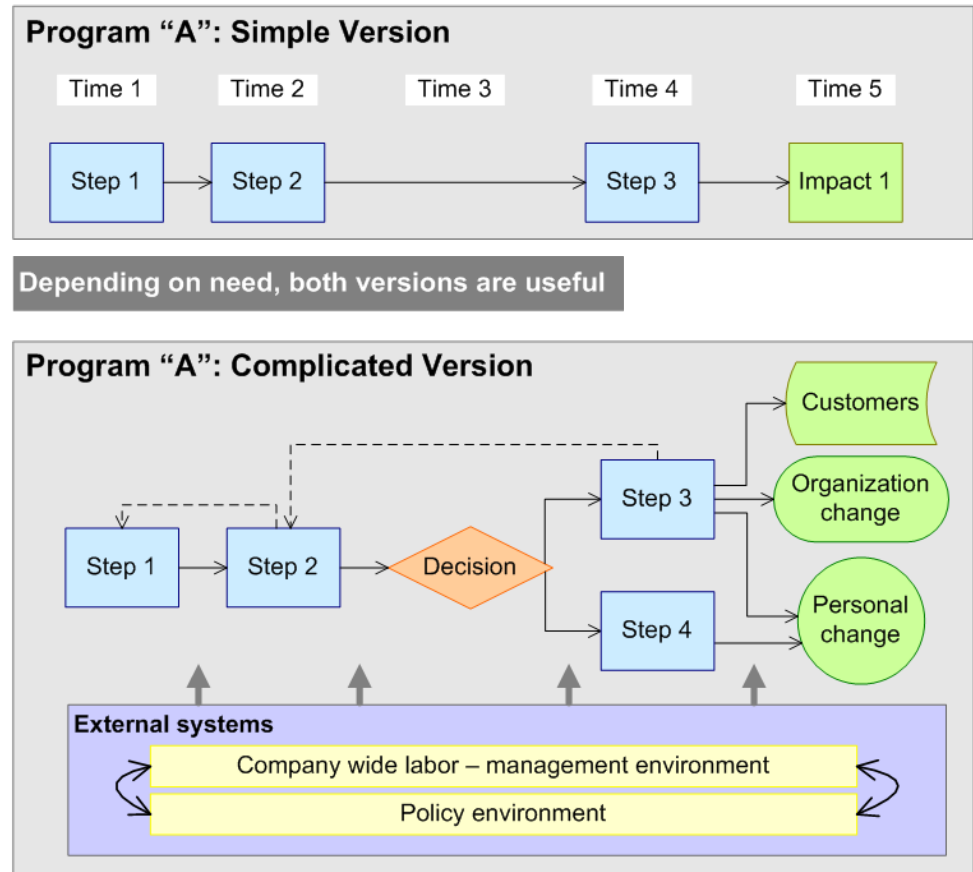
Operations	Activity	Output	Outcome	Impact
Legislation	Rulemaking	Rules	Reduced defects	Reduced fatalities
Funding	Investigation	Reports	Reduced failures	Less environmental harm
Industry	Inspection	Penalties		Less property loss
Industry standards	Enforcement	Information		

Advocacy



Depending on use, logic models can be simple or complicated

- Scale and complexity of program
- Diversity of information needed to design the evaluation
- Number of
 - Elements represented
 - Systems represented
 - Nested models of different scales
 - Feedback loops
- The same evaluation might need multiple versions, e.g.
 - Technical development vs.
 - Explanation to outsiders



What can be in a logic model?

- Feedback loops
- Verbal description
- Outside influences
- System boundaries
- Stakeholder priorities
- Timeline for observation
- Estimates of measurement feasibility
- Relationships among program elements
- Program content , process, and structure
- Guess as to whether parts of the model are correct
- Any other useful information

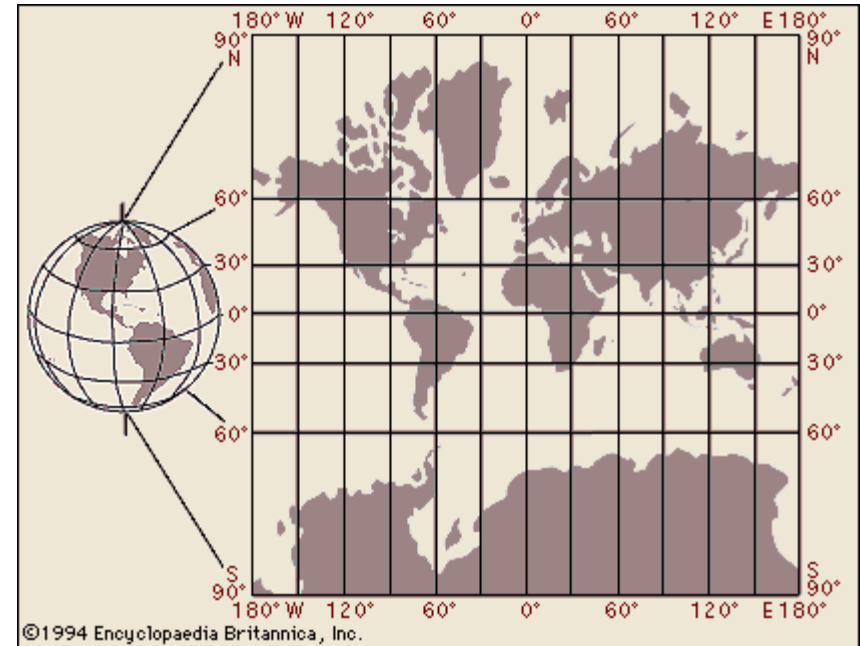
What kinds of relationships can a logic model show?

- 1 : 1
- 1 : many
- Many : many
- Probabilities
- Grouped, relationships unspecified
- Precedence
 - A before B
 - A & B simultaneously
 - Agnostic with respect to precedence

Different distortions for different reasons

Mercator

- Straight lines are rhumb lines, you can use the map to navigate.
- What is wrong?
- Land mass areas are distorted with increasing latitude



Hammer – Aitoff

- Correct land mass
- Problematic for navigation



Let's go back and look at your examples

- What got left out that might have been put in?
- Advantages and disadvantages of leaving it out.
- Any social science theory implicit in the model?
- As a practical matter how useful was the model in helping to do evaluation?
- How was it useful?

Part 2: Core of workshop

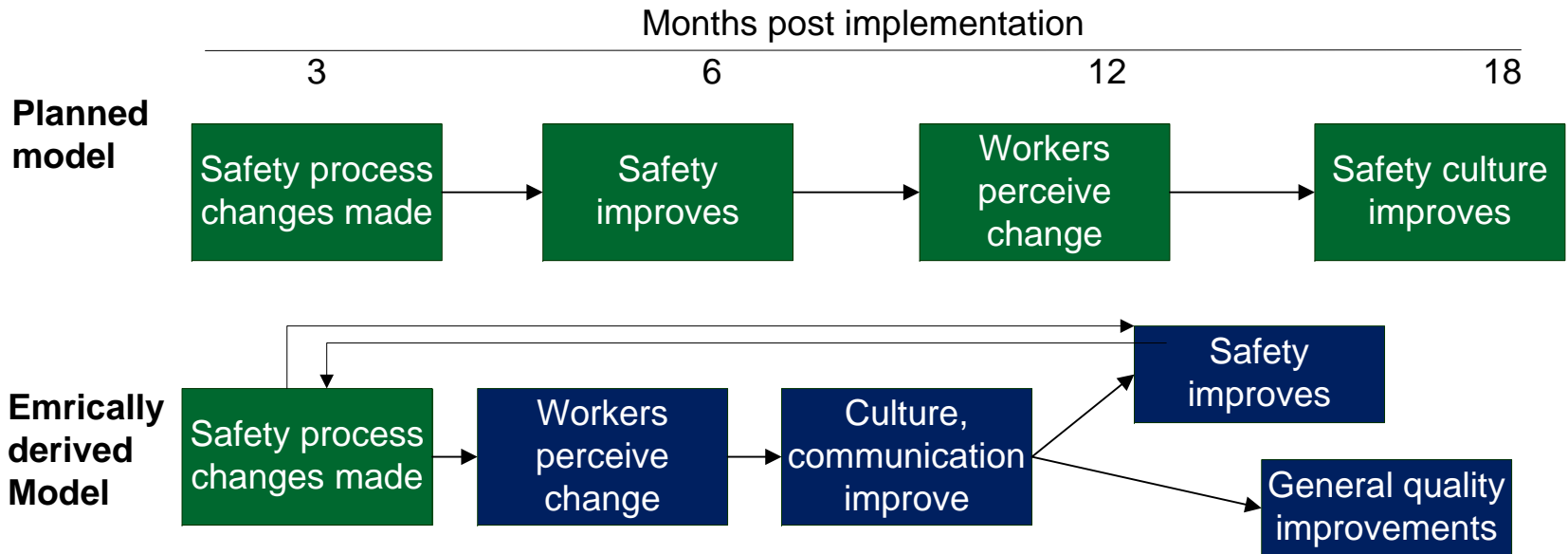
2.1

2.2 Implications of model form for unintended consequences

- Why are models always wrong?
- What are the ways in which models can be wrong?
- Why do they result in unintended consequences?
- How would you rate your model?

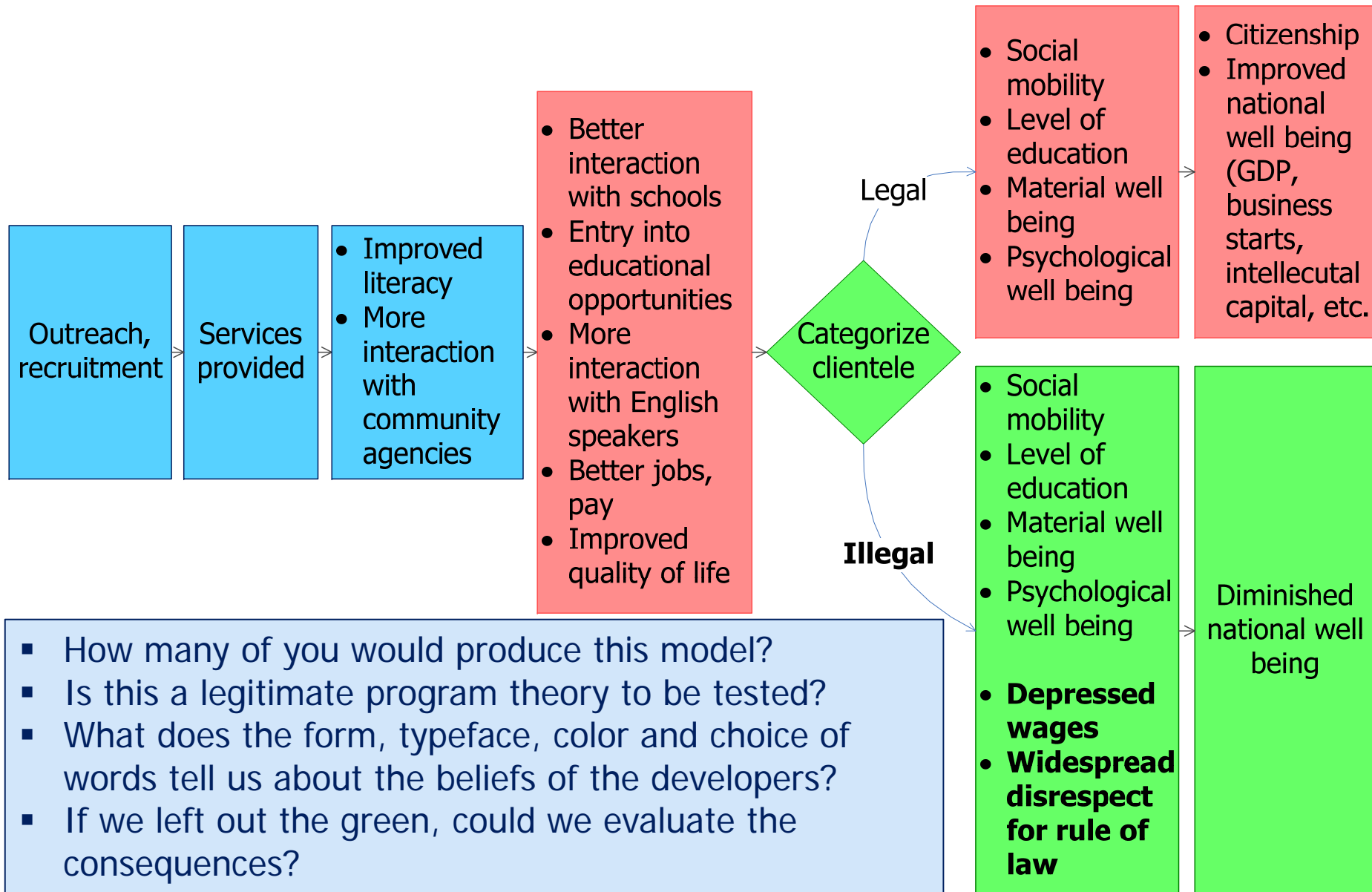
2.3

Different kinds of unexpected outcomes can combine

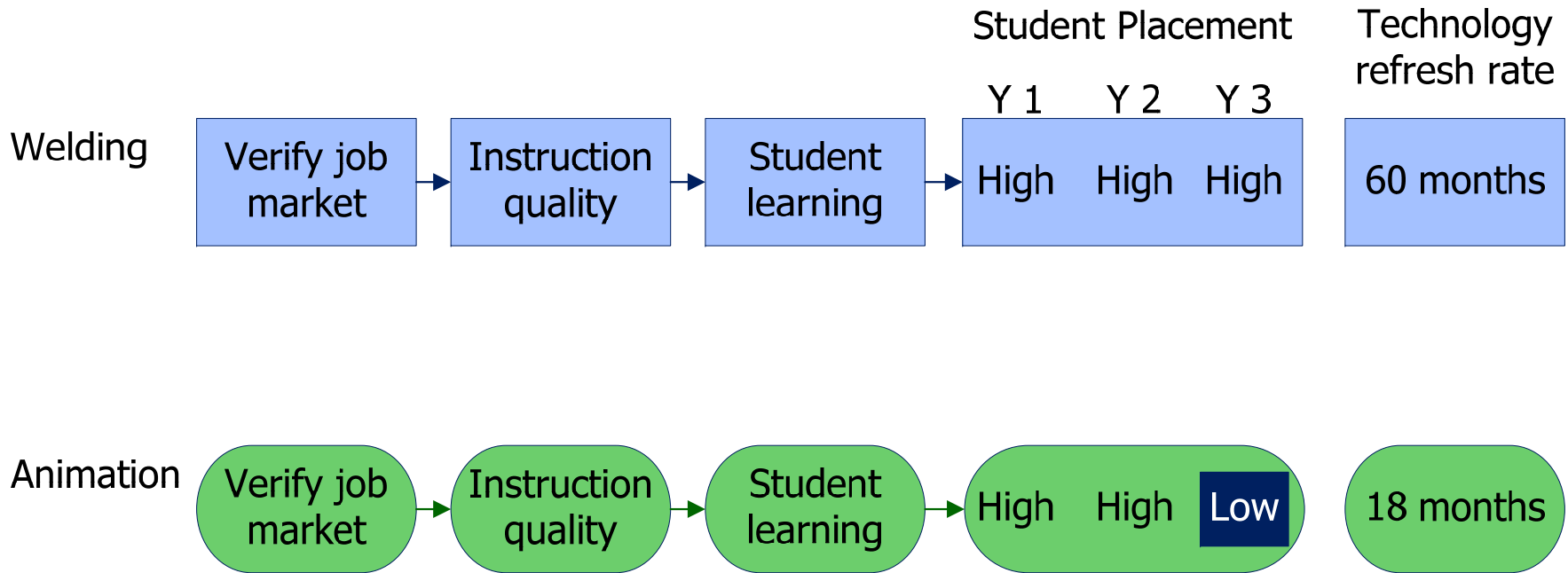


<ul style="list-style-type: none"> ▪ Timing ▪ Sequence ▪ Outcomes ▪ Unexpected feedback ▪ Causal paths 	<p>Impact starts earlier than expected</p> <p>Culture change before safety change</p> <p>Better communication leads to quality improvements</p> <p>Improved safety provides impetus for further safety change</p> <p>Culture to safety because of safety process change, <u>and also</u> more open communication between labor and management</p>
---	---

Ideology can impose blinders: Immigrant education

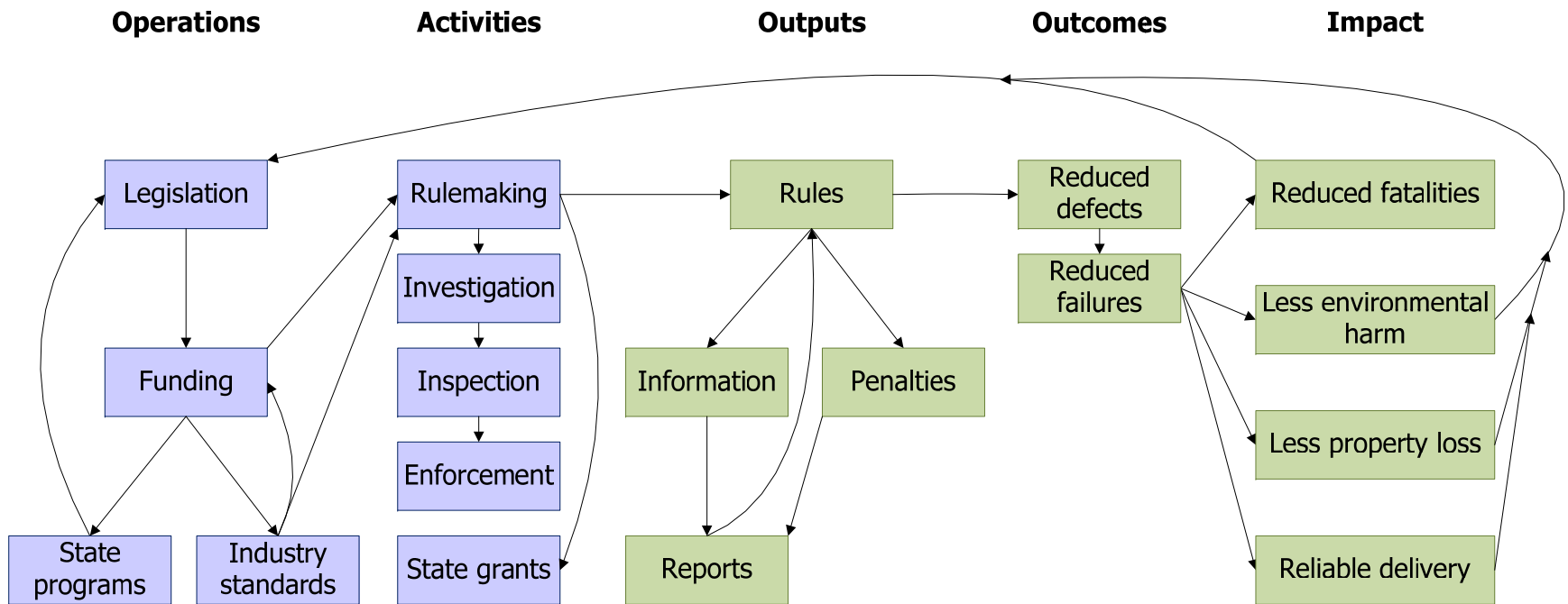


Logical elements can be omitted: Refresh rate missing from animation training



Over-specification. Everything is a hypothesis. We expect a lot more than we should.

Do we believe this....



Or...

This....

If stuff happens here



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



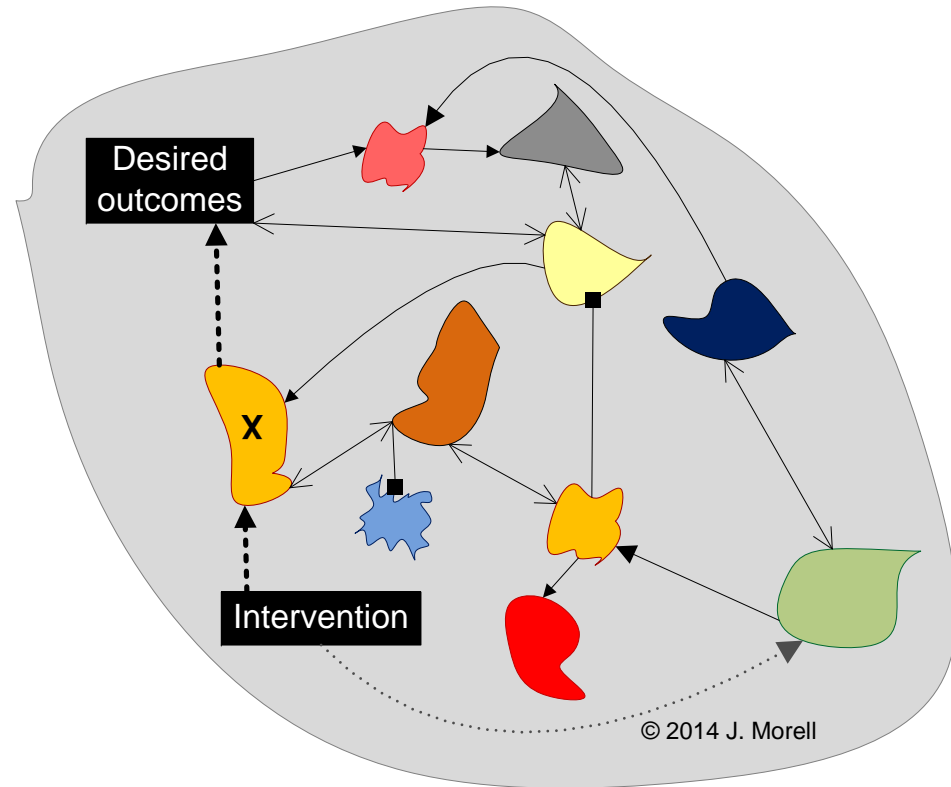
Stuff will happen here

Complex Systems: Do we really think we can do a surgical evaluation of X?

- Because a deterministic model cannot fully specify an open system, logic models are always incomplete approximations.

- Error potential increases with:
 - Length of causal chains
 - Number of feedback loops
 - Network richness (nodes:edges)
 - Accuracy of assumptions
 - Program's departure from previous known settings and solutions

- Small perturbation can often cause major change



A model might not include plausible alternate theories

Original program theory

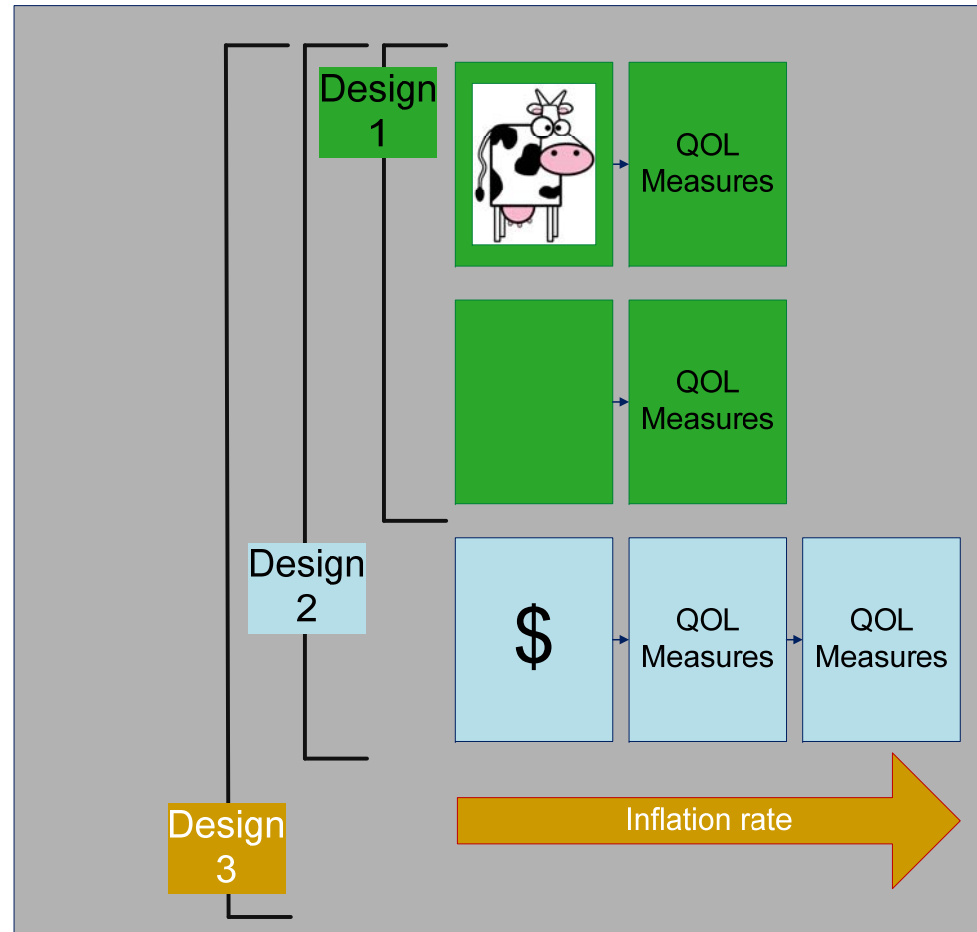
- Provide farming resources
- QOL improves in near term
- Longer term QOL change not different enough to be worth measuring.

Program theory #2:

- \$ to allow individual choice is better than assuming what people need.
- QOL may change in the long term.

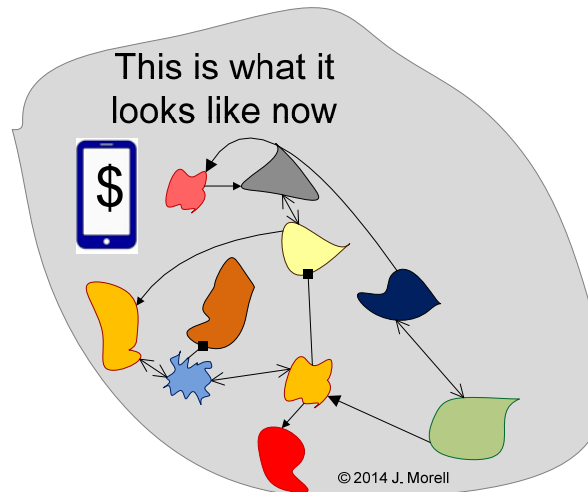
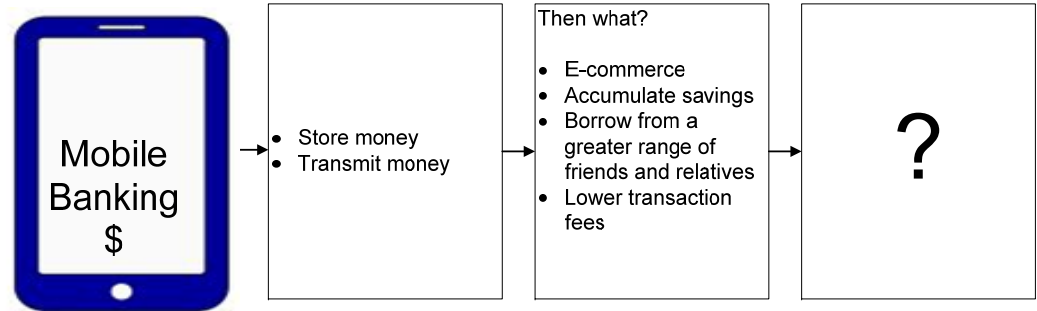
Program theory #3:

- #2 may work, but may have an undesirable consequence
- Too much money injected into economy will spur inflation



Sometimes no useful model is possible

- Models imply some ability to predict future events.
- Sometimes this may be true only at such a high level as to be useless



Both version of the model tell the same story. Which tells it better?

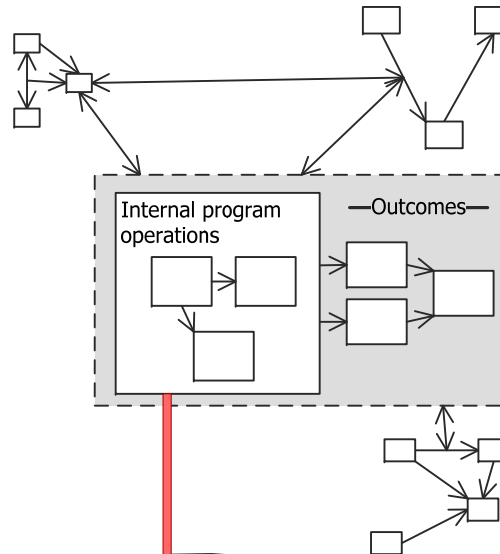
Because planners are constrained, so too are evaluators

- Any single organization has limited money, political capital, human capital, authority and power
- Narrow windows of opportunity
- Competition requires bold claims
- Resource owners have parochial interests
- Design expertise limited
- Collaboration across agency boundaries is very difficult/

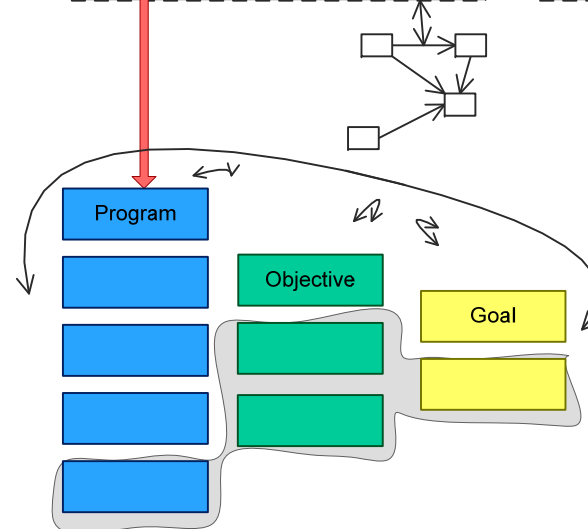
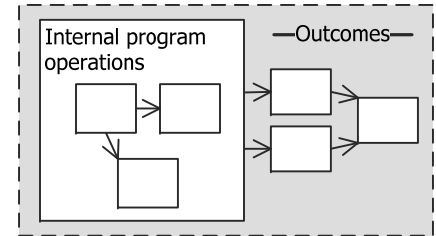
- Short term success is rewarded
- Partial solutions can accrue to major success over time
- Pursuing limited success with limited resources is justifiable.

- Result
- Narrow programs
- Simple program theories
- Small set of outcomes

If people are smart enough to know that the world looks like this



Why are they forced to design programs like this?



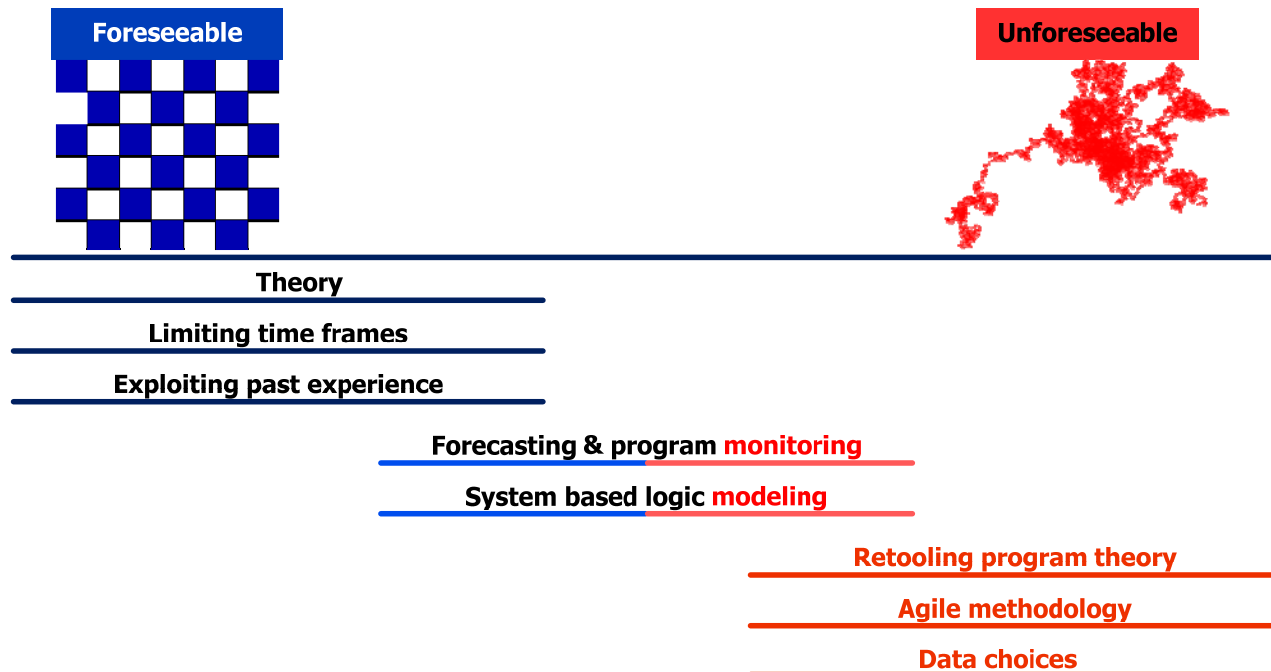
Part 2: Core of workshop

2.1

2.2

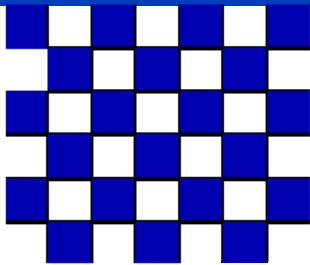
2.3 Designs for dealing with unintended consequences

- Now we know about models, their strengths, their weaknesses and the implications for unexpected outcomes.
- What can we do about it?
- Every single idea I am going to discuss will be familiar to you.
- What's new is how to apply them systematically
- There is no good overall solution but we can chip away at it.



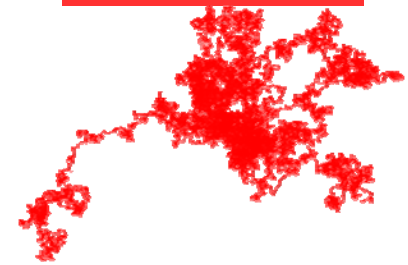
These methods are most useful early in evaluation life cycle

Foreseeable



- Get lucky
- Knowledge from stakeholders
- Good program theory
- Use research literature
- Use experts

Unforeseeable



- Complex system behavior makes prediction impossible no matter how clever we are.
PS – do not assume that complex systems are always unpredictable!

Theory

Limiting time frames

Exploiting past experience

Theory and Research

Planned

Program to implement a new, proven best practice

- 70% uses → much better results
- 10% uses → no change
- 20% uses → little worse

Use of best practice increases

Better outcomes

Findings

Program to implement a new, proven best practice

- 70% uses → much better results
- 10% uses → no change
- 20% uses → little worse

Minimal use

Unchanged outcomes

- Program designers did not account for research and theory about loss perceived as more aversive than gain is rewarding.
- If by chance a few failures cluster, innovation will slow.

Capitalizing on what we already know

Few programs are unique. There is always something to learn. Some examples:

- | | |
|--|---|
| ■ What happens to programs like mine in similar circumstances? | How do needle exchange and healthy eating programs fare at election time? |
| ■ What do we know about how programs like mine work? | Do threatening public service announcements encourage diabetics to monitor their blood sugar and control what they eat? |

Has anything like this ever happened to you?

Capitalizing on what we already know

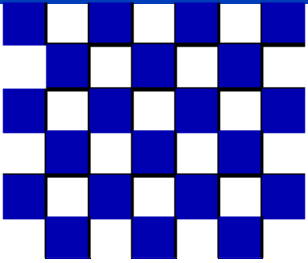
Few programs are unique. There is always something to learn. Some examples:

- What happens to programs like mine in similar circumstances? How do needle exchange and healthy eating programs fare at election time?
- What do we know about how programs like mine work? Do threatening public service announcements encourage diabetics to monitor their blood sugar and control what they eat?

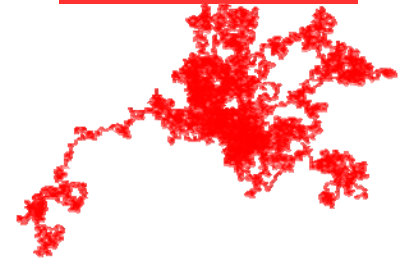
Has anything like this ever happened to you?

These methods are most useful for detecting leading indicators

Foreseeable



Unforeseeable



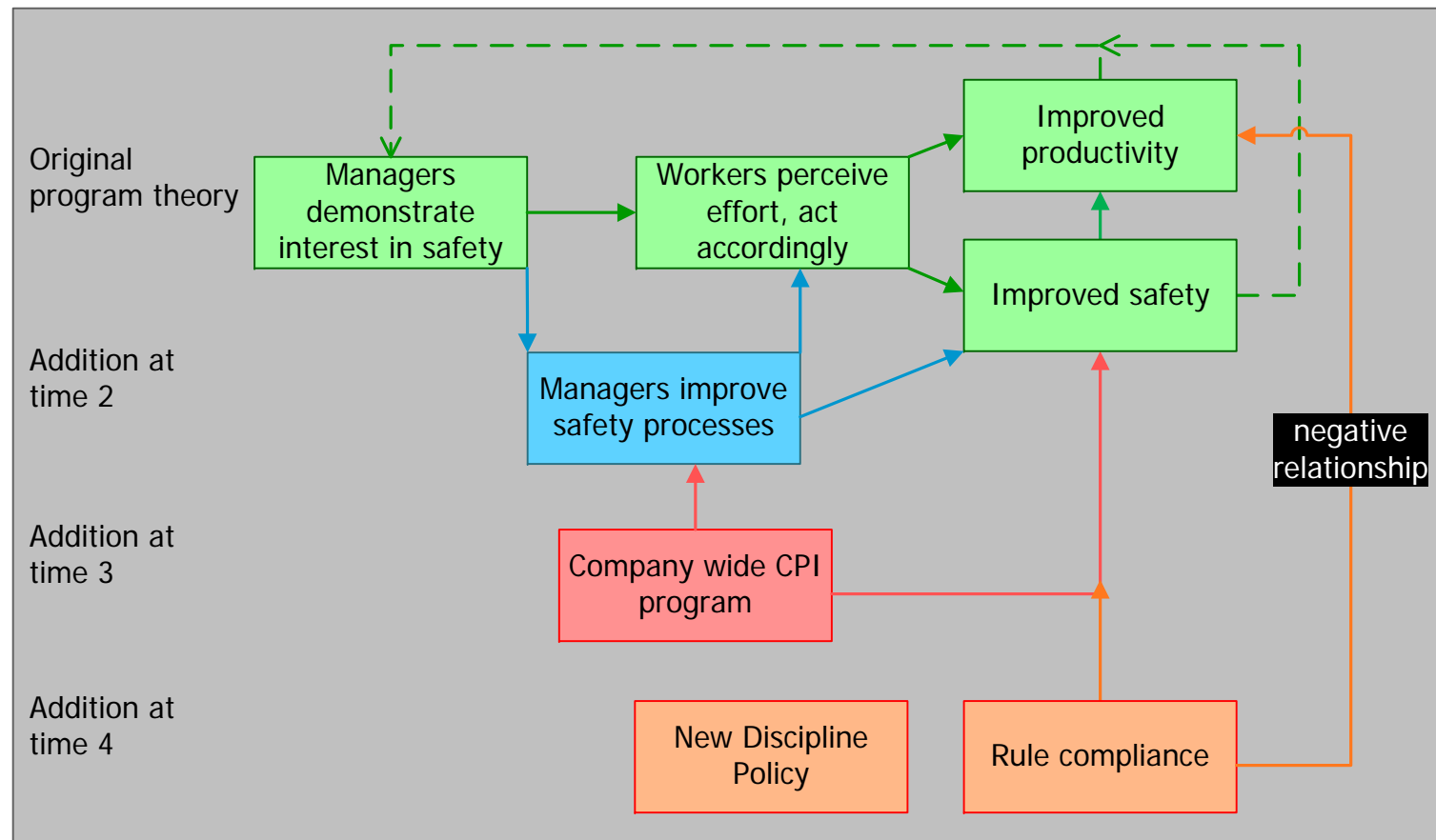
The trick is to do a little better than the Delphic oracle

Forecasting & program **monitoring**

System based logic **modeling**



Here is a program that changes over time: Outcomes, Methodology, Metrics



What would you do to keep informed of the changes?
If you knew, what would you do about it?

Use planning and monitoring techniques

What *is* the program? How has it evolved?

- Structure
- Function
- Positive and negative outcomes

In its present form, what assumptions underlie success?

- Which are critical?
- How robust or brittle?
- Indicators of failure?

What are the future states?

- Given where the program is now, what is the desired future?
- What are the likely futures?

What are the environmental conditions?

- Related processes and activities
- Needs of service recipients
- Needs of relevant stakeholders

Internal operations

- Intentions, plans of staff, and funders

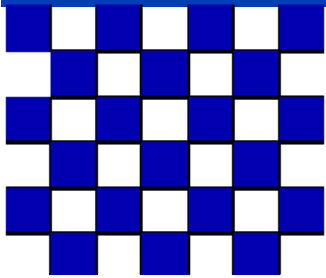
How to get the planning and monitoring information?

Be systematic and to stick to a schedule

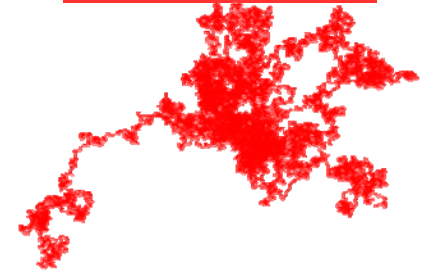
- Sampling:
 - Stakeholder groups – interviews
 - Archival data sources
 - All the environmental conditions that were identified
- Be sensitive that seemingly small changes may be significant

Agile Evaluation

Foreseeable



Unforeseeable



Data choices

Agile methodology

Retooling program theory

Data

Can the data be modified to meet new needs?

e.g.

- Validated scales vs. open ended questions
- Custom programming vs. standard lookup
- Structured teacher observations during class vs. casual assessment by visitors

Is gatekeeper approval needed?

e.g.

- OMB
- Air Force Survey Office
- Corporate VP

Are substitutes available without harming the intent of the evaluation?

e.g.

- Self report → clinical record
- Direct cost → total cost

Are substitutes practical?

- Collection burden increase
- Development cost to move to new methods
- Switching time relative to deadline for getting data
 - E.g. Clinical records vs. patient report

How to make an evaluation design agile?

- Partition
- Boundaries
- Dependencies
- Retool program theory
- Flexible vs. rigid design elements

Agile and brittle evaluation components: Example from a Safety Improvement Program

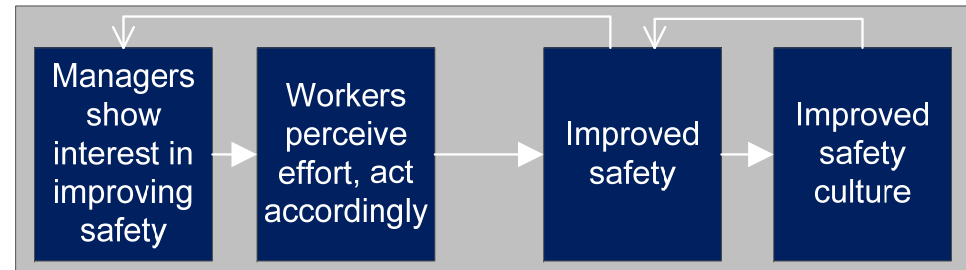
Methodology

- Two possible comparison groups
- Time series and cross sectional possibilities
- If any one comparison goes away others remain

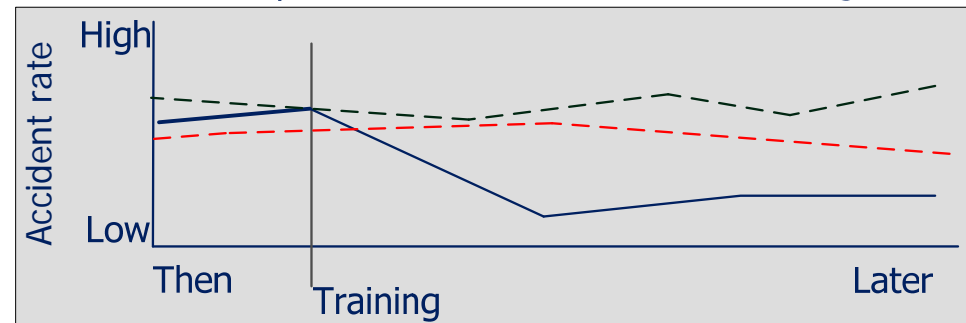
Data

- Develop, validate fixed-choice instruments for pre-post training assessments
- Interviews ½ way through training for course improvement
- Develop, validate fixed-choice instruments for pre-post training assessments
- Interviews ½ way through training for course improvement

Program theory



Data from an experimental / control time series design



What are the agile and brittle components?

Data: Formative	Data: Summative	Design	Implications for Agility
Validated instrument test training quality		2, beginning, end of training	<ul style="list-style-type: none"> Time, cost: difficult to change instrument Timing to training critical
Semi-structured questions: if/why managers change		1 half way through for course improvement.	<ul style="list-style-type: none"> Minimal effort to determine questions. Variation around midpoint OK.
	Validated safety culture scales	3, start, end, 6 months post	<ul style="list-style-type: none"> Time, \$, difficult to change instrument. First 2 timed to training. 3rd can move
	Interviews: why manager behavior affects safety	Keyed to occurrence of accidents.	<ul style="list-style-type: none"> Minimal time to determine questions. Synchronize with accidents
	Safety & accident stats	From company IT system	<ul style="list-style-type: none"> Available any time Not linked to training
		1- Control groups other parts of company	<ul style="list-style-type: none"> Difficult to implement. Considerable negotiation needed.
		2- Time series on accidents	<ul style="list-style-type: none"> Available from IT systems. Fallback if #1 disappears

Quick overview of material so far

- Evaluation (usually) benefits from having one or more models
- By their nature, models distort or miss real world conditions
- Because models are wrong, programs will often have unexpected outcomes
- There are ways to build evaluation that is capable of addressing unexpected outcomes.
- Methods are familiar. What is important is to be systematic and consider unintended consequences as one of the many factors that go into planning an evaluation
- There is no magic bullet, familiar chip away at the problem and end up much better off.

Optional choices

- 1 Jointly optimizing information density and readability when constructing models
- 2 Dependencies among models, metrics, and methodologies
- 3 Tactics for working with stakeholders

Color characteristics make a difference

Modality makes a big difference in color

Computer screen

Projection monitor

Use

Screen set to	Same color in print reads as
• Red 30	• Red 0
• Green 255	• Green 128
• Blue 131	• Blue 131

Read me	Read me	Read me
Read me	Read me	Read me

Color saturation can assure that differences show in B&W

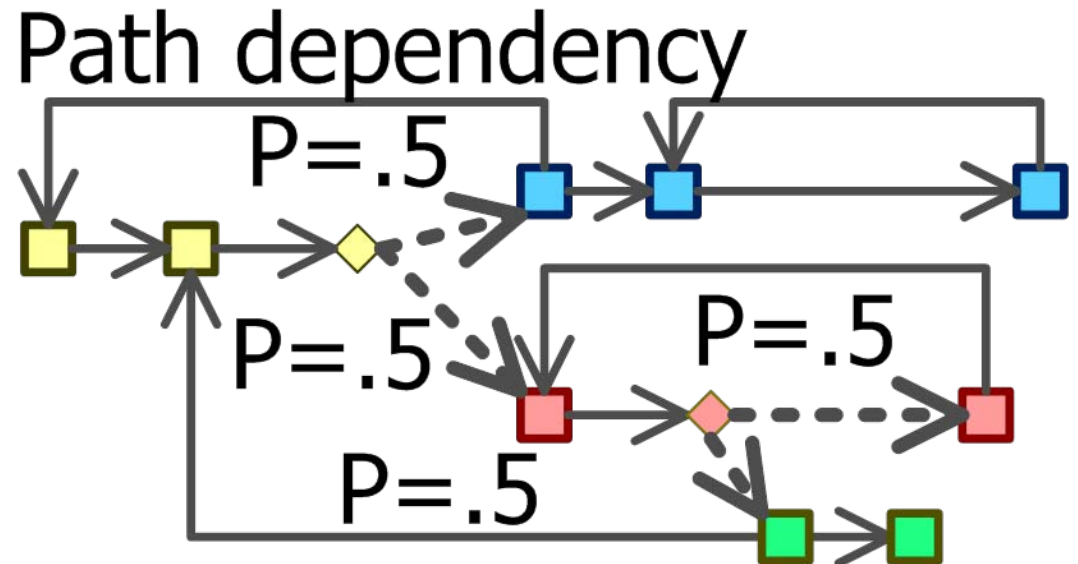
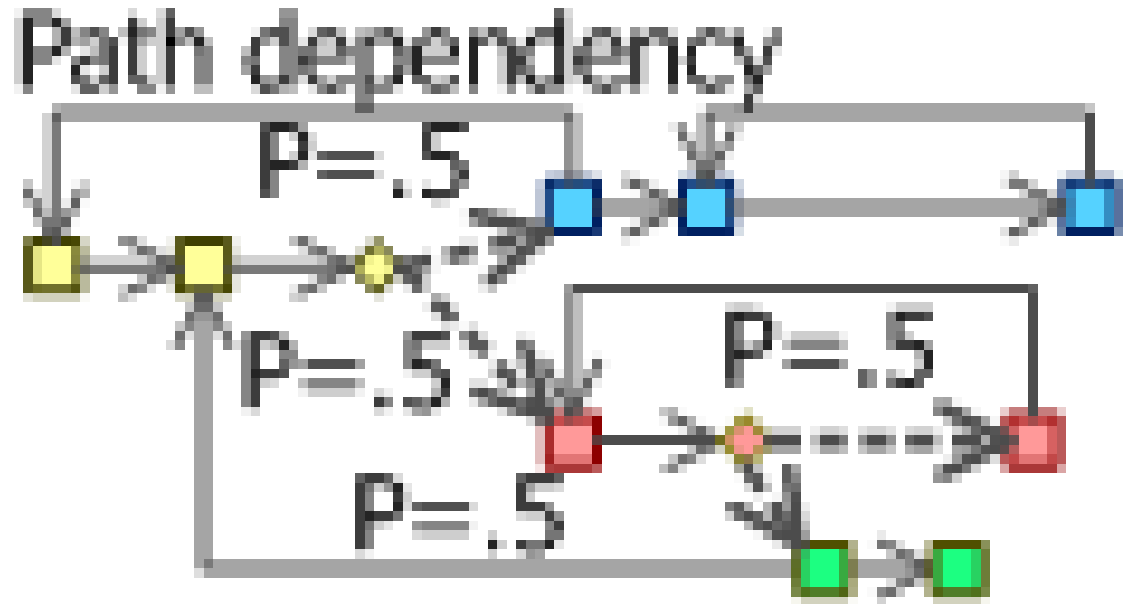
If screen color gets too dark, text is unreadable

File formats matter if you want to print large scale

1 x 2 original as a
bitmap

Use

1 x 2 original as a
vector graphic



Type characteristics make a difference

- 11 point
- Serif
- 0 line spacing
- Black lines

Use

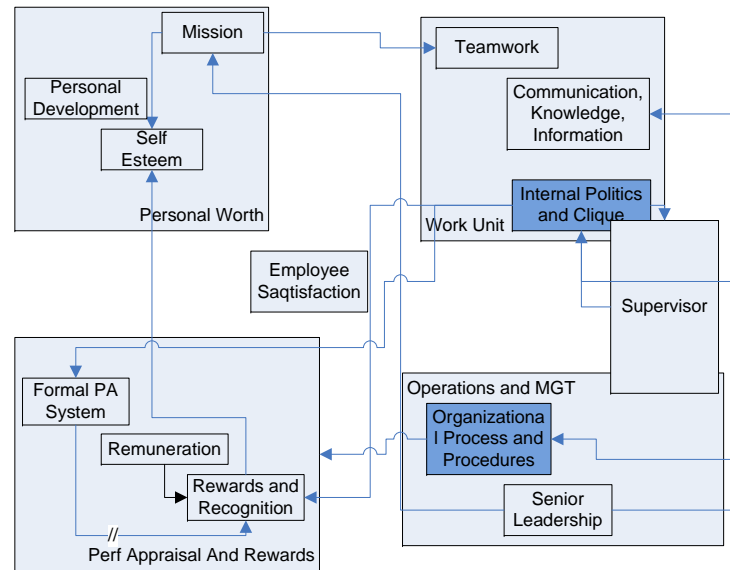
Operations	Activities	Outputs	Outcomes	Impact
Legislation	Rulemaking	Rules	Reduced defects	Reduced fatalities
Funding	Inspection	Reports	Reduced failures	Reduced industries
Industry	Enforcement	Penalties	Limited	Less environmental
Industry	Investigation	Information	propagation	harm
standards	State grants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

- 11 point
- Sans serif
- 2 point line spacing
- Gray lines

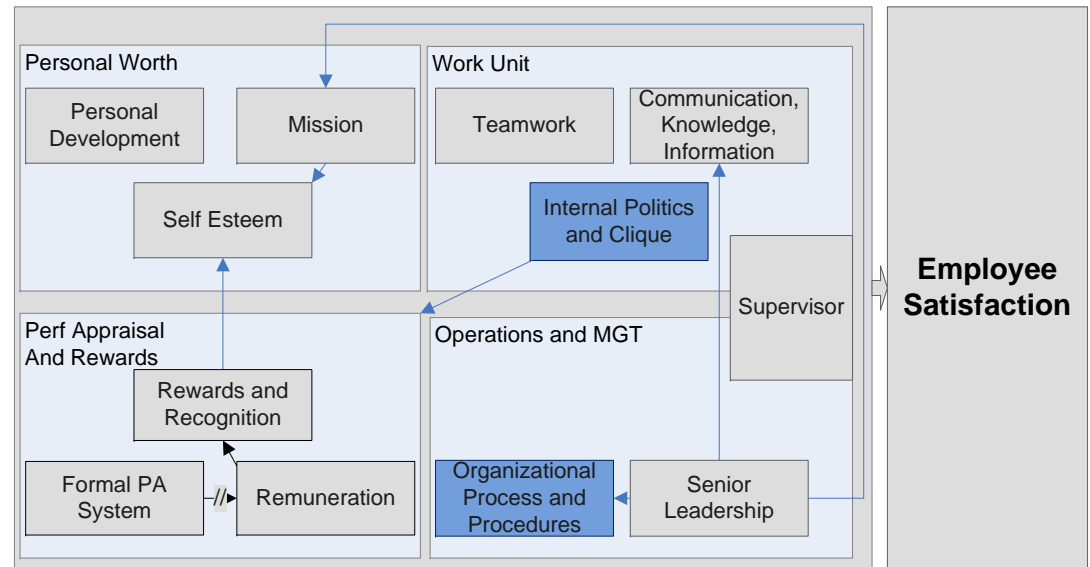
Operations	Activities	Outputs	Outcomes	Impact
Legislation	Rulemaking	Rules	Reduced defects	Reduced fatalities
Funding	Inspection	Reports	Reduced failures	Reduced industries
Industry	Enforcement	Penalties	Limited	Less environmental harm
Industry	Investigation	Information	propagation	Less property loss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

Two versions, two formats, same logic

Draft 1: deliberately done quickly to capture the logic



Draft 2: cleaned up for presentation



Optional choice #2

Dependencies among models, metrics, and methodologies

Optional choice #3: Tactics for working with stakeholders

Working with stakeholders

- Appreciate people's mixed motives for having logic models
- Besides stakeholders, use other inputs.
- Logic models are useful but not sufficient for knowledge transfer
- Respect what you know and stakeholders don't
- Prepare for unpleasant realities.
- Tactics for working with stakeholders
- Choose the right people and variety of people to work with.
- Get people to question assumptions
- Manage revisions
- Software choices depend on requirements

Appreciate people's mixed motives for having a logic model

Informed decision making

- Process
- Outcome
- Sustainability

Planning

- Especially true in the early stages of the program life cycle
- Working with evaluators to determine program theory, hidden assumptions, critical activities.
- Might be called “evaluation” but it’s really a planning exercise.

Advocacy

- Act of evaluation and/or findings will help keep my program going (even if I have to be selective and distort findings.)
- The fact that something called “evaluation” is being done implies a foundation of rational decision making that shields (hides?) advocacy from scrutiny.
- Often evaluators are not aware of the mix of modes they are operating in
- Not getting into a debate about legitimacy but lack of awareness can lead to trouble

Sources of input to logic model

Source	Strength	Weakness
Stakeholders	<ul style="list-style-type: none"> ▪ Deep appreciation of context ▪ Knowledge of program detail ▪ Vested interest in participation ▪ Sets groundwork for evaluation implementation 	<ul style="list-style-type: none"> ▪ Lack of perspective, may have strong + or – feelings ▪ Vested interest ▪ Not likely to have insight from comparable efforts ▪ Not likely to have insight from research literature
Critics	<ul style="list-style-type: none"> ▪ More complete / balanced model ▪ Alternate program theories 	<ul style="list-style-type: none"> ▪ Hard to recruit ▪ Those who are paying you might resist
Evaluation team	<ul style="list-style-type: none"> ▪ Experience with other programs ▪ Sensitivity to implications for methodology 	<ul style="list-style-type: none"> ▪ Lack of domain knowledge
Non-stakeholders familiar with similar programs, & research literature	<ul style="list-style-type: none"> ▪ Objective ▪ Knowledge not known to stakeholders 	<ul style="list-style-type: none"> ▪ Blind to context and specifics

Respect what you know and stakeholders don't, or are likely to forget

- Enthusiastic stakeholders can get carried away. The evaluation really does have a
 - Scope
 - Budget
 - Purpose
- Every element and relationship in a model is a hypothesis
 - Hypotheses can be wrong
 - Error piles up
 - Level of detail scope should reflect what we know
- Evaluation is more than just a logic model
 - Metrics
 - Methodology
 - Knowledge use plans and procedures

Appreciate unpleasant realities as you go in

- Because many logic models have a component of “advocacy” tension will lurk.
- There will always be resistance to including negative consequences no matter how integral they may be to achieving desirable outcomes.
- Types of negative outcomes to watch for:
 - Opportunity costs
 - Conflicts with other activities, systems, programs, etc.
 - Perverse effects, e.g. education for girls leads to social displacement
- Consensus may not be possible or needed, but people may not agree on which
- “Illusory agreement” is a constant possibility

Tactics for working with stakeholders

- Begin with a small group who already knows what a logic model is
 - Work out model to just below a very high level
 - Use draft to get feedback from a wider circle of stakeholders and experts
- Draw a rough model and send it off for feedback and approval.
 - Can be useful for mid-term corrections or to deal with unanticipated developments
 - Requires a good working relationship with stakeholders
- Chat about the program
 - Begin to sketch the logic they are verbalizing or implying.
 - Put burden on yourself – “This is what I understand you are telling me about the program. Did I get it right?”
- Depending on people and their experience with logic models it may be a good idea to begin with a large group

Group process choices for logic model development

	1:1 – Evaluator to Respondent	1: Many – Group Meeting
Face to face		
Phone, video, Internet		

Considerations for choice of tactics

- Time pressure
- Need for consensus vs. advice
- Potential for conflict among stakeholders
- Working relationships among group members
- Opportunity for multiple rounds of deliberation
- Power / status differential among stakeholders
- Degree of common understanding among group members

Here is an approach I like

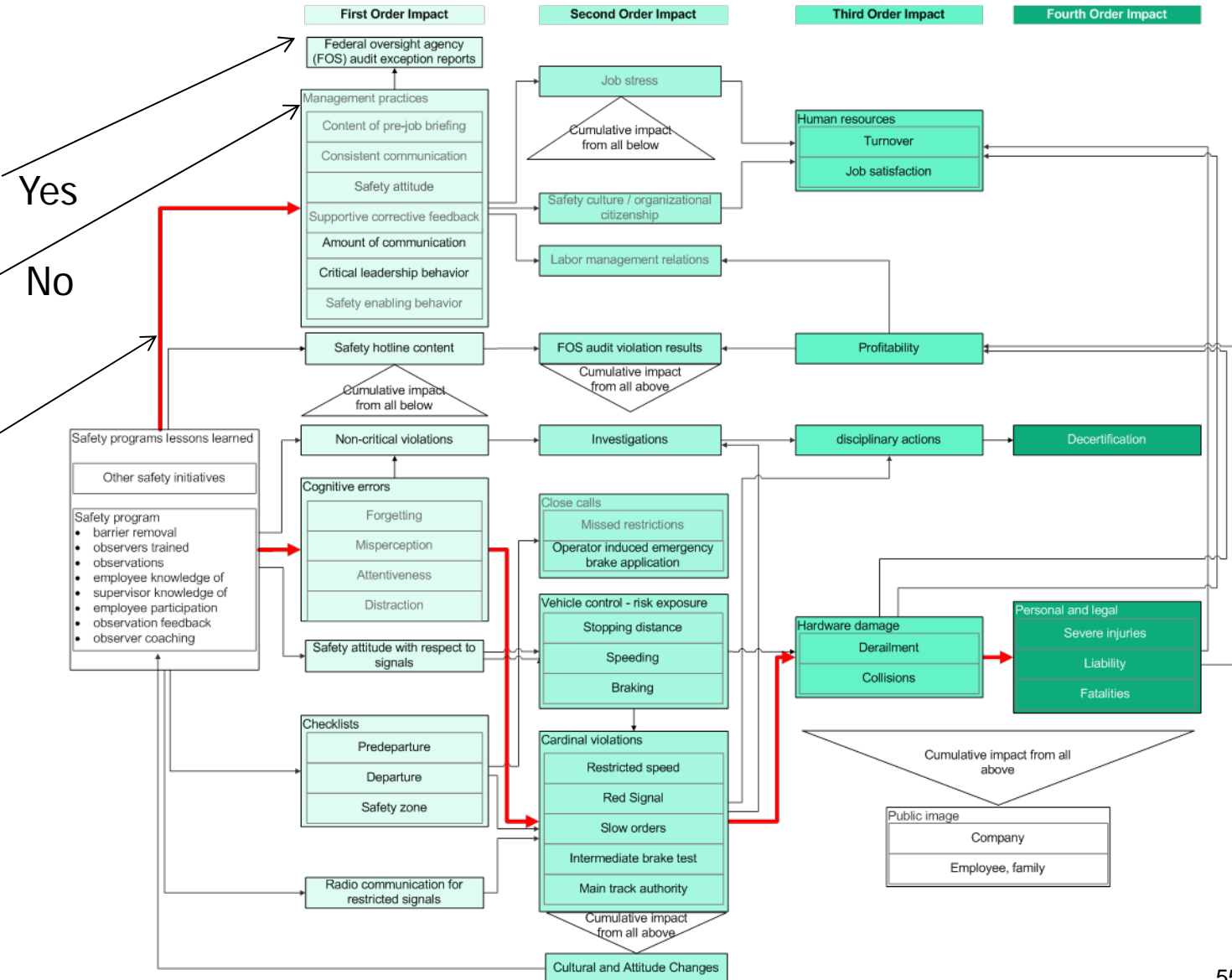
Complete but Overly Complicated Model

Step 1:
Build complete model

Step 2:
Can we measure all important elements?

Step 3:
How far can we get with what we can measure?

Remember to critique the visual clutter!

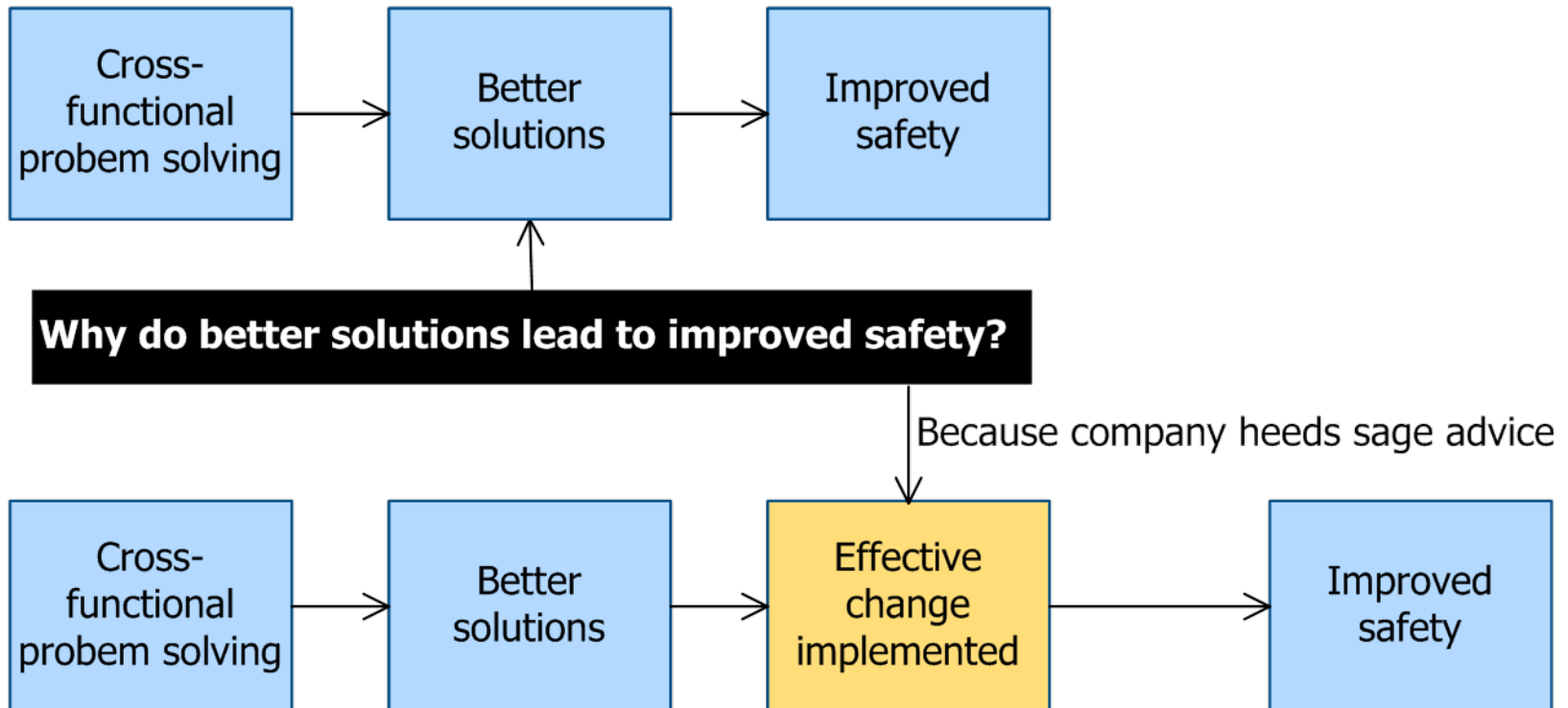


Choosing group members

- Who can influence program operations?
 - Implementation
 - Outcome
 - Sustainability
- Who can influence the evaluation?
 - Access to data
 - Integrity of the design
- Who can make use of the evaluation findings?
 - Same program in same setting
 - Same program in a wider range of settings
 - Other programs with similar objectives
- Values
 - Who has a right to influence what the evaluation measures?
- Operational
 - Given constraints of time and money, who should be involved?
 - Will candidates put in the work?
- Some stakeholders can be sampled, e.g. teachers,
- Some stakeholders are unique, e.g. minister of education

Get people to question assumptions

- Improves evaluation
 - Design and measurement
 - Customer expectations
- Depending on where the evaluation comes in program life cycle, may also improve program design



Get people to question assumptions

- 5 whys
- What does the research say?
- What do non-involved experts say?
- Push people to consider the program's environment/neighboring systems
 - What are they?
 - What do they do when the program starts to function or starts to have an impact?
- Does the level of detail and specificity only produce noise?
- What comes from an assumption based planning exercise?
- Half way through model development stop and ask:
 - Before we go any further let's look at what we have constructed and why. Do we really believe it?