Questions for proponents of theory-driven evaluation
In Appreciation

• Focus on useful/’good’ theory of change
• Learning is a process not just a product
• Empirical basis for a theory of change
• A richer vocabulary around theories of change
  – Capabilities, opportunities, motivation
  – Unpacking the black-box of interventions
• A science of causation
  – Assumptions, risks, mechanisms
  – Contribution analysis
• Richer description of how programs can work
• Knowledge translation description
• Highlights the complexity of the intervention
• Steps towards precision in theorizing
Guiding Questions

A Story from the World Health Organization

Evaluation of Dancing with Parkinson’s

Pushing theory-driven evaluations: Some questions

Summary
Some questions

- How does one test a theory of change?
- How does one test a link of the theory of change?
- How does one test key assumptions?
- What works for whom, under what contexts?
- How does one test if the assumptions will hold across contexts?

- Representation challenges;
- From representation to testing
- Understanding diversity
A STORY FROM THE WORLD HEALTH ORGANIZATION: Of Mice and Elephant
## LESSONS

<table>
<thead>
<tr>
<th>The importance of understanding the nature of connections</th>
<th>Issues of power</th>
</tr>
</thead>
<tbody>
<tr>
<td>The attribution/contribution problem</td>
<td>The inequity problem</td>
</tr>
</tbody>
</table>
AN EXAMPLE OF A REALIST EVALUATION: EVALUATION OF DANCING WITH PARKINSON’S
Parkinson’s disease (PD)

- PD is the second most common neurodegenerative disorder after Alzheimer’s disease.
- Causes a variety of physical symptoms: rigidity, tremor, slowness of movement, impaired balance, as well as “freezing” or the inability to initiate movement.
- Other non-motor symptoms include: depression, anxiety, cognitive impairment, problems swallowing, chewing, and speaking, masked facial expressions, fatigue, and sleep disturbances.
- These symptoms invariably lead to increased social isolation and lack of engagement with the world.
Why Dance with PD?

• Dance offers similar benefits as other movement forms: focus on balance, strength, and flexibility
• But **artistic elements of dance add value that other types of exercise do not**
• For Example:
  – **Music**: rhythmic and emotional stimuli help initiate movement more easily for people with PD
  – **Imagery**: assists in fostering movement and self-expression
  – **Social**: dance is learned, practiced, and enjoyed with others
  – **Cognitive**: new movement sequences, improvisation exercises engage and stimulate the mind
<table>
<thead>
<tr>
<th>Context</th>
<th>Mechanism</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals who danced regularly before symptoms of Parkinson’s disease</td>
<td>Dancing at a lowered capacity reminds them of how they have deteriorated</td>
<td>They avoid attending classes and therefore don’t get health benefits of participation</td>
</tr>
<tr>
<td>Individuals who have never danced regularly before DWP</td>
<td>While dancing they experience the feeling of freedom from their symptoms and relief from stress and depression</td>
<td>They feel better, have an incentive to attend classes regularly, and therefore gain health benefits</td>
</tr>
<tr>
<td>Individuals who live alone</td>
<td>Social connection with other participants, teachers and volunteers, physical contact, and creative and emotional expression</td>
<td>Deepened social connections, enhanced artistic expression, improved quality of life</td>
</tr>
<tr>
<td>Individuals with care-partner or caregiver</td>
<td>1+ hour that person with PD is engaged in dancing in a safe environment with others with PD</td>
<td>Enhanced care-partner/caregiver relief and support</td>
</tr>
</tbody>
</table>
Data: Focus on short term variations, linkages between DWP and outcomes, understanding of other drivers, temporal context, between and within-individual variations

1. Reversed Pre- and Post-survey design

- **End of Spring Classes**
  - Last week of June

- **Summer Break**
  - July to mid-Sept

- **Start of Autumn Classes**
  - Mid-September

2. Daily Journals

3. New Client Baseline Survey + Qualitative Longitudinal Research

- **Start of Autumn Classes**
  - From mid-September

- **Follow up**

**15 Weeks** June 22 - October 4
How are you doing today?

WEDNESDAY, JUNE 24       Journal entry time: __________

Please circle or check one:

Very Bad      Bad          Okay      Good      Very Good

Any reflections or comments are welcome below:

• An Example of a Single Individual:
  – Daily Entries: How are you doing today?
  – Responds “okay” 91% of the time, “good” 5% of the time, and “bad” 4% of the journal days
  – Rates the day as “okay” even on days with something to look forward to (e.g., dinner with friends, concerts, birthdays)
  – Notably, 3 of the five “good” days were DWP class days
GOOD
“Busy day! Attended my nephew’s graduation from law school, had an important business meeting in the PM, and attended a jazz concert in the evening.”

OK
“Had lunch with my best friend. Always cheers my spirit.”

OK
“Another full day at a conference at Harbourfront. Felt tremendous fatigue around 4pm.”

GOOD
“Good news from the oncologist - my PSA continues to be low.”

OK
“Looking forward to dinner with 3 of my cousins and my brother.”

OK
“My brother took me out to dinner on my birthday. Felt fatigued but otherwise ok.”

GOOD
“DWP begins today – looking forward.”

OK
“I went to afternoon concert at Koerner Hall with friends. Felt tired at the end.”

OK
“Long day - picked up by a friend who drove us to another friend who was having a party at her place in Caledon. Long, long day. Very fatigued.”

GOOD
“Had blood work done in the morning at Princess Margaret Hospital in preparation for my appointment next week with the radiation oncologist.”

OK
“Another full day at a conference at Harbourfront. Felt tremendous fatigue around 4pm.”

OK
“DWP begins today – looking forward.”
How has your energy level been in general over the past week?

<table>
<thead>
<tr>
<th>Very Low</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
<th>Very High</th>
</tr>
</thead>
</table>

How has your body felt in general over the past week?

<table>
<thead>
<tr>
<th>Very much worse than usual</th>
<th>Somewhat worse than usual</th>
<th>Usual</th>
<th>Somewhat better than usual</th>
<th>Much better than usual</th>
</tr>
</thead>
</table>

How would you describe your most frequent mood in the past week?

<table>
<thead>
<tr>
<th>Sad, Depressed</th>
<th>Worried, Anxious</th>
<th>Angry</th>
<th>Contented</th>
<th>Happy</th>
</tr>
</thead>
</table>

Other feeling/emotion:
SOME QUESTIONS FOR PROPONENTS OF THEORY-DRIVEN EVALUATION
1. A ‘good enough’ programme theory: How do we know that there is an initial theory of change? What is a good enough programme theory? How does such a ‘good’ theory aid implementation?

2. The dynamic expectations challenge: How does one build expectations of programme impacts? How does one understand the trajectory of impacts?

3. The knowledge translation problem: How does a programme build in knowledge of what should work for whom? How can knowledge of heterogeneous mechanisms be implemented?

4. The learning problem: What is the relationship between evaluation and learning?

5. The causality and complexity challenge: What are some of the challenges in making causal statements in very complex systems?

6. The methods problem: How can an application of design and methods help develop an ecology of evidence?
1. What is a good enough theory of change?

• “What are characteristics of a useful implementable theory? How does the theory of a complex intervention differ from a theory of a simple intervention?”

• “… in virtually every case, the theories of change that drove the initiatives and their evaluations were not really adequate to substantiate the connections between activities and outcomes.”

  – Auspos and Kubisch, 2004
• “Is it time to shift the balance in what we define as quality from an exclusive focus on empirical method (the extent to which authors have adhered to the accepted rules of controlled trials) to one that embraces theory (the extent to which a theoretical mechanism was explicitly defined and tested)?”

• The need for case studies on “what are useful theories of change? What makes a good theory of change?”
2) The Expectations Problem: The Anticipated Trajectory of Impacts

- “Fundamental to judging performance is having clear expectations of an anticipated trajectory of impact of programmes (Woolcock, 2009). It is not entirely clear how the knowledge of such a trajectory of impacts is arrived at, especially for complex interventions.”

- How does one develop a clear understanding of the anticipated trajectory of impacts?
• “Even a cursory reading of social theory, for example, would suggest that in fact the most likely shape of such projects’ functional form is a J-curve (that is, things get worse before they – hopefully, maybe – get better) or a step function (that is, long periods of stasis followed by a sudden rupture brought on by, say, an election or the reaching of a ‘tipping point’ in the adoption of a new fertilizer technology, in which prevailing norms and/or uptake by an influential local leader rapidly leads others to do likewise).”
“In short, the development profession strongly prefers to sell known, universal solutions with high, immediate and readily measurable impact rather than wrestle with ambiguous, context-specific problems that may not have (at least ex ante) a known or even knowable solution.”
3) The Knowledge Translation Problem

• How does a program build knowledge of what works for whom?

• “What programme mechanisms are needed for what contexts? The evaluation exploration in our view should not be restricted to the impact evaluation question (what works for whom?) but also should be focused on the knowledge translation question: what should work for whom?”
• How is such knowledge of heterogeneous mechanisms implemented in programs?

• How do we incorporate knowledge of such heterogeneity into our theories of change?
4) The Methods Problem: Developing an Ecology of Evidence

• A problem that also comes through reasonably clearly in the recent literature on evaluation methods is the recognition that ‘learning is not the monopoly of a single method’ (Rodrik, 2009).

• The lesson increasingly is on the types of evidence that are useful given the heterogeneous nature of the real world/complex systems.

  – For example Rodrik (2008: 5) makes this important point: “The ‘hard evidence’ from the randomized evaluation has to be supplemented with lots of soft evidence before it becomes usable.”
• For the most part the theory-driven evaluation field has not fully engaged with methods that can help explicate the programme theory more fully, nor has it clearly described how methods can be used to test linkages in the programme theory.
Some examples of methods

- **Network Analysis**  How does the programme operationalize networks of interventions?

- **Event Structure Analysis**  What is the impact of the event structure (for example, the interactions between partnering organisations) on outcomes over time?

- **Concept Mapping**  How do different groups of stakeholders conceptualise the intervention? Do these conceptualisations change over time?

- **System Dynamics**  What are the anticipated trajectories of the impacts of the overall policy intervention? How would changing key components of the policy intervention alter the expected trajectory of outcomes?
• **Observational Studies** (Propensity scoring techniques) In the absence of randomized designs, what alternative analytical techniques can help assess causal effects of interventions? What is the hidden bias in observational studies?

• **Methods of Developmental Trajectories** Does the same intervention have very different impacts over time for different groups of individuals (or communities)?

• **Respondent-Driven Sampling** How does the programme (and evaluation) reach individuals who are hard to reach and not found in a standard sampling frame?

• **Adaptive/Sequential Designs** How can a design help an intervention adapt/evolve from large complexity with many components to an intervention with a few ‘effective’ components?

• **Realist Synthesis** How does evidence synthesis help identify the mechanisms by which an intervention works?
Summary

- Improved language
- Heterogeneous Impacts: Connect Assumptions to Contexts
- Testable Assumptions
- Methodological Implications
- Prioritization
- The need for case studies
- Utility from the user perspective
- Generative causation
- Assumptions to Contexts