Designing Evaluations to Inform Action in New Settings

Calum Davey (and Inception-paper co-authors)

INTRO

"Use evidence!"

Case-specific
Same problem, many contexts
How, why, and where, as well as what 'works'

Aim: suggest ways to address the issue of learning more from evaluations

Interventions do not *have* effects; effects are afforded by the context

= they cannot 'work', only 'have worked'

Key idea: intervention-centred and context-centred thinking

Experiment: a semi-context-centred presentation

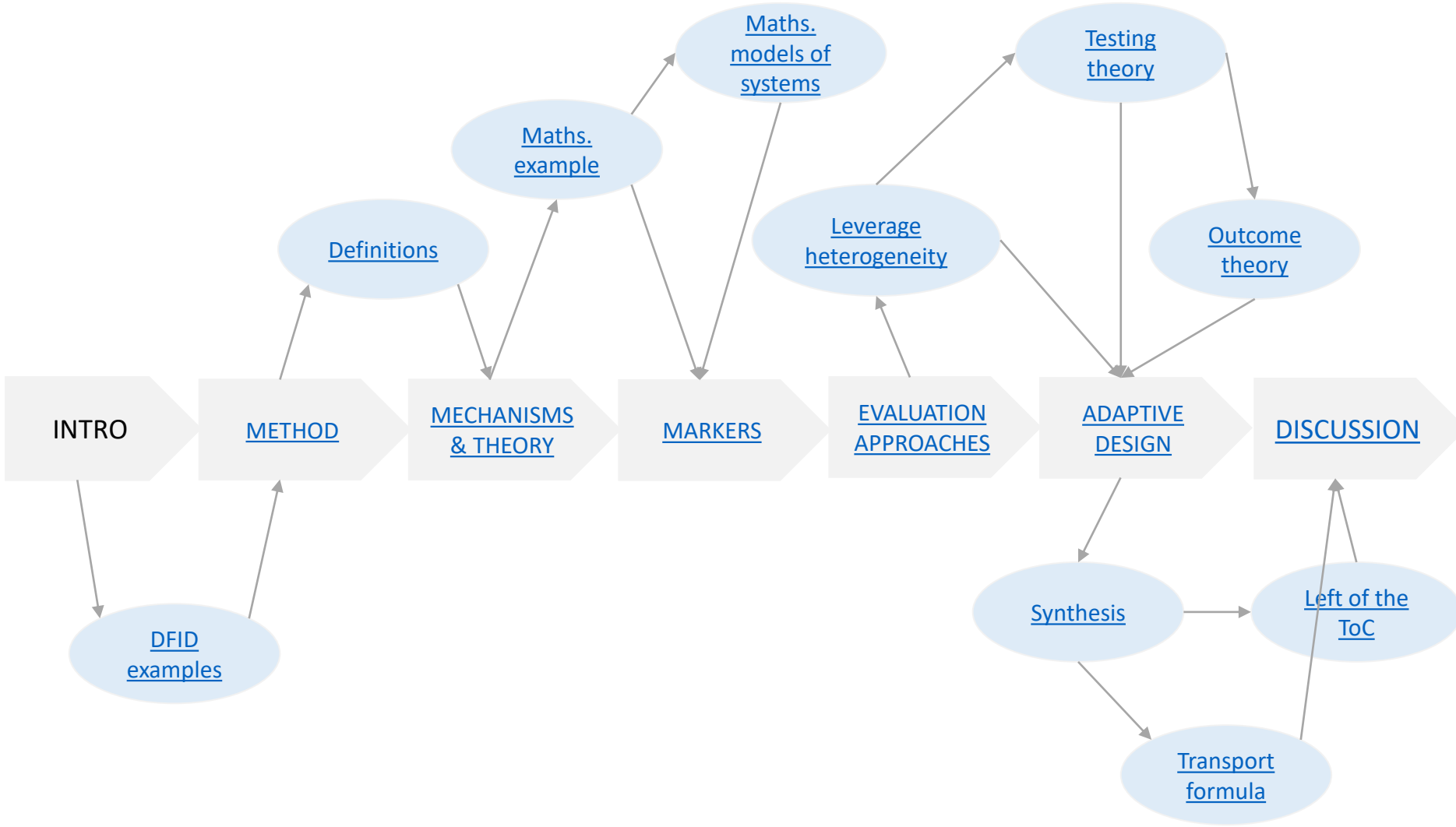
With a democratic process

But it won't *work*

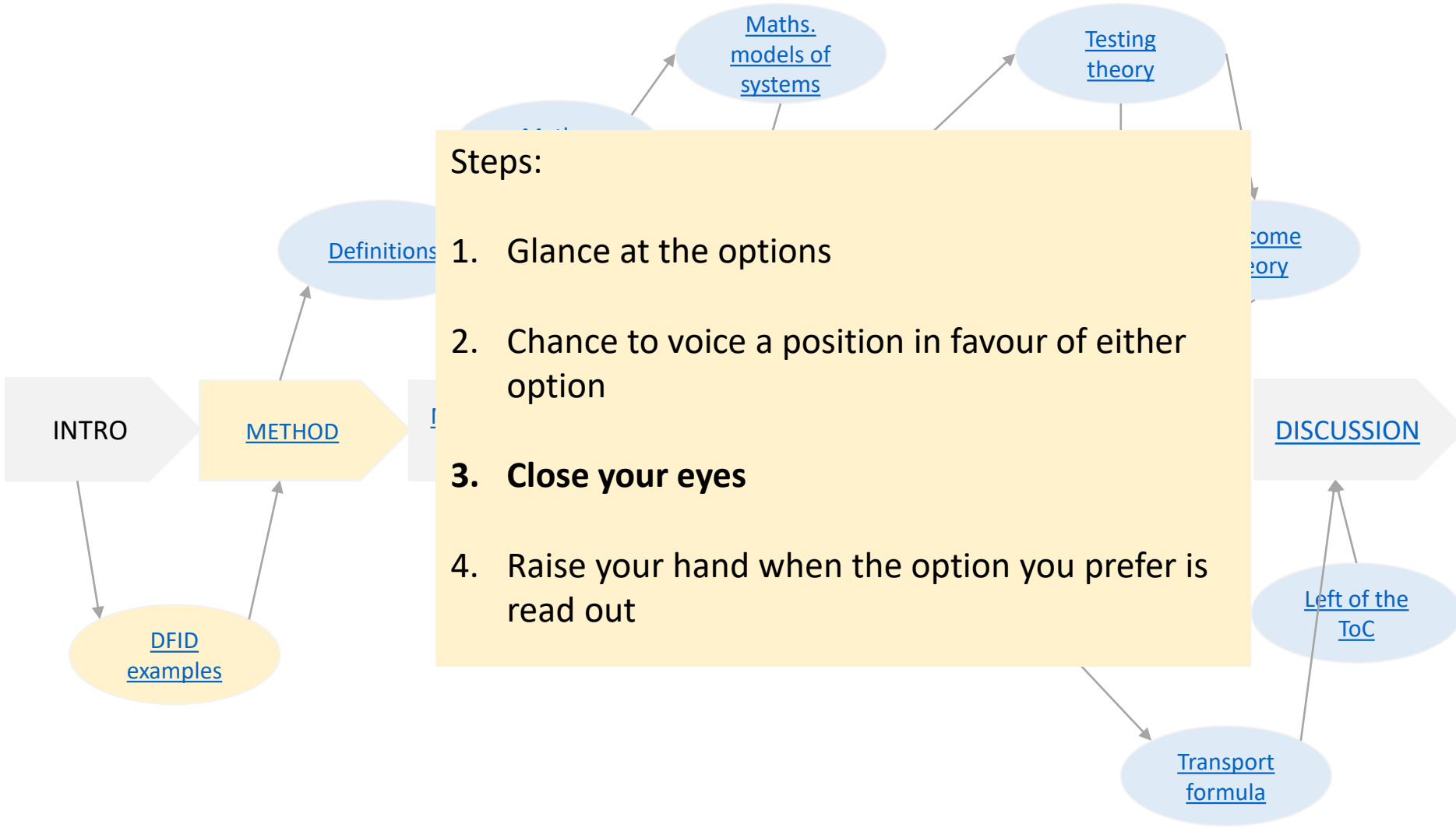
ROUTE MAP


INTRO
METHOD
MECHANISMS & THEORY
MARKERS
EVALUATION APPROACHES
ADAPTIVE DESIGN
DISCUSSION

ROUTE MAP


Maths. models of systems
Testing theory
Maths. example
Definitions
Leverage heterogeneity
Outcome theory
INTRO
METHOD
MECHANISMS & THEORY
MARKERS
EVALUATION APPROACHES
ADAPTIVE DESIGN
DISCUSSION
DFID examples
Synthesis
Left of the ToC
Transport formula

ROUTE MAP



Steps:

1. Glance at the options
2. Chance to voice a position in favour of either option
3. **Close your eyes**
4. Raise your hand when the option you prefer is read out

DFID Examples

How **best** can we learn from:

- Large-scale, multi-component initiatives to improve the education system in a single country
- Response to Ebola in West Africa for future outbreaks, outbreaks of other diseases, or more generally about how health promotion can be reconciled quickly with cultural norms and expectations?
- What can be learned from the peace-process in Northern Ireland that could be applicable in South Sudan?
- Mobile phone technology to change behaviours, both for future mobile-based interventions but also as a platform for understanding how habits can be changed efficiently?

METHODS

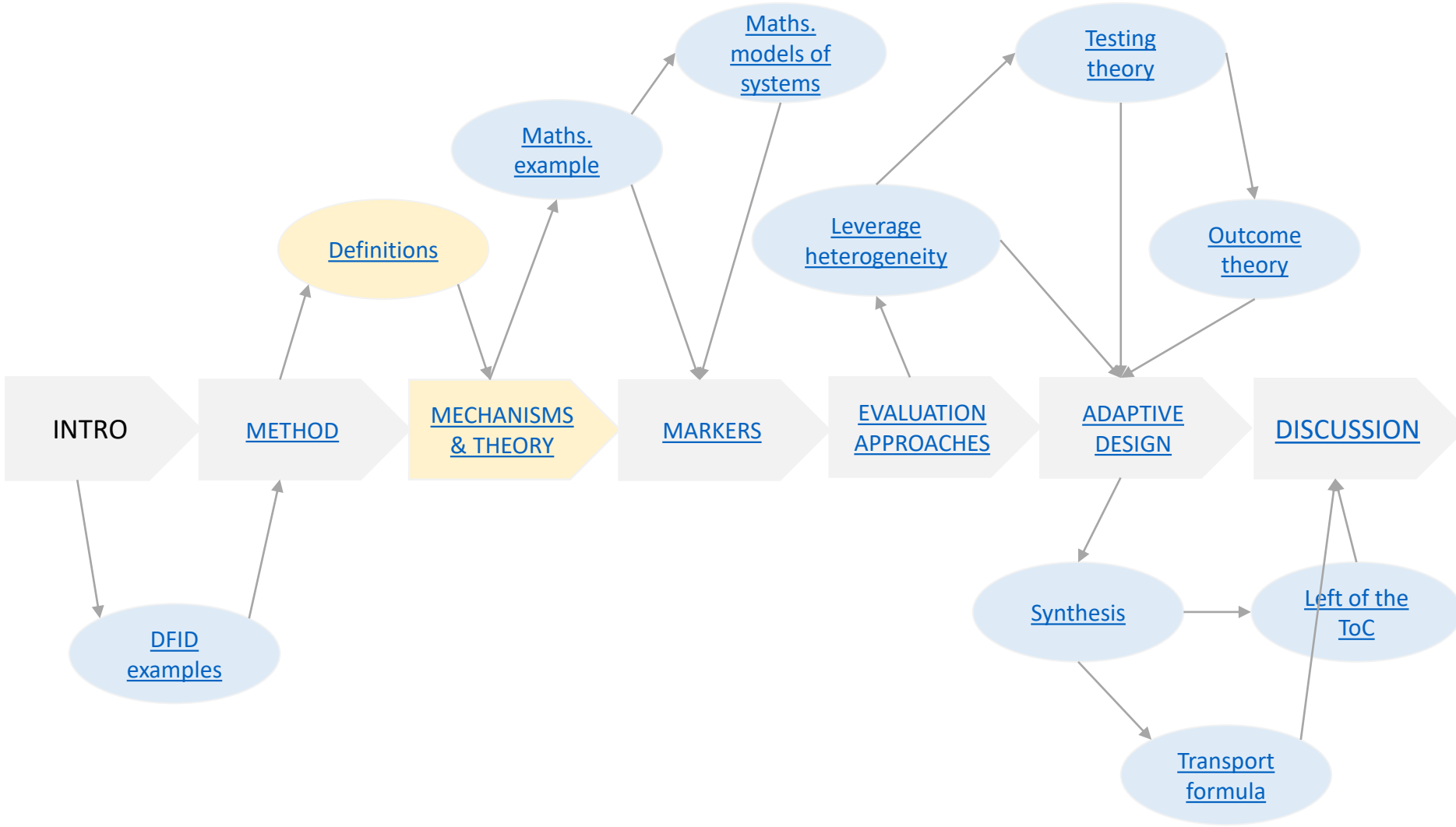
Co-authors	Discipline
Nancy Cartwright	Philosophy
Edoardo Masset	Economics
Macartan Humphreys	Political Science
Audrey Prost	Anthropology
David Gough	Evidence Synthesis
Sandy Oliver	Evidence Synthesis
Syreen Hassan	Public Health
James Hargreaves	Public Health
Chris Bonell	Public Health
Calum Davey	Public Health

- **Consultations**
- Centre for evaluation members (50)
- DFID Staff (1.5 hrs)
- Intellectual Leadership team (co-authors)

- **Review**
- Expert recommended papers (57)
- Backwards citation (84)
- Forwards citation (ongoing)

- **Writing workshop**

ROUTE MAP


Maths. models of systems
Testing theory
Maths. example
Definitions
Leverage heterogeneity
Outcome theory
INTRO
METHOD
MECHANISMS & THEORY
MARKERS
EVALUATION APPROACHES
ADAPTIVE DESIGN
DISCUSSION
DFID examples
Synthesis
Left of the ToC
Transport formula

Definitions

Concept	Definition	Context
Fidelity of form	Extent to which form of the intervention (activities, materials, delivery) maintained (or 'core'?)	Tension between remaining completely faithful to how an intervention was previously delivered (fidelity of form) and making implementation feasible through adaptation
Feasibility of implementation	Extent of the barriers and facilitators for delivering an intervention in a new setting	
Adaptation	Changes to an intervention in a new setting	
Fidelity of function	Extent to which the function of an intervention can be maintained in different settings	Interventions can be described in terms of how they function within a complex system
Interventions as events in systems	Interventions as events that interact with systems of people and institutions, no fixed form *or* function	
Generalisability	Extent to which claims about a sample can be applied to a population	>3 concepts used for extrapolating causal effects from an evaluation to other places, which differ primarily in the extent to which a particular target context is considered.
Transferability,	Extent to which claims made about one setting can be applied in another	
Transportability	(1) Extent to which an intervention can be moved from one place to another; (2) Extent causal inferences about one system can be used to make inferences in another	

Castro and Martinez (2004) The Cultural Adaptation of Prevention Interventions: Resolving Tensions Between Fidelity and Fit. *Prevention Science*

Hawe P, Shiell A and Riley T (2009) Theorising interventions as events in systems. In: *American Journal of Community Psychology*

Gardner F, Montgomery P and Knerr W (2016) Transporting Evidence-Based Parenting Programs for Child Problem Behavior (Age 3–10) Between Countries: Systematic Review and Meta-Analysis. *Journal of Clinical Child & Adolescent Psychology*

Burchett H, Umoquit M and Dobrow M (2011) How do we know when research from one setting can be useful in another? A review of external validity, applicability and transferability frameworks. *Journal of Health Services Research and Policy*.

MECHANISMS & THEORY

CMOs
Elster mechanisms
"Middle-range theory"
MRT, mechanisms, and laws

'Out of sight, out of mind'
'Absence makes the heart grow finder'

Is middle-range theory, the solution?

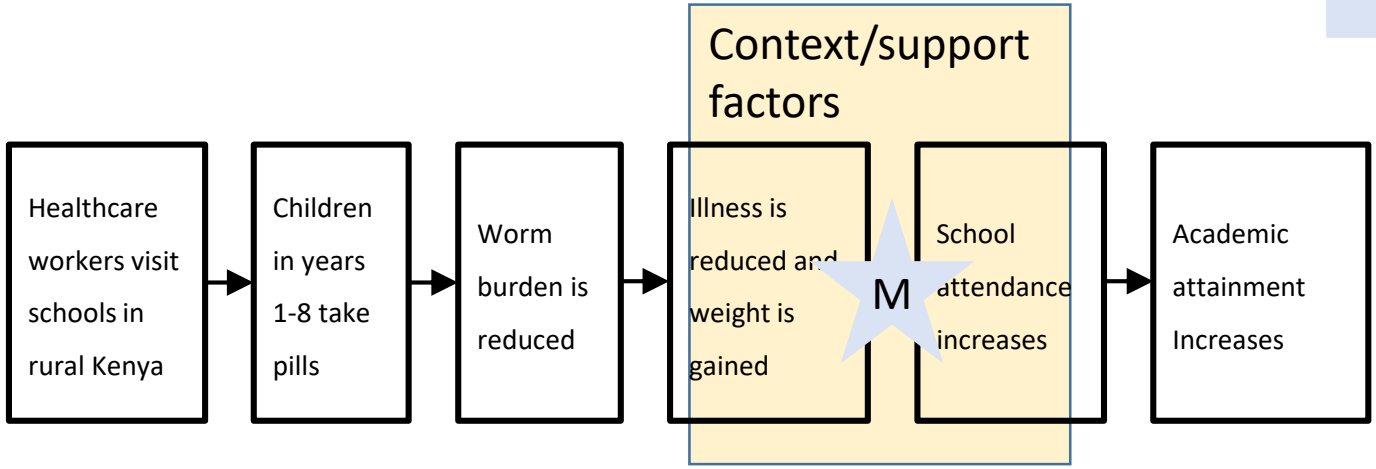
- Can't handle:
  - Many programme elements
  - Long causal chains
  - Interactions

Grand Theory
Structuration
Supply and demand

Mid-range theory
Role conflict
Tendency to cooperate
'entrepreneurs tend not to expand jobs if the available workers have been unemployed for a long period and have lost skills'

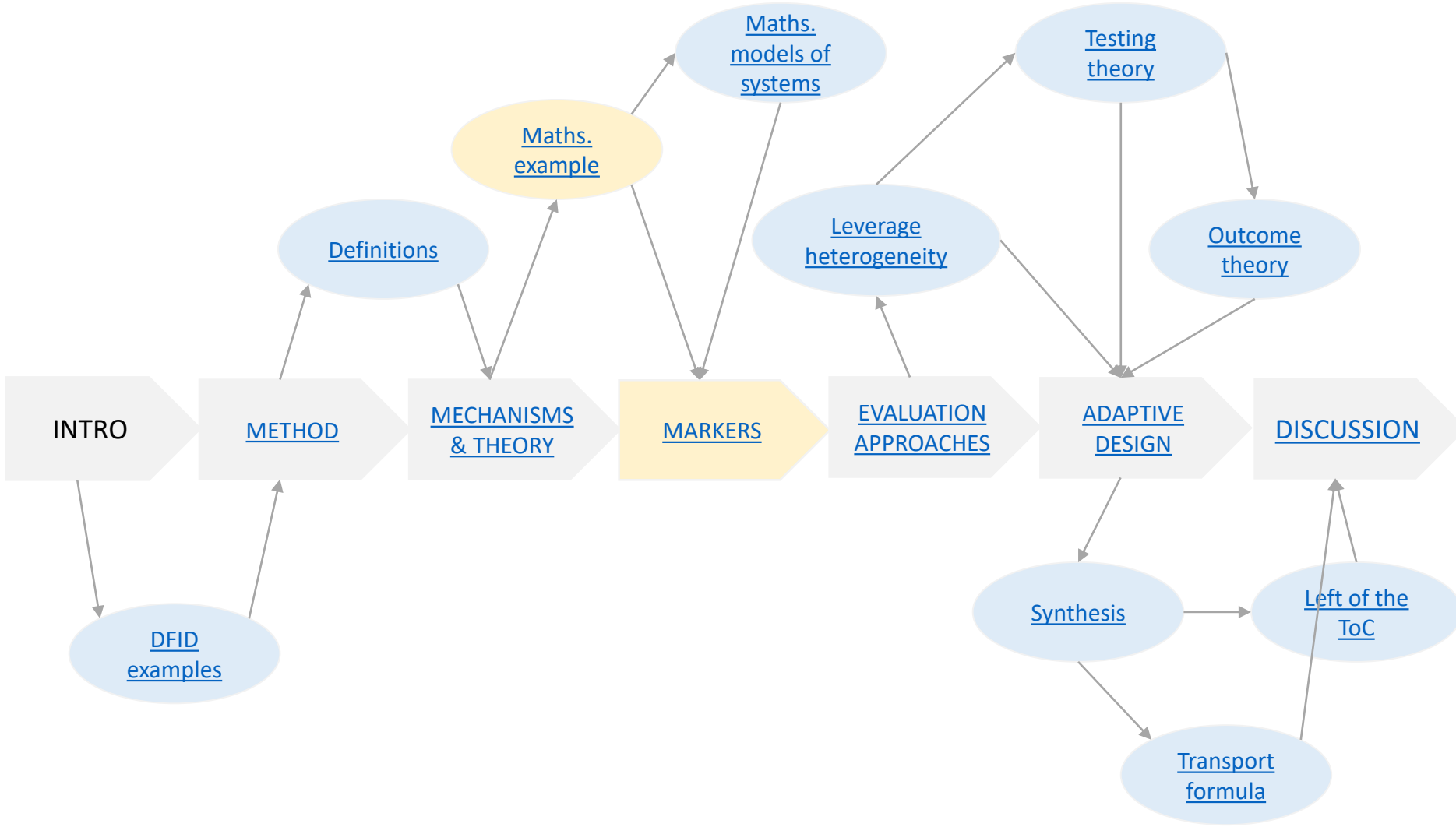
Programme theory
Context specific
Derived from more than one MRT

MRT and programme theory: example



Programmes are constructed, as manifestations of the various theories

ROUTE MAP


Maths. models of systems
Testing theory
Maths. example
Definitions
Leverage heterogeneity
Outcome theory
INTRO
METHOD
MECHANISMS & THEORY
MARKERS
EVALUATION APPROACHES
ADAPTIVE DESIGN
DISCUSSION
DFID examples
Synthesis
Left of the ToC
Transport formula

Maths example

Consider the following three theories over binary variables Y, X_1, X_2

Theory 1 ('grand theory'):

$$Y = X_1 X_2$$

Theory 2:

$Y = X_1$ in cases in which X_2=1

Theory 3:

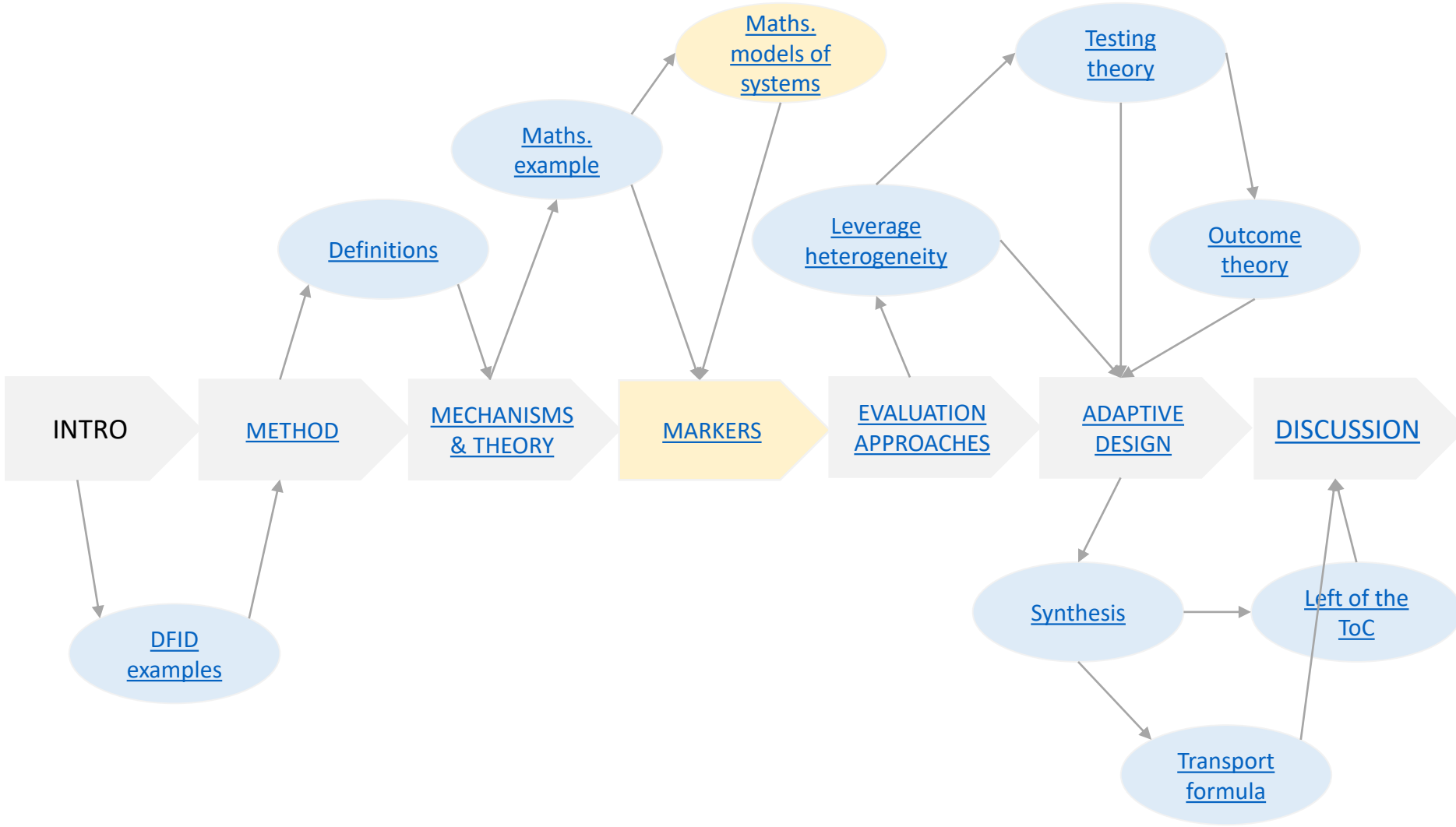
X_1=1 caused $Y = 1$

for individual *i* in context X_2=1

- Theory 1 provides a complete description of a simple process (e.g. children attend school when healthy and school available)
- Theory 2 is implied by Theory 1 -- that is, if Theory 1 is true so is Theory 2, but the converse does not hold (e.g. school attendance equals health status when schools are available)
- Theory 2 provides an incomplete description; here it is a *conditional* theory,
- Theory 3 is a specific claim about causal relations for an individual (e.g. being healthy caused a child to attend school when there was a school available) Theory 3 claims no generality.

- Within this set of theories Theory 2 is "middle-range"; it is implied by one theory and implies another. In a broader set of theories however Theory 1 might be midlevel, if for example it is implied by some Theory 0.

ROUTE MAP


Maths. models of systems
Testing theory
Maths. example
Definitions
Leverage heterogeneity
Outcome theory
INTRO
METHOD
MECHANISMS & THEORY
MARKERS
EVALUATION APPROACHES
ADAPTIVE DESIGN
DISCUSSION
DFID examples
Synthesis
Left of the ToC
Transport formula

Maths models of systems

Behavioural models (economics)

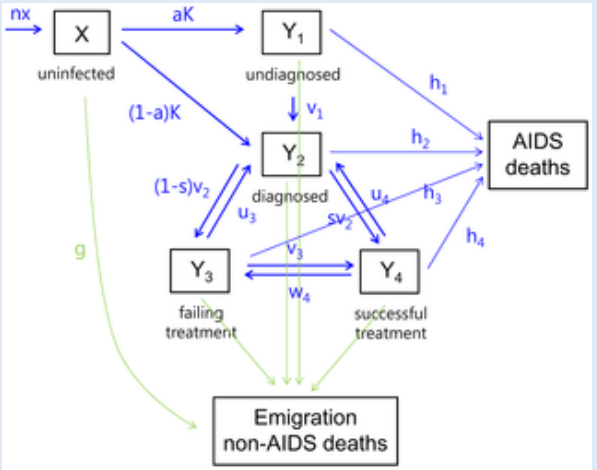
- Equations are used to describe the complex causal relations
- Combines 'economic' structure (deterministic econ theory) with statistics
- May imposes a solution by assuming maximisation of utility or profits or social welfare → make predictions

Both combine assumptions about *structure* with data

Both tend to be applied where implementation is unproblematic (e.g. taxes, vaccines)

BIQQ

- Promising method that used Bayesian methods to combine qualitative and quantitative data



Mathematical models

- Combine theory (e.g. R, R0) with data
- Present current understanding
- Usually some kind of agent or state model

- Rarely include social processes
- Common sort of question: 'what would happen to HIV incidence if we successfully treated STIs in this context?'

Deaton A (2009) Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development.

Garnett GP, Cousens S, Hallett TB, et al. (2011) Mathematical models in the evaluation of health programmes. *The Lancet*.

MARKERS

Context matters: CMOs, support factors

‘‘A short absence can do much good.’’
‘‘Absence lessens moderate passions and
intensifies great ones, as the wind blows out a candle but fans up a fire.’’

Voodoo that works

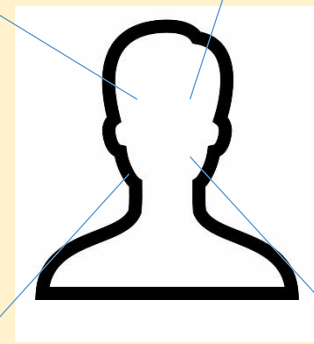
Lipsey: different models of bullying in primary and secondary schools

PolSci: democracies won’t go to war

Diagnosis

‘Support factors’ (e.g. food, comorbidities)

Short of breath, wasted, pale



Culture

Persistent cough (signs, symptoms)

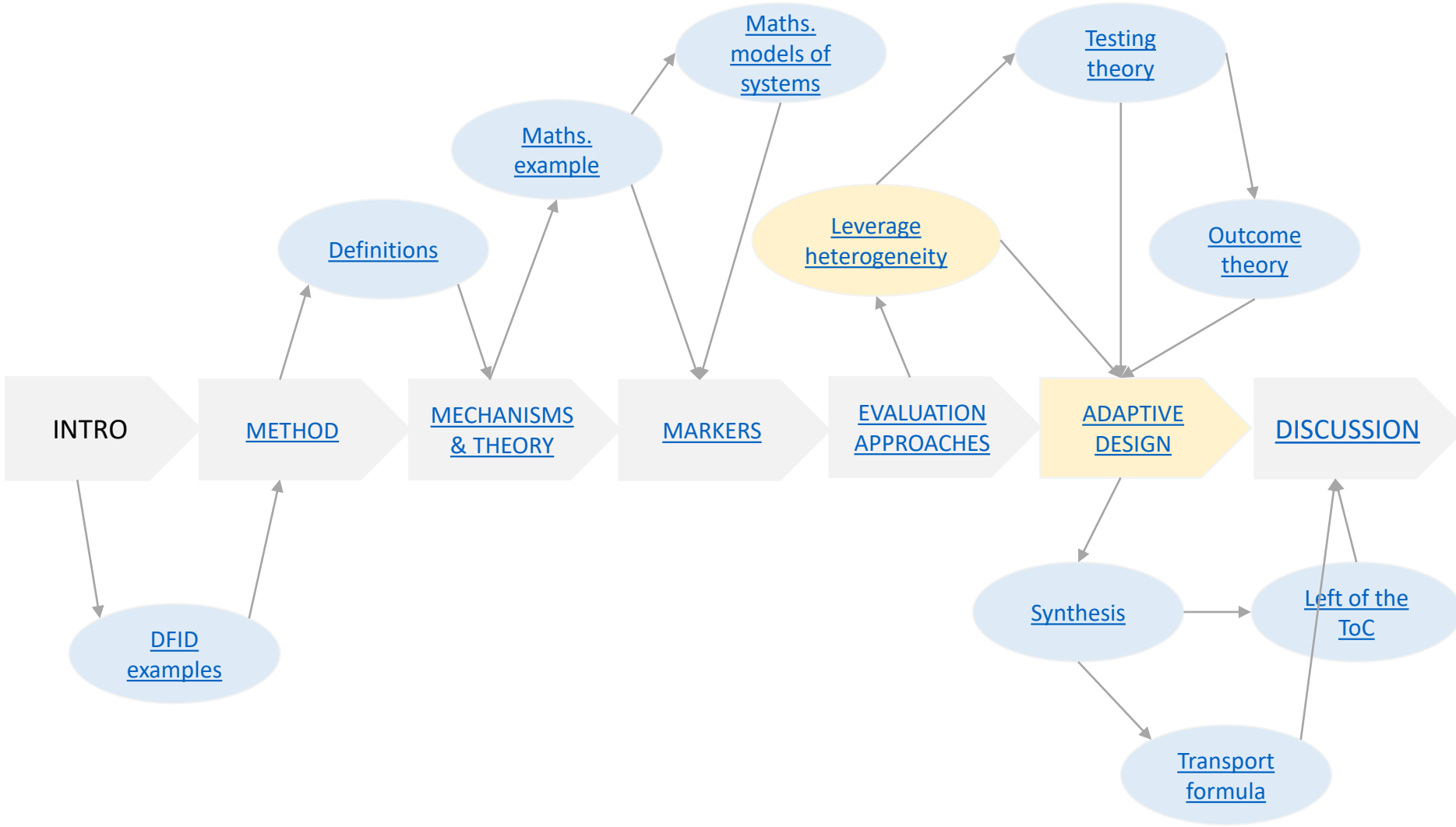
Lipsey MW (2009) The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and Offenders*

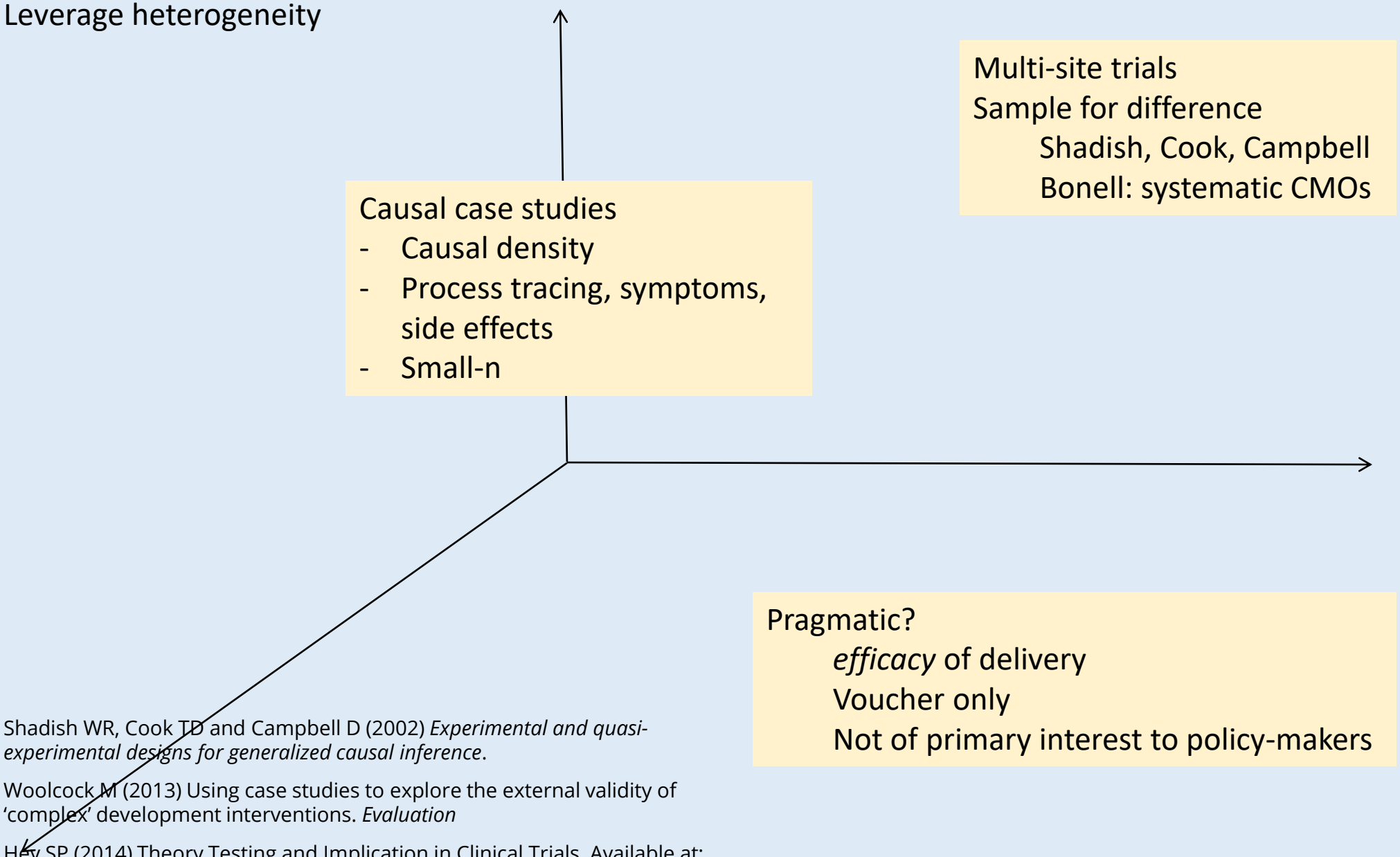
Strevens M (2012) Ceteris paribus hedges: Causal voodoo that works. *Journal of Philosophy*

EVALUATION APPROACHES

Approach	Summary
Test theories not interventions	Orienting evaluations to accumulate knowledge that refines theory, rather than testing the effects of interventions.
Integrated mixed-methods process evaluation	Gathering data on multiple elements: components of the intervention; implementation; mechanisms (mediators) and effect of context (moderators); representativeness of samples; risk factors; features of target place; practitioner experiences.
Leverage heterogeneity	Conducting multi-site and pragmatic trials to test assumptions in multiple contexts.
	Using case studies to identify conditions where diverse outcomes are observed focusing on context, implementation differences, and “trajectories of change” to predict whether replication or scale up is possible.

ROUTE MAP


Maths. models of systems
Testing theory
Maths. example
Definitions
Leverage heterogeneity
Outcome theory
INTRO
METHOD
MECHANISMS & THEORY
MARKERS
EVALUATION APPROACHES
ADAPTIVE DESIGN
DISCUSSION
DFID examples
Synthesis
Left of the ToC
Transport formula



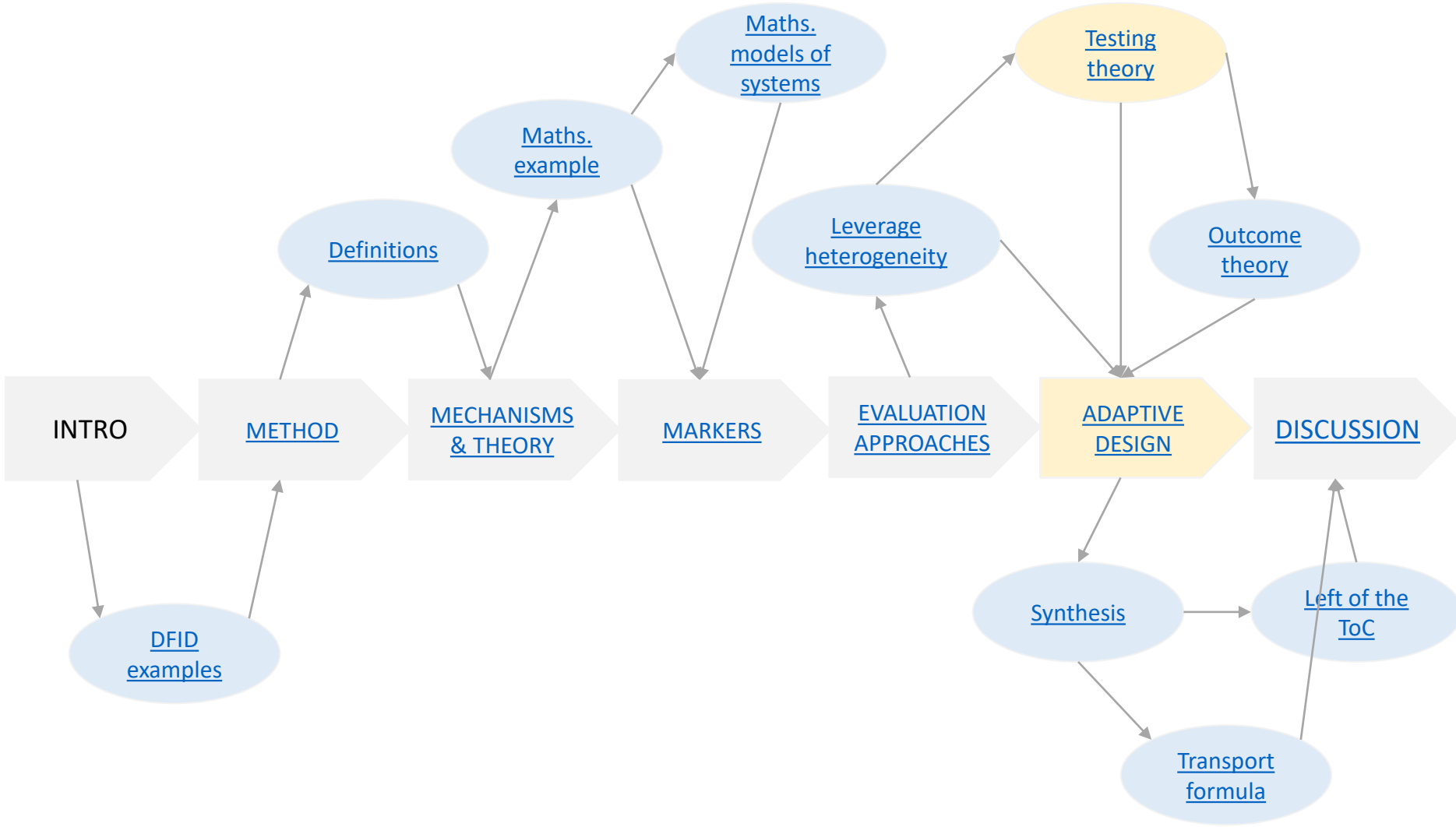
Shadish WR, Cook TD and Campbell D (2002) *Experimental and quasi-experimental designs for generalized causal inference*.

Woolcock M (2013) Using case studies to explore the external validity of 'complex' development interventions. *Evaluation*

Hey SP (2014) Theory Testing and Implication in Clinical Trials. Available at: http://philsci-archive.pitt.edu/11045/1/biomarker_theory_testing-psa.pdf

Cartwright N (2017) Single case causes: What is evidence and why. In: *Philosophy of Science in Practice*.

ROUTE MAP


Maths. models of systems
Testing theory
Maths. example
Definitions
Leverage heterogeneity
Outcome theory
INTRO
METHOD
MECHANISMS & THEORY
MARKERS
EVALUATION APPROACHES
ADAPTIVE DESIGN
DISCUSSION
DFID examples
Synthesis
Left of the ToC
Transport formula

Testing theory

Interventions are imperfect tests
Thinking against oneself

Bonell and bullying:

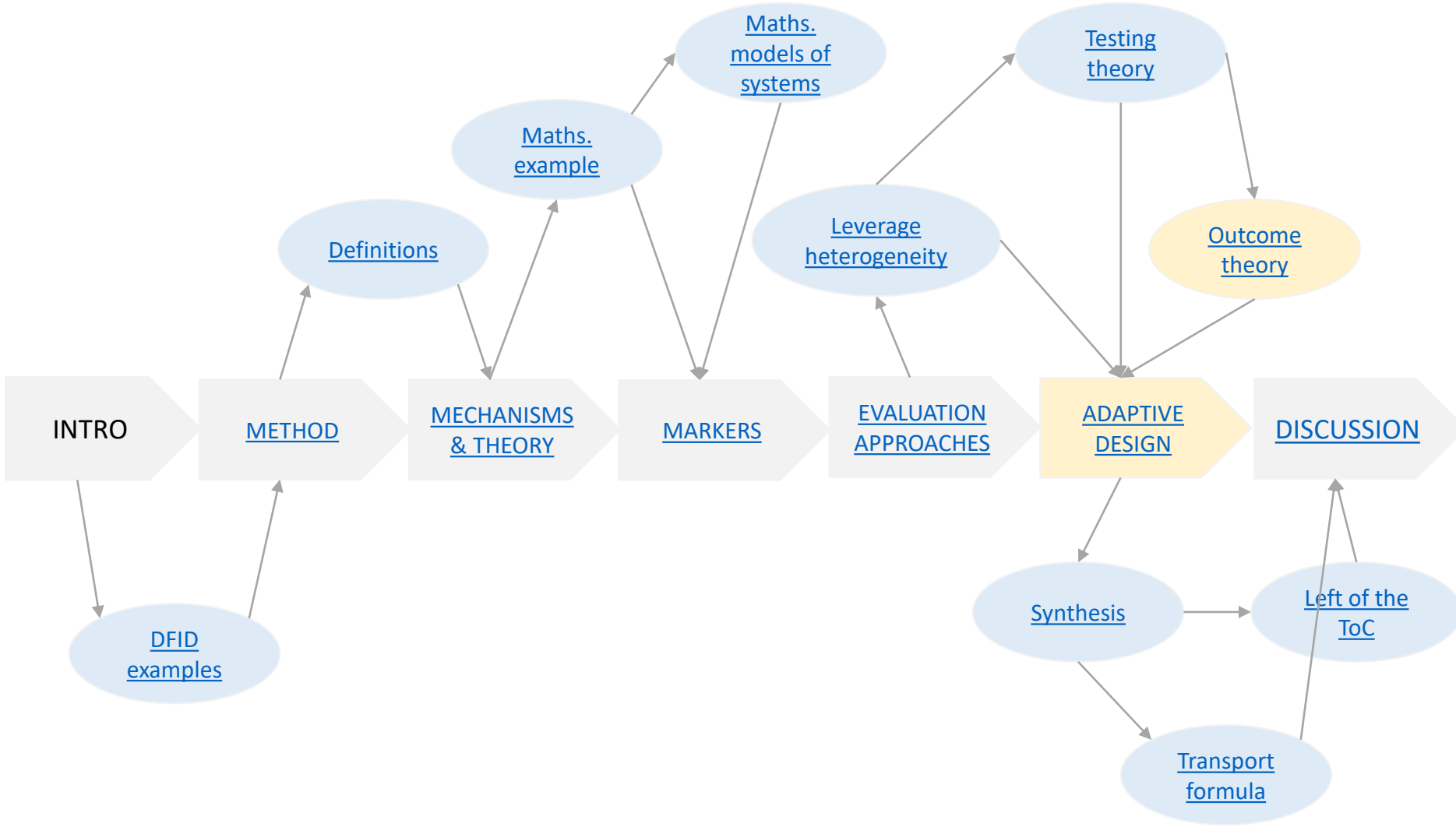
Underpowered analysis
Potentially confounded mediator analysis

Glennerster and JPAL:

The J-PAL framework begins with the question, ‘what is the disaggregated theory behind the program?’, where ‘theory’ is described as ways of simplifying the world to help make and test predictions. The framework then focuses on the local conditions (and whether the theory is likely to apply), the evidence available to support the behaviour-change underpinning the intervention theory, and considers the evidence that the programme can actually be implemented in this setting. With reference to cases from various contexts, the authors argue that superficial contextual differences might not be important when mechanisms travel and that learning about theory will require synthesis methods that interrogate the mechanisms behind the interventions.

Blattman C (2016) Why what works is the wrong question - Evaluating ideas not programs. URL: https://chrisblattman.com/2016/07/19/14411/

ROUTE MAP


Maths. models of systems
Testing theory
Maths. example
Definitions
Leverage heterogeneity
Outcome theory
INTRO
METHOD
MECHANISMS & THEORY
MARKERS
EVALUATION APPROACHES
ADAPTIVE DESIGN
DISCUSSION
DFID examples
Synthesis
Left of the ToC
Transport formula

Outcome theory

Using evaluation to ask: why does the outcome occur at all?

Randomisation = exogenous variation

- IV

Burchett, Ghana

- Local understanding of problems
- Efficacy studies

Stephen Birch: policy makers care about how health comes about

- Shift focus from providers and services
- To populations and problems

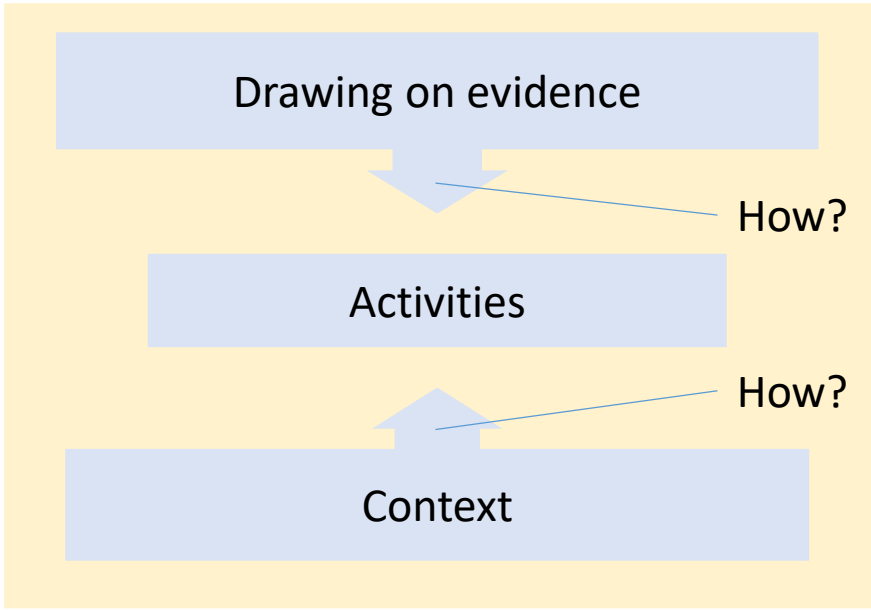
Burchett HED, Mayhew SH, Lavis JN, et al. (2015) The usefulness of different types of health research: Perspectives from a low-income country. *Evidence and Policy*

Birch S (1997) As a matter of fact: evidence-based decision-making unplugged. *Health Economics*

ADAPTIVE DESIGN

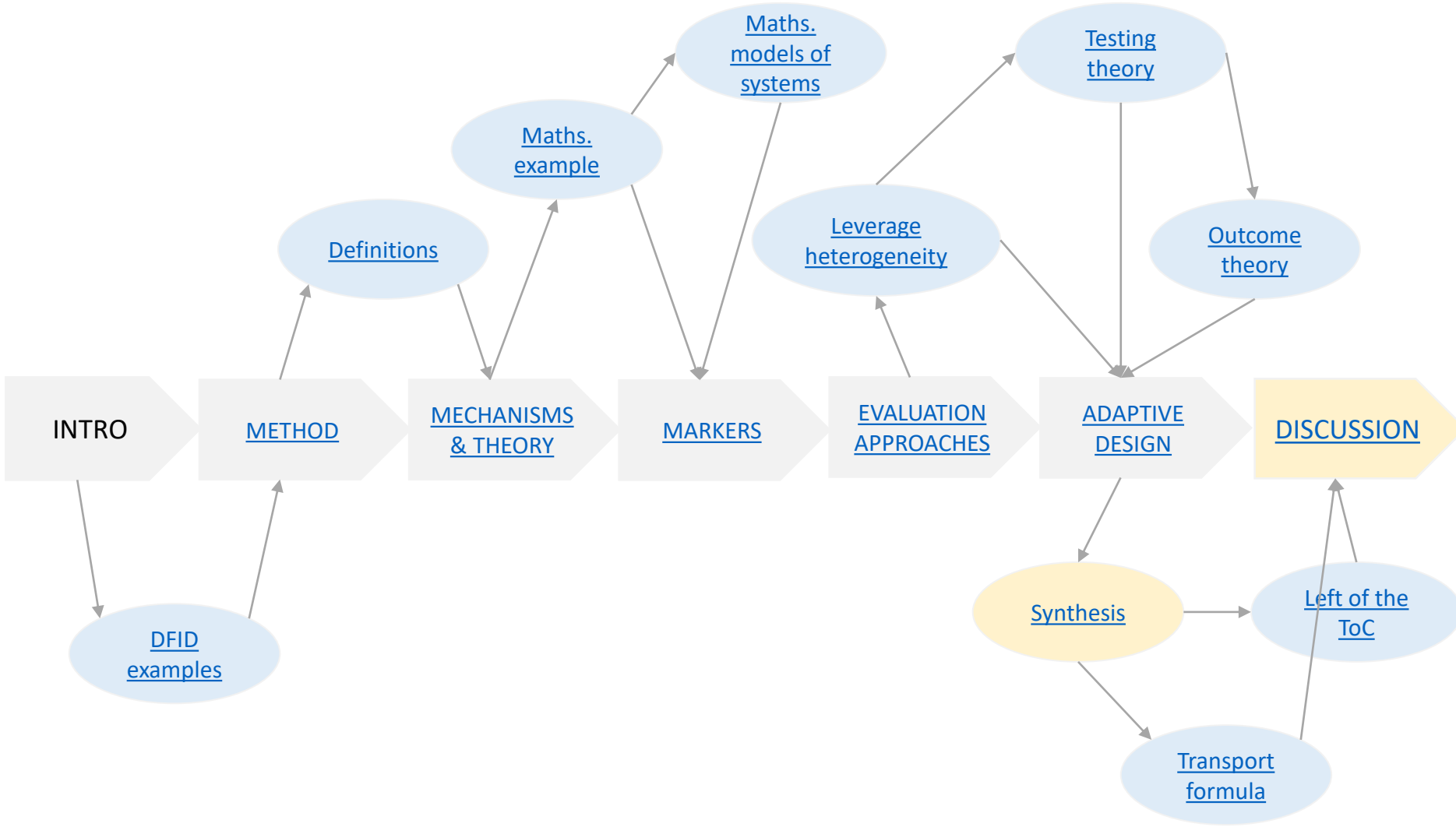
The ‘intervention’ *is the response* to context
Formative?
Democratic?

‘the intervention’ =



Examples
Women’s groups
Adaptive management

ROUTE MAP

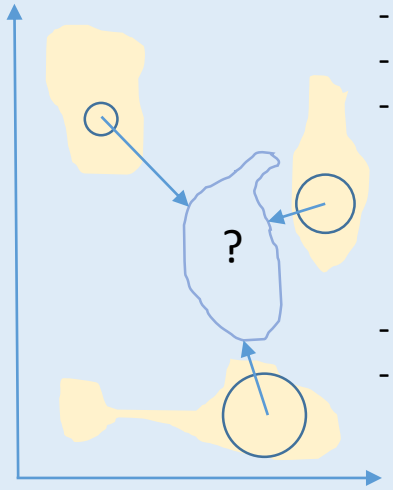

Maths. models of systems
Testing theory
Maths. example
Definitions
Leverage heterogeneity
Outcome theory
INTRO
METHOD
MECHANISMS & THEORY
MARKERS
EVALUATION APPROACHES
ADAPTIVE DESIGN
DISCUSSION
DFID examples
Synthesis
Left of the ToC
Transport formula

Synthesis

Quantitative

Qualitative

Interpolation



- Meta-analysis
- Meta-regression
- Re-weighting
  - Blinder-Oaxaca decomposition
  - matching methods
  - machine learning
- Overlap needed
- Large contextual factors difficult (often personal only) – theory?

Wang et al (2005) suggested that a 'list of attributes that may impact on applicability and transferability can be developed, based on knowledge of the proposed intervention. Then the applicability and transferability of the intervention to the local setting can be rated, and given a score, based on knowledge of the local setting'

Burchett – none of the frameworks work

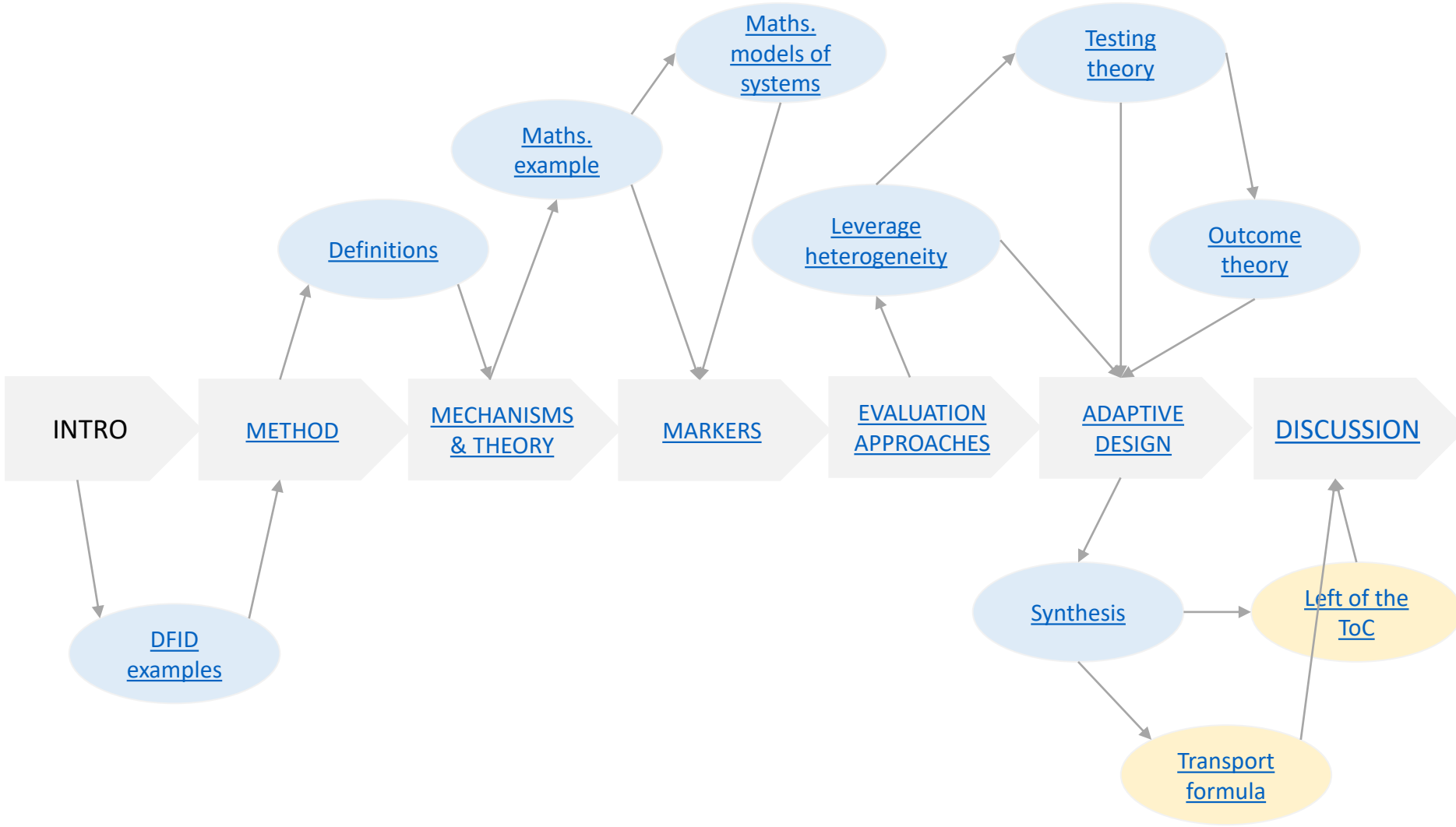
Extrapolation

- Structural models (economics)
  - Tuition fees & wages
- Infectious-disease models
- Testing theories in RCTs
- General equilibrium models
- Bayesian integration

Design focus

- Design strategy for "hybrid prevention programmes" that "build in" adaptation to enhance programme fit (Castro)
- ADAPT-ITT model (Wingood and DiClemente)
  - Eight phases, including formative research

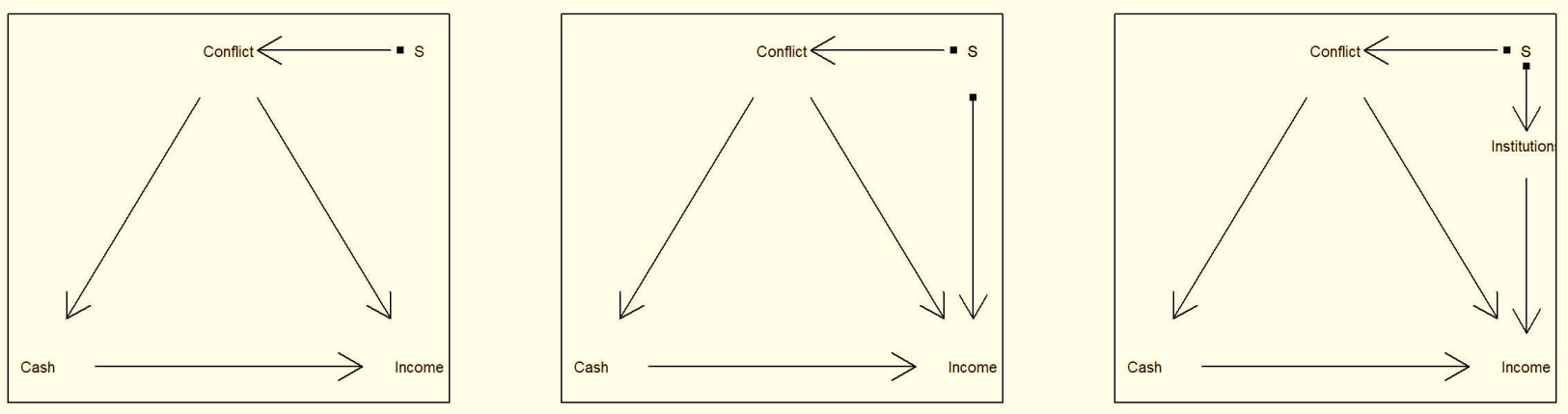
ROUTE MAP


Maths. models of systems
Testing theory
Maths. example
Definitions
Leverage heterogeneity
Outcome theory
INTRO
METHOD
MECHANISMS & THEORY
MARKERS
EVALUATION APPROACHES
ADAPTIVE DESIGN
DISCUSSION
DFID examples
Synthesis
Left of the ToC
Transport formula

Transport formula

When are causal inferences estimable from both the statistical information available and the causal information transferred from the experiments?

Use ‘selection diagrams’, extensions of DAGS; *S* depicts the things that differ



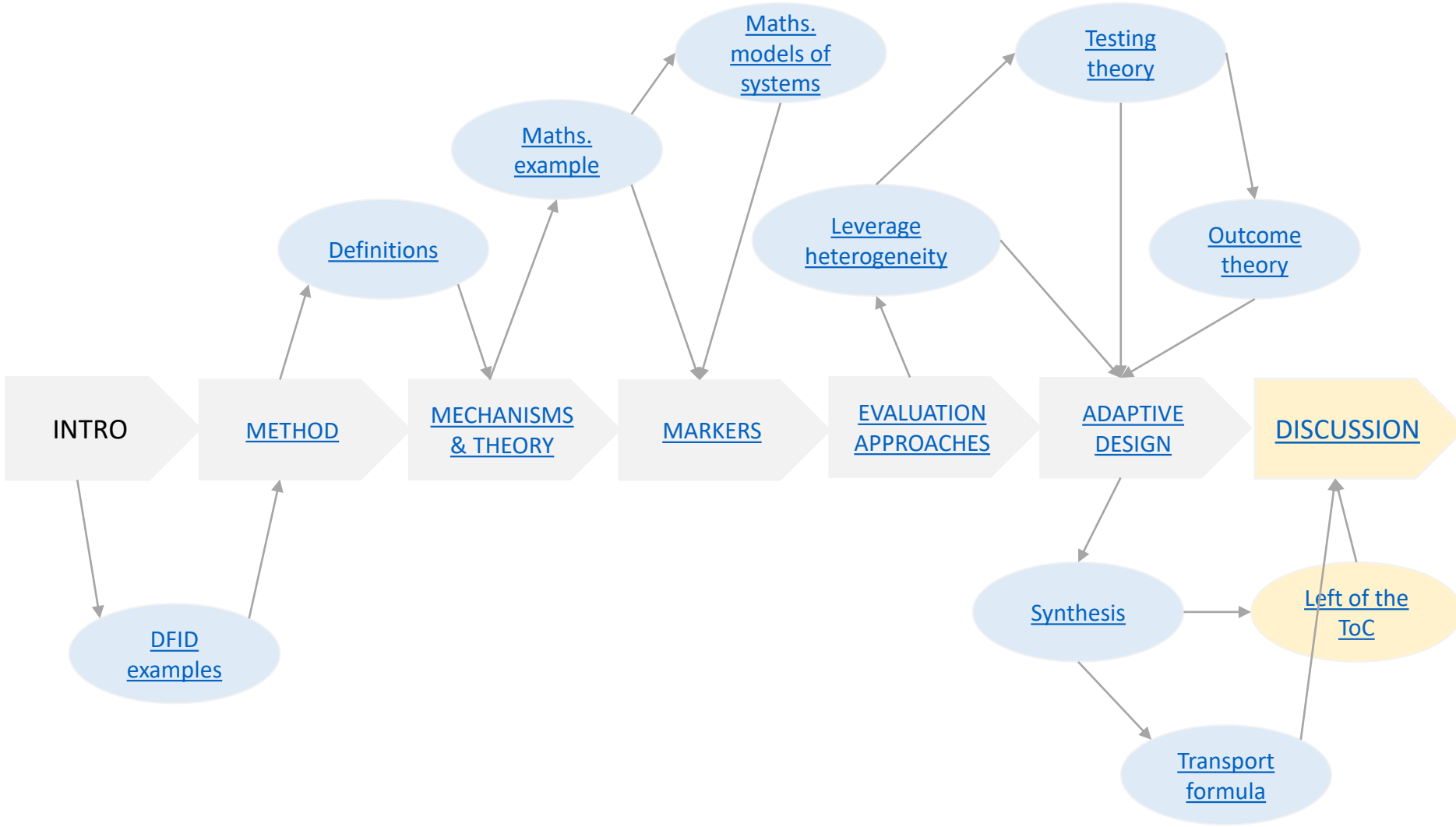
Requires considerable confidence in one’s understanding of the causal model in both populations in order to specify the key ways in which they differ.

Requires a minimal compatibility between graphs.

Nodes are assumed to be identical in different contexts, which is a challenge for construct validity, and may require that latent variables are added

Bareinboim E and Pearl J (2013) A General Algorithm for Deciding Transportability of Experimental Results. *Journal of Causal Inference* 1(1). DOI: 10.1515/jci-2012-0004

ROUTE MAP


Maths. models of systems
Testing theory
Maths. example
Definitions
Leverage heterogeneity
Outcome theory
INTRO
METHOD
MECHANISMS & THEORY
MARKERS
EVALUATION APPROACHES
ADAPTIVE DESIGN
DISCUSSION
DFID examples
Synthesis
Left of the ToC
Transport formula

Left of the ToC

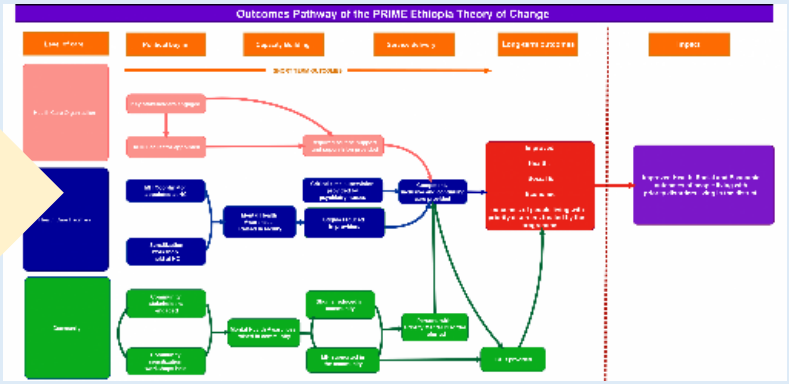
Reminder: most things don't 'work'

Contexts vary a lot

When things don't work, are we looking in the right place?

EEF example

Decision / design processes



Contextual information

How did this ToC, this intervention, get applied here in the first place?

Why do we think of the 'intervention' as starting *after* the design is finished?

DISCUSSION

Summary

- Interventions are based on theories, and evaluations can help strengthen them
- Need to embrace heterogeneity
- Don't be scared of assumptions (=theory)
- Need to think about how to respond to context in evidence-based way

Context-centred delivery

- Not sure if it 'worked' (counterfactual = normal delivery)
- Not everyone got what they wanted
- Was voting a good way? How would better ways be found? And evaluated?

Dissent!

More science? Scientists would say that

Sociology of science

China -> Tanzania

Design processes

Recommendations

Gap map of theories
Theory hubs
Theory-focused synthesis
Putting theory in ToC
Theory-testing in evaluations
Think about markers