Exploring the possible: a deep dive into the sea of options for Qualitative Impact Evaluation (QIE)

Barbara Befani, PhD

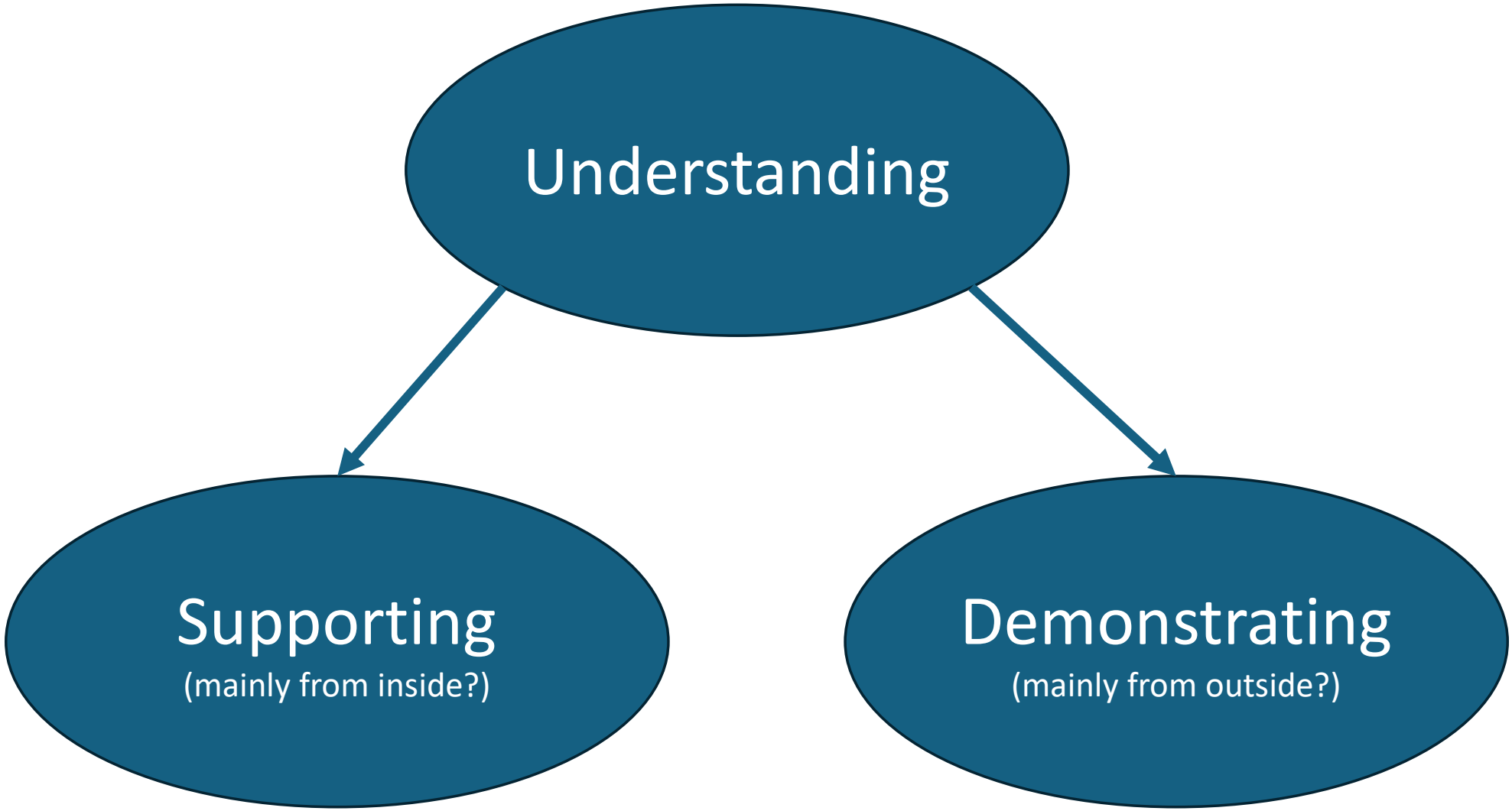
National Centre for Social Research (UK)

Danish Evaluation Society Conference “Perspectives and Approaches to Impact Evaluation”

Middelfart, 12th September 2025

Outline

- Three themes: Understanding, Supporting, Demonstrating Impact
- How evaluation (QIE) can foster and document meaningful, long-term, real, long-lasting (Transformational? Systemic?) change
- The process of evaluation design
- Why do we want to evaluate impact? [preferences]
  - Values / Principles
    - Moral (ethics, social justice, transparency)
    - Scientific (validity, credibility, transparency)
  - Motivation / Purpose
    - Increasingly higher demands for transparency and accountability
    - Continuous learning for continuous improvement
  - Aims & objectives, evaluation questions
- What can we know about impact? [constraints]
  - Epistemology, data availability, applicability, feasibility, requirements
- Choice of evaluation questions, approaches, methods (combinations)

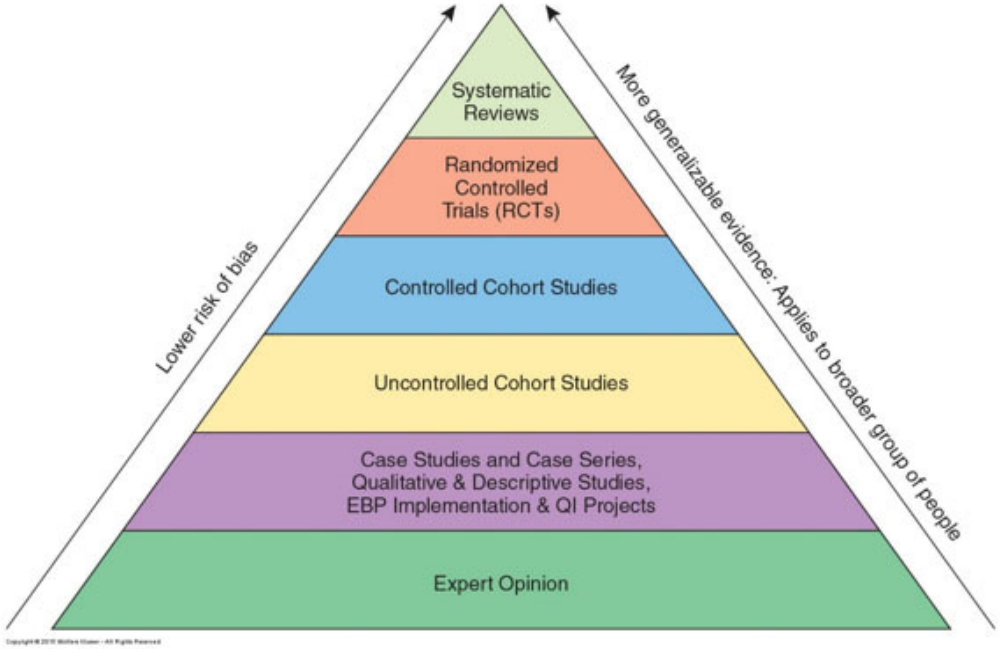

Understanding
Supporting
(mainly from inside?)
Demonstrating
(mainly from outside?)

Understanding, Supporting, Demonstrating

- Demonstrating
  - Ideally placed “outside” of the intervention
  - External “unbiased” expert
- Supporting
  - Ideally placed “inside” the intervention
  - Stakeholder able to “drive” intervention progress from inside
- Understanding Impact
  - Relevant and needed both inside and outside the intervention

Hierarchy of evidence

- I usually start these talks defensively, in reaction to the popularity of the hierarchy of evidence
- Saying there is no one best option that works all the time
- Maryland Scale



Level 1: Correlation between a prevention programme and a measure of crime at one point in time

Level 2: Measures of crime before and after the programme, with no comparable control condition

Level 3: Measures of crime before and after the programme in experimental and comparable control units, controlling for other variables that influence crime

Level 4: Measures of crime before and after the programme in multiple experimental and control units, controlling for other variables that influence crime

Level 5: Random assignment of programme and control conditions to units

(Endless) list of possible design options

- Explanation: Contribution Analysis, Realist Evaluation, Process Tracing, Qualitative Comparative Analysis (QCA)
- Inclusion: Outcome Harvesting, Outcome Mapping, Most Significant Change, other participatory approaches
- Systems and networks: Participatory Systems Mapping, Bayesian Belief Networks, Agent-Based Modelling, Social Network Analysis, Causal Loop Diagrams, Systems Dynamics, etc.
- Net effect: RCTs, various forms of QEDs
- Efficiency: cost-benefit, cost-effectiveness analysis
- These are just “approaches”

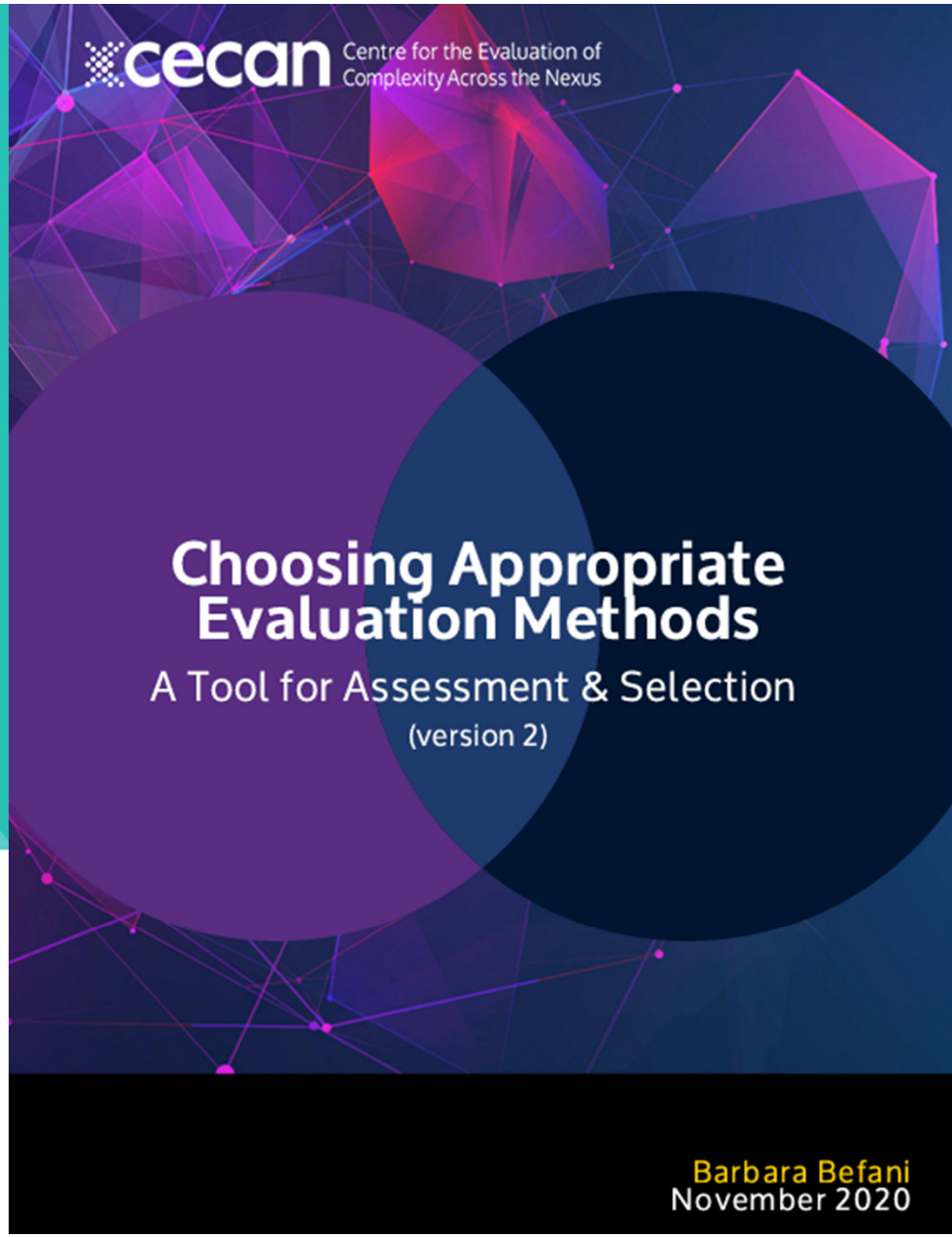
Evaluation of International Development Interventions

An Overview of Approaches and Methods

Jos Vaessen, Sebastian Lemire, and Barbara Befani

Independent Evaluation Group

November 2020



How do we choose?

- Appropriate choice

- Choice is an action, coherent with / mediated by our:
  - Preferences [motivations, purpose, values, principles]
  - Opportunities / constraints
  - Beliefs / capabilities

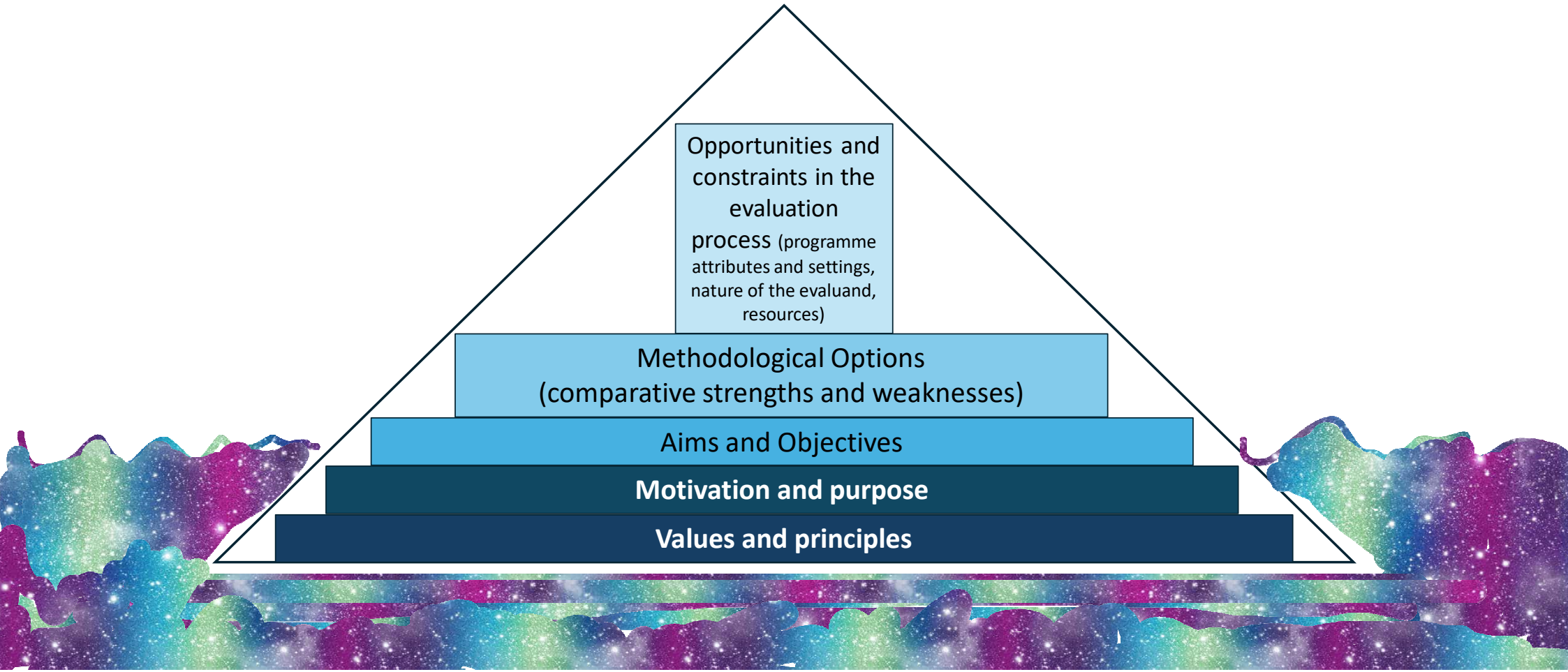
- Why do we want to evaluate impact?
  - Good starting point

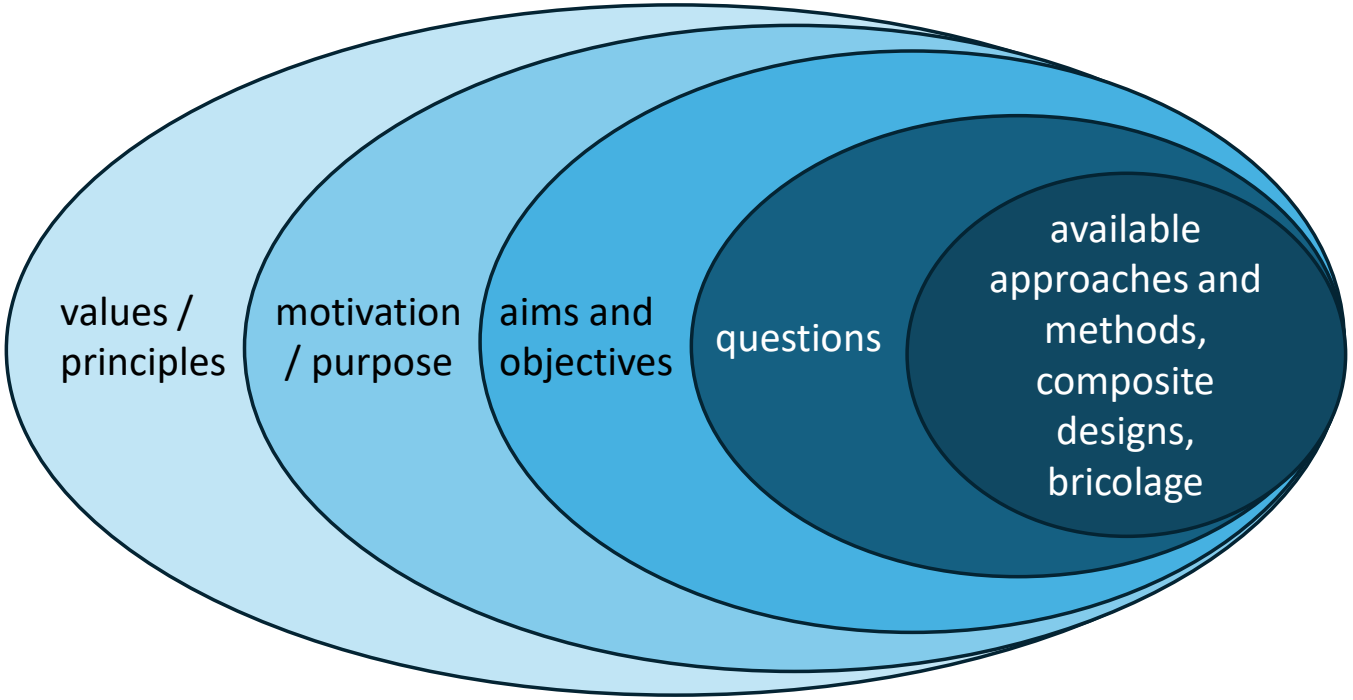
Why do we want to evaluate impact?

- What is our motivation or purpose with a specific impact evaluation?
- Is the purpose consistent with the values and principles we want to uphold?

- Aims and objectives of the evaluation
- Methodological options
  - With their own strengths and weaknesses
  - Different abilities to reach aims and objectives, serve motivations and purpose
  - Differently align with values and principles
- Opportunities and constraints of the evaluation process
  - Nature of the evaluand, programme attributes
  - Resources
  - Relevant for matching with methodological requirements and determining feasibility

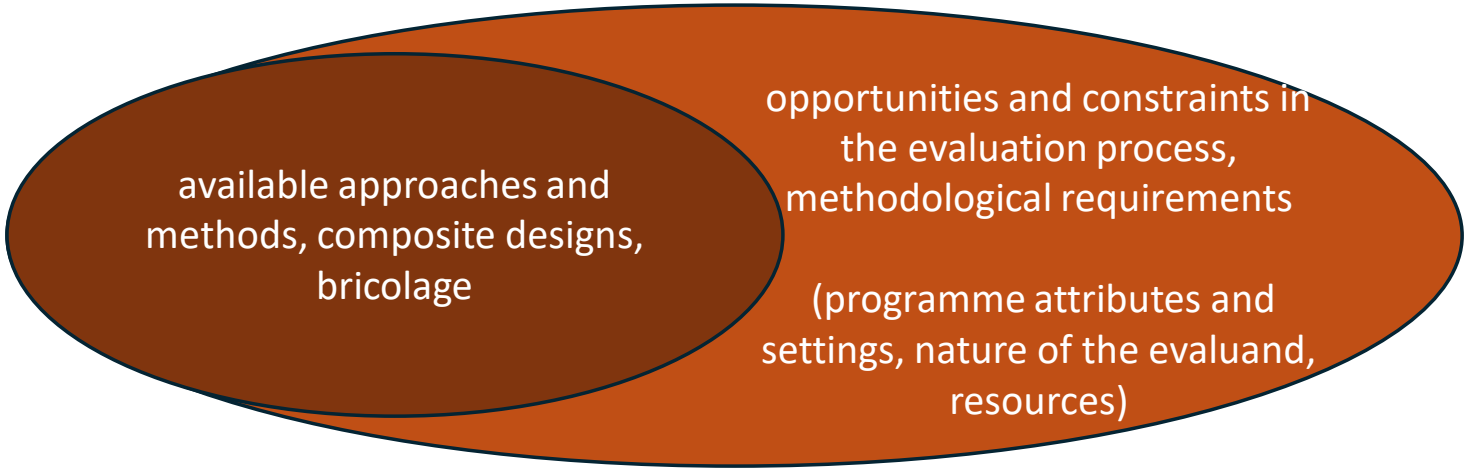
The iceberg (submerged pyramid?) model of evaluation





preferences

constraints



Values and principles

- Ethicality
- Usability
- Integrity
- Social Justice
- Flexibility and adaptability (need?)
- Relevance
- Validity (construct)
- Credibility
- Reliability
- Transparency
- Transferability

Grounding values in the evaluation design

- Ethicality
  - E.g. in data collection and how participants are engaged
- Usability
  - Report language and formatting; dissemination; engagement of stakeholders and ownership
- Flexibility, adaptability (could also be a need)
  - in evaluation planning and adjustment opportunities
- Findings:
  - Rigour?
    - Credibility
    - Reliability
  - Validity (construct)
  - Transferability
- Transparency on methods and approaches and the rationale for their choice
- Integrity
- Relevance for specific stakeholders

Tradeoff between usability (supporting impact) and credibility (demonstrating impact)?

- Distance between evaluator and intervention
- Close: stronger support, better usability, more bias, less credibility?
- Far: weaker support, worse usability, less bias, more credibility?

- Magic potion is transparency
- Highly participatory processes ("closeness" to the intervention) can achieve credible / reliable findings when there is full transparency on the process
  - When the (co-construction, co-investigation) process is documented accurately and comprehensively

Motivation and purpose

- Accountability
  - E.g. assessing efficiency and value for money
  - Benchmarking against similar programmes
  - Informing decisions on (dis)continuity of programme
    - Quantitative Impact, sustainability
- Learning
  - Scaling up a pilot programme
  - Improving effectiveness and achieving greater impact (next time) and / or sustainability
    - Qualitative Impact
- Inclusion and representation
  - Empowering stakeholders, taking their advice, fostering ownership
- Understanding
  - Usually improved / deeper understanding (e.g. of evaluand / object of evaluation) is a pre-requisite for achieving most other goals

More specific: aims and objectives

- Description of the evaluand (as yet unknown reality)
  - in iterative processes where evaluands have emergent properties
  - E.g. “plans” with not clearly defined “actions”
- Assessing any of the DAC criteria
  - Efficiency, effectiveness, impact, coherence sustainability, relevance

- Providing advice and recommendations
- Gathering (different) perspectives on specific topics / themes
- Establishing / exploring / investigating causal links
  - E.g. the role of an intervention in achieving an outcome

- Not mutually exclusive!

Evaluation questions (1/2)

- Answering Qs is one typical specific goal!
- In IE questions are mostly causal
  - Grounded on causal inference frameworks

- Understanding impact
  - How was impact generated (generative causality)
    - How did the intervention make a difference?
    - What was it in it that made a difference?
    - How did the intervention (and other factors) contribute to the outcome?
    - What role did the intervention play in achieving the outcome (or other outcomes, positive or negative)
  - What factors supported the achievement of the outcome?
  - What factors were necessary (or sufficient) to achieve the outcome?

Evaluation questions (2/2)

- Supporting impact (learning and improvement, sustainability)
  - How can the intervention be improved / made sustainable?
  - Which factors are responsible for (poor) performance / sustainability?
  - (generative causality)

- Demonstrating impact (accountability)
  - What was the net effect? (sequential causality, Mill's methods)
  - What is our confidence level on the contribution made by the intervention?
  - How significant / relevant was the intervention's contribution?

Methodological options: comparative strengths and weaknesses

- Explanation
  - Contribution Analysis, Realist Evaluation, Process Tracing, Qualitative Comparative Analysis (QCA)
- Participation
  - Outcome Harvesting, Outcome Mapping, Most Significant Change, Participatory Evaluation
- Systems and networks
  - Participatory Systems Mapping, Bayesian Belief Networks, Agent-Based Modelling, Social Network Analysis, Causal Loop Diagrams
- Net effects
  - RCTs and QEDs
- Efficiency
  - cost-benefit, cost-effectiveness analysis
- Comparison typically less obvious within-group?

Methodological options: comparative strengths and weaknesses

- Explanation
  - QCA (and RS) are cross-case; RE, PT within case; CA mostly within case but flexible
  - PT focuses on strength of evidence; CA on comprehensiveness; RE on depth of explanation
- Relevance
  - OM ex-ante, OH ex-post, both immediate outcomes; MSC on ranking, participation on inclusion of diverse perspectives
- Systems and networks
  - Root causes, complexity, emergence, interdependence; PSM relevance & relationships: Bayesian Belief Networks on prediction, Agent-Based Modelling for micro-macro emergence, Social Network Analysis on relationships, Causal Loop Diagrams on non-linearity
- Net effects
  - RCTs require randomisation, QEDs different ways of creating control groups

Data collection and analysis techniques

- Traditionally relatively structured list of options
  - Qualitative interviews, focus groups, Delphi methods, surveys, desk reviews
  - Computer-assisted thematic analysis, statistical analysis
- It is becoming more varied and creative
  - Big data, social media content analysis
  - Photography and video-assisted tools
  - Artificial Intelligence
- Ever-growing list of options!
- Co-design becoming increasingly relevant
  - Higher need to set boundaries and make choices that are not obvious or dependent on “objective” expertise

Methods as fuzzy objects

- Toolkit? Tools are clearly defined ("crisp" definitions)
  - You can't use half a hammer but maybe only side of a screw?
- Methods are ambiguous in the way they are defined
  - "plastic" tools that can be moulded, reshaped
- For example, what is Contribution Analysis exactly?
  - Not fully clear, it has evolved over time
- Different versions of QCA and Process Tracing as well
- Realist principles vs. using CMO configurations
- Multiple variations in systems-based approaches

- Bespoke designs entail picking-n-mixing features of different methods
  - Metaphor: bricolage

Broad Methodology (original cloth)	Feature used (piece of cloth used)	Evaluation questions?	Aim / Objective it helped achieve	Purpose / goal the feature helped achieve	Values upheld?	Downsides if any
Contribution Analysis	Alternative explanations					
Contribution Analysis	Inclusion of risks & assumptions					
Process Tracing	Consideration of both confirmatory and disconfirmatory evidence			Mitigating confirmation bias	Credibility	
Bayesian Process Tracing	Bayesian Updating			Measuring confidence	Transparency, credibility	

Broad Methodology	Feature not used	Requirements not met?	Incompatible with programme attributes / settings?	Incompatible with Nature of the evaluand?	Too resource-intensive?	Loss with exclusion?
Outcome Mapping	[not used at all]	Immediate outcomes				

Importance of co-construction in co-design

- It's difficult for one single person / expert to understand all relevant information that should play a role into the choice
  - including stakeholders' preferences (not just properties of methods and approaches)
- Participatory process involving experts and stakeholders with clear ideas on
  - the purpose of the evaluation
  - the values it should embody
  - opportunities offered by methodologies
  - limitations and requirements of methodologies

Summary: (co-)creating an evaluation design

- Align what you want [preferences]
  - Motivation / purpose, values / principles
  - E.g.: long-term, meaningful, sustainable, systemic, transformational change

- With what you need [constraints]
  - to achieve your goals, in line with your purpose and values / principles
  - Requirements, fit with context / settings, resource limitations, nature of evaluand
    - about feasibility
  - Flexibility / adaptability could also be a need

- Sufficient knowledge of a wide range of options
- Assessment of these options
  - Their ability to:
  - reach aims and objectives, serve relevant purposes, align with values and principles
  - Fit with programme / evaluation context, resource limitations, nature of evaluand

Conclusions

- Think about your motivation / purpose with impact evaluation
- Gain / expand knowledge on options and their strengths / weaknesses
- Ensure chosen option(s) fit with your values (as well as your preferences ^ and constraints)
- (Co-)create a bespoke design