Symposium on the Future of Evaluation

Farewell to Evaluation:
'From Verdicts to Explanations'

Ray Pawson
Northumbria
2025

1. Farewell to verdicts and the accreditation model.
2. Farewell to the commissioning model and atomised inquiry.
3. Enter explanation – abstraction and cumulative inquiry.
4. Demolish the evaluation research 'silos': concentrate on programme theory commonalities.
5. Prioritise synthesis over evaluation: apply retrospective evidence prospectively.
6. Begin at the beginning: Shift resources from ex-post to ex-ante.
7. Widen focus from 'programmes' to the whole policy apparatus.
8. Incorporate institutional history: learning from continuing trial and continuous error.
9. Attempt the impossible: confront the 'wicked problems'.

In the beginning : All hail the accreditation model

The **Great Society**. A massive set of post-war domestic programs in the US aimed at social betterment across the welfare and heath system.

This created the **What Works** agenda. Each program was put to research, with the aim of identifying and accrediting only effective interventions.

Various **Research and Evaluation Clearinghouses** were so tasked, pulling together evidence from trials, quasi-experiments and meta-analysis.

Amidst the optimism doubts began to emerge about evaluation's ability to identify intervention winners and turf out the losers. Rossi's (1987) **Iron Law of Evaluation:** "The expected value of any net impact assessment of any large-scale social program is zero."

Enter Complexity and 'System States'

Programmes are complex, adaptive, self-transformative systems inserted into institutions or services or societies which are also complex, adaptive, self-transformative institutions.

'Effects are system states' (Byrne 2024)

Outcomes of programmes thus depend on action of systems within systems, which changes the evaluative challenge. It might be put:

What is it about a programme that works, for whom, in what circumstances, in what respects, with what resources, at what cost, with what political/public support, with what sustainably, designed and funded by whom, implemented by whom, in the presence of other programmes / reforms, etc., etc.'

Effectiveness depends on the interaction of all of these conditions and more. Let us call this: *Multiple Condition Causation (MCC)*
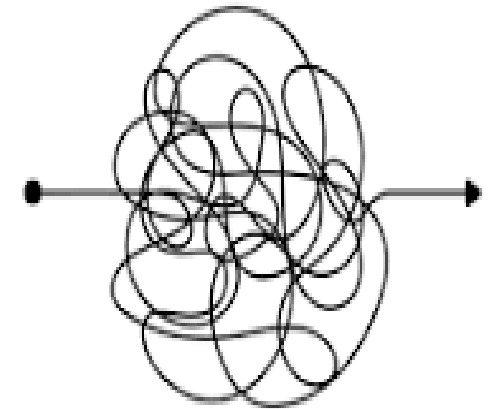
Farewell to the accreditation model

The effectiveness of any intervention thus depends on dozens of caveats and conditions.

Result: Conventional evaluation research is a 'leaky bucket'.

- No single case evaluation can capture all of these conditions. To do so would require omni-surveillance (everything, everywhere, more-than-once).

- Single case evaluations provide an account of some necessary conditions, leaving others behind. This limitation applies to all methods and strategies.

- Results derived from the single case will not apply to others, which inevitably carry a different configuration of system states. And thus we bid … farewell to accreditation

Farewell to the business model (a.k.a. groundhog day)

Evaluation for Hire

But business remains largely unchanged. Key agencies launch programmes ceaselessly, with the stock requirement (and funding) that each one should put to out to 'evaluation'.

This one-intervention-one-evaluation model:

Problems often noted include inconsistencies in chosen methods - the cash nexus, short time frames, arms-length involvement, etc.

The key problem:

Inquiry becomes atomised. Each evaluation starts from scratch. The evidence base consists of thousands of disconnected studies. There is little mutual learning. Studies run in parallel rather than sequentially.

Even before the ink has dried on your most recent final report, along comes another ITT.
GROUNDHOG DAY

Evaluation Research: The Sum Total

As a result of incessant change to interventions and their settings PLUS the ceaseless commissioning of new programmes, the evidence base of evaluation research consists of a BABEL of disconnected, non-transferable demi-regularities.

Enter explanation: Evaluation research as cumulative science

Popper formulates the logic of cumulative scientific inquiry as follows. PS stands for 'problem situation', TT stands for 'tentative theory', and EE stands for 'error elimination'.

$$PS_1 \rightarrow TT_1 \rightarrow EE_1 \rightarrow PS_2 \rightarrow TT_2 \rightarrow EE_2 \rightarrow PS_3 \dots\dots\dots\dots\dots$$
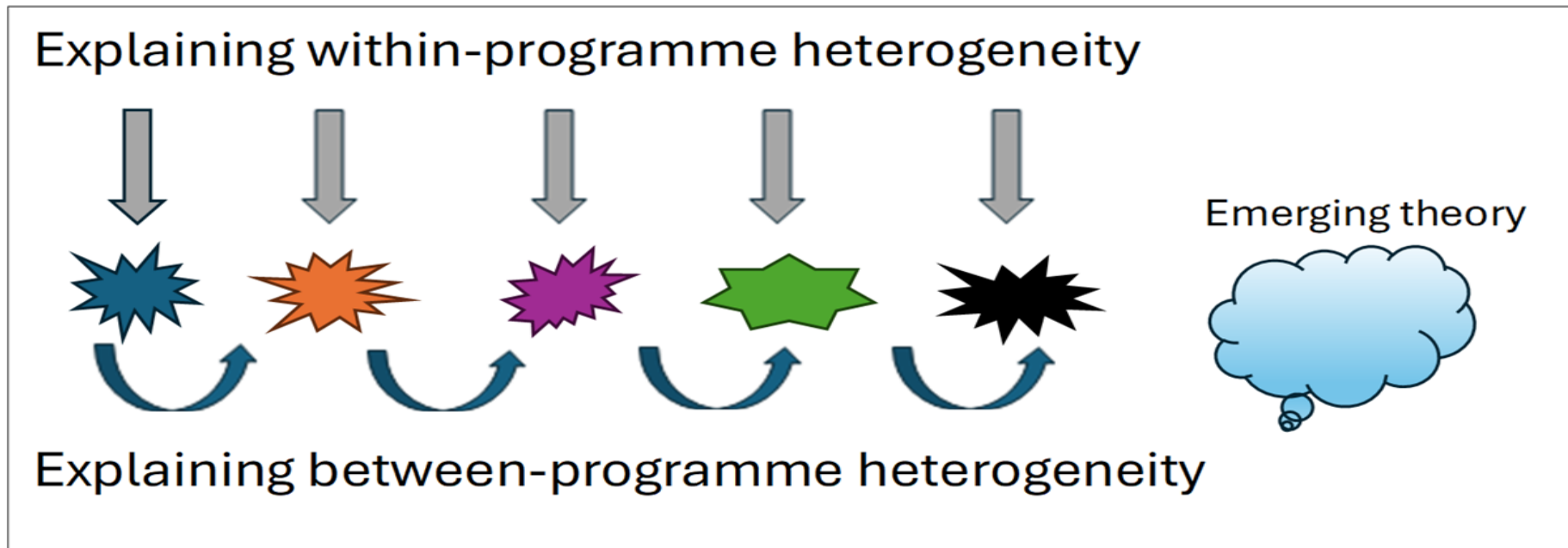
Abolish the Evaluation Silos

Due to narrow imagination of policy makers and their limited resources, variants of the same programmes are tried, tried and tried again. Programmes are limitless. Programme theories are not.

This inadvertent repetition is recognised in the heroic maxim proposing that there are only underline three types of programme theory – 'carrots', 'sticks' and 'sermons'. (Bemelmans-Videc et al)

E.G. Carrot or incentivisation theory. UK incentives are offered to insulate homes, to improve security, to quit smoking, to increase waste recycling, to join educational provision, to encourage independent living for disabled people, etc., etc.

Abolish the evaluation silos. The same programme theories and their successes and failures are evident in quite different domains >>>> health, education, crime, *environment, etc.*

Cumulative inquiry begins with the process of ABSTRACTION.

Prioritise synthesis over evaluation
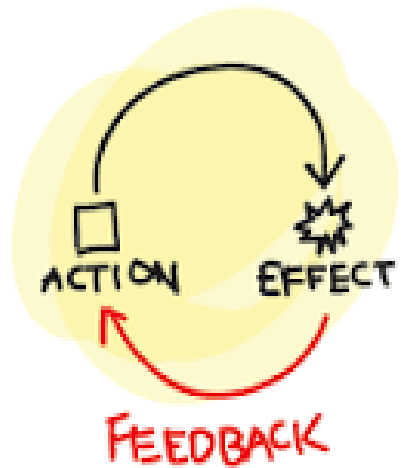
- An antidote to groundhog day. Programmes are always variants of something that has been attempted before.

- Instead of reinventing (and re-evaluating) the wheel, a feedback or learning loop can be created retrospectively, by synthesising the evidence from previous inquiries on the same family of programmes.

- Synthesis should be *explanatory* not *enumerative*. It should build theories of within programme and between programme heterogeneity.

- Feedback can then be applied in the development of any new programme. The ideal research strategy is probably – **explanatory synthesis + focussed evaluation.**
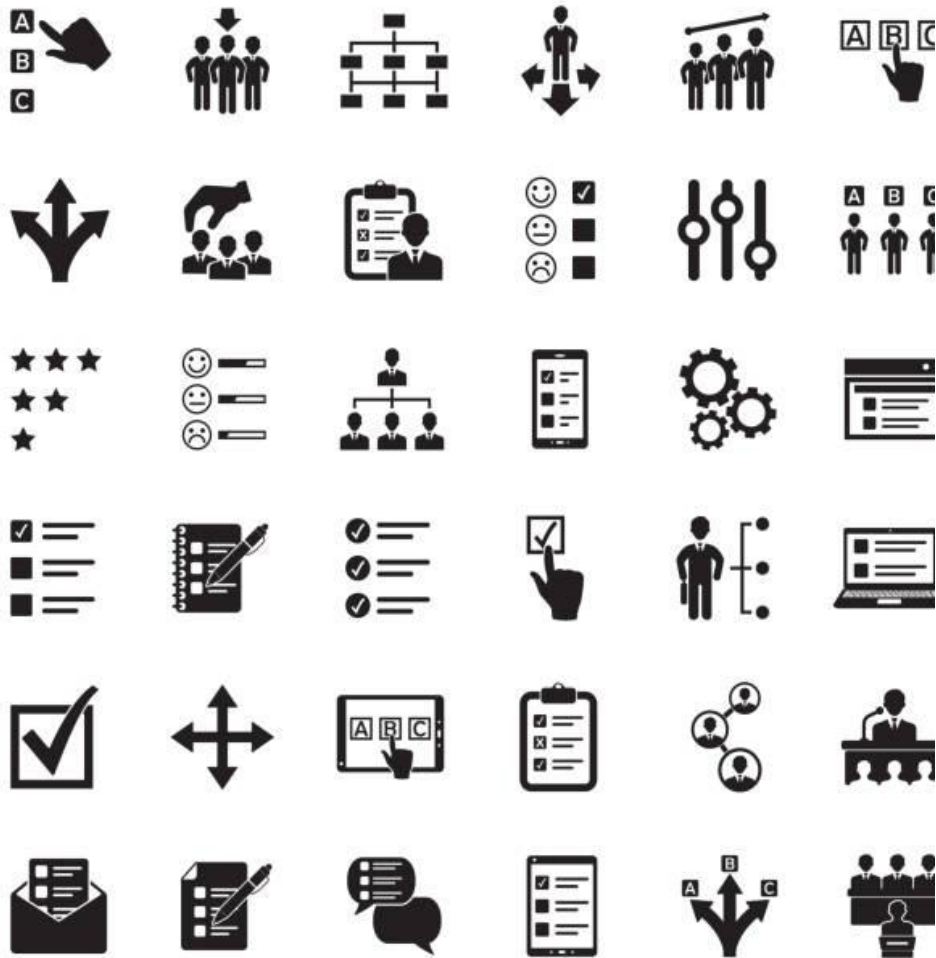
Evaluation research is too closely aligned with 'program evaluation'.

Other policy instruments (regulation, legislation, tribunals, management reforms, public inquiries, fiscal instruments, capacity building, work routines, etc.) are just as interesting, potentially more powerful, and yield to the MCC caveats -'what works, for whom in what circumstances, in what respects, at what cost, with what sustainably, how implemented, ETC>ETC>ETC.

The research method remains the same. Lay out the policy assumptions (theory) in detail and put each one to research.

Shift resources from 'ex-post' to 'ex-ante'.

'Thought experiments' are more feasible than real experiments and potentially more useful. Ex-ante evaluation occurs in the form of 'scoping studies', 'front-end analysis', 'policy scrutiny' as well as feedback from 'research synthesis'.

Programmes are somewhat febrile. They follow on from crises and failures. The working model is to 'do something', pass over responsibility for its implementation, the details of which are settled 'on the hoof'. Ex-ante consideration is thus limited.

A better model? *Legislation* is not easily reversible and thus undergoes line-by-line, parliamentary debate and scrutiny before ending up on the statute book. The rationale is to squirrel out potential loopholes, unforeseen barriers and unintended consequence in order to refine legislation *before laws are passed*.

Incorporate institutional history:

'History doesn't repeat itself, but it rhymes'
Mark Twain

Key policy agencies grapple with the same enduring issues, making endless tweaks and adaptations. Think of 'welfare reforms' or 'heath service re-organisations', etc.

Rather than evaluating only current initiatives, tracking institutional histories will uncover stubborn background causes, unintended consequences, unforeseen errors and partial victories.

Within-institution synthesis reduces heterogeneity and can build explanations by 'learning from mistakes.'

Attempt the impossible: confront 'wicked-problems'.

Definition: Key stakeholders disagree fundamentally about the nature of the problem and thus about any potential solution. The problem is beyond the control or responsibility of any single agency. Policy solutions are unlikely to turn on bigger cash injections or on untried, undreamt-of interventions.
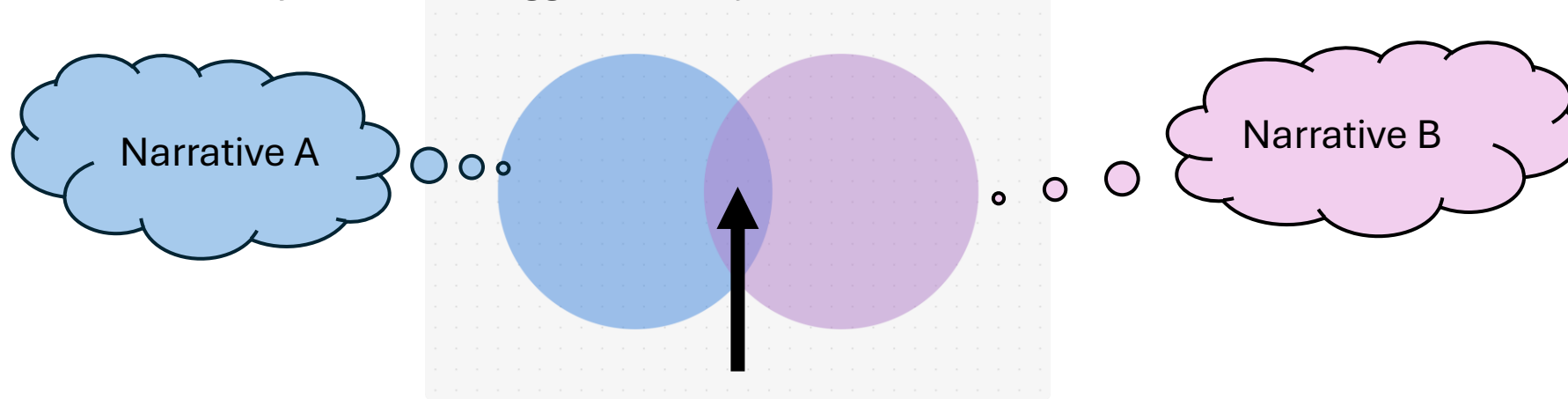
Solution: 'Small wins' framework. Attempt accommodation by gradually 'changing the narrative'. Research examines different stakeholder perspectives, probing *for whom, in what circumstances, in what respects, with what support, with what longevity, that some* of the rival beliefs may be brought together.