

# Evaluating complex programmes

*Erik Arnold*

*Technopolis, KTH and MIoIR*

*Stockholm 14 February 2018*

---

## Road map

---

- The problem
- Complexity
- Transitions – the extreme end of the problem
- Evaluation perspectives
- Where do we go from here?
  
- Discussion

## Road map

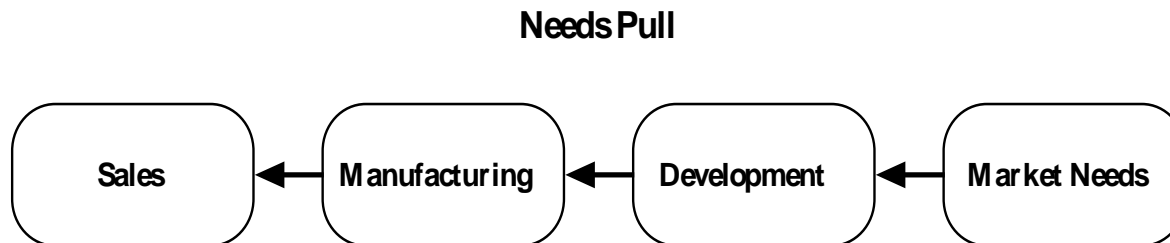
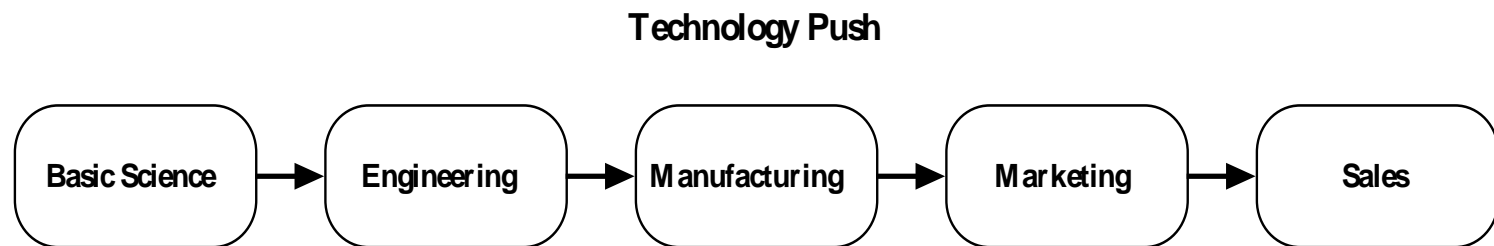
---

- **The problem**
- Complexity
- Transitions – the extreme end of the problem
- Evaluation perspectives
- Where do we go from here?
  
- Discussion

## Why are we interested in how to evaluate ‘complex’ programmes?

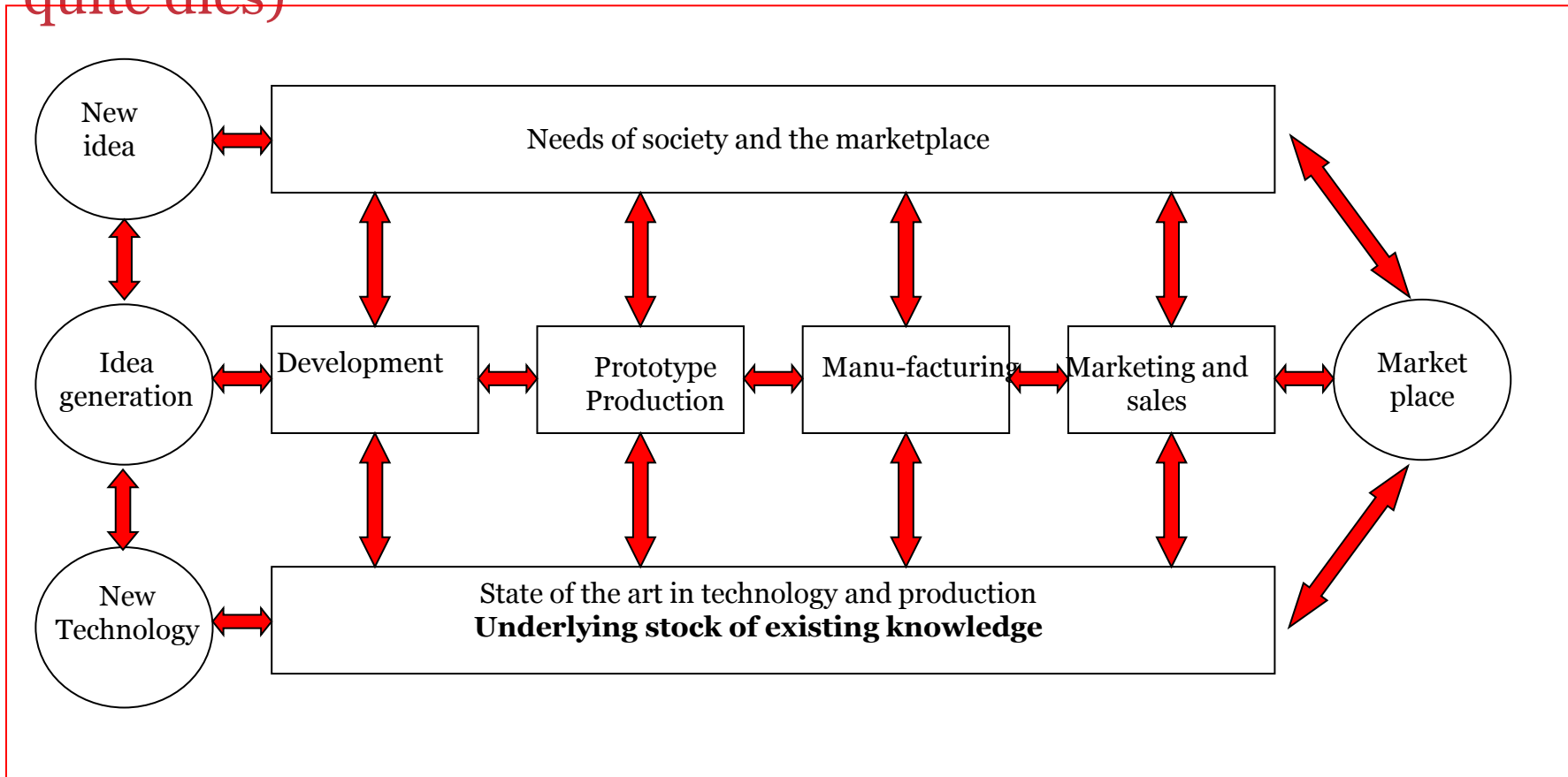
- There is a resurgence of high-level national strategies for industry, enabling technologies and innovation
- Following the Lund Declaration, policymakers are increasingly interested in addressing ‘societal challenges’
  - *Cross-sectoral and interdisciplinary in nature*
  - *Large-scale and requiring wide societal engagement*
- These and other large-scale interventions tend to involve multiple ministries and agencies – therefore we need common evaluation strategies and framework
- They tend to be dynamic and to involve learning
  - *This makes governance more complex*
  - *And means we have to think more explicitly about evaluation governance and how evaluation ties into the evolution of the intervention*

We get to this place after a history of fascination with linear innovation models ...

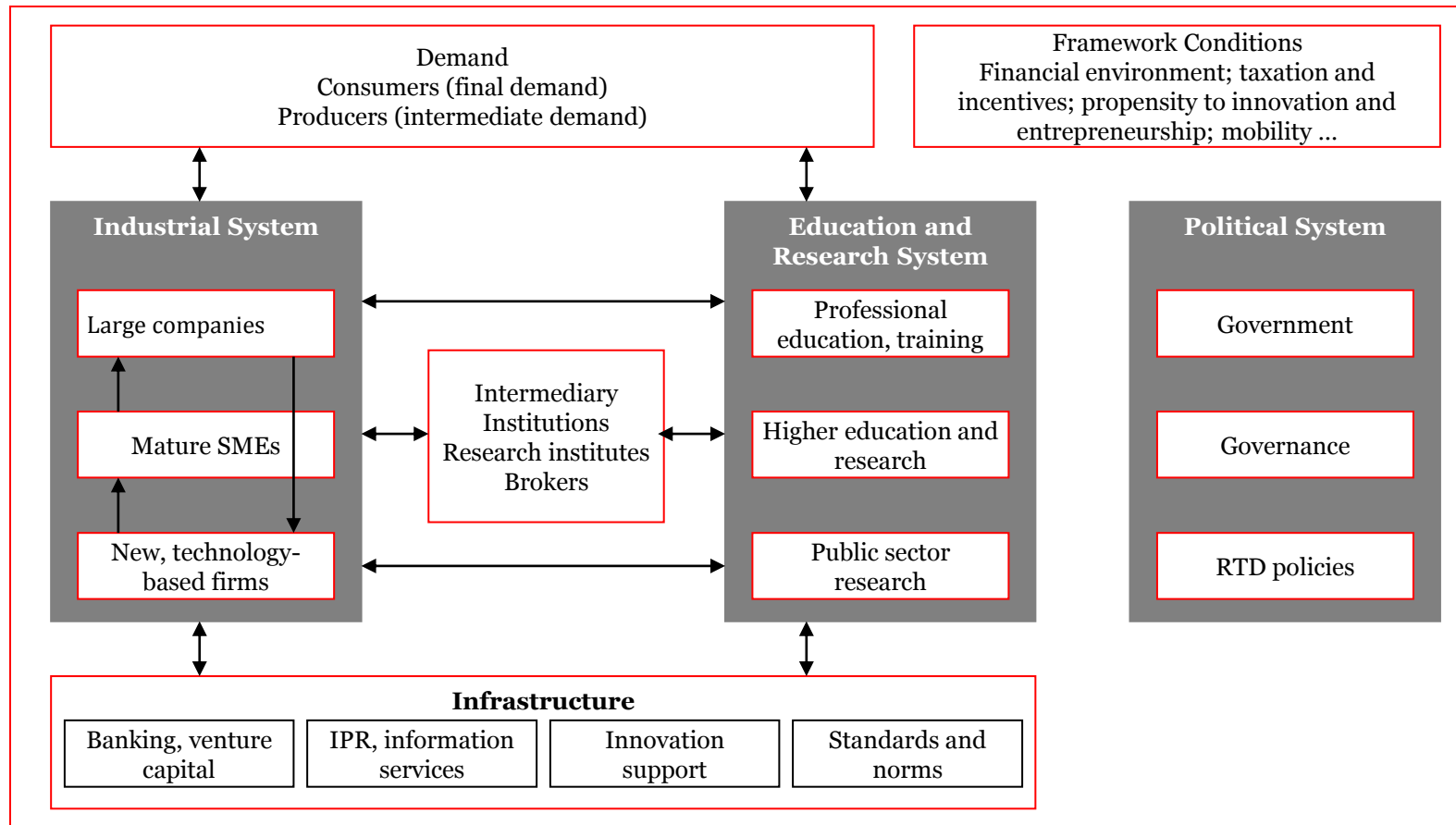


technopolis<sub>|group|</sub>

... which have largely been rejected in favour of more complex, systemic ones (though the old linear model never quite dies)



# The complexity of innovation drives us to think in terms of National Innovation Systems: here's a structural view

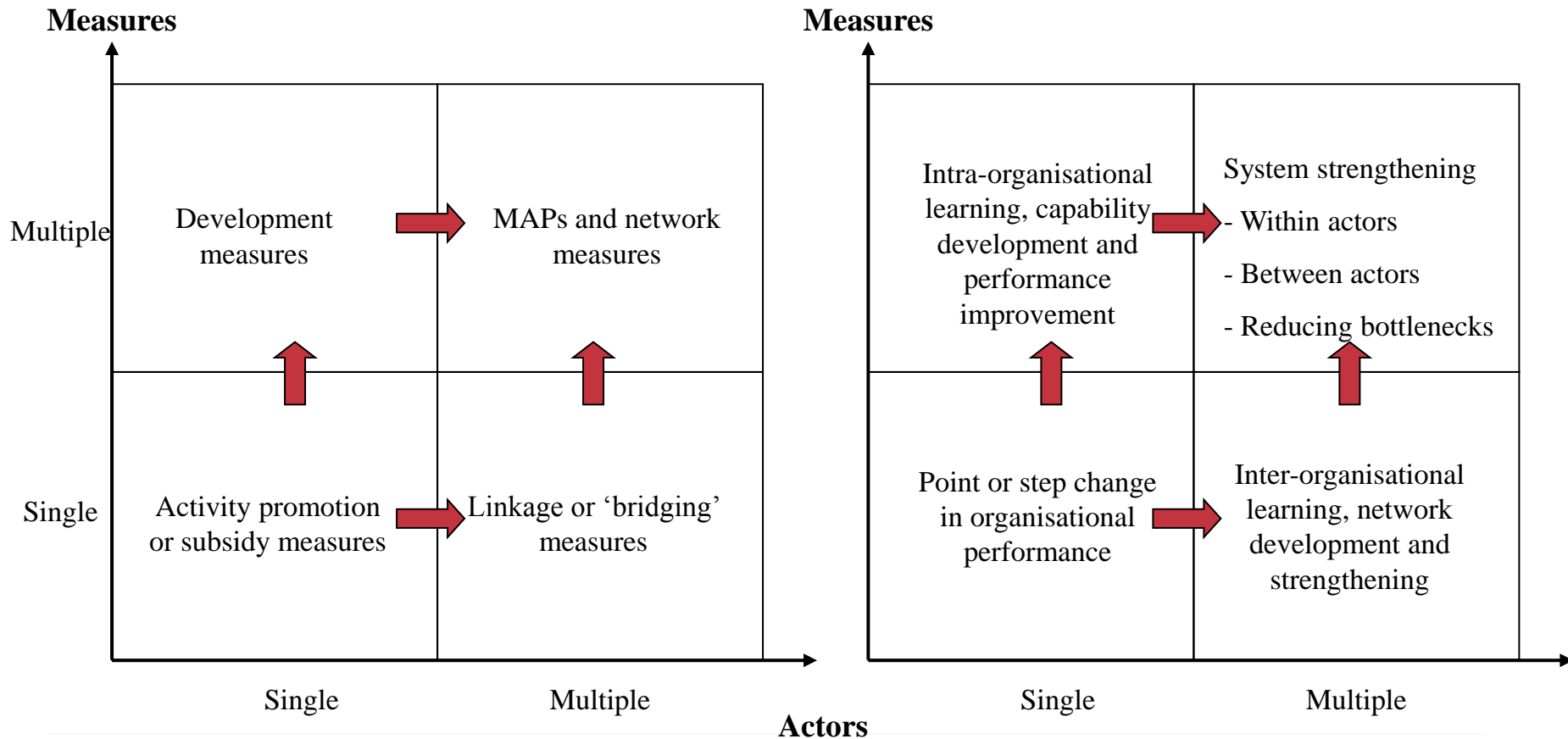


## These ideas interact with how policymakers act

- 1950s/60s, ‘science push’ innovation policies focusing on research
- These expanded to include technology-push ‘grands projets’
  - *Some successful, like French atomic power, Airbus, often relying on ‘development pairs’ where the state controlled supply and demand*
  - *Others disastrous like Concorde or the Plan d’Action pour la Filière Electronique’, which ignored demand and existing market power*
- 1970s, growing understanding of the centrality of producer-user relations in innovation
  - *SAPPHO (1972), Lennart Elg (IVA) + others in the STU debate, von Hippel (1976), Mowery & Rosenberg (1979)*
- 1980s – ‘national’ technology programmes that partly misunderstood the Japanese model (Alvey, ESPRIT, IT4 ... )
- Since then, a growing aversion to ‘picking winners’ – refocusing on clusters and ecosystems (implying a need for reflexivity)



Since the 1980s, interventions have become more complicated, evolving into 'Multi-measure, multi-actor' (MAP) programmes



## Three generations of innovation system governance – sedimentary layers in institutions and policy

- Post-WW2 ‘blind delegation’ to the scientific community based on the linear model (Bush)
  - *Disconnect between research from innovation*
- ‘Science policy’ (OECD) and eventually ‘innovation systems’. Innovation policy as industry policy
  - *Requires a holistic approach with growing focus on coordination across ministries and sectors and on institutional performance*
- ‘Societal challenges’ whose resolution requires various degrees of transition between socio-technical systems
  - *Engagement of more stakeholders (many from outside the innovation policy sphere) to create consensus about directions of travel and enable implementation*

## Three generations of 'failure' justifications for intervention

---

### Market failure - often about basic research

- Indivisibility
- Inappropriability
- Uncertainty

• Nelson, 1959, Arrow, 1962

### Systems failure - mostly about inadequate performance

- Capability
- Institutional
- Network (including lock-in failures)
- Framework

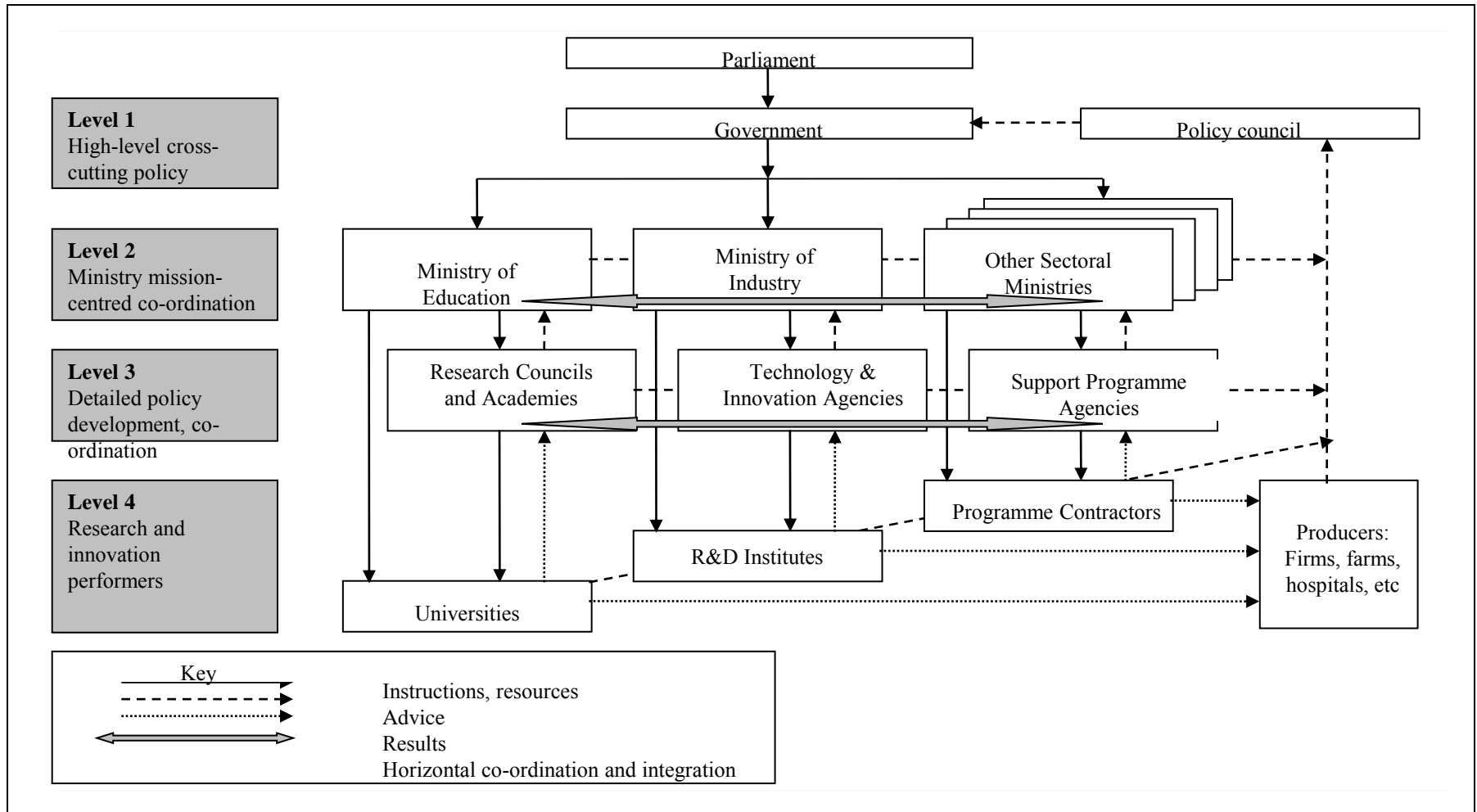
• Smith, Arnold, many others ...

### Transition failure - mostly about inadequate performance

- Directionality
- Demand articulation
- Policy coordination
- Reflexivity

• Weber & Rohracher, 2012

# Coordination mechanisms from second-generation governance are probably not up to the job in the third generation



## Road map

---

- The problem
- **Complexity**
- Transitions – the extreme end of the problem
- Evaluation perspectives
- Where do we go from here?
  
- Discussion

## Complexity – interventions may be complicated; the systems on which they operate can be complex

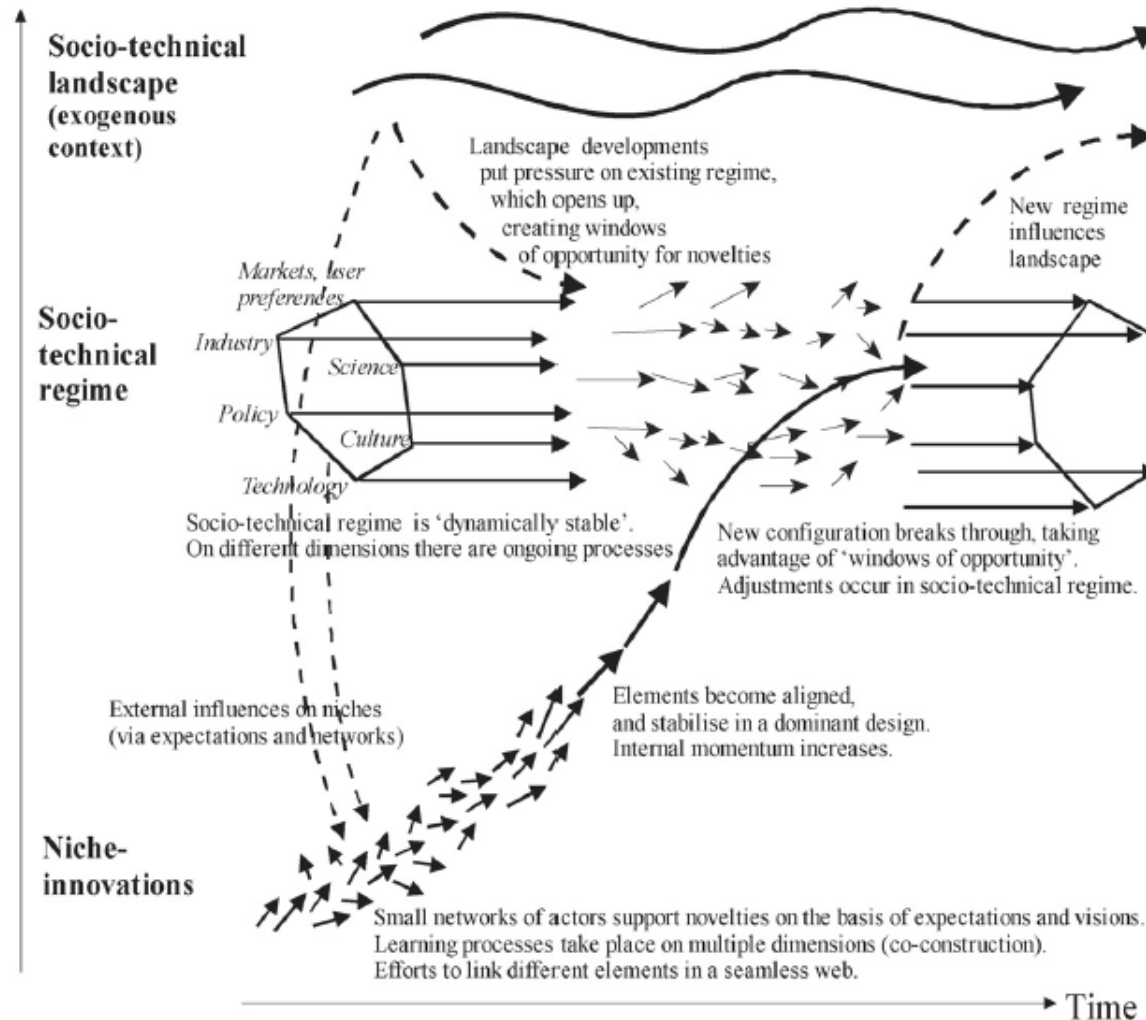
- Roots in the early history of computing
  - *Norbert Wiener, Cybernetics – Control and Communications in the Animal and the Machine. New York: John Wiley, 1948*
  - *Ludwig von Bertalanffy, General Systems Theory, New York: George Braziller, 1968*
  - *Also worth reading: Lars Ingelstam, System – att tänka över samhälle och teknik, Eskilstuna: Statens Energimyndighet, 2002*
- Complexity repeatedly pops up as an issue in policy and social science – but hasn't (yet?) made much of a difference there
- Key concepts
  - *Non-reductionism*
  - *Feedback leading to systemic change*
  - *Emergent properties*
- So: 'complex' is not the same as 'complicated'

## Road map

---

- The problem
- Complexity
- **Transitions – the extreme end of the problem**
- Evaluation perspectives
- Where do we go from here?
  
- Discussion

## Transition literature often uses a multi-level framework



Geels and Schott, 2007



## Barriers to systems innovation/transitions

- Good summary by OECD/Geels (2015)
  - *Over-reliance on market failure rationales*
  - *Short-term political processes (election cycles)*
  - *Fragmented, multi-layered institutions, governance structures and processes*
  - *Technological trajectories and lock-in*
  - *Market power and political clout of incumbents*
  - *Lack of customer acceptance and adoption*
  - *Institutional inertia and path dependency*
- Also important
  - *Absolute costs of change and long periods before obtaining RoI*
  - *Uncertainty and risks associated with disruptive innovation*

## Policy implications

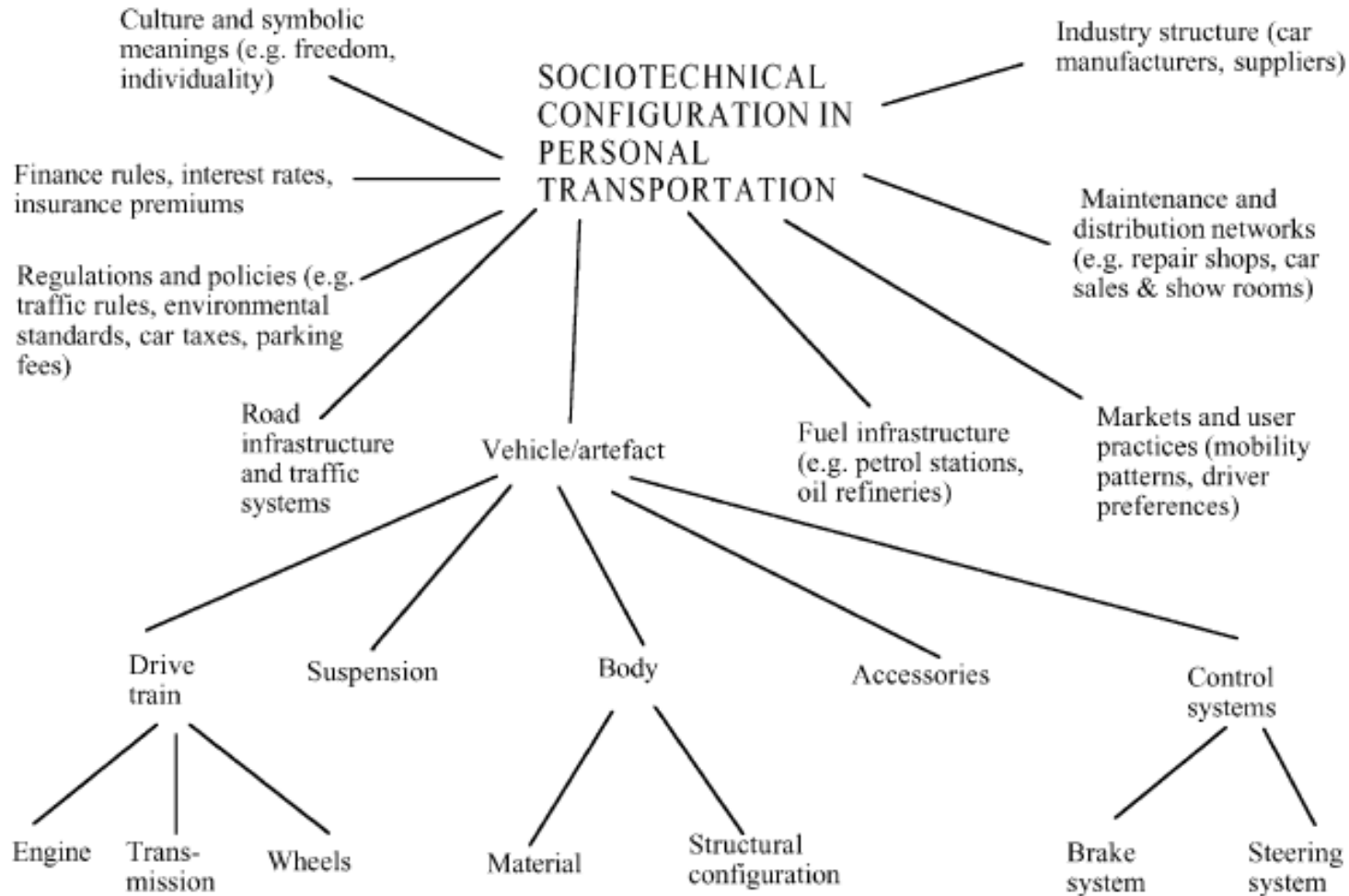
- Understand the systemic nature of the problem, the role and feasibility of architectural change
- Need clear focus from political and administrative levels
- Create shared visions and consensus among stakeholders
- Figure out how to manage and overcome (deliberate) resistance, including by building social capital behind the transition
- Develop change agency and coordination capacities in the administration
- Intensify the collection and analysis of strategic intelligence
- Develop a transition strategy
  - *For example, challenge a dominant design*
  - *Then put in place the system elements needed to support new ones (OECD, 2015)*

## Dimensions of sociotechnical regimes relevant in transitions

---

- Technology
- User practices and application domains (markets)
- Symbolic meaning of technology
- Infrastructure (e.g. physical, knowledge)
- Industry structure
- Policy
- Techno-scientific knowledge (Geels, 2002)

## Changing socio-technical regime involves more than we tackle in conventional R&I or innovation systems policy



Geels, 2002

Fig. 1. Elements from the sociotechnical configuration in transportation.

## Transitions evoke **functions** in Technological Innovation Systems that we don't normally include in innovation policy

1. Entrepreneurial activities: entrepreneurs realise the potential of new knowledge networks and markets.
2. Knowledge development: knowledge is developed by learning and R&D.
3. Knowledge diffusion through networks: it is essential to exchange information in networks. Not only within the R&D setting, but also between R&D, government, competitors and the market. Policies can be adjusted to the latest technology and R&D agendas can be modified.
4. Guidance of the search: guidance is needed because the resources are almost always limited. Guidance is also needed from a social perspective. The society has to adjust itself, or needs to be adjusted, to the new technology/innovation.
5. Market formation: a new technology often has difficulties in competing with established technologies. This issue can be addressed by the formation of temporary niches.
6. Resources mobilisation: both financial and human capital are needed as inputs to activities within the innovation system.
7. Creation of legitimacy/counteracting resistance to change: the technology has to become part of the incumbent regime or even overthrow it. (Hekkert et al, 2007)

## How we operationalised this in the OECD Finland innovation review: A vision to coordinate and prioritise

- A high-visibility national visioning exercise with whole-of-government commitment
  - *Defining and addressing the societal challenges that provide innovation and growth opportunities for Finland*
  - *Building on Finland's strong record in foresight and governance*
- Broad engagement across sectors and parts of society: 'we are all in one boat'
- A wide-ranging public process, guided by foresighters, road mappers and government and supported by analysis of how the Finnish system could support alternative strategies
- Generating wide commitment to a set of priorities – while not ignoring the continuing need for parts of the innovation system to be governed using first- and second-generation techniques
- Link global societal challenges to industrial renewal and business opportunities.

## Changes needed in policy: time to go on the offensive

- Reactive → Proactive
- Retrenchment → Supporting R&I-driven growth
- Fragmented → Systemic
  - *Involving all relevant actors*
  - *No important gaps, eg strategic research*
- Silo'ed → Co-ordinated
- R&I actor focused → Societal, platforms, networks
- Incremental → Radical

## Changes **not** needed in policy

---

- Ignoring existing assets and comparative advantages in favour of green fields
  - Abandoning aspects of policy from earlier governance generations that provide the foundations for growth
    - *cp Tekes, Academy, VTT*
  - Abandoning systemic policy in favour of simple ‘either/or’ solutions
-



## Use PPPs to guide the trajectory and implementation for each challenge

---

- Trigger PPPs involving many stakeholder groups through competitive processes, not top down
  - Develop Strategic Research and Implementation Agendas in the context of the wider societal changes needed in each case
  - Build on experience to evolve a functioning model
    - *National experience in bio-economy, healthcare and SHOKs*
    - *International experience such as Sweden's Strategic Innovation Areas*
    - *Experiment in mainstream policy formation – perhaps invite SITRA to support with further policy experiments*
  - Take great care with governance: PPPs bring many of the risks we associate with principal-agent relations
-

## Road map

---

- The problem
- Complexity
- Transitions – the extreme end of the problem
- **Evaluation perspectives**
- Where do we go from here?
  
- Discussion

## Most of the serious evaluation community uses a ‘realist’ perspective

- Realism asserts that both the material and social worlds are ‘real’ in the sense that they cause effects
- All enquiry involves seeing the world through particular theoretical ‘lenses’ so there is no ‘final’ truth or knowledge
- Social systems are open systems. Hence, a programme interacts with its context and its systemic role has to be considered; the boundaries of the system to be evaluated are not ‘given’ but must be chosen by the evaluator; and the relevant systems and boundaries may change over time
- Causation results from the interaction of intervention and context. (The role of the context may be hard to observe without comparing similar interventions in different contexts.)
- Context affects which impact mechanisms operate and whether they operate

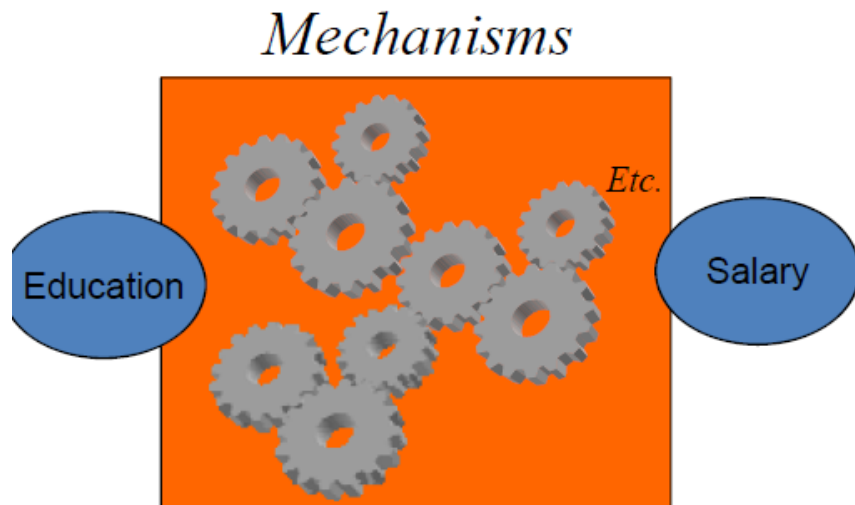
## Realist Evaluation

---

- Experiments can identify the mean effect but this is rarely, if ever, evenly produced.
  - “*what works, for whom, and in what circumstances?*” and even better, to also help us to understand “why?”
  - Mechanism + Context = Outcome
  - Realist approach better when main purpose of evaluation is “*informing the development of policy and practice*”.
  - See Ray Pawson and Nick Tilley, *Realistic Evaluation*, London: Sage, 1997 – Selling point: the only funny book in the evaluation literature
-

## Generative causation:

- the study of mechanisms that lead to causation



What happened after education was provided and how did it lead to salary?  
For whom, under what circumstances?  
How, why so?

Realist inquiry opens the ‘black box’ of program implementation

-- key to determining root causes of program outcomes.

Develop “**Context-Mechanism-Outcome (CMO)**” Configurations.

## What's happening in the mainstream evaluation literature?

- Endless incremental changes to evaluation tools, branding and territory-marking with little sense of much progress (Zzzz...)
- Interest in systems and complexity,
  - *cp Patton, M. Q. (2011). Developmental evaluation: Applying complexity concepts to enhance innovation and use. New York: Guildford Press.*
- Growing focus on mixed methods and triangulation
- Need for better stakeholder involvement – especially in complex systems where there is learning and the intervention logic evolves
- More interest in participative evaluation to generate social legitimacy
- Multi-level intervention governance implies interaction with evaluators as well as stakeholders at different evels
- Evaluation governance is therefore more important

## Where's the action in evaluating impacts?

- Following the money
  - *Computable general equilibrium (CGE)*
  - *Production functions*
  - *(Micro)econometrics, control group analysis*
  - *Cost-benefit analysis*
  - *Randomised Control Trials*
- Understanding impacts
  - *Tracing*
  - *Surveys, interviews*
  - *Case studies*
- Focusing on particular outcomes
  - *Human capital*
  - *Scientometrics*
  - *Altmetrics and webometrics*
  - *Social Network Analysis*
- **Impact assessment for performance measurement**
  - *Eg university performance-based research funding systems*

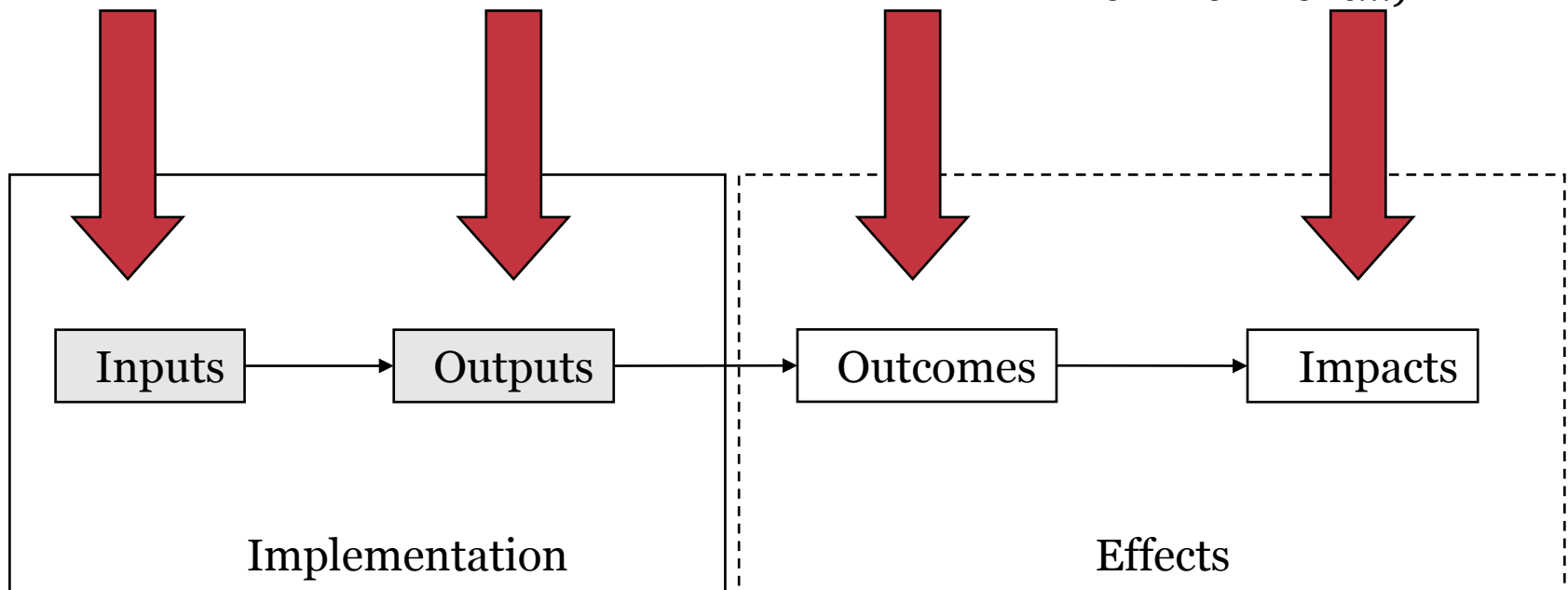
## We use theory-based evaluation, which tests the predictions that the programme designers originally made

Finance,  
organisation,  
legal framework

Results directly  
produced using  
the inputs

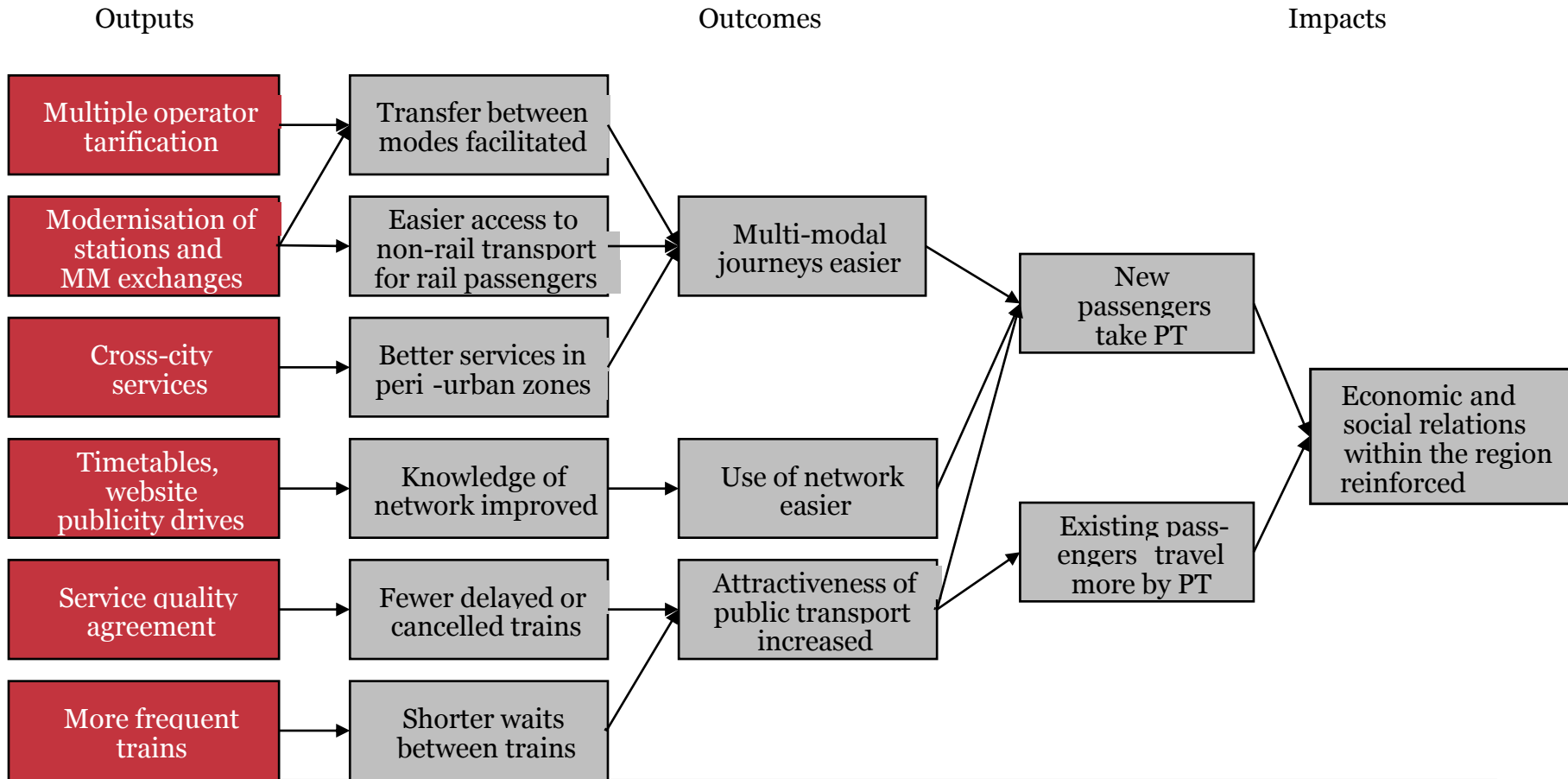
Consequential  
effects on  
beneficiaries

Longer-term  
effects on 'indirect'  
beneficiaries (society, the  
economy, the  
environment...)

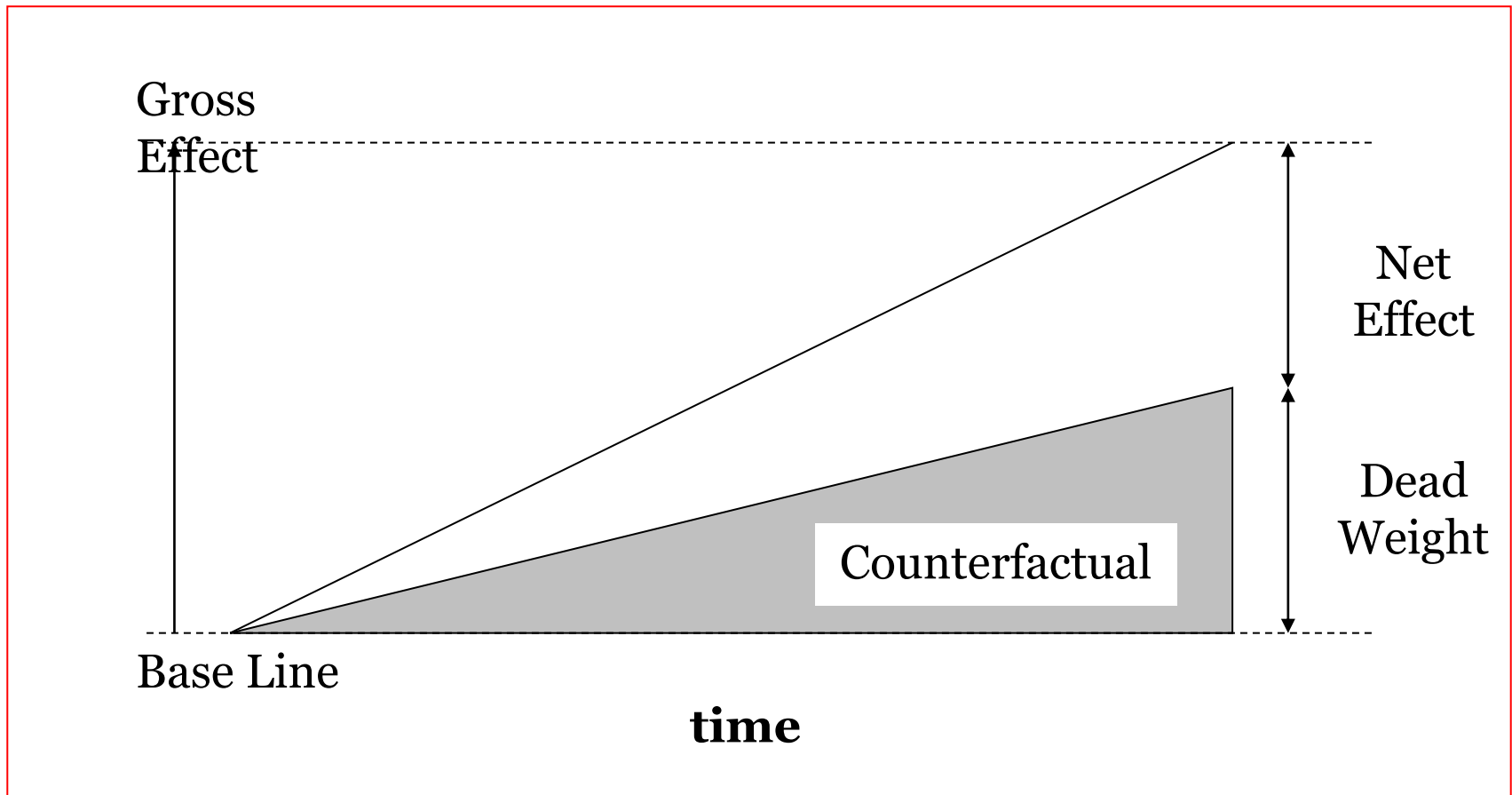




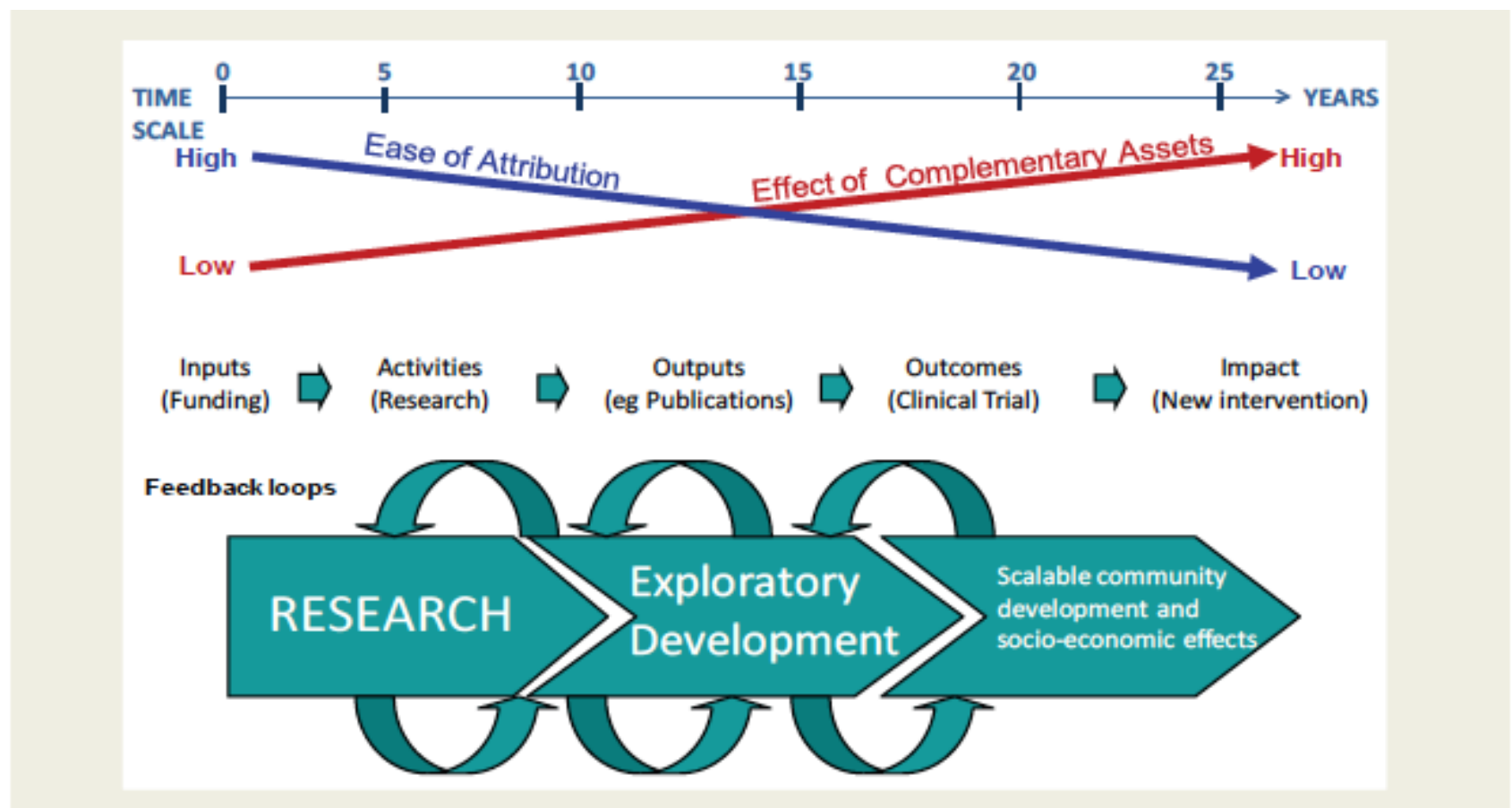
# Complexity challenges traditional theory-based evaluation because reality changes during the intervention



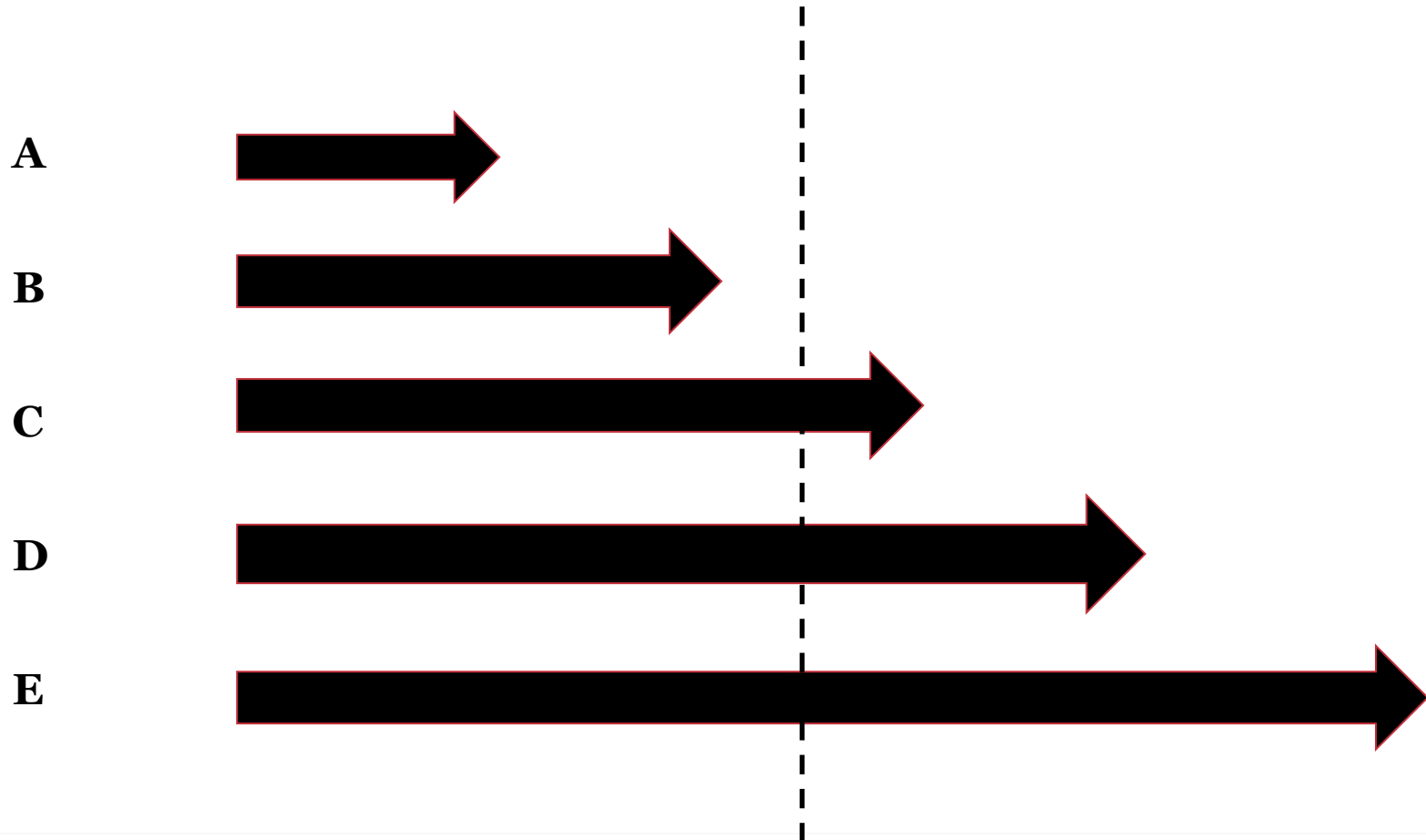
## The philosopher's stone: net effect



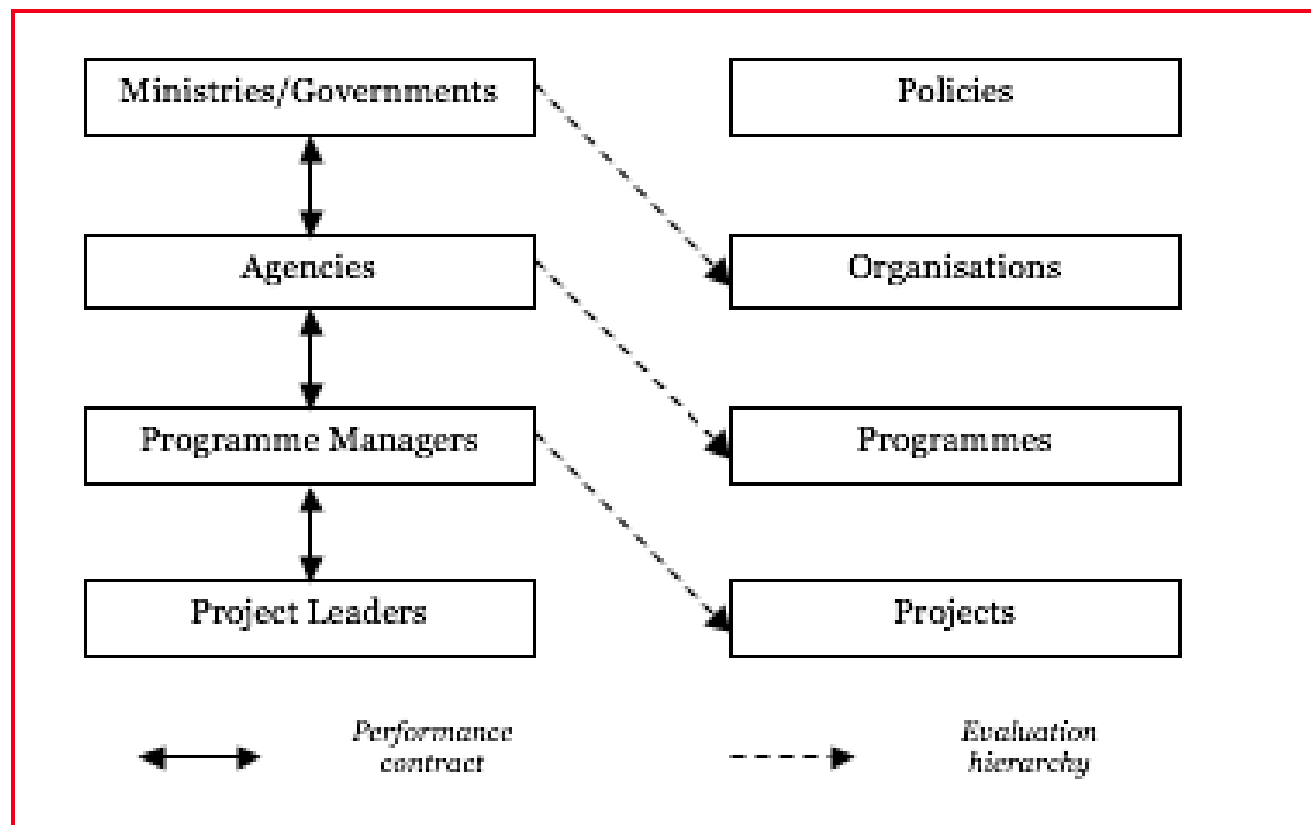
## The joy of attribution ...



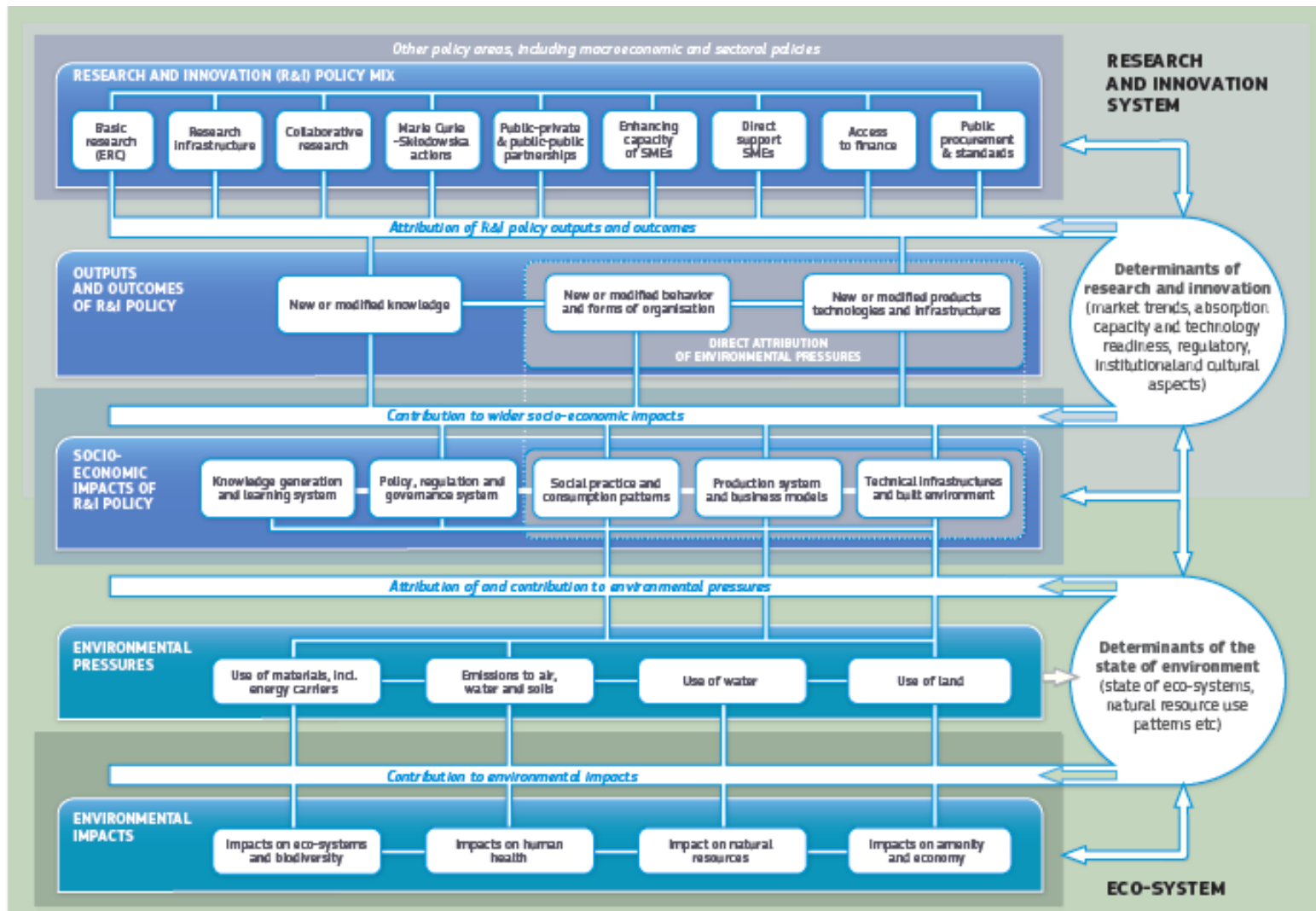
As in so much of life, timing can be everything ... big strategies and transformations are inherently long term





‘Waterfall principle’ in evaluation, so no-one evaluates the politicians ;-). For big things, we need to go beyond this



# We can extend theory-based evaluation beyond R&I. Miedzinski et al: Cumulative Policy Impact Assessment



## We can use the I-O-O-I model to think about evaluation at multiple levels – connecting with transition governance

Objectives	Inputs	Activities	Outputs	Outcomes	Impacts
<b>Strategy</b> Improve system performance by increasing and balancing sub-system performance	Strategic intelligence Governance/steering	Studies, monitoring, evaluation Influence on budget	Innovation strategy Other policy documents	Improved performance by agencies Improved performance by NIS sub-systems	Improved NIS systemic performance Contributions to overall welfare, quality of life
<b>Organisation</b> Improve performance of sub-system(s)	Strategic intelligence Management Budget 	Programmes	Improved performance by specific beneficiary sub-groups	Improved sub-system performance	Contributions to improved NIS systemic performance
<b>Programme</b> Improve performance of sub-system components	Programme management Money 	Projects	Knowledge <ul style="list-style-type: none"> <li>• For beneficiaries</li> <li>• Public goods</li> </ul>	Improved performance by specific beneficiary sub-groups	Contributions to improved NIS sub-system performance

Diagrammatic connections: Red arrows point from 'Strategic intelligence' to 'Management' and from 'Programme management' to 'Money'. Black arrows point from 'Improved performance by specific beneficiary sub-groups' to 'Improved sub-system performance', and from 'Improved performance by specific beneficiary sub-groups' to 'Contributions to improved NIS sub-system performance'. Black arrows also point from 'Improved sub-system performance' to 'Contributions to improved NIS systemic performance'.

## Generic Evaluation Issues at Three Levels

	<b>Relevance</b>	<b>Efficiency</b>	<b>Effectiveness</b>	<b>Utility</b>	<b>Sustainability</b>
<b>Strategy</b>	Do the objectives of the strategy reflect national needs?	Has the strategy been developed in an efficient way and with high quality?	Is the strategy deployed efficiently and effectively? Are sub-goals being achieved?	Have increases in the performance of the NIS satisfied the national needs originally identified?	Are improvements in the performance of the NIS based on structural changes so that they are likely to be permanent?
<b>Organisation</b>	Do the objectives of the organisation correspond to the needs of the sector with which it deals?	Does the organisation design and implement programmes that work in efficient ways? Does it spend the right amount in administration?	Do programmes reach their goals and increase sub-system performance? What is the overall effect of the agency, over and above programme goals?	What are the effects of the organisation on the overall performance of the sub-system (sector) that it addresses?	Are improvements structural in nature? Have needs changed?
<b>Programme</b>	Do programme goals match identified needs?	Does the programme meet its objectives in a cost-efficient manner?	To what extent does the programme meet its goals, especially in relation to beneficiaries?	Does the programme solve the problem it was intended to address?	Is this a permanent solution? Have needs changed?



## Road map

---

- The problem
- Complexity
- Transitions – the extreme end of the problem
- Evaluation perspectives
- **Where do we go from here?**
  
- Discussion

## Some clues ... but the literature is pretty thin on ideas for evaluating big innovation strategies and transitions

### Methods

- Reflexive monitoring – which does what it says but isn't evaluation
- Applied Systemic Programme Evaluation Framework- focusing on learning and interaction rather than pre-defined impact
- Contribution analysis – abandons attribution and looks for contribution – but is massively labour-intensive

### Approaches

- Process focus
- Flexibility and adaptiveness
- Timing – getting closer to real-time evaluation
- Redefining the role of the evaluator – 'getting your hands dirty'
- Learning-focused evaluation

## What will we need to do?

- Long-term evaluation strategies with governance linked to interventions
- Evaluation frameworks spanning multiple sub-interventions, organisations and levels
- Use of prospective as well as retrospective analysis
- Address the fuzzy boundary between evaluation and programming
- Multiple methods, triangulation – more participative, learning evaluation
- Avoid ‘capture’ of evaluators by beneficiaries or policymakers over time
- Secure adequate evaluation budget and independence of evaluators, especially in PPPs, where there is a principal-agent problem

- The problem
- Complexity
- Transitions – the extreme end of the problem
- Evaluation perspectives
- Where do we go from here?
- Discussion – på Svenska!!