technopolis |group|

How should we evaluate complex programmes for innovation and socio-technical transitions?

technopolis|group|

How should we evaluate complex programmes for innovation and socio-technical transitions?

technopolis |group| June 2018

Erik Arnold

Tomas Åström

Charlotte Glass

Marika De Scalzi

Summary

This report to the Swedish Agency for Growth Analysis responds to the requirement – partly driven by the need to address the so-called 'societal challenges' – to evaluate much larger and more complex research and innovation programmes than we have tackled in the past. These comprise, on the one hand, complex innovation programmes such as national innovation strategies and, on the other, programmes that address the need to make transitions in socio-technical systems, such as removing fossil fuels from the electricity supply system. Together, we refer to these as complex innovation and transition programmes (CITPs).

Our report is based on reviews of the evaluation literature and the literature on socio-technical and sustainability transitions. In addition, we have done case studies of three CITPs in order to understand some of the emerging experience with these and interviewed a number of relevant Swedish policymakers.

Traditionally, we govern research and innovation partly by asking the scientific community itself to set the agenda for some of the work it does and partly by providing funding that deliberately links research and innovation to societal needs. These governance modes will not go away.

CITPs are large, long-lasting and require involvement across many parts of government as well as stakeholders (such as consumers) who have not traditionally been involved in research and innovation policy. To handle them, we must find a new mode of governance that covers many ministries but also many stakeholder groups. Typically, CITPs have an overall strategy but multiple actors are involved in implementation, so they tend to need multi-level governance and a multi-level, independent evaluation system that stands outside the governance structure keeping the CITP accountable to society while also providing feedback on how to improve the programme and its implementation.

There are well developed evaluation methods for tackling simple interventions that have a linear logic but it is clear that we need a more sophisticated approach with CITPs. These are 'complex' in the sense that they deal with systems of people who learn and change their behaviour during the course of the programme. Many CITPs 'experiment' with different ways to do implementation. So evaluation methods must now be able to cope with understanding causality in a changing reality with learning and sometimes changing goals. Some currently-used and generally quantitative evaluation techniques tell us little about the mechanisms of change. Some newer evaluation approaches focus on giving feedback to programme managers and stakeholders rather than to society. We need a 'realist' approach that develops theories of change to explain why programmes succeed or fail, uses mixed methods in order to be robust and undertakes both of the twin tasks of public policy evaluation: providing accountability to government and society; and feedback to those working with the programme to help them improve.

The community that researches 'transitions' today does not share this view but focuses on providing feedback. It has developed a rich conceptual understanding of transitions and their management that increasingly is beginning to have an empirical base. But we argue here that this understanding has to be made accountable to government if it is to work in the interests of society and not just the interests of programme managers and beneficiaries.

Evidence from the three cases studied in this report – the UK Climate Change Act, 2008, the Netherlands Top Sector Policy and the German High-Tech Strategy – is largely consistent with what we see in the broader literature. Large, complex programmes can only be effective if there is longer-term political commitment. They need to be connected to a multi-level governance that stretches across government and wide range of external stakeholders, including those with social authority. Programmes require a management organisation that has legitimacy and funding. An independent, multi-level evaluation is needed, providing accountability as well as information for learning. Strategies need periodically to be reviewed and revised in a transparent way, taking account of evaluation and other evidence.

Policymakers want robust evaluations that shed light on the systemic effects of interventions as well as their more direct effects. It is not clear how to develop methods to do this, especially as many evaluation customers are risk averse and therefore unwilling to fund experimental work. The problem becomes more acute when large and complex programmes are considered. There is a need to be realistic about what various evaluation techniques can and cannot do and about ways in which they may be used to complement each other. It may be necessary to bring together different evaluation communities and evaluation methods in order to improve practice.

We conclude that an evaluation approach for CITPs should: satisfy the normal needs of public policy evaluation to provide both accountability and advice, answering to society through government; avoid creating new institutions but use ad hoc and temporary structures tailored to the individual evaluation and overlaid on existing organisational responsibilities; explore the mechanics of causation, focusing in a cost-effective way on those components that appear critical; develop existing evaluation approaches so that they can become iterative and evolve as the programme evolves and endeavour to link systemic or macro evaluation with the evaluation of individual sub-programmes at the micro level. Methods development and experimentation will be needed.

Where a CITP can be managed within a single agency (or even a small number of agencies) there is every reason to use Sweden's well-functioning existing structures. On the infrequent occasions when this is not the case, we envisage a system of governance and evaluation that would involve

- The government designating a point in its structure to which the CITP would ultimately answer
- Creating a temporary implementation 'platform' to design and govern the programme, in which a wide range of agencies (or other agents such as cities) and other stakeholders are involved but which has staff resources and a leadership team of its own, able to lead the programme. The platform would coordinate individual interventions managed by existing agencies (or combinations thereof), using their own budgets, resulting in a hierarchical structure of interventions with a reporting structure leading back up to the platform
- The platform would be responsible for revising the strategy periodically, informed by evaluation not only at the overall level but also at the level of sub-interventions, as well as by consultation with stakeholders and other relevant sources of expert knowledge
- There should be a separate evaluation platform answering to government, not to the implementation platform, and coordinating evaluation at the different levels needed. This enables accountability as well as objectivity. It should devise and implement an evaluation strategy in discussion with government, the implantation platform and the implementing agencies. It should then do (or more probably outsource, using relevant and independent expertise) evaluation at the strategy level on a cycle consistent with the (re)planning cycle of the programme
- Evaluations at the intervention level should be organised by the implementing agencies but must be reported to the evaluation platform and be partially aligned with needs defined buy that platform. This, combined with the use of agency on the operational side, should ensure not only the local relevance of the intervention evaluations but also their ability to feed into the higher-level strategic evaluations.

Table of Contents

Tables

Figures

1 Introduction[1]

We live in interesting times for innovation policy. In addition to the traditional repertoire of programmes and instruments, we can see in recent years the emergence of two important new ways of doing innovation policy. Both aim to make big changes in the way entire systems (national or technological innovation systems, energy systems, healthcare systems and more) operate and perform.

One involves large, **complex innovation programmes** like the German High-Tech Strategy that try to tackle innovation in a holistic way. Policymakers increasingly take a more systemic perspective than before on innovation, growth and development. This in turn has led to increased interest in running large-scale policy interventions such as national industry or innovation programmes spanning the responsibilities of multiple ministries and reaching across several sector of society.

The other and more radically new type of programmes are those that address **socio-technical transitions** , some of which are sustainability transitions, in the sense that failing to make the needed changes could make our species and planet unsustainable. The most visible are programmes addressing climate change. To do this, we need not only to change technology but also human behaviour – hence we describe the needed change as a 'socio-technical' rather than simply a technological transition. Programmes addressing transitions can require even greater coordination across different parts of society (including stakeholders well outside what we have historically regarded as the research and innovation community on the demand as well as the supply side) than complex innovation programmes, creating larger-scale and longer-term programmes that span more actors than before. This applies both horizontally across different sectors of society and also vertically, with for example different activities aimed at climate change mitigation taking place at international, national and city levels.

More broadly, society, policymakers and politicians alike are concerned about the evolving set of issues that we discuss as 'societal challenges': not only climate and energy supply but also questions such as human health and ageing, food security and the health of the biosphere, transport, social inclusion and security. To varying degrees, these require socio-technical transitions.

Individual complex innovation and socio-technical transition programmes inhabit a spectrum of complexity. We can think of complex innovation programmes as one end of the spectrum and, perhaps, climate change at the other, with various other societal challenges in between. Their structure and governance have important similarities, as do the ways in which they need to be monitored and evaluated. For simplicity, we will refer in this paper to the programmes along this spectrum as **complex innovation and transition programmes (CITPs)**.

This report to the Swedish Agency for Growth Analysis discusses how to evaluate CITPs. Evaluating such large and complex programmes is new territory. We are used to evaluating simple programmes involving few actors and with fairly simple goals, though we are not very advanced at relating their effects to changes at the level of whole innovation systems or social systems more generally. This report aims to develop some principles for how to evaluate CITPs and thereby to contribute to the work of the Agency for Growth Analysis as well as, we hope, to that of the policymaking community more broadly.

'Complexity' is key to the challenges that policymakers and evaluators alike face in using CITPs. In principle we can distinguish between complicated and complex systems (though in practice we sometimes find that systems we thought were merely complicated turned out to be complex). The distinction is based on complexity theory. A complicated system may have many components that interact in ways that are not always obvious (as any frustrated PC user knows). The big challenge with such systems tends to be understanding how all the different bits fit together and interact. However,

[1] We gratefully acknowledge the contributions of Lars Bager-Sjögren, Katharine E Barker, Enrico Deiaco, Jakob Edler, Fraser McLeod and Peter Svensson to the development of this paper through comments and discussion. The paper nonetheless is the sole responsibility of the authors. Its contents should not be read as necessarily reflecting the views of the people listed or, indeed, of the Swedish Agency for Growth Analysis.

CITPs tend to be complex in the sense that they contain feedback loops and social actors that together mean we cannot always predict their future behaviour: they evolve over time. Often it is not possible to understand them only by adding up the properties of their sub-systems. Influences at higher levels also affect the way they work. CITPs are complex because they address socio-technical systems: that is, systems that encompass people as social beings on the one hand and technologies on the other. (And we should here remember that technologies are themselves social products, so they already embody the characteristics of particular societies, social and economic systems.)

The terminology relating to large complex programmes is not standardised. While we distinguish between **complex innovation programmes** and **socio-technical transition programmes** or **transition programmes** and when we discuss both types together we refer to **complex innovation and transition programmes** (**CITPs**), the reader should be aware that there are other terminologies.

We anticipate that our readers will come from different parts of the policymaking, evaluation, transitions and academic communities and have different needs for explanation. We have therefore tried in each Chapter to provide sufficient explanation to make the paper useful to a wide range of readers.

The structure of this report is as follows. In Chapter 2, we look at changes in the way society has regulated and governed its relation with research in recent decades. This explains how and why is necessary for the state to intervene in research and innovation and discusses the different types of policy instruments we have used to do so over time, culminating in the type of complex programmes in which we are interested. That interest spans two distinct literatures and two different ways of thinking about the role of evaluation, so we need to discuss them one by one before bringing the experience from each to bear on our overall problem. Therefore we explore in Chapter 3 what the literature on evaluation has to say about evaluating large or systemic programmes before moving on to that literature on socio-technical and sustainability transitions in Chapter 4, which considers how to manage and evaluate programmes that aim to change the way socio-technical systems work. This is the most extreme end of the spectrum of complexity that policymakers and evaluators need to consider in CITPs – so if we can understand how to do the most difficult things we should also be able to understand how to deal with large, complex programmes in general.

Next, in Chapter 5, we report on three international cases of CITPs, so that we get a sense of how their design and evaluation have been done in the past: the implementation of the UK Climate Change Act of 2008, which is a socio-technical transition programme; and the Dutch Top-Sector policy and German High-Tech Strategy, which are complex innovation programmes. These are summarised in the main report, while the full cases are presented at the Appendix.

We interviewed seven policymakers across relevant parts of the Swedish system. Chapter 6 explains the way they see the challenges addressed in this report and some of the issues they see for design and evaluation. Their concerns offer a useful challenge and we try to address that as well as other problems identified here in finally drawing some lessons for future design, management and evaluation of large complex programmes in Sweden in Chapter 7.

2 Research and innovation policymaking and governance

This Chapter discusses the different ways the relationship between research and society has been governed in recent years, arguing that the need to address the societal challenges requires a new form of governance able to deal with large and complex programmes affecting many sectors of society. Each governance generation is justified by the need for different types of government intervention and has been accompanied by the evolution of new and more complicated innovation policy instruments. This development culminates in a need for large programmes such as CITPs and may entail a need for multi-level governance.

2.1 Three generations of research and innovation governance

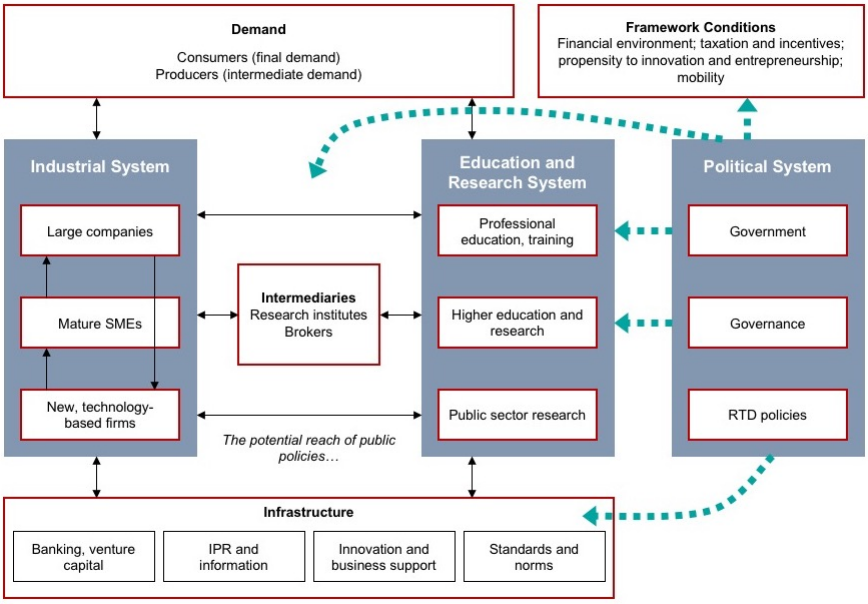
Our focus on CITPs underlines the fact that the relationship between science and society is changing. We can distinguish three generations of research and innovation governance since the Second World War, each tied to a different view of the 'social contract' between science and society. It is perhaps best to view these as generations as sedimentary layers, rather than as complete changes in direction. Today, all three co-exist – each arguably providing a perspective that is relevant to different parts of research and innovation of policy.

The manifesto for the first generation social contract in the post-War years was *Science, the Endless Frontier* (Bush, 1945), when the relationship between society and science was based on 'blind delegation' (Braun, 2003). At that time, we generally believed in the so-called linear model of innovation – the idea that if society funds science and leaves scientists to decide what research to do (often 'basic' research), their work will drive innovation and economic development. We still do this in the way we fund the research councils, perhaps no longer because we believe that the linear model explains the social effects of research but because we understand that we never know enough to plan everything. Therefore, while a lot of research funding today is provided with specific social ends in mind, we would be foolish not also to fund the creativity of the research community. We also do this because basic research plays a large role in scientific education – even for people who end up doing applied research or work outside the research sector – thus providing a huge contribution to society.

The second generation was triggered by a backlash that started in the 1960s, partly led by the OECD, arguing that we needed 'science policy' to direct research towards national, socially-determined goals. It underpinned a period of great technological optimism in the 1960s and 1970s, party driven by great projects like the race to put a man on the moon. New institutions (such as STU – now VINNOVA – in Sweden and the Science Policy Research Unit in the UK) were created in 1968 and others in other countries during the next few years to fund and study socially-relevant research, alongside the continuing work of the (basic) research councils. The following few years were the first time that a significant amount of systematic research was done on the relationship between research, innovation and society. An early realisation was the huge importance of relations between suppliers and users within industry (Rothwell, et al., 1974) but also in the links between science and industry (von Hippel, 1976). The growing understanding of the complexity of the innovation process led initially to the rejection of simple innovation models such as the linear model and then successively to the idea of research and innovation taking place in 'national innovation systems' (Freeman, 1987) (Lundvall, 1992) (Nelson, 1993). This idea emphasis the interdependence of different parts of the system, the fact that people and companies make imperfect judgements because they have 'bounded rationality' (Simon, 1945) and therefore that institutions and learning – not just research – are key to successful performance.

Figure 1 shows one of a number of available representations of a national innovation system, emphasising the structure and main components of such a system and illustrating its comprehensive nature.

Figure 1 A National Innovation Systems Heuristic



Source: (Kuhlmann & Arnold, 2001)

From about 2000, it was increasingly recognised that if innovation systems were as complex as suggested in Figure 1, policy needed to be holistic. Making policy across a system with so many components involved many actors so there was a need for a national coordination and governance system that spanned many if not all ministries as well as other organisations (OECD, 1999). The discussion focused on the role of government and various parts of the state to a greater extent than on other actors and the idea that governance had a 'vertical' component (as between ministries and their agencies) as well as a 'horizontal' component (as among ministries) took root (Kuhlmann, et al., 1999) (Arnold, Kuhlmann, & van der Meulen, 2001) (Edler, Kuhlmann, & Smits, 2003).

Since about 2000, in the third generation, there has been growing concern about the societal challenges. Key triggers at the EU level included a 'manifesto' published in Nature (Georghiou, 2008) and the 'Lund Declaration'[2], from an innovation conference under the Swedish presidency of the EU, which triggered the inclusion of the societal challenges in the EU Framework Programme. Some of these societal challenges are seen as existential for mankind and are therefore reminiscent of similar concerns from the past (Malthus, 1798) (Meadows, Meadows, Randers, & Behrens III, 1972). These were dismissed as premature in the past (Cole, Freeman, Jahoda, & Pavitt, 1973) but are being taken seriously this time in the face of evidence that they already have visible and negative effects. Tackling these challenges generally involves overturning existing technologies, structures and practices in socio-technical systems. Interventions therefore involve complexity at the same time as they need a wider and more difficult form of governance and collective action even than the challenges posed in innovation system governance.

In the post-War period, therefore, we largely relied on the idea that scientists know best about what research should be done and that – if given generous funding and left to their own devices – the results of their work would turn out to be socially useful. We still fund basic research in this way. From the 1960s, society started to make more demands for specific types of results, especially those useful to industrial development, so we developed ways to prioritise particular research themes in addition to funding 'free' or basic research. Over time we have learnt that coordinating different parts of the research and innovation system so as to produce successful innovation is complicated and requires coordination across government, science industry and wider society. We are still a long way from finishing the task of devising the best governance and evaluation structures to do this.

[2] This appears not to have been formally published, but is widely available, including from the Swedish Research Council http://www.vr.se/download/18.7dac901212646d84fd38000336/

The societal challenges exacerbate these difficulties. Now we have to deal with the 'quadruple helix' of government, science, industry and wider society and address the need for far-reaching changes across the responsibilities of different ministries and in the ways many different parts of society function. This means we now have to innovate new styles of governance and evaluation.

2.2 Rationales for intervention

These changing ideas about the relationship between society on the one hand and research and innovation in the other mean that each governance generation justifies policy intervention based on a different set of 'failures' that government needs to correct (Table 1). In the first generation, intervention in research was justified by the idea of 'market failure' (Nelson, 1959) (Arrow, 1962). Since it is hard for capitalists to generate and monopolise knowledge (especially more basic science) in order to make a profit out of it, it makes sense for the state to fund research. The spill-overs from the research to the rest of society (or the 'social returns') easily justify society's investment. The second generation recognises that, if innovation systems are interconnected but imperfect, there is a need for society to invest in combating 'system failures', in order to help the system work better. Discussion of 'failures' in relation to the third generation has only just begun. Those identified so far in the third generation are essentially government failures. The failures we associate with the first two governance generations have strong empirical and theoretical bases. Those of the third generation are more hypothetical and will undoubtedly be further developed. Nonetheless, it is useful to view them at least as prototypes for the ways we will eventually describe third generation 'failure' justifications.

Each of the three generations of intervention thus aims to rectify different 'failures' in the research and innovation system, starting with the market failure that means private firms invest a socially sub-optimal amount in research, going on to address failures within the way our research and innovation systems work and culminating in the need to find ways to reach beyond market forces and intervene to address the societal challenges. Current policy needs to tackle all three types of failure.

2.3 Developments in research and innovation policy instruments

The massive expansion of research in the post-War years involved a mixture of traditional 'mission' policies (health, defence, and so on) and a deliberate focus on 'basic' research, often through the creation and expansion of bodies such as research councils and national science foundations, whose main purpose was to fund researcher-initiated research without any consideration of its societal relevance, based on the idea that the scientists know best what to do. The basic research funders made little effort to influence the directions of research or the structure of the research community. Funding tended to focus on individual beneficiaries and to be provided through small grants. International science cooperations based on expensive infrastructure such as CERN started to appear in this period but there was little national funding aiming to create large research groups or centres.

The emergence of science policy and a greater desire for society to govern science was paralleled in the 1960s by a surge in technological optimism with large technology projects such as the Moon race and the Concorde supersonic airliner. Experience was mixed. The Moon race was an expensive technological and diplomatic triumph that built substantial technological capacity and capability, especially in the USA. Concorde was an expensive economic disaster that proved to be a first stage leading towards the creation of Airbus Industrie.

Many of the early second-generation funding schemes were binary, linking a single company with a single researcher or research group. But by the 1980s, policymakers had acquired a strong aversion to 'picking winners', preferring where possible to promote clusters or networks of technologies and companies in the hope that one or more robust winners would emerge through competition. Influenced by Japanese state R&D programmes, large and complicated 'pre-competitive, collaborative' programmes began to appear in the 1980s, especially in information technology. An early example was the ESPRIT programme, which became the first building-block in the EU Framework Programme (Arnold & Guy, 1986).

Table 1 Overview of failures in three governance generations

	Type of failure	Failure mechanism
Market failures	Information asymmetries	Uncertainty about outcomes and short time horizon of private investors lead to undersupply of funding for R&D.
	Knowledge spill-over	Public good character of knowledge and leakage of knowledge leas to socially sub-optimal investment in (basic) research and development.
	Externalisation of costs	The possibility to externalise costs leads to innovations that can damage the environment or other social agents.
	Over-exploitation of commons	Public resources are over-used in the absence of institutional rules that limit their exploitation (tragedy of the commons).
Structural system failures	Infrastructure failure	Lack of physical and knowledge infrastructures due to large scale, long time horizon of operation and ultimately too low return on investment for private investors.
	Institutional failures	Hard institutional failure: absence, excess or shortcomings of formal institutions such as laws, regulations, and standards (in particular regarding IPR and investment) create an unfavourable environment for innovation. Soft institutional failure: informal institutions (eg social norms and values, culture, entrepreneurial spirit, trust, risk-taking) that hinder innovation.
	Interaction or network failure	Strong network failure: intensive cooperation in closely tied networks leads to lock-in into established trajectories and a lack of infusion of new ideas, due to too inward-looking behaviour, lack of week ties to third actors and dependence on dominant partners. Weak network failure: too limited interaction and knowledge exchange with other actors inhibits exploitation of complementary sources of knowledge and processes of interactive learning.
	Capabilities failure	Lack of appropriate competencies and resources at actor and firm level prevent the access to new knowledge, and lead to an inability to adapt to changing circumstances, to open us novel opportunities, and to switch from an old to a new technological trajectory.
Transformat-ional system failure	Directionality failure	Lack of shared vision regarding the goal and direction of the transformation process: inability of collective coordination of distributed agents involved in shaping systemic change; lack of targeted funding for research, development and demonstration projects and infrastructures to establish corridors of acceptable development paths.
	Demand articulation failure	Insufficient spaces for anticipating and learning about user needs to enable the uptake of innovations by users. Absence of orienting and stimulating signals from public demand. Lack of demand-articulating competencies
	Policy coordination failure	Lack of multi-level policy coordination across different systemic levels (eg regional-national-European) or between technological and sectoral systems; lack of horizontal coordination between research technology and innovation policies on the one hand and sectoral policies (eg transport, energy, agriculture) on the other; lack of vertical coordination between ministries and implementing agencies leads to a deviation between strategic intentions and operational implementation of policies; no coherence between public policies and private sector institutions; no temporal coordination resulting in mismatches related to the timing of interventions by different actors.
	Reflexivity failure	Insufficient ability of the system to monitor, anticipate and involve actors in processes of self-governance; lack of distributed reflexive arrangements to connect different discursive spheres, provide spaces for experimentation and leaning; no adaptive policy portfolios to keep options open and deal with uncertainty

Source: (Weber & Rohracher, 2012)

The Swedish national counterpart was the IT4 programme (Arnold & Guy, 1992) These programmes involved many organisations, consortia rather than binary cooperation and pursued multiple themes and technical goals. In Sweden, Norway and Finland a particularly strong 'technology programme' tradition appeared that involved industrial and academic stakeholders in planning and governance, while leaving the state in overall control. These programmes involve multiple actors and multiple measures (Baumann, et al., 2004) but are essentially focused on researchers and innovators.

The period since about 2000 has seen the emergence of instruments that are larger, longer term or have wider scope than multiple-actor programmes. Up to this point, research and innovation policy focused on the 'supply side', in the sense of the research-performing organisations and industry. The re-emergence of demand-side policies such as innovation procurement, establishing lead markets and using standards to stimulate innovation involved integrating actors on the demand side into innovation policy (Edler, Geourghiou, Blind, & Uyarra, 2012).

These larger instruments tend to varying degrees to be complex, involving more stakeholders, involving the demand side, spanning the responsibility of more ministries and potentially bringing about more radical changes. They also tend to be multi-level, typically having a strategic level and then one or more lower levels, inhabited by various sub-programmes or interventions. Their governance and evaluation need similarly to be multi-level and often have to contend with the long life of the programme and the need to make adjustments to its goals or the way it achieves them during its life. In a growing number of cases, they involve delegating aspects of programme design and implementation to public-private partnerships (PPPs) or public-public partnerships (P2Ps) among different state organisations. Principal-agent problems can arise as a result, however, so PPPs and P2Ps need careful governance to ensure that the beneficiaries act in the national interest and not just in their own (Luukkonen, Arnold, & Martínez Riera, 2016).

Examples of such complex programmes include: the UK Climate Change Act, 2008, which aims to ensure the UK meets its carbon emissions obligations; the Netherlands Top Sector Policy, aiming to promote innovation and competitiveness across nine sectors of the economy; Germany's High-Tech Strategy for transforming industrial competitiveness; and Chile's National Innovation Strategy for Competitiveness, spanning actions across multiple ministries and sectors of society – all of which are discussed in this report. More widely, CITPs belong in this category.

Thus, research and innovation policy instruments have evolved through the three governance generations from simple grant funding for individual researchers through bilateral research-industry collaborations to increasingly large-scale and complicated funding programmes involving many actors at once. At the extreme, large and complex programmes such as CITPs emerge to address major systemic shifts in national competitiveness and in order to tackle socio-technical transitions.

2.4 Multi-level governance and CITPs

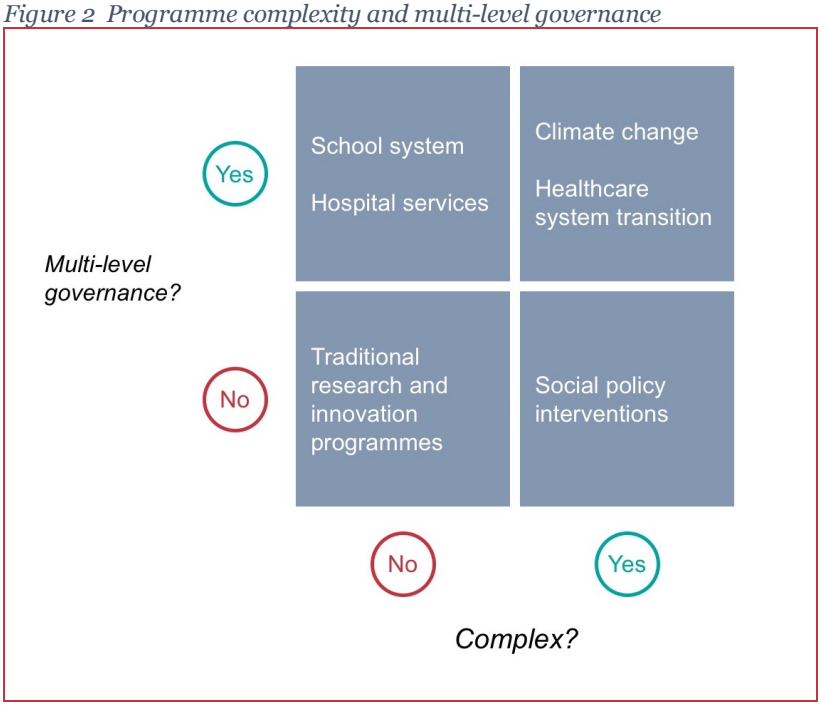
As policy interventions become larger and more complicated and involve more actors, coordination becomes increasingly important – both horizontally across organisations working at the same hierarchical level (such as ministries) and vertically, as between ministries and their agencies. Some form of multi-level governance that involves decision-making, analysis or monitoring at more than one level tends to become necessary. But there also are many situations where multi-level governance is inappropriate.

Figure 1 shows when multi-level governance appears to be necessary.

- Traditional innovation programmes, for example the provision of grants or loans, are simple to deliver, address situations that are normally not complex and are typically managed by an agency such as a research council or innovation agency. Multi-level governance is not needed
- Systems to deliver school or hospital services operate within the domains of individual ministries. They tend to have multi-level governance – at the level of the entire system and at the individual organisational level – but managing them or introducing simple innovations like curriculum changes or new treatment regimes does not involve complexity or call for wide social involvement
- Social policy interventions – for example changes to welfare benefit schemes – notoriously involve intervening in complex social systems. However, this is done within the sphere of a single ministry or agency and a narrow set of stakeholders so multi-level governance is not needed
- CITPs, on the other hand, span multiple levels and involve a wide range of stakeholders. Frequently, multi-level governance is appropriate. For example, climate change programmes fit into a structure

of multi-level governance that connects them to national commitments on CO_2 emissions and ultimately up to international treaties. A radical restructuring of healthcare delivery, for example through massive use of e-Health technologies and remote monitoring to keep patients out of hospital for more of the time, would require coordination with other agencies such as those dealing with welfare, social work and even labour market as well as many other stakeholder groups. It would also require both central and de-centralised decision-making and therefore imply a need for multi-level governance

Figure 2 Programme complexity and multi-level governance



The idea of multi-level governance has mainly been discussed in the EU context since the mid-1990s (Marks, Scharpf, Schmitter, & Streek, 1996) (Hooghe, 1996) and continues to evolve (Kuhlmann, 2001). In some areas, policy decisions have been centralised to the EU level – where they emerge from processes of negotiation among stakeholders and the Commission rather than being imposed from above. In other areas such as research and innovation, policy is made in parallel at multiple levels including the European, national and regional levels.

Since the declaration of the European Research Area (ERA) (Busquin, 2000), the Commission has additionally sought to coordinate national research and innovation policies, initially through the 'Open Method of Coordination' that used performance indicators as a way to 'name and shame' the laggards, and subsequently through various efforts at joint programming with and among Member States. In parallel, the regional development directorate has sought to coordinate innovation and other policies at the regional level using Structural Funds. The result is a patchwork of overlapping strategies and programmes at different levels, with no-one in overall charge but constant evolution as a result of negotiations in the policy arenas located at different levels.

More recent studies emphasise that EU multi-level governance operates through many models that evolve via negotiations between the EU (essentially the Commission) and Member States (Sabel & Zeitlin, 2008) (Tömmel & Verdun, 2009). The EU now exerts three kinds of governance pressure on Member States (Stame, 2008): the Community method, via decisions of the European Council of Ministers; the Intergovernmental method, based normally on unanimous agreement in the European Council of Ministers about matters that are in the competence of the Member State; and the Open method of coordination.

Implementing a multi-level governance system relies on the presence of adequate strategic intelligence and capacity at all levels of the governance hierarchy (Kuhlmann, et al., 1999). 'Strategic intelligence' literally means the analysis and information needed to inform the development and implementation of strategy. In the context of research and innovation policy, this means the intelligent combination of data, analysis, specific studies such as evaluations and foresight exercises needed to design and implement policy at all levels. Thus Gollata and Newig (2017) point out that while the EU uses a comprehensive system of multi-level governance for its air quality directive, in Germany a disjunction has arisen between the national and sub-national governance levels (where there are 137 air quality action plans), driven by a lack of strategy and planning capacity at the sub-national level.

EU multi-level governance typically relies on an 'objectives hierarchy' often with a division of labour among different agents expected to intervene in order to reach the objectives. This hierarchy is taken into account via multi-layered arrangements for commissioning evaluations and standardised evaluation guidelines proposing ways to tackle causality (Stame, 2008). However, EU multi-level governance appears to have a recurring difficulty in coordinating below the EU and national levels. Kaiser and Prange (2004) argue that the EU's open method of coordination has largely failed in innovation policy because innovation policies are defined at multiple levels while the open method of coordination only operates at the EU and national policy levels. Member States' governance of innovation policy is itself is not well coordinated below the national level in many countries. The diverse nature of national innovation systems both in terms of structure and performance in fact requires innovation policies to be nationally and regionally nuanced and context-specific, posing problems both for benchmarking and for the relevance of the policy developed only at the EU level.

Experience with the EU Framework Programme (FP) – a massive CITP – suggests that it exerts two forces on the pattern of research undertaken at the national level. Some countries (especially those that lack strong national research systems)tend to ape FP priorities (Langfeldt, Godø, Gornitzka, & Kaloudis, 2012). In other cases, including Sweden and Norway, the FP strengthens centrifugal forces in the national research system, effectively creating opportunities for those whose interests lie outside the national agenda (Arnold, et al., 2011) (Langfeldt, Godø, Gornitzka, & Kaloudis, 2012). More broadly, the FP clearly exerts 'soft power' over national and continent-wide research and innovation strategies, without exerting any degree of compulsion (Laredo, 1995) (Arnold, 2012).

Multi-level governance can be more voluntaristic in what is sometimes known as 'network governance'. Hertting and Vedung (2012) describe the rise of network governance as being driven by (a) the 'institutional complexity' of governing multiple hierarchies of organisations (such as ministries, their agencies and their beneficiaries) and (b) the 'substantive complexity' of the domains upon which they operate, where the actions of one organisation impinge upon the actions of others in systems with complex properties. These complexities in turn drive a need for network governance. The voluntarism limits its power to steer, so it tends to focus on the development of the network, hence "The accountability issue is the Achilles heel of network governance." (Hertting & Vedung, 2012)

This discussion suggests that while it may be necessary for operating large and complex programmes, it is difficult for the centre to exert tight control, especially over time as aspects of the system change. It is important for there to be 'distributed strategic intelligence' present at each level in the governance hierarchy in order to adapt to local circumstances and compensate for the fact that central planning is unable to cope with the many detailed aspects that must be handled at lower levels in the hierarchy. The designers of large and complex programmes therefore face a difficult challenge in devising a multi-level governance that is strong enough to maintain the integrity of the programme but flexible enough to allow variations in implementation and adaptation over time while still maintaining accountability.

CITPs tend to need multi-level governance, coordinating the implementation of strategy but also providing the local strategic intelligence and flexibility to adapt design to local needs. Such governance seems more powerful and more likely to focus on societal goals if there a degree of central power or compulsion. Even then, however, there is a risk that the higher levels lose touch with the lower ones. But networked governance without strong central authority and links to government easily evolves to follow the goals of the network stakeholders rather than societal goals.

3 The evaluation literature

This Chapter explores what the mainstream evaluation[3] literature has to say about complex interventions. It has almost nothing to say specifically about CITPs – that discussion takes place in the transitions literature, which we describe in Chapter 4.

We begin by considering the role of evaluation in public policy – which is already well understood among public policymakers but which is different from the understanding of evaluation in transitions. We discuss three currently important ways of doing policy evaluation– positivist, constructivist and realist – taking a position in favour of a realist approach, which we believe is the most objective and robust and the one likely to produce results of use in policymaking. We go on to discuss two major new concerns in the evaluation literature that are relevant to our subject: how to deal with the complexity that poses increasing challenges to evaluators; and the emergence of 'developmental evaluation' – a constructivist school that is useful for providing feedback to intervention managers but which does not hold them accountable to society. Finally, we discuss multi-level evaluation because evaluation has to match the needs of multi-level governance, while remaining independent of it.

3.1 The role of evaluation in policy governance

Evaluation can be done for a number of reasons. The kind of evaluation with which we are concerned here typically aims to provide evidence about the relevance, efficiency and effectiveness of a public policy intervention and advice about how to improve it and future policy, if possible. In evaluation jargon, therefore, it is both 'summative' (summing up what happened) and 'formative' (educating those in charge of the intervention and helping them shape future ones).

The immediate audience and usually the client for the evaluation is normally part of the state apparatus. But this is accountable to the government whose policies gave rise to the intervention and ultimately to society (including, but not only, taxpayers), which provided the resources to undertake the intervention and expects a return on its investment.

Typically, evaluation can be done at three points: ahead of the intervention (e*x ante* evaluation, providing a kind of a 'design review' by trying to predict the likely effects of the intervention); during the intervention (*in itinere* or mid-term evaluation, providing interim evidence on relevance, efficiency and effectiveness and often advice that is intended to help improve the intervention during its life); or after the intervention (*ex post* evaluation, looking back and summing up the evidence on relevance, efficiency and effectiveness and sometimes providing advice about future interventions).

Evaluation is needed because governments are monopolies and in order to provide accountability to taxpayers and more widely to citizens. There is no 'market' that can judge their performance so evaluation tries to simulate the kind of criticism that companies get from the market. Many governments still believe that they should base their policies on objective evidence rather than lobbying from interest groups, fake news or ideologically-based fiction. So we evaluate public policy also in order to provide that evidence. The way we do evaluation needs to live up to these requirements.

3.2 Why we need a realist approach to evaluation

In evaluation as in much of social science, it is rare for theories or schools of thought to succeed each other in the way that we are used to seeing in the 'hard' sciences, where it seems meaningful to talk about 'progress'. Rather, in evaluation, competing schools of thought tend to proliferate and it is rarely possible to do 'crucial experiments' that let us declare one to be better than the others.

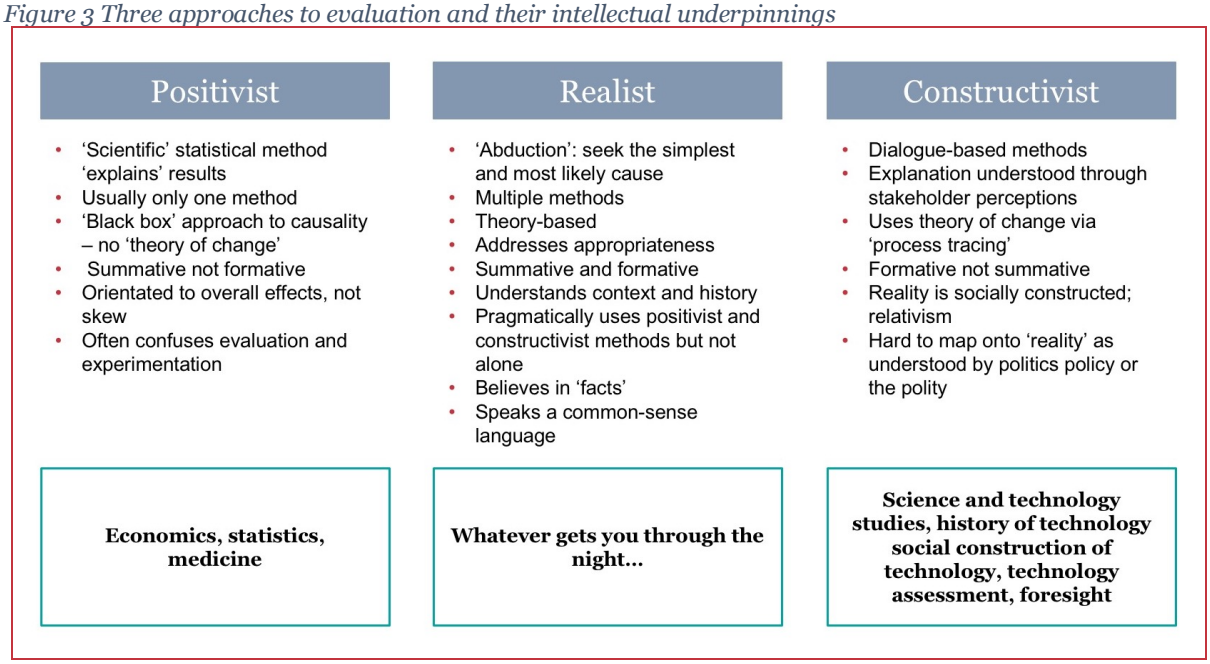
Different evaluation schools may be useful in different situations. Our question, then, is: what approach is most relevant to evaluating CITPs, where we want to use evaluation in support not only of learning but also of judging the relevance, efficiency and effectiveness of public policy as well as providing accountability to the taxpayer and society more broadly? We situate ourselves in the mainstream of

[3] We do not discuss 'evaluation' in the sense of assessing proposals, for example for research grants

policy evaluation and take a 'realist' approach to our subject (Pawson & Tilley, 1997) (Pawson, 2013). We do so for a pragmatic reason: namely that it lets us deliver evidence about both the effects of interventions and their causation mechanisms that is meaningful to policymakers and the public and which can be used to learn how to improve policy interventions in the public interest.

Figure 3 sketches the three evaluation approaches we discuss here, as well as their intellectual underpinnings.

Figure 3 Three approaches to evaluation and their intellectual underpinnings



A lot of evaluation is 'positivist' in character (Pawson & Tilley, 1997). Positivist evaluation largely uses experimental techniques (theory testing/deduction, experimental and quasi-experimental designs, even randomised control trials) to establish results that are generalisable. These 'black box' approaches compare inputs with outputs but pay little attention to the mechanisms of causation. By ignoring these mechanisms, they effectively treat widely different intervention types as equivalent[4].

In practice most positivist evaluations use a single evaluation tool, so they are not well placed to handle 'emergence' or changes in the behavioural characteristics of the systems they study. The randomised control trial (RCT) is regarded among positivist evaluators as the 'gold standard'[5] tool. RCTs are useful in looking for the effects of an homogenous treatment on a relatively large and homogenous population – for example, to understand the effects of a medicine in a clinical trial. Typically, we are interested to know whether the medicine is effective in an acceptably high proportion of the treatment group to make it worthwhile licensing for routine medical use. A simple innovation policy intervention such as providing small 'innovation vouchers' to SMEs turns out to be more complicated because the population turns out to be non-homogenous: there is a large variation in companies' capabilities and therefore their ability to make use of such vouchers. A lot of creative work has gone into tackling such problems – for

[4] For example, focusing on the economic rate of return to different interventions treats them as equivalent because they are only measured in one dimension. Clearly, interventions to do basic research, build industrial R&D capacity or support the commercialisation of universities' inventions have very different effects in the innovation system and may even be inter-dependent. Focusing narrowly on return on investment obscures these differences and implies that the interventions are substitutes – which clearly, given their different functions, is clearly not the case

[5] Evaluators with a sense of humour and of history might recall the economic damage that Winston Churchill did to the UK economy by adopting the gold standard in 1925

example by using propensity score matching in effect to try to compare pairs of treated and non-treated firms with similar characteristics.

In research and innovation policy, we know that most research and innovation grants lead to little that is of much importance – but that a minority generates exceptionally important results that typically more than justify the investment in all the grants put together. Here we need ways to understand the skew – a problem to which more creative effort is being devoted in quantitative analysis but that is often not taken into account in evaluation. Some positivist evaluation tools are designed to answer forward-looking questions such as "Will this intervention work in general?" as opposed to backwards-looking evaluation questions such as "How well did this intervention work in particular?" They may have to be twisted considerably in order to be used in a backwards-looking context by constructing 'quasi-experiments' or looking for 'natural experiments'. Since such methods do not explain the mechanics of causation and are difficult to use well, they are best used in combination with others.

Constructivist evaluation, in contrast, searches for meaning through techniques such as qualitative design, narrative analysis and interpretation of the perspectives of different stakeholders involved in the intervention. Through these, a picture of reality is constructed that is based upon perceptions, largely of stakeholders and beneficiaries. This applies in dialogue-based evaluation and in the new school of 'developmental evaluation' and provides a socially constructed view of reality that is relativist – and not necessarily similar to the view of others in society whose perspective could be equally valid or with what we would commonly regard as the 'facts'. The results of constructivist evaluation, therefore, cannot be mapped onto policy in a clear and simple way, even though they may provide considerable insight.

The realist position is intermediate. It accepts that there are 'facts' in a more objective and reproducible sense than that facts are socially constructed. It tries to establish both what the effects of an intervention are and the mechanisms through which they occur through a process of retroduction (abduction): that is, by seeking the simplest and most likely explanation of causation.

Three other things come along with the realist position. The first is the use of theory-based evaluation (Weiss, 1987) (Chen & Rossi, 1989) (Chen, 1990), which uses the goals of the intervention to construct a set of logical steps via which the intervention is expected to lead to outcomes and impacts. This provides the evaluator with a way round the 'black box' of causation in positivist approaches (Stame, 2004) by providing testable hypotheses about how causes lead to effects.

The second is a response to the problem that theory-based approaches encourage the evaluator to ignore things that do not form part of the theory. That response is to try to take deliberate account of the context outside the intervention, wittily summed up by Pawson and Tilley (1997) in what they call the "basic realist formula":

Mechanism + context = outcome

The third is the use of mixed methods, in the belief that while any individual method has weaknesses, these can be mitigated by using several and triangulating among the results they produce. Realist evaluation may use positivist or constructivist techniques as individual evaluation tools, but only within a mixed-methods approach. Thus, Blamey et al (2012) show that it is possible to use RCTs within a theory-based evaluation by addressing them to what the evaluators (from previous research) believed to be key links in the evaluation logic. Hawkins (2016) similarly claims that RCTs can be one of the tools used in a realist evaluation but that attempting to use an RCT on its own does not produce an evaluation that can be regarded as 'scientific'. Realist evaluation is willing to use dialogue-based methods prevalent from constructivist evaluation in order to understand perceptions because realists believe that it is not the interventions themselves but the way in which they are perceived that determines outcomes (Pawson, 2013).

Realist evaluation can be seen as part of the movement for 'evidence-based policy' of the past thirty years. However, people who advocate evidence-based policy often use evidence-based medicine as an exemplar. This is problematic. First, it involves a caricature, because in fact evidence-based medicine often does not let research evidence overrule clinical experience by default (Oliver & Pearce, 2017). It

does not only use positivist evaluation approaches such as RCTs but actively seeks out the mechanisms of disease (Pawson, 2018). Second, even if evidence-based medicine did solely rely on 'gold standard' experiments, mapping this exclusive focus onto evidence-based policy would ignore people's experience and values as well as most of the machinery of decision-making when evidence-based policy needs to address power, actors and context (Pawson, 2013).

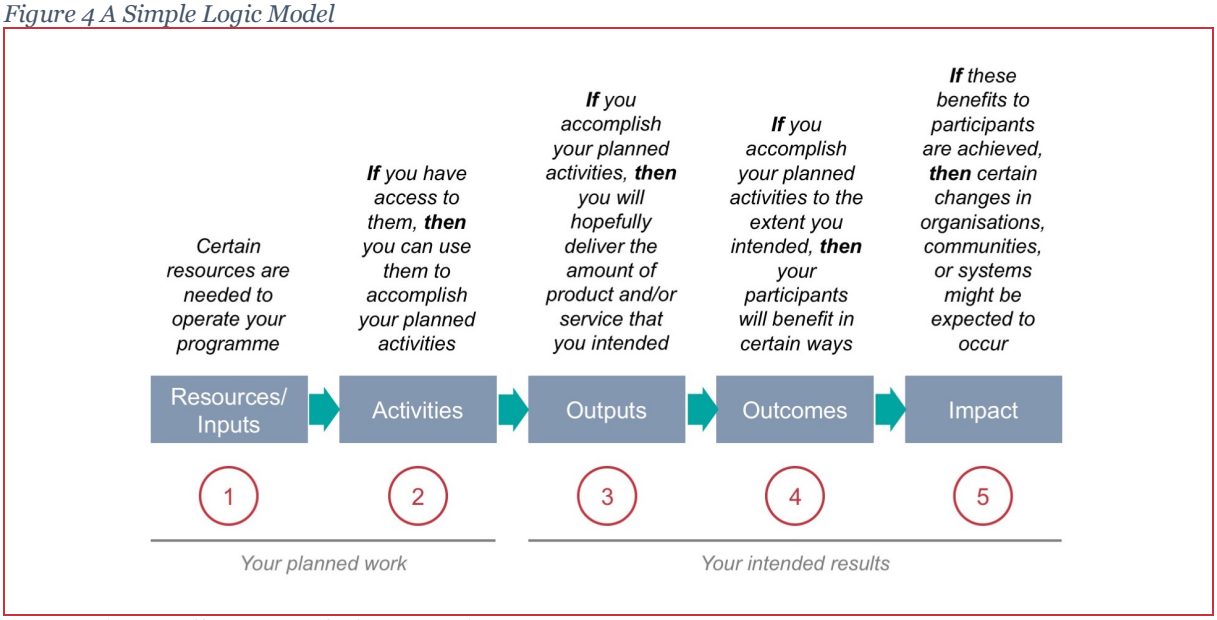
Not all approaches, then, produce evaluations that conform to the traditional expectations expectation that evaluation provides robust and independent analysis of the relevance, efficiency and effectiveness of interventions. Of the three alternative approaches, only the *realist* approach is in that sense fit for purpose.

3.3 Theory-based evaluation and intervention logic

If we accept the realist idea that evaluation should not only try to judge *whether* but also explain *how* interventions cause effects, then we have to do theory-based evaluation. Sadly, there is a jungle of terminology associated with designing a theory that explains how an intervention is intended to lead to impact[6]. The terminologies are not standardised sufficiently to remain stable over time. We summarise here what we understand as a widely-used European terminology.

Theory-based evaluation (as well as intervention design) relies on building a logic model for the intervention. By logic model, we mean an explanation of *how* the resources and activities invested in the intervention are expected to lead to the desired outcomes and impact in society. Usually these models are presented as logic *charts*. (A chart documents a model or theory.) Figure 4 shows a very simple template for a logic model. It is the business of the intervention designer and the evaluator to detail this out in a way that is specific to the intervention. Formal logic models such as logical frameworks tend to focus on the activities within the intervention and the logical steps that lead from these to outputs, outcomes and impacts. It is possible to associate indicators with each step in the logic but they do not make explicit the causal mechanisms that connect the steps in the logic – or the boxes in Figure 4 – together.

Figure 4 A Simple Logic Model



Source: (WK Kellogg Foundation, 2004)

A key distinction is between theories of change (Weiss, 1987) (Chen, 1990) (Chen & Rossi, 1989) and other kinds of logic model. Theories of change try to explain *why* the various steps in a logic model lead to changes in subsequent ones. They may explore multiple possible paths to impact, both those inside

[6] Theory of change, programme theory, intervention logic, logical framework, logic diagram – the terminology seems endless

and those outside the intervention. Their explanatory power is said to make it easier to devise useful indicators and to establish IF … THEN … logic to specify the conditions under which events at one point in the diagram trigger events at the next stage (Clark & Andersson, 2004). Major evaluation clients such as the UK DFID increasingly emphasise a desire for theory-of-change based evaluation (Vogel, 2012).

The theory of change approach is central to the relatively new tradition of 'contribution analysis' in evaluation (Mayne, 2012). This rejects the idea that it is possible properly to solve the so-called 'attribution problem' – that is, to show that an intervention was wholly responsible for particular effects in society, or if it was partly responsible to say how big that part was. Rather, contribution analysis builds detailed theories of change that are then tested empirically to show whether interventions *contributed* to changes. However, while contribution analysis is intellectually elegant and has found some applications in health and welfare evaluation, it is not widely used – mainly because the level of detailed investigation required is prohibitively expensive (Delahais & Toulemonde, 2012).

In practice, the use of logic models has been refined over time. Some versions are now very explicit about the need to identify the assumptions being made in moving from one step in the logic to the next as well as the assumptions made about things happening *outside* the intervention (ie in the context). A good example is Goal-Oriented Project Planning GOPP – originally *Zielorientierte Projektplanung* or ZOPP – which was developed by the German aid agency GTZ. It starts with an analysis phase that generates a theory of change and then uses USAID's Logical Framework approach for project planning. In effect, good practice in the use of theories of change and logic models has tended to converge and practitioners try to enrich their logic models using some principles from theory of change. Another approach is to reserve the detailed and costly analysis needed to use a theory of change for critical steps in a logic model while doing less ambitions analysis at the other steps (Delahais & Toulemonde, 2012).

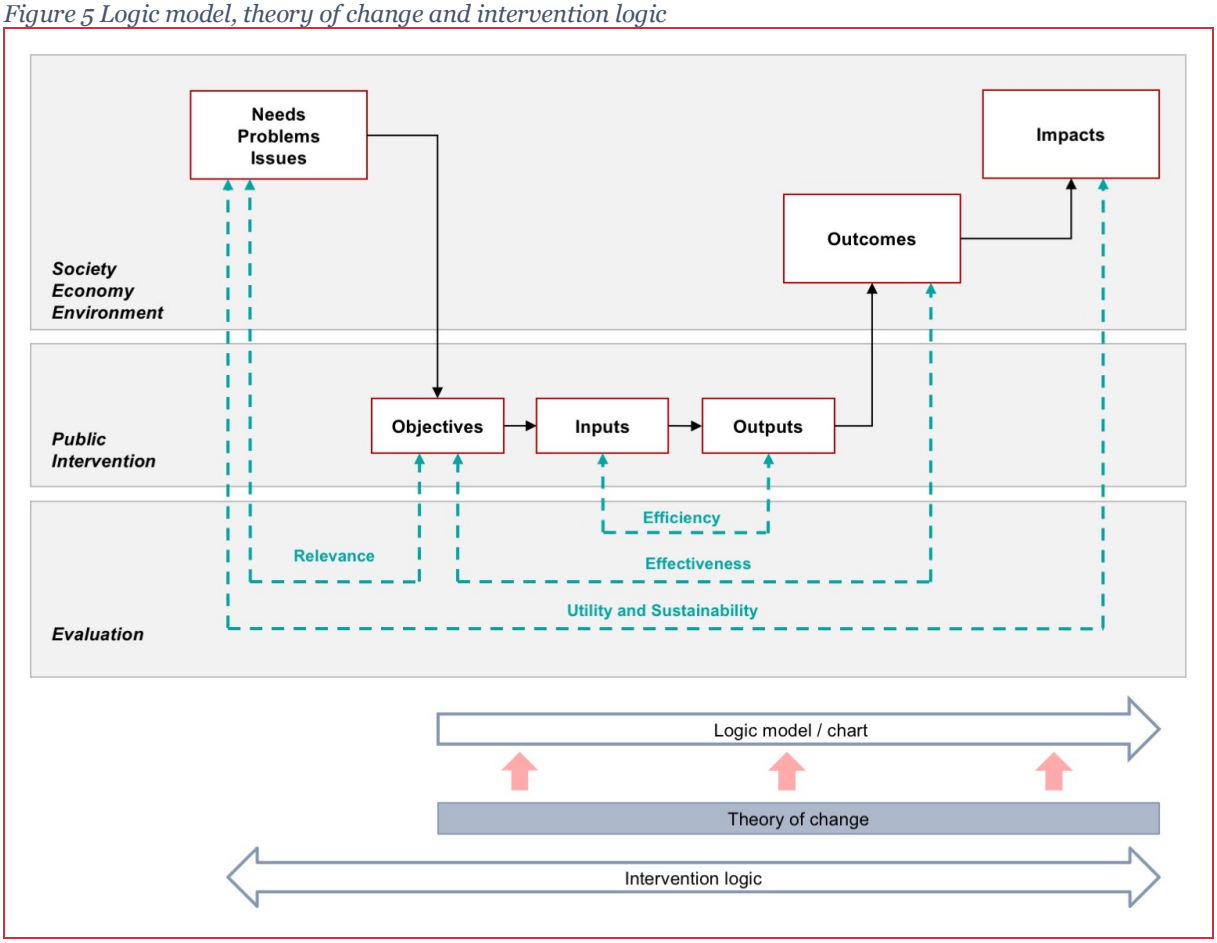
In the EU, the term 'intervention logic' is often used in connection with theory-based intervention and evaluation. Nowhere is the meaning of the term authoritatively defined. We understand it as involving

- An explicit problem analysis and an analysis of why the intervention is an appropriate policy tool for addressing the problem
- An underlying theory of change
- A logic model that explains how the theory of change is to be implemented

The distinctive aspect of an intervention logic is the inclusion of the first step, which involves querying whether the intervention is actually a good policy decision as opposed simply to testing how well and effectively that decision has been implemented.

Theory-based methods for understanding causation mechanisms can be demand very high levels of evaluative effort if they explore in detail how every step in a causal chain works. There is some consensus in the evaluation community that overly simple models are not very useful – especially for learning – so prescriptions for how to build more sophisticated theories and becoming more demanding. However, the large amount of effort needed to explore detailed theories of change implies that the evaluation designer needs to decide when this effort is justified and when a quicker approach is adequate. Figure 5 illustrates the differences in scope between intervention logic on the one hand and logic models and theories of change on the other.

Figure 5 Logic model, theory of change and intervention logic



The evolution of practice in theory-based evaluation, then, leads to the use of what we might think of a 'rich logic models', namely ones that go at least part of the way from explaining *how* interventions lead to effects, in order to address the question of *why* they do so. It has also extended from the traditional evaluation focus on analysing how well interventions reach their goals to asking whether interventions themselves as well as their goals are the right ones.

Thus, theory-based methods for understanding causation mechanisms can demand very high levels of evaluative effort if they explore in detail how every step in a causal chain works. There is some consensus in the evaluation community that very simple models are not very useful – especially for learning – so prescriptions for how to build more sophisticated theories and becoming more demanding. However, the large amount of effort needed to explore detailed theories of change implies that the evaluation designer needs to decide when this effort is justified and when a simpler approach is adequate.

3.4 Dealing with complexity

In recent years, the evaluation literature shows a growing interest in complexity (Caffrey & Munro, 2017) (Walton, 2016), giving rise to increased attention to phenomena such as emergence, local rationality, socio-technical systems and feedback for learning. This interest includes growing attention to 'wicked' problems, which resist solution because of interdependencies and their social complexity and which can involve multiple policy domains (Walton, 2016). A considerable increase in the number of journal articles about complexity in evaluation has taken place especially since 2006, though there are grounds for questioning the extent to which this increase adds much to existing knowledge beyond semantics (Gerrits & Verveij, 2015).

There are increasing efforts to do evaluation which is 'complexity-aware', evident for example in USAID evaluation and monitoring guidelines (Britt, 2016) (Douthwaite, Mayne, McDougall, & Paz-Ybarnegaray, 2017). But in practice this awareness amounts to a minor sophistication of the idea of 'context' (Pawson & Tilley, 1997), recognising that the context contains interdependencies, is dynamic and is liable to change its behaviour. On the one hand, this has far-reaching effects in driving interest in constructivist evaluation approaches such as developmental evaluation (Patton, 2010). But on the other hand, it leaves evaluation practice tackling complex situations in a piecemeal way, exploring the interactions between individual programme activities and their immediate context, but rarely if ever taking on the (enormous) challenge posed by the prospect of approaching the whole or large parts of the systems within which interventions operate. Pawson holds that it is in fact not possible to evaluate the whole of a system – only parts of it, so that "complexity confronts evaluation with a never-ending challenge that cannot be completed" (Pawson, 2013).

By and large, members of the evaluation community's interest in complexity is not triggered by the discussion about transitions but because they find themselves increasingly evaluating projects and programmes they understand as being complex, notably in the fields of development (Hummelbrunner, 2011) and social policy (Caffrey & Munro, 2017). While there is discussion of complexity in health evaluation, it tends to refer to complicated interventions or situations rather than to complex ones and hence to recommend traditional, positivist methods (Medical Research Council, 2006) (Academy Health, 2017).

Some writers distinguish between complex interventions and complex situations in which there is intervention (Walton, 2016). Consistent with normal usage, they tend to class systems and interventions with many components as complicated and those that have feedback loops, showing non-linearity and emergent outcomes as complex (Table 2).

Table 2 Complicated and Complex Aspects of Intervention

Aspect	Simple version	Not-simple version	Challenges for evaluation	Suggested label
1. Governance and implementation	Single organisation	Multiple agencies, often interdisciplinary and cross-jurisdictional	More work required to negotiate agreement about evaluation parameters and to achieve effective data and collection analysis	Complicated
2. Simultaneous causal strands	Single causal strand	Multiple simultaneous causal strands	Effective programmes may need to optimise several causal paths, no just one; evaluation should both document and support this	Complicated
3. Alternative causal strands	Universal mechanism	Different causal mechanisms operating in different contexts	Replication of an effective programme may depend on understanding the context that supports it. The counter-factual argument may be inappropriate when there are alternative ways to achieve the outcome	Complicated
4. Non-linearity and disproportionate outcomes	Linear causality, proportional impact	Recursive, with feedback loops	A small initial effect may lead to a large ultimate effect through a reinforcing loop or critical tipping point	Complex
5. Emergent outcomes	Pre-identified outcomes	Emergent outcomes	Specific measures may not be able to be developed in advance, making pre- and post-comparisons difficult	Complex

Source: (Rogers, 2008)

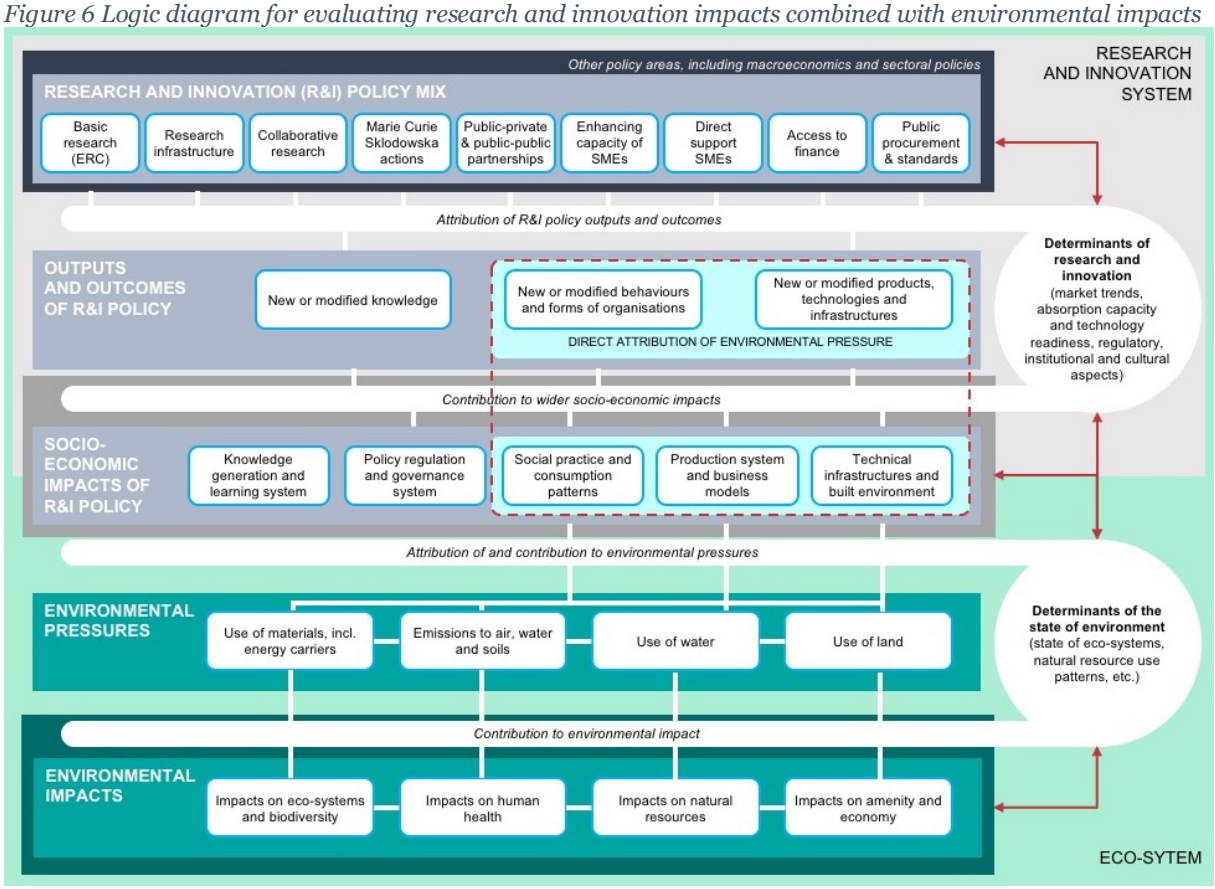
Patton (Patton, 2010) makes a useful distinction between social and technical aspects of complication and complexity. He suggests that where there is a high level of social agreement about the agenda and a high level of technical certainty in relation to the intervention, it is relatively simple to evaluate it. Where there are low levels of social agreement and technical certainty, the context is chaotic, making both intervention and evaluation very difficult. In between are zones where complexity derives from a balance

between social and technical complication or complexity. Intervention and evaluation in these areas may be to varying degrees difficult, but they are tractable.

Complexity introduces a big problem for theory-based intervention design and evaluation: since the system involved changes its characteristics over time, the intervention logic and its components do not necessarily remain stable. Indeed, they may have to be changed in order to adapt to the changes in reality. It even follows that some goals may have to change. The kind of linear intervention logics traditionally used in evaluation assume stability in the intervention and its context, represent a single theory of change and do not necessarily reflect the viewpoints of multiple stakeholders (Barnes, Sullivan, & Matke, 2004). Only a relatively narrow range of interventions is suitable for results-oriented management based on a simple logic model, ie where goals can be agreed and quantified, progress can be reliably determined and where both activities and the results of those activities can readily be observed (Wholey, 2003).

Miedzinski et al (2013) have taken an interesting step towards extending the scope of theory-based evaluation to cover both traditional research and innovation and environmental issues. As Figure 6 shows, this is achieved by first analysing the overall effects of research and innovation intervention and then subjecting these to a form of environmental impact evaluation. While this approach is helpful in explicitly incorporating environmental effects into the evaluation of innovation interventions, it remains essentially linear and does not address the 'complexity' problem.

Figure 6 Logic diagram for evaluating research and innovation impacts combined with environmental impacts

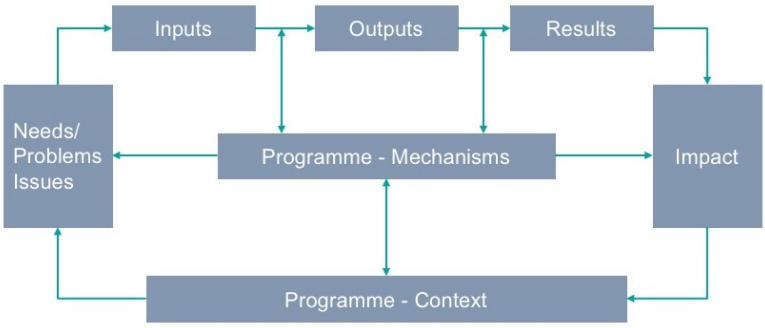


(Miedzinski, et al., 2013)

Davies (2005) proposes that logic models should be represented via networks of actors in order to cope with non-linear patterns and that networks can be analysed at a range of different levels in order to explore different phenomena. "Dealing with complicated and complex interventions implies iterative logic models, which are changed through the process of learning in programme implementation"

(Rogers, 2008). The way to deal with complex interventions is to incorporate feedback loops into the intervention logics and then to update the model based on learning and experience acquired through monitoring and evaluation. Hence it is argued that there is a need to move away from linear intervention logics towards 'circular' ones (Figure 7).

Figure 7 Schematic for a circular intervention logic to address complexity



Source: (Hummelbrunner, 2011)

As with the jungle of terminology surrounding various aspects of intervention logic, what practising evaluators do about complexity turns out to be something of a mess. In a review of 46 peer-reviewed publications using complexity ideas in evaluation, Walton (2016) identified two groups of themes: the fact that complexity affected how to define the purpose and scope of evaluations; and the need to adopt different methods – participatory, case-study and case-comparison, and multiple-method designs – in order to tackle complexity. Evaluators tried to respond to clients' increasing demand for them to tackle complexity but were hampered by inconsistent definitions of complexity, limited budgets, time and data requirement of methods and the dominance of existing approaches. The expectations of evaluation funders and users were being met with limited practitioner knowledge and limitations in complexity-informed methods.

Another recent study (Gates, 2017) of eight 'seasoned' evaluators' showed that they chose systems approaches in the hope of addressing emergence and system-level change while focusing more on implementation than outcomes and increasing the influence of those evaluated over the objectives of the evaluation, rather than letting these wholly be determined by the evaluation customer. However, their use of systems ideas involved a wide divergence of evaluation approaches, methods, terminology and even what is meant by systems approaches.

The evaluation literature offers no radical solution to the need to handle complexity. The key novelty is the realisation that the theory of change or the intervention logic can become subject to iteration in response to changes in the way the complex system works. Outside the domain of energy systems modelling, there appears to be no recourse to systems analysis.

Pawson suggests that a degree of modesty and pragmatism is appropriate in the face of complexity.

- *Stare it in the face* – map out the complexity as far as possible, looking for alternative theories but recognising that tackling the whole system will require infinite resources
- *Concentrate your fire* – concentrate resources on those parts of the system that seem vital to the success of the intervention and make do with light monitoring elsewhere
- *Go back to the future* – incorporate not only formative and summative elements in the evaluation but also design it so that it can contribute to future meta-analysis and policy development
- *Stand on others' shoulders* – where elements of the intervention have been well tested in evaluations of similar schemes, rely on the programme theories involved rather than repeating the work and focus on aspects that gave not been tested. (This requires 'organisational memory' among evaluation clients as well as evaluators.)

- *Criss and cross* – compare with at how similar interventions work in different contexts, such as different implementation agents or arrangements, in order to learn whether these differences are important
- *Remember your job* – useful evaluation shows how to think through intervention design as well as to judge its effectiveness, providing 'enlightenment' about how to do things better but not aiming to provide a general theory of public policy (Pawson, 2002)

Thus, theory-based evaluation methods traditionally use linear logic models or theories of change. Complexity poses a challenge because it means there can be feedback loops leading to changes in the complex system. In that case, the causality becomes non-linear. Evaluators find themselves increasingly working in complex situations – not only in large complex programmes but also in much smaller interventions, for example in economic development or social policy. The complexity can be social – but it can also be technical. As a result, evaluators need to work with logic models that contain feedback or are circular. Evaluation needs to be done in a staged way that can cope with such changes in logic while still maintaining their accountability to wider society and not becoming captive to the needs or views of the stakeholders involved.

Leading practitioners in evaluation in practice understand different things by 'complexity' and confront it in different ways. There is no clearly good practice that we can use here. Evaluators need to recognise that evaluating entire socio-technical systems is close to impossible. It is better to make use of what is already known about the performance of the intervention being studied and similar ones, focus on the parts of the intervention that seem most crucial to success, and compare similar interventions in different circumstances in order to make it possible to address complexity with limited evaluation resources.

3.5 Development evaluation and the growth of constructivism

Michael Quinn Patton's book *Developmental Evaluation* (2010) is much discussed in the literature. It has important similarities with the approach taken to evaluation in the transitions literature (Chapter 4). Patton promotes 'developmental evaluation' as a new approach, focusing on supporting stakeholders involved in interventions. He enumerates eight essential principles, which in summary are

1. *Developmental principle* – support what is being developed in the intervention by identifying the nature and patterns of development
2. *Evaluation rigour principle* – ask probing evaluation questions, use appropriate evaluation methods and use evidence to ground the evaluation
3. *Utilisation-focused principle* – focus on intended use by intended users
4. *Innovation niche principle* – show how the change processes involved in the intervention and their results lead to innovation and adaptation, which is the particular focus of developmental evaluation
5. *Complexity perspective principle* – use complexity principles to make sense of the problems being addressed and provide guidance about how to change the intervention in response to emergent characteristics
6. *Systems thinking principle* – think through interrelationships, perspectives, boundaries and other social aspects of the intervention and the context
7. *Co-creation principle* – develop the innovation (intervention) and evaluation together – interwoven, interdependent, iterative and co-created – so the evaluation becomes part of the change process
8. *Timely feedback principle* – Provide feedback to the intervention as it is needed and developed, rather than sticking to an arbitrary delivery schedule (Patton, 2016)

In effect, the developmental evaluator becomes a part of the intervention team, whose task is to provide strategic intelligence. The 'reality' that developmental evaluation constructs is that of the intervention team and the stakeholders, including the beneficiaries. This is similar to the role ascribed to evaluation in the transition literature, where the role of evaluation is to provide information support to the transition team.

It is of course fully legitimate for those who run an intervention to hire consulting support. Clearly there are cases, as in supporting community investments in support of the community itself, where the techniques of developmental evaluation have relevance (Kelly, 2010). However, as Patton makes clear elsewhere (Table 3), developmental evaluation is not on the side of society or government. There are other approaches to complexity in the evaluation literature which take the more conventional position that evaluation is the servant of society, not of stakeholders or beneficiaries. Hence the suggestion in Table 3that changing the task of evaluation from supporting policy and the social interest to helping the intervention managers and stakeholders cannot be driven by complexity. From the perspectives of public policy and realist evaluation alike, in developmental evaluation the lunatics take over the asylum. It shares this characteristic with the way evaluation is understood in most of the transitions literature.

Table 3 Comparison between traditional and developmental evaluation

Traditional evaluations	**Complexity-aware, development evaluations**
Render definitive judgements of success or failure	Provide feedback, generate learning, support direction or affirm changes in direction
Measure success against predetermined goals	Develop new measures and monitoring mechanisms as goals emerge and evolve
Position the evaluator outside to assure independence and objectivity	Position evaluations as an internal, team function integrated into action and ongoing interpretive processes
Design the evaluation based on linear cause effect logic models	Design the evaluation to capture system dynamics, interdependencies and emerging interconnections
Aim to produce generalisable findings across time and space	Aim to produce context-specific understandings that inform ongoing innovation
Accountability focused on and directed to external authorities and funders	Accountability focused on learning and responding to what is unfolding
Evaluator controls the evaluation and determines the design based on their perspective of what is important	Evaluator collaborates in the change efforts to design a process that matches philosophically and organisationally
Evaluation engenders fear of failure	Evaluation feeds hunger for learning

Source: (Patton, 2006)

A new interest in the evaluation literature is thus 'developmental evaluation'. This aims to tackle complexity but does so via the perceptions and the interests of the intervention managers and stakeholders. It provides them with formative information to help them learn and make adjustments to the intervention. Development evaluation sets out to make itself accountable to the intervention stakeholders but not to society. It is therefore not alone appropriate for policy evaluation in the conventional sense.

3.6 Multi-level governance and evaluation

CITPs involve multiple stakeholders inside and outside government and the state. Complexity means that their goals may be unstable, their strategies may need modification, and the consensus about these things needs to be renewed throughout the process. Their governance needs to have both vertical and horizontal elements so it is inherently multi-level.

The need to cope with complexity both in the interventions and in the contexts where the interventions operate, to coordinate and govern both horizontally and vertically within the wider innovation system while respecting the division of labour among different actors and levels leads naturally towards multi-level governance and evaluation. A review of large-scale public sector reforms in Denmark confirms the usefulness of a realist evaluation approach using programme theory. The multi-level perspective in the analysis and management of transitions pushes the same way (Weber & Rohracher, 2012). Large-scale reforms, however, need an hierarchical (or multi-level) evaluation treatment (Pedersen & Rieper, 2008).

Discussions of how to evaluate CITPs are conspicuously absent from the evaluation literature. There is a discussion about 'megaprojects' in the sense of large and complex infrastructure construction projects and the 'pathologies' such as cost overruns from which they suffer (Lehtonen, 2014) but that offers little to the present analysis. Megaprojects tend to have a single client and to be managed by a single contractor, potentially presenting management problems where complex supply chains have to be managed. But there is a single, stable goal whose degree of fulfilment is easy to understand.

3.6.1 Evaluation systems

A conceivable response to the need for multi-level evaluation is the use of a permanent national evaluation system. However, few if any countries have a comprehensive 'evaluation system' in place, covering all aspects of state intervention. Rather, those evaluation systems that exist are focused on particular domains (Thijs & Segone, 2012). Leeuw and Furubo (2008) offer four criteria for identifying evaluation systems (which, in their definition, must be carried out by organisations and institutions, not individuals).

- A distinctive epistemological perspective
- Organisational responsibility
- Permanence or history in the activities involved
- A focus on the intended use of evaluations

In practice, such systems typically focus on performance monitoring, audit, inspection and oversight, accreditation and performance evaluation. Fosse Hansen (2013) also discusses evaluation systems, which she describes as 'systemic evaluation governance' (SEG) and views as a form of hierarchical evaluation governance. These may be

- Inter-organisational SEG, via benchmarking, ranking and rating – essentially comparing the performance of organisations such as schools
- Accreditation SEG – where a single authority certifies that organisations achieve an adequate level of performance
- Systems and process SEG – which judges (and may certificate) the processes organisations use to reach a standard set of goals
- Evidence-based SEG – where a single authority collects and meta-evaluates evaluation evidence

The first three appear where an authority such as an education ministry manages multiple equivalent organisations such as schools. Some of the examples Fosse Hansen gives of evidence-based SEGs are open systems, such as the Cochrane reviews which tackle global sources of medical evidence. But she also points to the Danish education clearing house, which focuses on the Danish school system and is therefore a closed system. Evidence-based SEG can therefore be used to learn lessons within the Danish school system or to learn lessons for Danish schools with the benefit of international experience. Her earlier work on review methodologies (Fosse Hansen & Rieper, 2009) suggests that for at least some purposes the closed version will be needed, in order to cope with the specificities of the Danish school system. In practice an open system will be more difficult to organise and is likely to need access to an evidence database that has wider scope. (In the case of schools, this might be organised through the federation of national systems for school evaluation governance.)

The institutionalisation of SEGs appears to inhibit dynamism.

1. Evaluation systems produce largely routinised information relevant for day-to-day practices and single-loop learning but of little relevance for fundamental reassessments and double-loop learning
2. Evaluation systems are largely used as providers of (procedural) assurance
3. Evaluation systems produce information that confirms rather than questions policies
4. Evaluation systems breed evaluation (systems) (Leeuw & Furubo, 2008)

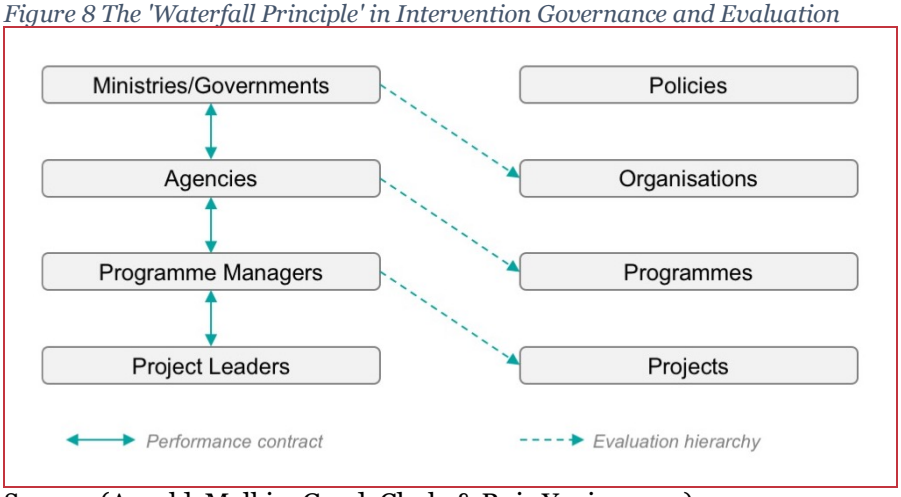
It is difficult to see this kind of highly structured governance system being applicable to complex programmes. First, such complex programmes generally do not deal with well-defined systems (like the school system) with components that are comparable with each other. Second, a pre-existing SEG is

unlikely to have a way to react to the effects of complexity such as changes in strategic direction and goals based on learning. Third, their complex nature implies that the intervention and evaluation strategies must be custom-designed, and therefore do not necessarily fit well within a pre-ordained structure.

3.6.2 Multi-level governance and complex programmes

Traditional state innovation system governance is multi-level, in the sense that there is a vertical division of labour down a chain of principal/agent relationships. Normally, each hierarchical level evaluates the level below it (Figure 8), so there is no coherent evaluation picture of how the entire system of intervention operates. In most countries individual ministries and their agencies inhabit separate 'silos', among which there is little horizontal coordination about evaluation. Hence, evaluations produce little systemic information.

Figure 8 The 'Waterfall Principle' in Intervention Governance and Evaluation



Source: (Arnold, Malkin, Good, Clark, & Ruiz Yaniz, 2009)

If, as the literature suggests, both multi-level governance and multi-level evaluation are needed in order to manage transitions and presumably other CITPs, a key missing practice involves combining the two. There are examples of multi-level evaluation, as in the evaluation of the Finnish innovation system where parallel evaluations of different system components were undertaken and accompanied by an evaluation at the system level (Veugelers & al, 2009). But it is hard to find cases of programmatic interventions being both governed and evaluated in a coordinated way at multiple levels. As with multi-level governance, most of the evidence comes from EU practice.

The European Commission organises evaluations at the EU level, asking conventional evaluation questions about relevance, efficiency, effectiveness, impact and – crucially – European Added Value. It leaves lower-level evaluation to the Member States. Where EU money is spent under the control of the Member States or of their constituent regions, the EC requires that evaluations be conducted at the Member States' expense but using guidelines determined by the Commission. Its distance from the level of national, reginal and local implementation means that the 'downward' reach of evaluation is limited and opportunities to do meta-analysis such as realist synthesis are rarely taken (Stame, 2008). By inference, since the EU level cannot determine what happens at national and regional levels, it has little interest in conducting the kinds of meta-evaluation that could better explain causality and support learning at these levels. This suggests more generally that multi-level governance and evaluation of complex national programmes needs to pay attention to the lower as well as the higher levels.

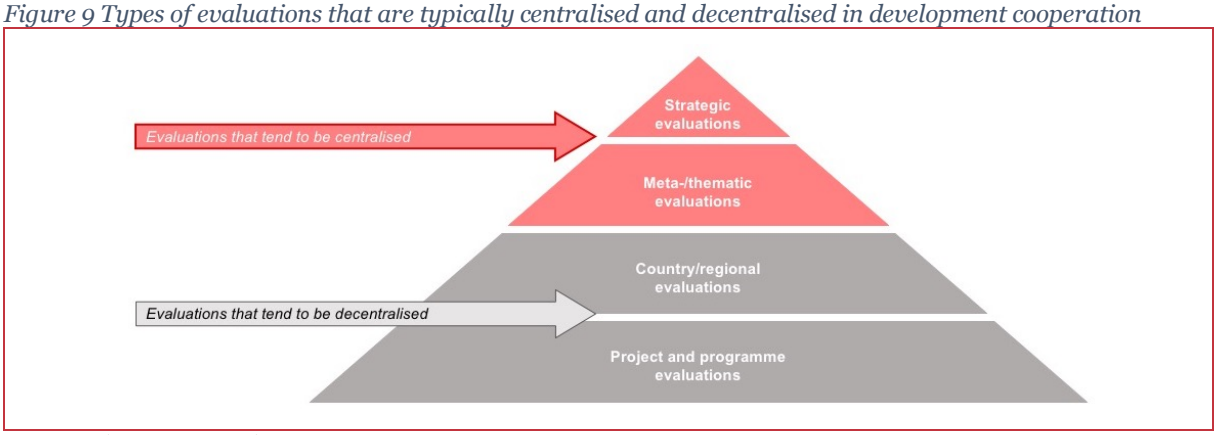
Stame (2004) argues that EU multi-level governance arrangements fail to generate evaluation strategies with adequate vertical sharing of strategic intelligence between principals and agents, while a standard set of guidelines is not sufficient to deal with complexity. If evaluation focuses solely on reaching goals in the objectives hierarchy it is possible to reach conclusions about the effectiveness of the programme but unless evaluation is also interested in the intervention instruments and their mechanisms of

operation it will be unable to explain why it succeeded or failed. In the latter case, the evaluation effectively involves many separate instrument-level intervention logics but is unable to establish relative efficiency or best practice.

In contrast to the EU, Nordic research cooperation is small-scale voluntaristic and bottom-up, despite being coordinated by the Nordic institutions. It tends to reinforce existing positions of strength rather than change structures or priorities (Arnold & Carlberg, 2009) (Tõnurist & Kattel, 2016). Evaluation is done at the Nordic level but national-level evaluation of Nordic cooperation is not prioritised – presumably because the resources allocated by individual Nordic states are so small that their evaluation at national level does not seem worthwhile.

International multi-level evaluation in the development sector appears to suffer similar problems at the lower level to those experienced in EU evaluation and governance. The OECD Development Assistance Committee (DAC) network on development evaluation recently surveyed 25 country-members about their evaluation systems (OECD, 2016). A general pattern was the use of the kind of hierarchy shown in Figure 9, with strategic and meta-thematic evaluations tending to be contracted or carried out centrally, leaving country, programme and project evaluations to be handled in the beneficiary countries. One consequence was, according to the respondents, poor communication between the central and decentralised evaluation functions so that the centre to some degree lacks understanding of the implementation of its policies.

Figure 9 Types of evaluations that are typically centralised and decentralised in development cooperation



Source: (OECD, 2016)

Programmes involving multi-level governance therefore seem to suffer from a double 'evaluation deficit': on the one hand, under-use of lower-level evaluations; on the other, impact assessments are unable to explain why and how outcomes occurred.

We have been able to identify three examples of evaluation approaches to the kind of complex programmes considered in this paper. None has been tested via implementation or been designed to cope with complexity.

Hanberger (2013) proposes an (untested) evaluation hierarchy that focuses on the intervention but takes no account of the agency of the organisations involved. He stresses the need to use theory-based evaluation, in order to make the process of intervention explicit and clarify opportunities for improvement.

- Type I monitoring and evaluation, which largely produces 'streams' (Stame, 2006) of evaluative information such as indicators, that are routinely produced rather than being tied to specific ex-post evaluations
- Type II comprises stand-alone studies of programmes or other entities within the overall intervention

- Type III attempts to synthesis information from the other levels to make judgements about the intervention as a whole

Magro and Wilson suggest a modular way to evaluate a policy mix within a multi-level governance system spanning the EU, national, regional and local levels (Magro & Wilson, 2013). They recognise that the innovation policy mix will have a number of different components, which will be implemented at more than one level of the governance hierarchy though not all instruments will be used at all levels. They suggest defining the boundary of the policy system to be investigated, then evaluating each component of the policy mix separately across the various levels, concluding with a meta-analysis and "holistic evaluation". A key simplification here is that by looking at the policy instruments (and their rationales) one by one, it is possible in each case to treat all the other instruments as part of their context. A drawback is that this means the interaction of the policy instruments will be poorly understood, yielding little understanding of complexity. Unfortunately, this idea is tested as a 'thought experiment' but not in practice.

The third example of an approach to large-scale multi-level evaluation (Arnold, Malkin, Good, Clark, & Ruiz Yaniz, 2009) is also untested in practice. It was developed to evaluate the innovation strategy of Chile's National Council for Innovation for Competitiveness (CNIC), whose funding and remit were drastically reduced shortly after the approach was proposed, owing to a change of the national presidency. Hence, the evaluation was never done. A theory-based approach was used across three hierarchical levels – corresponding to: the overall innovation strategy, in the hands of the CNIC; the agencies that operationalise the strategy via programmes; and the level of the individual programmes. By using organisations as the middle level, this evaluation approach addressed the capacity development within the state that is needed to deliver the innovation strategy. Table 4 provides a general summary. It follows that generic evaluation questions of can be articulated differently for each of the three levels of intervention: strategy; organisation/agency; and programme (Table 4, Table 5). This involves applying a logic model, which is inherently blind to unintended effects. In practice, therefore, questions about unintended and perverse effects need also to be included with the issues of effectiveness and efficiency when designing evaluations. A second danger with using this kind of linear programme theory is that it is implicitly static and needs to be modified in order to take complexity into account.

Table 4 Theory-based evaluation of Chile's national innovation strategy

Objectives	Inputs	Activities	Outputs	Outcomes	Impacts
Strategy 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
Organisation 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
Programme Improve performance of sub-system components	Programme management Money	Projects	Knowledge • For beneficiaries • Public goods	Improved performance by specific beneficiary sub-groups	Contributions to improved NIS sub-system performance

Note: NIS means national innovation system in the sense of all the components that are involved in innovation processes – not just the state organisations but also industry, education, banking and finance systems, libraries, patent systems etc etc

Table 5 Generic evaluation issues for Chile's national innovation strategy at three levels

	Relevance	Efficiency	Effectiveness	Utility	Sustainability
Strategy	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?
Organisation	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?
Programme	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?

Multi-level evaluation implies a need to synthesise the results of lower-level studies in order to provide an overall assessment – a process that the literature on multi-level governance across international borders suggests tends to fail in relation to sub-national components of the intervention, owing to a failure of coordination.

One approach to generating the top-level evaluation is to define common evaluation questions and require that these be addressed in the lower-level studies, as was recently done by the European Commission in a meta-evaluation of 'Article 185' initiatives (Meyer-Krahmer, Nauwelaars, Galetta, Santos, & Hunter, 2017). This leads to loss of understanding of the individual lower-level interventions.

A second is effectively to dictate the terms of a nested evaluation covering the entire intervention, which is a massive undertaking that also inhibits the opportunities for learning at the lower levels.

A third is to use meta-evaluation, review or systematic review to evaluate the intervention as a whole. Here, however, differences in review methodologies can have serious implications for the type of analysis and conclusions that can be produced (Fosse Hansen & Rieper, 2009). Medical techniques for systematic review such as those used in the Cochrane review focus on positivist evaluation tools, where RCTs are seen as the most reliable, aiming to produce conclusions that are universally valid. Statistical meta-analysis tends to be preferred to synthesis. In contrast, those in health, social welfare and education (where RCTs are more difficult to apply, but where the disciplines involved make greater use of qualitative methods) are more likely to prefer synthesis and to focus on results in the country or field where they will be applied to policy, since these disciplines recognise that context has a significant influence on events.

It is also possible to use systematic review techniques to incorporate learning from other interventions into the analysis, thereby working towards a more widely applicable understanding of what works where – and more strongly supporting future intervention design (Pawson, 2002) (Edler, Ebersberger, & Lo, 2008).

3.6.3 Implications

If large-scale and complex interventions need multi-level governance, then they also need multi-level evaluation. It is tempting to consider setting up a national evaluation system covering a wide range of topics, so that it large programmes could be treated within the set of parameters it defines. The fact that no country does this is perhaps a warning signal. Where such national systems exist they focus on quality control and accreditation and tackle stable systems with relatively homogenous components, such as schools, so they do not offer very much for evaluating interventions that are fine-tuned to specific circumstances and may change over time.

The main experience with multi-level evaluation comes from the EU. As with our discussion of multi-level governance, a major problem is that the strategic-level evaluations are out of touch with the lower-level ones, so they understand little about the mechanisms that cause change at 'grass roots' level. This seems to be partly because higher-level evaluations do not concern themselves with information from the lower levels and partly because different people and organisations plan and do the evaluations at high and low levels, often in ways that are not compatible. Some writers suggest evaluating individual, lower-level interventions and then in some way adding them up. Others propose a multi-level logic model, connecting the strategic, organisational and operational levels together. But none of these has been tried in practice. Nor has there been much work on meta-evaluation or systematic review to bring the different levels of the programme together. Here, there is a clear need for evaluators and programme designers to innovate.

4 Socio-technical transitions

As we indicated in our introduction, we regard the management of socio-technical transitions as the extreme end of a spectrum of programme complexity. In this section, we explore what we can learn from the transitions literature – both for transition management and for the management and evaluation of CITPs more generally.

We begin by discussing the nature of transitions in more detail than was possible in Section 2.1. We introduce the idea of 'technological innovation systems' that historically has been important in developing an understanding of transitions then move on to the 'multi-level perspective' on transitions commonly used in the literature, which provides a view not only on how transitions happen but also on how they might be triggered and governed. We look at barriers to transitions and what is known in principle about how to manage them. The literature has a lot more to say about principles than about practice, so next we have a short section that discusses empirical evidence about transition management. We then look at transition governance and how evaluation fits into it.

4.1 The idea of transition

Transitions are important in this document because many of the societal challenges seem likely to require them and because they represent an extreme in the complexity of interventions. Our concern here is with **socio-technical** transitions: namely, transitions that are both technological and societal in character. Involving the way society functions in addition to technology makes them inherently complex because they involve human behaviour. Complex systems contain feedback loops, evolving and changing their character over time, generating 'emergent' properties that could not necessarily have been predicted at the outset.

> *The more disruptive a socio-technical change, the more uncertain and uncommon it will be (hence reducing the potential for control). This is because of the greater scope and depth of the changes required, because of the more complex set of countervailing factors (losses, unforeseen costs of adjustment, trade-offs, resistance), and because of new problems that emerge in the complex process of systems change. (Turnheim, et al., 2015)*

While it is useful to develop heuristics (such as Figure 10 and Figure 11) about how they operate, each transition is inherently unique. It is therefore necessary to tackle the individual specificities of each.

Several intellectual traditions feed into the current interest in transitions. We can think of them as forming three strands.

One strand comes from economics and the economic study of technical change. 'Technological transitions' are conceptualised as changes from one technological configuration to another – for example, transitioning from combustion-engine based cars to electric ones. If we want to promote such transitions, a key issue is how to understand (and manage) the inertia that tends to prevent it from happening. Nelson and Winter explain inertia via 'technological regimes'. Incremental innovation triggers evolution in the technological regime and can often be absorbed without disturbing the logic of the regime itself. However, more radical changes can be disruptive and require learning right across industry, society and government including by users. "Policies and institutions also play a role, as well as infrastructures, cultural discourse or maintenance networks." (Nelson & Winter, 1982) Economics has also linked some transitions to Kondratieff long waves – rather irregular business cycles of fifty years or more that appear to be triggered by radical technical change and its spread across the economy (Freeman, 1996) (Rip & Kemp, 1998). This strand tends to focus on economics and research and innovation policy, paying limited attention to wider societal actors and impacts. It feeds into the idea of 'technological innovation systems', discussed below.

A second strand emerges from the history of technology, which tries to understand past radical shifts in dominant technologies – such as the transition from sailing ships to steam (Geels & Schot, 2007). This

strand has tended to become dominated by efforts to show how to create and manage energy transitions, in the process becoming more prospective than retrospective. This has the advantage that it is policy relevant and the disadvantage that there is limited evidence on which to prescribe what practices should be adopted. Bussels et al distinguish between the sociotechnical approach to transitions that takes historical inspiration and a future-orientated approach that emphasises governance and foresight (Bussels, Happaerts, & Bruininckx, 2013). It might be more accurate to say that many people who used to take the former approach now also take the latter. This strand is associated with science and technology studies (STS), considering the societal effects of technology as well as the processes by which technology is socially constructed (SCOT). Criticism and societal engagement are important themes through activities like technology assessment, foresight and democratic participation, and this is very evident in the approaches taken to transition governance, emphasising wide participation and stakeholder involvement.

A third strand is the tradition of quantitative energy systems modelling – an area where Sweden has historically been strong, not least as a result of continuous funding by the Swedish Energy Agency and its predecessors – generally focusing on techno-economic variables rather than the social and political aspects of shifting between systems. This tends to take a purely technical and economic view.

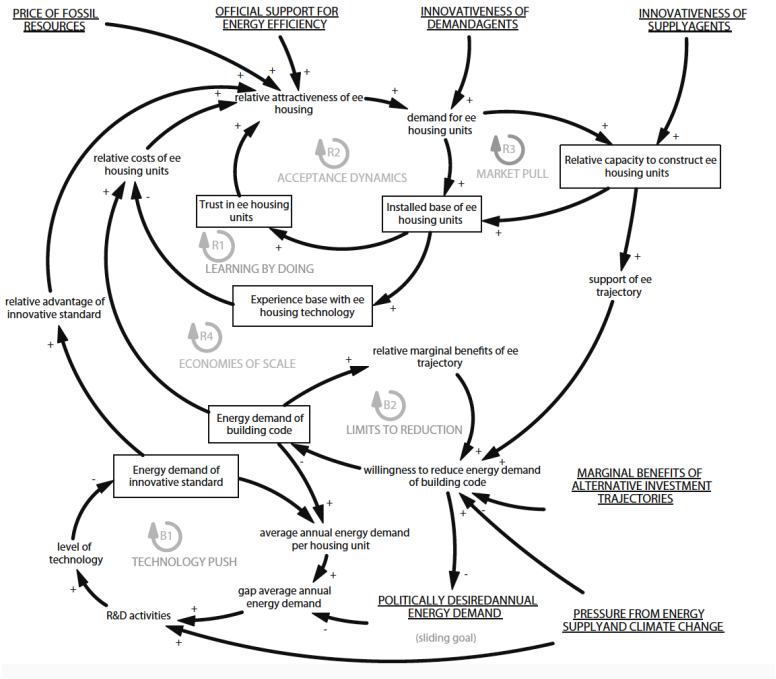
In a similar vein to the three 'strands' described, a recent review classifies perspectives on energy transitions into three categories: techno-economic, focusing on the economics of resource allocation and use; socio-technical, focusing on the history of technological change and addressing path-dependency and other lock-ins; and political, focusing on government, policy and the state. It suggests that these perspectives reflect different arenas of change and action – that events within these arenas can be independently generated yet affect the other arenas, so that an adequate analysis needs to address all three (Cherp, Vinichenko, Jewell, Brutschin, & Sovacool, 2018).

4.1.1 Technological Innovation Systems

The idea of innovation systems emerged from the techno-economic tradition discussed above A particular adaptation of the idea – a technological innovation system (TIS) – is sometimes used in the transitions literature in order to focus on a specific transition. A TIS can be defined as "a network of agents interacting in the economic/industrial area under a particular institutional infrastructure … and involved in the generation, diffusion, and utilisation of technology." (Carlsson & Stankiewicz, 1991) It is, in effect, a technologically specific sub-set of an innovation system and focuses on the industrial system rather than broader society.

As with innovation systems the effects of a TIS cannot be understood using a 'variance' approach, ie looking at how parameters co-vary, but has to be understood through processes and events occurring in particular orders (Hekkert, Suurs, Negro, Kuhlmann, & Smits, 2007). Analysis therefore involves tracing processes (Pettigrew, 1990) – in effect establishing a theory of causality in the form of a logic diagram, as is commonly done in mainstream evaluation or by process-tracing. Ulli-Beer offers a good example of process-tracing – which she calls 'process theorising' (Figure 10) – to explain the factors and processes that have played a role in the diffusion of energy-efficient housing designs in Switzerland. This offers a theory of how a specific transition has taken place. Evaluators would recognise it as a logic chart.

Figure 10 'Process theory' for the diffusion of energy-efficient housing in Switzerland



Source: (Ulli-Beer, 2013)

A further contribution has been to complement the structural view of an innovation system shown in Figure 1 with a functional perspective. Edquist & Johnson (1997) introduced the idea that particular functions need to be fulfilled by an innovation system. Influential papers by Jacobsson and colleagues combined the TIS perspective with the idea of a sectoral innovation system (Jacobsson & Jonsson, 2000) and went on to consider the importance of various functions in generating change in TIS (Jacobsson & Bergek, 2004), (Bergek, Jacobsson, Carlsson, & Lindmark, 2006). Building on the Swedish work, Hekkert et al suggest that there are seven functions that are necessary for a TIS to become established (or, therefore, for it to displace an earlier TIS).

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, Suurs, Negro, Kuhlmann, & Smits, 2007)

This idea has become influential in efforts to design transition management more generally

Thus, we need to understand about managing socio-technical transitions, because they will be needed in order to address at least some of the societal challenges. Economists have tried to explain how such transitions work in terms of past transitions in 'technological regimes' – in effect the systems of

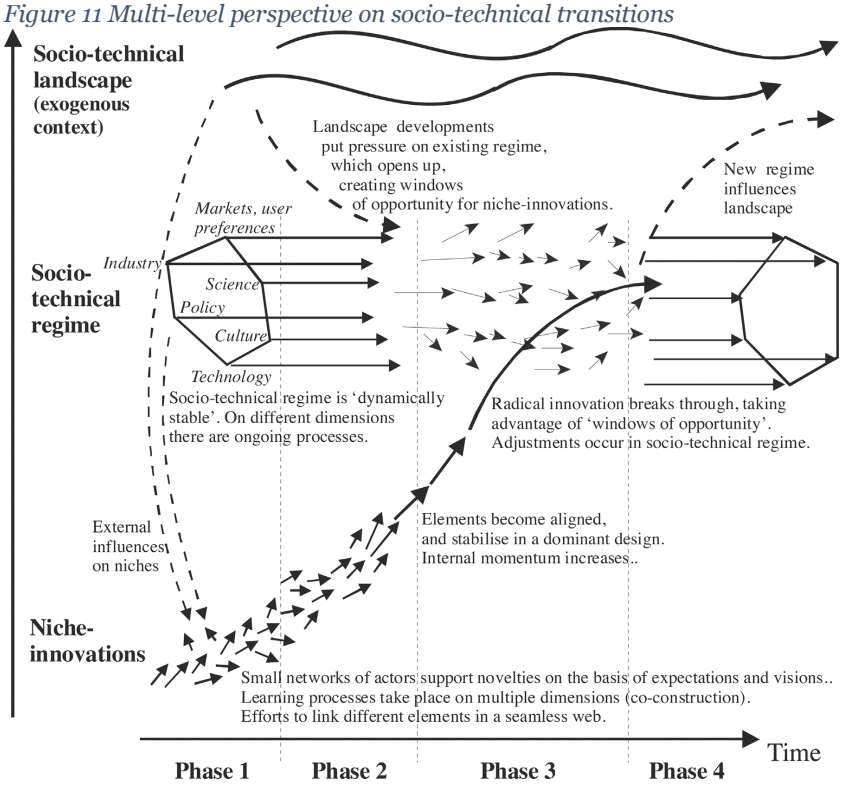
innovation that cluster around individual classes of technology, impeding or fostering the introduction of disruptive technologies. Historians of technology have taken a broader approach to transitions – such as that between sailing ships and steam – that takes account of a wider range of social factors and actors. Those who now work on 'transitions management' emphasise the broader approach and the need to involve a wide range of stakeholders from many parts of society in achieving transitions, such as removing fossil fuels from the electricity production and distribution system.

4.2 The multi-level perspective and how transitions happen

Much of the discussion about transitions now uses a multi-level perspective (MLP) into which the idea of a TIS is effectively subsumed. The MLP was developed in the period when transitions studies were mainly backwards-looking but is now used in connection with forward-looking attempts to manage transitions as well. It describes in a generic way the context in which transitions take place and has three levels. That of Geels (2002) (2010) (2018) is widely employed. (Figure 11 shows the most recent version.) Together, the two lower levels in the MLP appear to make up what other authors call a technological innovation system.

The top level in the MLP is the 'landscape', consisting of the broad social, technical, infrastructural, cultural and political environment and other exogenous or external factors that form a context for the socio-technical regime. This is the level at which social and political pressures to address societal challenges emerge.

Figure 11 Multi-level perspective on socio-technical transitions



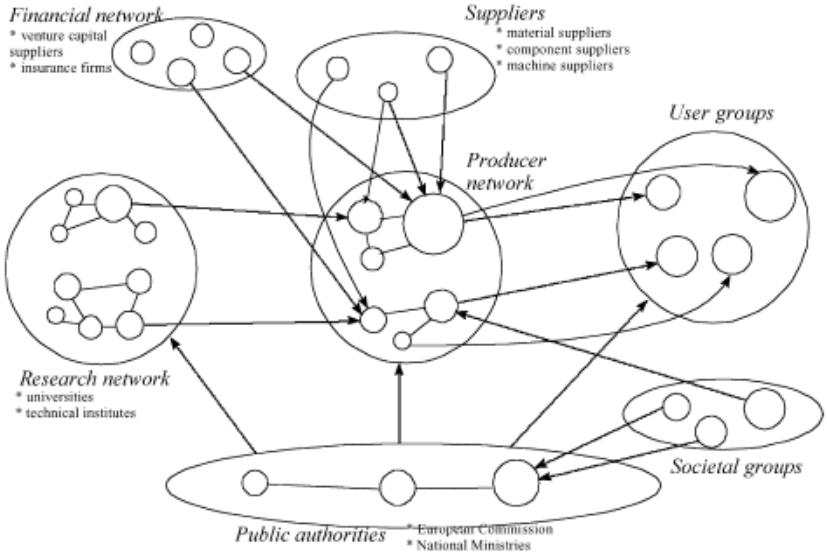
Source: (Geels F. W., 2018)

The socio-technical regime is the specific 'world' within which transitions take place and comprises: technology, user practices and application domains (markets), symbolic meaning of technology, infrastructure, industry structure, policy and techno-scientific knowledge (Geels F. W., 2002). Niche innovations – whether incremental or more transformative – contribute to the development of the regime (Figure 12).

> *In a nutshell, radical innovations emerge in peripheral niches in phase 1, and stabilize and enter small market niches in phase 2. Breakthrough in phase 3 depends on niche-internal drivers such as price/performance improvements, scale and learning economies, the development of complementary technologies and infrastructures, positive cultural discourses, and support from powerful actors. But diffusion also depends on external windows of opportunity, due to regime destabilisation because of landscape pressures or persistent internal problems. Regime transformation occurs in phase 4, including adjustments in infrastructures, policies, lifestyles and views on normality. (Geels F. W., 2018)*

The socio-technical regime contains seven groups of actors (Figure 12). A key difference with the innovation systems approach is the explicit inclusions of 'societal groups' as significant actors in the regime, rather than as users or consumers.

Figure 12 The multi-actor network involved in sociotechnical regimes



Source: (Geels F. W., 2002)

The MLP has been criticised for over-simplification of the levels and their roles in transitions (Berkhout, Smith, & Stirling, 2004) (Markard & Truffer, 2007) and for undervaluing agency (Smith, Stirling, & Berkhout, 2005) (Shove & Walker, 2007) and politics (Genus & Coles, 2008) (Meadowcroft, 2009). Rauschmayer, Bauler, & Schäpke (2013) argue that the role of individuals in sustainability transitions has been understated and needs better exploration. Svensson & Nikoleris (2018) also argue that the simplification involved in the MLP removes the idea of complexity, referring to Bhaskar's (1997) distinction between 'open' and 'closed' systems. An open system such as a sociotechnical system involves multiple mechanisms and therefore can change both the system itself and its context. Only closed systems have regularity and predictability – and the argument is that the MLP tends to treat systems as closed.

The early version of the MLP (Geels F. W., 2002) tends to rely on the idea that the major pathway to transitions is via niche innovation, which eventually destabilises the socio-technical regime. This is understandable in that historical studies of socio-technical transitions are largely looking at changes that occurred more or less spontaneously, in response to innovation. In the context of a growing interest in *managing* transitions so that the impulse for transition comes from policy rather than spontaneous innovation (Genus & Coles, 2008), this was criticised as being over-simple (Berkhout, Smith, & Stirling, 2004) (Smith, Stirling, & Berkhout, 2005).

As governments increasingly prioritise sustainability transitions (and others stemming from the societal challenges) we would expect the importance of policy and politics as well as the landscape more broadly to increase. Our concern with transitions in this paper is essentially with such landscape- or policy-driven transitions rather than those that occur more spontaneously.

Moving from observation of past transitions to management of future ones makes it especially important to identify which *functions* – such as knowledge development, market formation, creation of legitimacy – need to be performed in order to make the transition possible.

4.3 From description to prescription: how to manage transitions

Geels offers a useful list of barriers to transitions (OECD, 2015). These (and presumably others) need to be understood if successful transition management programmes are to be put in place.

- Over-reliance on market failure rationales
- Short-term political processes (election cycles)
- Fragmented, multi-layered institutions, governance structures and processes
- Technological trajectories and lock-ins
- Market power and political clout of incumbents
- Lack of customer acceptance and adoption
- Institutional inertia and path dependency

Given the uniqueness of individual transitions, no such list is likely to be complete. Other issues that may be important include: large absolute costs of change to both the state and to business; long periods before obtaining a return on investment, in some cases; how to manage economic and other types of stranded assets such as capital equipment and infrastructure left behind from the previous sociotechnical regime; the more general risks and uncertainties associated with disruptive innovation.

Elsewhere, Geels offers four ways to overcome the barriers

- Head-on fights or struggles against the regime, exploiting economic arguments, mobilising networks and political power
- Using 'windows of opportunity' to have clever fights
- Circumventing existing regimes rather than competing directly with them, for example by opening up new markets or piggy-backing on growing ones
- Using new configurations of the new and the old in order gradually to displace the existing regime (Geels F. W., 2014)

Government can support changes in sociotechnical systems by being an 'alignment actor' and not just a regulator (Rip & Kemp, 1998). Transition tactics can include establishing 'front runners', establishing a transition agenda or road map, conducting transition experiments and transition monitoring (Wittmayer & Loorbach, 2016). Haxeltine et al (2013) argue that there is a need to mobilise social as well as economic innovation in transition processes. The transition policy mix is likely to need elements of 'creative destruction', hastening the demise of old sociotechnical systems as well as supporting the development of new ones (Kivimaa & Kern, 2016).

In line with the idea in the multi-level perspective that transitions grow out of niches, 'strategic niche management' was early proposed as a way to foster transitions (Kemp, Schot, & Hoogma, 1998) (Ramos-Mejía, Franco-Garcia, & Jauregui-Becker, 2017). Later, as the focus of interest changed from understanding past transitions to managing future ones, a revised approach of 'transitions management' appeared, shifting the focus from the niches towards managing the whole transition (Loorbach & van Raak, 2006). Transitions management involves "interactive and selective participatory stakeholder searching processes aimed at learning and experimenting" (Grin, Rotmans, & Schot, 2010).

The operational model for transitions management was visualized as a cycle and the policy tool developed for this was the transition arena. This cycle had the following components.

1. *Structure the problem in question and establish & organise the transition arena*

2. *Develop a transition agenda, a vision of sustainability development and derive the necessary transition paths*
3. *Establish and carry out transition experiments and mobilize the resulting transition networks*
4. *Monitor, evaluate and learn lessons from the transition experiments and, based on these, make adjustments in the vision, agenda and coalitions" (Loorbach & van Raak, 2006)*

Reflexive activities are important in steering the transition (Loorbach D. , 2010)

The shift from historical analysis to transition management (TM) has created a need to consider: Who does the management? The conventional response is a 'transition arena' (Loorbach & Rotmans, 2010) where 'front-runner' stakeholders – understood as those with a strong commitment to the transition (Wittmayer, Frantzesjaji, van Steenbergen, Roorda, & Henneman, 2012) – act under the guidance of a 'transition team'. Meadowcroft (2009) accuses TM of political naivety, failing to take into account either the continuities or the discontinuities among political processes or personalities. How those involved in the transition team get political legitimacy is not clear. While (Sendzimir, Magnuszewski, Balogh, & Vári, 2006) among others argued that political considerations should deliberately be excluded from TM, others point out that this is neither practical nor legitimate (Hendriks & Grin, 2007) (Shove & Walker, 2007). It is clear that TM needs to establish political and democratic legitimacy by linking into the political level (Voß & Bornemann, 2011) and ensuring not only that relevant stakeholders are empowered but also that others are not dis-empowered (Hölscher, Wittmayer, Avelino, & Giezen, 2017).

Ehnert et al (2018) argue that in the past agency has been under-emphasised in both the MLP and the technological innovation systems literature. Similarly, politics have been given little weight (Meadowcroft, 2009) (Shove & Walker, 2007) (Smith, Stirling, & Berkhout, 2005). However, as Ehnert et al (2018) point out, more recent studies have put both in a much more central position (Avelino & Wittmayer, 2015) (Hess, 2014) (Raven, Kern, Verhees, & Smith) (Geels F. W., 2014) (Markard, Suter, & Ingold, 2016) (Normann, 2015). As a result the field has started to move away from the 'governance without government' approach of the earlier transitions management literature and towards the idea that governments have to play key roles in transitions but can only do so in partnership with research, business and wider society (Wittmayer & Loorbach, 2016).

Thus, most writing about transitions now involves a multi-level perspective. The central level is the 'socio-technical regime', which broadens the earlier idea of technological regimes to take better account of wider society. This is the battleground between alternative socio-technical systems. It is fed from below by a flow of innovations – some of which are absorbed as incremental changes and some of which have potential to trigger transitions. Above is the 'landscape', comprising politics, broader society and the broader context, such as infrastructures.

Intellectually, the transitions writers faced an important challenge in moving from historical analysis of transitions to specifying how to trigger them. Past transitions have generally happed rather spontaneously as individual innovations perturb the socio-technical regime. Addressing the societal challenges means that instead of transitions being triggered bottom-up, we need to find out how to drive them top-down on the basis of policy. However, the sociotechnical transitions literature has been heavily criticised for failing to take account of politics, policy, agency and the ability of individuals to trigger changes – perhaps precisely the ingredients that are needed in moving from observing to acting.

4.4 Experience with transition management

When Loorbach and van Raak (2006) complained that "The current [transition management] literature is highly conceptual, and so far the practical experiences with the prescriptive model have barely been documented or published" there was considerable justice in their criticism. Since then, however, some evidence has started to build up about managed transitions.

Much of the recent empirical work in the field has focused on cities (perhaps because they are large enough to be complex but small enough to be tractable). There is also evidence that large world cities

have political ambitions to participate in socio-technical transitions (Hodson & Marvin, 2010), as have numbers of medium-sized cities (Emilianoff, 2014). This tends to mean that the national and international levels of multi-level governance have been under-studied in relation to transitions (Ehnert, et al., 2018).

Against the background of an expectation that experiments should become embedded in transition management practice, climate change programmes at city level are often described as 'experiments' of various kinds.

- Testing hypotheses
- Electing designs that work
- Learning by doing
- Opening up alternatives
- Fostering alternatives in protected spaces
- 'Performing reality', in the sense of providing legitimation for political proposals (Turnheim, Kivimaa, & Berkhout, 2018)

Research on climate change management in cities indicates that while (as the innovation literature more generally points out) cities have a high propensity to innovate, the actors with an interest in climate change mitigation are many, fragmented and overlapping as well as not all sharing the same interests. Private sector organisations are increasingly important in policy implementation, as well as in pursuing their own interests (Hughes, Chu, & Mason, 2018). Cities are more likely to act in the presence of a higher-level signal about the need to act and the presence of a higher-level strategy. A national or state-level initiative is the stronger form of signal but voluntaristic arrangements can also be influential in the absence of such a signal. Similarly at lower governance levels, there is a need for freedom to adapt the initiative to local circumstances. These examples underline the need for multi-level governance with strategic intelligence and decision-making power at each level. In that sense they reinforce the conclusions of work in governance of innovation systems for the past 20 years or so, cp (Kuhlmann, et al., 1999) that indicate the importance of coordination, the limits to top-down steering and the need for subsidiarity and local strategic intelligence in the multi-level governance of innovation systems.

- The Baltimore climate change and disaster preparation plan of 2013 led to significant changes at city and municipal level, but efforts to involve individual citizens and change their consciousness and behaviour proved ineffective. Stakeholders who were not adequately involved in the governance proved hard to influence (Sarzynski, 2018). As with any intervention, it is therefore important to identify and include stakeholders whose engagement is necessary for the policy to succeed
- Cities in turn are part of a wider national and international system of climate change governance. In Europe, the Council of Ministers signed the 2008 Climate and Energy Package and 2014 Climate and Energy Framework. This central signal – coming from both the EU and the national level – significantly encouraged German cities to act. In contrast, the European Commission's voluntaristic Covenant of Mayors had a limited effect (Kemmerzell, 2018) – probably because it sat next to a stronger signal from the Council of Ministers and the EU Member States
- However, in the absence of a national signal in the USA (owing to the efforts of the Trump administration to renege on the commitments of the Obama administration), strong but voluntary inter-state agreements such as the Regional Greenhouse Gas Initiative and the Western Climate Initiative have increased the likelihood of local administrations establishing climate change initiatives (Homsy, 2018). Within California, the regional climate planning law (SB) 375 offers a voluntary framework for cities to act. It was taken up strongly by larger cities but less by smaller ones with fewer resources (Boswell & Mason, 2018)
- The case of the Montreal Waste for Climate Change Mitigation initiative points to the importance of subsidiarity in governance. While many municipalities participated in the initiative, a clear success factor was that they had the freedom to design the details of the intervention locally, so as to tune it to local structures and needs (Bourgeois & Hughes, 2018)

Another study of five cases of city transitions in The Netherlands and Japan found that management was being done largely in ways consistent with the principles of transitions management.

- Transition management as creating spaces for interaction
  - A transition team comprising leading stakeholders who drive the transition agenda
  - A transition arena – a place in which the transition can be debated and influenced
- Transition management as challenging the status quo
  - Generating new ideas through (1) visions and perspectives, (2) increased understanding of problems and solutions, and (3) insights and understanding about the nature of the process and (4) the nature of the knowledge and understanding created
  - Establishing new practices in the form of projects and activities
  - Establishing new social relations and new actors (Wittmayer J. M., 2016)

A study drawing lessons from transition management in the UK, Netherlands and Germany (Laes, Gorissen, & Nevens, 2014) pointed to needs for

- 'Commitment devices' that build commitment to the need for transition beyond individual election cycles
- Breaking long-term commitments into near- and medium-term incentives and changes that can be monitored and measured
- Effective coordination processes across multiple actors
- Resolving inconsistencies and incompatibilities among intended pathways or niche plans
- Securing both technological, social and systems innovations needed to implement the transition pathway
- Governments to promote change in a number of ways
  - In promoting and stimulating local or regional initiatives (drawing on resources of regional/ local identification);
  - In stimulating wider involvement in the actual practice of conducting transition experiments;
  - In stimulating the actual practice of building coalitions around technological options for transition paths (e.g., the German RES policy design)
  - In allowing stakeholder input into the formulation of concrete measures for initiating a transition;
  - In setting up multi-stakeholder processes for developing scenarios, pathways, and visions.
- Reflexivity in monitoring and governance

Another study found that local-level climate change mitigation interventions unsupported by binding regulation are less likely to succeed than those that are (Keskitalo, Juhola, Baron, Fyhn, & Klein, 2016).

The presence of advocacy coalitions (Sabatier, 1988) can affect transition policies to such an extent that these coalitions can even be considered to be part of the governance structure – driving either stability as in Swiss policy (Markard, Suter, & Ingold, 2016) or change as among US cities adopting transitions strategies (Boswell & Mason, 2018). Governance initiatives can arise in the absence of official action through voluntary standards-setting, as has been the case in aspects of sustainable coffee production (Manning & Reinecke, 2016).

Resistance by incumbent regime actors acting alone, in coalitions with other incumbents or in partnership with other stakeholders, can provide a powerful brake on change. Incumbents are especially powerful when they partner with policy or political actors. In this sense, politics is central to understanding transitions and also to how to destabilise existing regimes (Geels F. W., 2014). Political coalitions working through election processes can also influence sustainability transitions (Hess, 2014). More generally, power relations among different actors or groups of actors are important in determining outcomes (Avelino & Wittmayer, 2015).

These studies suggest that practical transition management is strengthened by the legitimacy and regulatory power of government. Preferably, government should lead, though where government fails to act voluntary coalitions can nonetheless to some degree coordinate transitions. They confirm the need

for 'arenas' where transition management can be negotiated, discussed and coordinated as well as for 'reflexivity' in the sense of self-analysis and adaptation of strategy based on learning from experience. It is important to involve the relevant stakeholders, especially those who can impede the transition: actor coalitions can either enable or prevent transitions. The misalignment between the shortness of election cycles and the long periods of time during which transitions must be managed means that discontinuities among the people in government, policymakers and policy can be disruptive factors.

However, a common thread among the cities is disconnection from the national and international governance levels, so the scope of the transitions handled is regionally limited. City initiatives tend to be more effective if they succeed in connecting to the higher governance levels.

Moving from description to prescription thus means we have to understand the barriers that need to be overcome in order to enable transitions. These include short-term political processes, fragmented institutions, lock-ins of various kinds (including societal attitudes and habits) and often the active resistance of incumbent producers and users, who feel they stand to lose from change. Managing transitions involved finding ways to overcome these barriers and will involve the 'creative destruction' of old structures as well as the creation of new ones. Much of the discussion about transitions management involves the creation of a multi-level network governance but in effect involves 'governance without government' and has only just started to consider who appoints and legitimises the governance or how it is connected to the interests of society as a whole rather than the stakeholders with an interest in the transition, or even who funds it.

4.5 Transition governance and evaluation

Despite the recent increased interest in politics and agency, most of the transitions literature is remarkably silent about how to operationalise governance and to connect it to government. It does not ask how to choose the specific transitions that should be selected for intervention. Nor does it say how the transition governance is to be appointed or on what basis it obtains its legitimacy. Nor does it consider the relationship between the transition governance and government or how it obtains budget and other resources such as staff and institutional capabilities. Nor does it address its accountability. It is difficult to avoid an accusation of political naivety (Meadowcroft, 2009).

Sustainability transitions present a number of governance challenges due to (1) the multiple scales, geographies and temporalities of transformational processes, (2) uncertainties associated with radical innovation and the limits of prediction, (3) the interplay between the inertia of existing socio-technical systems and the emergence of novelty, (4) the problem of shaping innovation in relation to multiple social objectives and public goods, and (5) contested perspectives about the governance of complex processes of social, economic and technical change. Actively shaping sustainability 'transitions in the making' needs analytical approaches that take such difficulties and related challenges into account (Weber & Rohracher, 2012). This challenges traditional research and innovation governance, not least because of the need to coordinate a broader set of stakeholders than we are used to doing in policy. As a consequence, it also demands that – while overall objectives such as sustainability transitions may be stable over long periods – the strategies and tactics used to reach them should be allowed to change over time, in response to learning in the governance system.

Loorbach provides a good synthesis of starting-points for transition governance, reflecting the assumptions of transition management writers. In summary

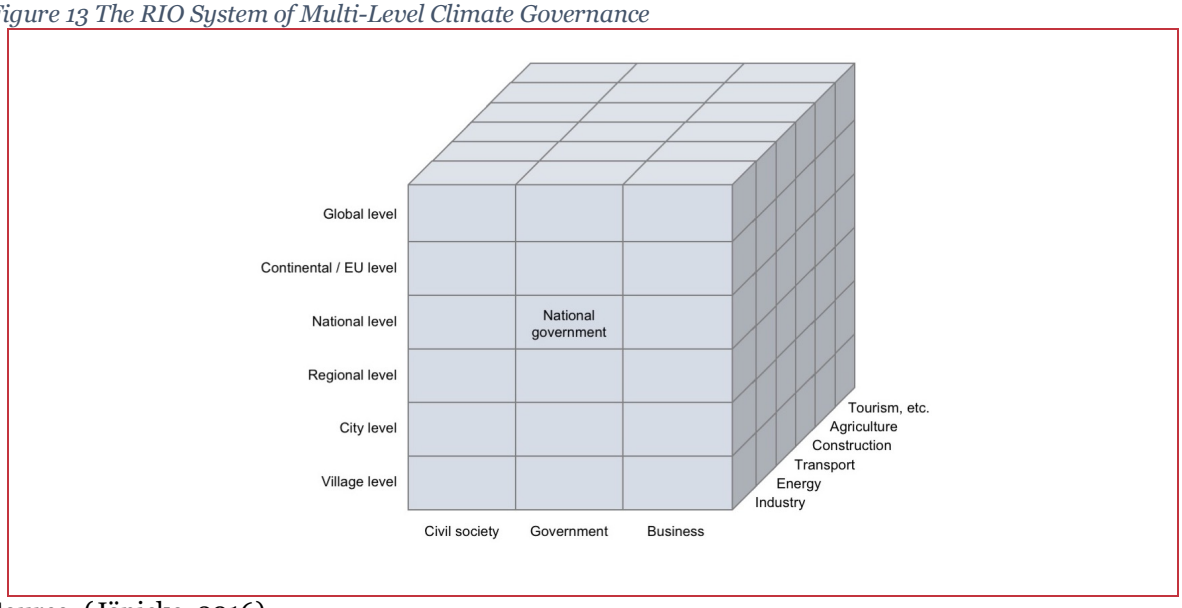
1. Change processes need to be designed using an understanding of how the wider system works
2. Transitions happen over long periods – at least 25 years – so there has to be interaction between short-term processes and the longer-term strategy, connecting short-term to long-term goals and using forecasting and scenario analysis
3. Setting specific objectives is at odds with the complexity of the wider system, so objectives must be flexible and subject to revision

4. The type of intervention used depends on the state of the system. Where systems are in crisis, it is possible to encourage rapid and radical changes. Where they are stable, the intervention may have to focus on developing alternatives that could be used to provoke crisis
5. Transition can be accelerated by exploiting or provoking disequilibria. Equilibria tend to be stagnant but are hard to upset
6. Innovation can be promoted by created protected spaces in which to develop alternatives that can lead to alternative regimes
7. Steering or governance mechanisms are themselves part of the systems they address, so steering from 'outside' the system is impossible. These mechanisms are part of the way the system operates and must be designed as such
8. Governance needs to focus on social learning, understanding the diversity of actors and their perspectives and developing a wide range of options
9. Wide stakeholder interaction is needed to learn how to intervene and to get support for the intervention, but only those actors with specific roles in the transition should be involved in the intervention itself
10. Complexity governance has to have the same ability to adapt as complex systems themselves. It needs to take an open approach, accepting uncertainties and surprises and adopting measures that tend to make permanent desirable aspects of the transition (Loorbach D. A., 2007)

While multi-level governance has always been implicit in representations such as the multi-level perspective, the transitions literature is now reconnecting with the idea that multi-level governance means the involvement of politics, power and the distribution of agency across multiple levels and hence the need to use multi-level governance in order to manage them, with a corresponding hierarchy of not only governance but also evaluation responsibilities (Turnheim, et al., 2015) (Turnheim, Kivimaa, & Berkhout, 2018). The way this plays out tends to be specific to individual transitions.

To the extent that transitions initiatives relate to global or societal challenges, this makes extreme demands of multi-level governance systems. Figure 13 illustrates the scope of the RIO system of governance: from the global to the village level; across civil society, government and business; and across various sectors of the economy. It has both a 'vertical', spatial component and 'horizontal' components across society and the economy.

Figure 13 The RIO System of Multi-Level Climate Governance



Source: (Jänicke, 2016)

The global system of climate governance may be the only one that operates comprehensively from the global down to the local level. Other transitions appear to be more locally governed – perhaps up to the continental level (as is evident in recent EU and Member State initiatives to transition out of fossil-fuelled road vehicles).

Governance of a transition initiative itself, then, is context dependent. A key part of the context is how it fits into larger, multi-level systems of governance. (Ehnert, et al., 2018) show that the way multilevel governance works in sustainability transitions depends significantly upon the degree to which national governance is central- or decentralised and whether the country is organised in a federal or a unitary way. Multi-level governance can mean that new requirements are overlaid on existing systems, for example in flood risk mitigation. Where there are existing, well-developed systems for flood mitigation, these new requirements can generally be more easily accommodated (Dale, Burch, Robinson, & Strashok, 2018). In other words, transition governance has to be adapted to the larger governance context.

Evaluation and governance are somewhat entangled in the transitions literature. Because the concept of accountability in the sense of an independent assessment of performance against objectives, cost efficiency and impact is largely absent, the purpose of evaluation and monitoring becomes to assist in learning by the transition management in order to allow it to modify its tactics as it goes along. From a policy perspective, this is a major weakness. It is difficult in practice to justify and fund interventions unless they are accountable through government to society.

Many writers eg (Voß & Bornemann, 2011) (Loorbach D. , 2010) emphasise the importance of 'reflexive governance' in transitions. By this is meant a governance that is able to experiment, learn and change course appropriately during the course of an intervention. The sub-interventions need to be regarded as experiments, their results analysed and fed back into strategy development (Turnheim, et al., 2015). Just as in 'development evaluation', 'evaluation' is focused on the generation of strategic intelligence to inform the detail of strategy development and implementation rather than the objectives we normally associate with evaluation: understanding and judging the appropriateness, relevance, efficiency and impact of an intervention in order to provide accountability to the government, taxpayers and society more broadly. In evaluation jargon, the transitions community tends to treat evaluation only as formative and not as summative. Evaluation is used solely for 'mid-course corrections' for the benefit of the stakeholders involved, so it tends also towards constructivism, which we regard as problematic in that it focuses the definition of 'goodness' on the processes and purposes of the stakeholders rather than on society as a whole and it generates opportunities for the principal-agent problems of moral hazard and adverse selection to arise (Braun, 1993) (van der Meulen, 1998).

Bussels, Happaerts, & Bruininckx (2013) argue that monitoring and evaluation of sustainability transitions initiatives is needed because interventions face the problem of emergence and therefore need to change tactics during their lifetimes. By implication, many groups of people will need to learn, including the transition managers and stakeholders but also the policymaking and political levels. The authors argue that this rules out a traditional 'logic diagram' approach, which has been widely criticised as static, cp also Befani et al, (2015), and hence results-based evaluation in general. A process- rather than results-focused approach is needed (cp Loorbach, 2007). A practical approach is 'reflexive monitoring'. Approaches include the Interactive Learning and Action approach (Reeger, 2010) and "reflexive process monitoring" (Van Mierlo et al 2010). Like Bressers' (2011) applied systemic programme evaluation' (ASPE) approach, these reject a focus on output and instead address learning how to improve processes. Bos, Hofman, & Kuhlmann (2016) have developed an assessment method for system innovation and transition that essentially maps and diagnoses the performance of an intervention as a transition intervention, rather than producing summative information about its wider effects. Interesting as these approaches are, however, they take us away from the need to provide evidence to policymakers: 'we had a good process and learnt things' does little to address the need for accountability in government interventions.

The self-referential nature of the literature extends to the kind of evaluation criteria seen as relevant. While the 'market' and 'system' failure associated with the former generations of innovation governance

discussed in Chapter 2 are properties respectively of capitalist economies and of innovation systems, transition 'failures' are actually government or governance failures.

One of the few attempts to devise a generic transition experiment evaluation scheme (Luederitz, 2017) uses a conventional input-process-output- outcomes framework and produces an impressive checklist of desired characteristics, which amounts to a synthesis of much thinking about transition experiments and governance. It does not generate a specific intervention logic for the transition. This means it primarily evaluates against a generic transitions management framework rather than focusing on the rationale or appropriateness of the intervention itself or its effects in society.

Turnheim et al (2015) offer an evaluation approach with a wider scope (Table 6), arguing that: quantitative systems modelling is the most appropriate approach for evaluation at the level of the overall socio-technical system; traditional socio-technical analysis is appropriate at the level of the regime, exploring more detailed aspects of the interaction between the transition initiative and components of that regime; and initiative-based learning is the best technique for exploring interventions at the level of niches.

Table 6 Overview of the strengths and weaknesses of three approaches to transitions evaluation

Approach	Strengths	Weaknesses
Quantitative systems modelling	Robust & highly formalised research methods Consistent analysis of complex systems Attention to system interactions (eg sectors) Attention to problem interactions Synthetic analysis of multiple options Links policy goals to required physical changes Ability to calculate effects of policy options on transition pathways Simple and coherent policy advice	Oversimplification of social realities, little attention to actors and behaviours (politics, power struggles, beliefs, strategies) Limited scope for changing economic, social and institutional rule sets Over-reliance on economic mechanisms Limited attention to implementation process
Socio-technical analysis	Fine-grained analysis and understanding Attention to different levels and temporalities Attention to relevant socio-technical dimensions Attention to multiple actors and behaviour types Analysis of institutions and changing 'rules of the game' (including shared cognitions and norms) Attention to inertia of existing systems Policy advice sheds light on uncertainties	Mainly descriptive (qualitative case studies) Qualified generalisation (context-specific, pattern-based, multiple and changing forms causal mechanisms) Limited forward orientation to political targets Policy advice focuses on general strategies (patterns) rather than instrumentality
Initiative-based learning	Analyses and/or engages in real-world initiatives as experiments Attention to local level and implementation Attention to actor-relevant dimensions (behaviour, legitimacy, learning, inclusion, etc.) Relevance to stakeholders and practitioners Policy advice is rooted in practice	Limited methodological standardisation Often context-specific and short-term oriented Limited attention to wider structural contexts Difficulty to generalise lessons for entire transitions

Source: (Turnheim, et al., 2015)

They propose an integration of the three approaches, with each focused at a different level. This three-level approach echoes the multi-level perspective – but its highest level is in effect a technical sub-set of the landscape – isolating the approach from government, politics, policy and the wider aspects of the landscape.

Thus, a major problem with much of the transitions management literature has been that it is conceptual rather than empirical. In the last few years, however there have been a growing number of studies of

climate transition management in cities. These suggest that practical transition management is strengthened by the legitimacy and regulatory power of government. Preferably, government should lead, though where government fails to act voluntary coalitions can nonetheless to some degree coordinate transitions. They confirm the need for 'arenas' where transition management can be negotiated, discussed and coordinated as well as for 'reflexivity' in the sense of self-analysis and adaptation of strategy based on learning from experience. It is important to involve the relevant stakeholders, especially those who can impede the transition: actor coalitions can either enable or prevent transitions. The misalignment between the shortness of election cycles and the long periods of time during which transitions must be managed means that discontinuities among the people in government, policymakers and policy can be disruptive factors. However, a common thread among the cities is disconnection from the national and international governance levels, so the scope of the transitions handled is regionally limited. City initiatives tend to be more effective if they succeed in connecting to the higher governance levels.

The transitions literature has yet adequately to tackle the integration of transition management into wider (especially national) systems of government, policymaking and governance. The response to the peculiar qualities of transitions has been to withdraw from these social and political realities and to avoid the central issue of accountability. Evaluation is the transitions literature serves transition management, not society. This not sustainable. It is clear that the need for feedback and iterative strategy development is greater in transitions than in many kinds of government intervention and that traditional, linear logic models need to become more iterative. This means that evaluation methods have to evolve, not that evaluation should abandon its role in providing accountability to the government and society. A key part of that evolution is the need to develop better multi-level evaluation models that address the difficulties of communication between governance and evaluation at the different levels.

5 Three case studies

This chapter analyses three case studies of the management and evaluation of large, complex programmes. The full cases are shown at the Appendix.

- The UK Climate Change Act, 2008 implements the Climate and Energy Package agreed by the European Council of Ministers that year, which established structures for the management and evaluation of the UK's progress towards meeting reduced carbon emission targets
- The Netherlands Top Sector Policy built on extensive social consultation and established nine large public-private partnerships, aiming to improve the competitiveness of the nine largest and most progressive sectors of the economy
- Germany's High-Tech Strategy initially federated existing efforts to encourage high-tech industrial innovation in an economy that was strongly based on existing and medium-tech technologies. Over time, it has moved to develop a strategy of its own and to initiating and coordinating interventions

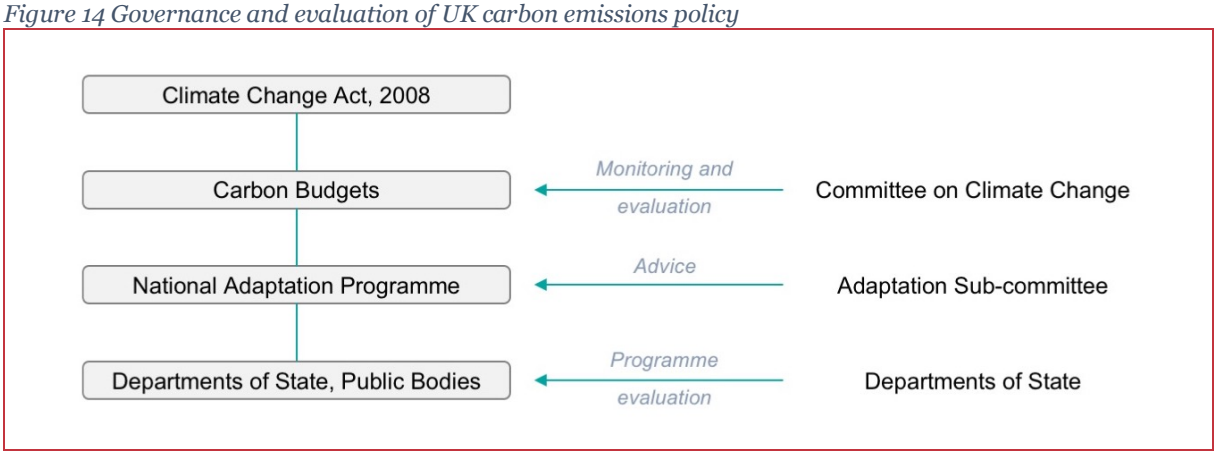
The Climate Change Act is an example of an intervention aiming to manage a socio-technical sustainability transition. The other two are examples of large, complex innovation programmes.

5.1 UK Climate Change Act

The Climate Change Act requires the government by 2050 to achieve an 80% reduction in UK CO_2 emissions compared with those of 1980. It was the first legislation of its kind worldwide and aims to manage the UK contribution to the needed global transition away from technologies that emit CO_2 and therefore cause climate change. It regulates carbon emission budgets, defined for successive five-year periods. The UK government is accountable to Parliament and is open to judicial review should it fail to comply. Judicial review is a legal process through which the courts establish the legality of an action of a public body and can impose requirements and penalties on that body.

As Figure 14 illustrates, the Act mandates the achievement of the five-yearly carbon budgets. These are to be achieved through the National Adaptation Programme, whose design is coordinated by the Department for Business, Energy and Industrial Strategy (BEIS) and the Department for Environmental and Rural Affairs (Defra), in cooperation with other departments of state. Various departments of state and other public bodies are then responsible for implementing the components of the National Adaptation Programme.

Figure 14 Governance and evaluation of UK carbon emissions policy



The Act created an independent Committee on Climate Change, answering to Parliament, that monitors and evaluates progress against the carbon budgets and effectively holds to government to account. Its Adaptation Sub-committee provides advice on the adequacy of the National Adaptation Programme for

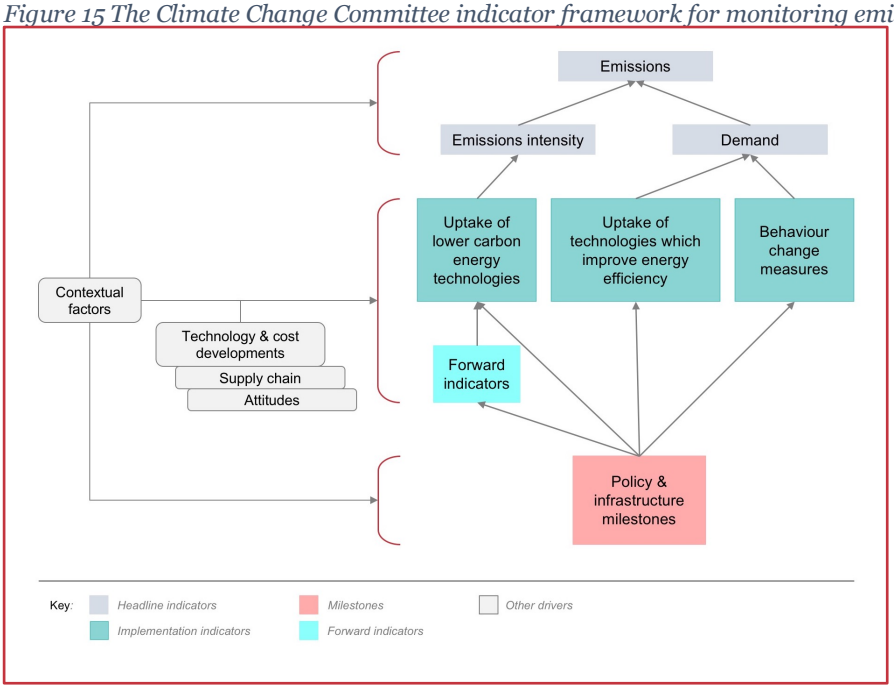
achieving the carbon budgets, so the Committee does summative evaluation and the sub-committee does formative evaluation. Individual departments of state separately evaluate components of the National Adaptation programme for which they are responsible, in both summative and formative terms.

The UK has a strong evaluation culture. The Treasury (finance ministry) both sets evaluation guidelines for government and encourages the use of evaluation and the development of methods. BEIS inherits a particularly strong planning and evaluation culture. In response to the scepticism of Prime Minister Thatcher about government intervention in the 1980s, the former Department of Industry required all interventions to provide so-called ROAME statements. These set out the Rationale, Objectives, Assessment criteria, Monitoring and Evaluation arrangements for the programme and were accompanied by the creation of a well-staffed evaluation department. BEIS maintains an evaluation strategy, which is updated every two years.

Climate change mitigation enjoys wide political and public support in the UK, especially since the publication of the 'Economics of Climate Change' report (Stern, 2008). The government has published high-level plans for reaching the carbon budget targets across each five-year period up to 2032. Each five-year cycle begins with a climate change risk assessment and results in a specific National Adaptation Programme.

While the Climate Change Committee monitors emissions, its five-yearly evaluations consider the mechanisms through which emissions are produced, so it takes a theory-based approach (summarised at a high level in Figure 15), focusing on seven sectors of society and tracking individual pathways to reduced emissions, rather than only focusing on emissions indicators. Its reports therefore provide feedback into the further development of the National Adaptation Programme.

Figure 15 The Climate Change Committee indicator framework for monitoring emissions reduction



Source: Committee on Climate Change (2008) Meeting Carbon Budgets – the need for a step change.

The Adaptation Sub-committee itself produces an evaluation of the National Adaptation Programme every two years allowing tactical changes to be made in the plans while maintaining accountability to Parliament. It uses an 'adaptation preparedness ladder' covering the capacity for change, decision-making and action in each of five priority areas. The Sub-committee aims to use a systems approach, recognising that a linear intervention logic is not adequate to address such a complex system and therefore recognising the need for frequent revisions.

Many of the evaluation and monitoring parameters are technical in character and in many cases it has not been possible to obtain the needed data. There is continual dialogue between The Climate Change Committee and its staff on the one hand and relevant stakeholders, partly because this allows the Committee to obtain data and information that is can use to triangulate against its own data.

Success factors in the implementation of the Climate Change Act include

- Broad political commitment
- A legislative framework that binds governments and applies across successive administrations
- A combination of feedback and summative evaluation
- The power to coordinate across multiple departments of state and public bodies
- The provision of significant staff resources to the Committee
- The independent position of the Climate Change Committee, which makes it difficult for the government to exert pressure on it to report more favourably than would be accurate
- The fact that reporting is in the public domain

Despite these advantages, there remains a problem that it is hard to link every detail of the National Adaptation Programme and related policies to the Carbon budget. So far, the UK has performed well in reducing carbon emissions but much of that has been achieved through the closure of the last coal mines. Despite the growing use of renewables in UK electricity production, the next steps may involve tackling more significant social and technological barriers.

5.2 The Netherlands Top Sectors

In the first decade of this century, the Netherlands Ministry of Economic Affairs (EZ) ran a number of sector-focused innovation programmes, involving subsidy instruments inter alia to promote academic-industrial cooperation. The Top Sector Policy was introduced to replace them in 2011 by the van Rutter I government, which had been elected on a conservative platform that included corporation tax cuts and sought to avoid subsidy schemes. The Top Sector approach was intended to improve Dutch performance in existing sectors of industry. It is a large complex innovation programme but does not involve a socio-technical transition. Its implementation was based on public-private partnerships (PPPs). It was accompanied by wider business deregulation and efforts to reduce state bureaucracy. This change in direction compared with previous innovation policy has to be understood as party-politically driven.

From 2010-16, the Top Sectors were intended to promote the competitiveness of the sectors. Since then, they have been asked to focus on the societal challenges defined by the EU, while still pursuing sectoral competitiveness.

Evaluation culture in The Netherlands, and especially in EZ, is strong. As in Sweden, evaluation and policy studies have a long history and are understood as necessary to open and democratic government. Regulations and laws require periodic evaluation of policy, especially policy instruments that involve subsidy.

The nine Top Sectors[7] were chosen after consultation with industry and the research sector. Each Top Sector is a self-organising PPP run by Top Team assisted by a secretariat and an advisory board, both comprising people from industry, academia, government and wider society. It establishes smaller consortia called Top Sector Knowledge and Innovation (TKIs). Each Top Sector has to organise activities to do with knowledge and innovation, internationalisation and human capital development. They developed individual strategies, based on very widespread consultation across industry, academia, government and wider society but then applied a top-down management approach to ensure that implementation was done.

[7] Horticulture and propagation materials, Agri-food, Water, Life sciences and health, Chemicals, High tech, Energy, Logistics, Creative industries

The Top Sectors run pre-competitive research projects and support company innovation activities (typically through loans rather than subsidies). They receive little new government funding but almost one billion Euro were diverted from the budgets of various departments of state, the national research funding council NWO and the Royal Academy of Arts and Sciences KNAW to serve the Top Sectors. Industry has to pay its own way when it participates in projects, but receives a bonus of 30% (formerly 25%) of its contribution from the government.

The Top Sectors were intended to reduce the fragmentation of innovation policy and trigger more of a whole-of-government approach to innovation policy but did not evidently include strategic measures to reach these ends. Rather, the scientific and industrial officers of the various Top Teams meet periodically to coordinate at the level of individual projects and devices as well as institutions so, for example, more than one Top Sector might together coordinate projects to do with medical devices.

Shortly after the Top Sector policy was developed, EZ commissioned a study to define evaluation methods pertinent to its policies. Although intended to address the Top Sectors, the resulting report recommended exclusively the use of positivist statistical and econometric techniques suitable for interventions (such as innovation vouchers) with large numbers of small-scale projects (Theewes, 2012), leaving EZ with no strategy for evaluating the Top Sectors.

EZ produces annual monitoring data for the Top Sectors (and other parts of the economy) in a series of Progress Reports on Enterprise Policy. In 2014, the Dutch Advisory Council for Science, Technology and Innovation (AWTI) undertook a review of the Top Sector approach. It acknowledged that it was still too early to attempt to measure any tangible outcomes, as such the review largely concerned the delivery model of the approach and a review of the policies at a meta-level (AWTI, 2014). It complained of mission drift among the Top Sectors towards addressing societal challenges and while it argued that the policy had increased dynamism in the relevant fields, it was unable to use the available indicators to attribute changes in the indicators to the policy.

Dialogic's evaluation (Janssen, et al., 2017) addressed the situation by taking a qualitative approach, focused on Hekkert et al's (2007) seven functions of a technological innovation system – assessing the achievement of each and analysing the effectiveness of the instruments used in order to achieve them. (Perhaps it is no coincidence that Hekkert was a member of the steering group for the evaluation.) It was based on extensive document review and almost 100 semi-structured interviews. It concluded that the policy had implemented a shift away from subsidy-based policies, reducing fragmentation and improving networking in some sectors. However, it had not greatly encouraged radical innovation. Some sectors had make great strides in trade facilitation and human capital development but the links with government were unclear – for example, government had not been able to link innovative procurement to the Top Sector approach. A key problem had been the government's 'hands off' policy, leaving the implementation to the industrial and research actors and therefore failing to generate an overall national strategy or direction. The details of the policy were opaque, so it was difficult to organise at the national as opposed to the sector level.

PPPs can suffer particularly from principal-agent problems, where the agent may act in her own interest rather than in the interest of the principal. Dialogic's evaluation (Janssen, et al., 2017) makes it clear that several of the Top Sector interventions tended to bolster existing actors and technologies rather than fostering new ones, essentially because the Top Sectors were dominated by industrial incumbents.

In contrast to the UK case, therefore, the Top Sectors policy was unclear in its socio-economic (as opposed to its ideological) goals. There was some monitoring but no real multi-level governance, leading to fragmentation. The evaluation strategy having foundered by initially using a study group intent on promoting methods inappropriate to the intervention, it was left to Dialogic to rescue the evaluation – largely without using the wealth of monitoring data that had been collected. The next steps appear unclear. The lessons are more evident.

- Over-decentralisation of a national innovation strategy leads to fragmentation and loss of accountability

- Structured governance is needed to generate an overall strategy and ensure convergence between the objectives of the state and the PPPs involved. A particular issue is how to prevent incumbents' interests from constraining the rate of innovation
- An evaluation strategy needs to be designed specifically to deal with the type of intervention involved and should be in place from the outset. The monitoring apparatus needs to be designed based on that strategy
- The spotty evaluation pattern means that evaluation can provide little support or steering to the government or the stakeholders during the course of a long-run intervention, which almost inevitably will have to change some of its tactics as it goes along

5.3 The German High-Tech Strategy

The High-Tech Strategy (HTS) is a government-wide innovation strategy published in Germany in August 2006 to promote research and innovation, for a total government investment of €50bn (Foreign & Commonwealth Office, 2018). The aim of the strategy is to secure economic, cultural and social prosperity (BMBF, 2010), rather than a specific socio-technical transition. The HTS is led and primarily funded by the Federal Ministry of Education and Research (BMBF) but involves coordination with and funding from ten other ministries. It has evolved considerably over the more than ten years since it was established.

Germany has a medium-strong evaluation culture. There is an obligation to evaluate individual funding instruments, which tends to be respected at least in the case of the larger ones. The fact that this obligation is tied to individual instruments means that there was for a long time no impulse to evaluate the Strategy as a whole – something which is only now coming onto the agenda. On the other hand, Germany has a strong tradition of using advisory bodies such as the Wissenschaftsrat to carry out evaluation and monitoring as well as to give advice. The HTS has benefited from two bodies, which have performed this function.

The HTS has evolved through three phases.

- Phase I (2006-2009) focussed on key technologies and lead markets. The aim was to create conditions in Germany to enable researchers and organizations to gain leading positions in markets that are both technologically advanced and likely to grow in importance in the future (Allen, 2009).
- Phase II (2010-2013) prioritised the use of key technologies to overcome major societal challenges.
- Phase III (2014-2020) of the High-Tech Strategy was called Innovations for Germany. This was developed as a broad innovation strategy with new topics and new instruments enabling innovation being added. Originally planned as a four-year effort, it has been extended to 2020 and now incorporates five 'missions' that cut across sectors and ministry responsibilities
  - Priority challenges with regard to value creation and quality of life
  - Networking and transfer
  - The pace of innovation in industry
  - Innovation-friendly framework
  - Transparency and participation.

In the early years, the strategy was criticised for lacking originality and dismissed as a repackaging of existing efforts. However, as demonstrated by the shifting focus of each of the phases, the strategy has evolved and attracted greater government financial commitment.

While the BMBF provides the primary management of the HTS, it has been supported throughout by a coordination with other ministries. There is a proposal on the table to formalise this coordination by creating a Federal Committee of State Secretaries for the High-Tech Strategy.

The HTS has been overseen and supported by two external bodies since the inception. In 2006-8, it was overseen by the Council for Innovation and Growth, the government's advisory body. This committee

was replaced in 2008 by the Commission of Experts for Research and Innovation (EFI), which has since monitored, reported and advised on the HTS via its annual report to the government on research, innovation and technological performance in Germany as part of its broad task of monitoring and advising on the development of the national research and innovation system.

The other advisory body– the Industry-Science Research Alliance – was dedicated specifically to supporting the HTS. It comprised representatives from the industrial and science sectors and worked to provide support for both the implementation and continued development of the strategy between 2006 and 2013. The Alliance was replaced in Phase III by the High-Tech Forum, also functioning as an advisory body. This meets three times a year and is made up of civil-society representatives and stakeholders from academia and business. It advises on the Strategy's development of the new thematic areas and on the implementation of suitable related programmes, integrating its work with the contributions of the other existing committees.[8] The High-Tech Forum reports to the Federal Government. The inputs it provides significantly influence the implementation and further development of the HTS, the fact that it was established so recently limits its active work to only two years, unless the HTS is further extended. Industry has played a central role in both the Alliance and the Forum, both in advising on the HTS and in lobbying for its continuation across successive governments.

While many individual instruments used have been evaluated, the development of the High-Tech Strategy is marked by the absence of any overall evaluation of the strategy or the full scope of its activities. As late as 2015, the EFI made recommendation for the development of an evaluation strategy, covering the indicators, success criteria and evaluation methods in addition to a clear for integrating the evaluation outcomes into policy (EFI, 2015). The review even went so far as to stipulate the inclusion of such a framework within the design process of further research and innovation policy measures.

The one external 'evaluation' at the level of the strategy that we can identify (Heimer, et al., 2011) is effectively a monitoring exercise and was perceived as supporting the federal ministries in assessing the design and the structure of the HTS up to 2011. 'Are the challenges framed in the right way?' was one example of question characterising the evaluation, which studies the innovation system for each thematic areas, looking to identify drivers and barriers. It did not tackle the classical evaluation questions of efficiency and effectiveness so it does not provide information for accountability. Rather, it effectively contributes to the design and management of the HTS.

While it is possible to point to significant improvements in the innovation performance of German industry during the period of the HTS, there is no evaluation evidence that allows these improvements to be attributed to the HTS.

Key lessons from the HTS include

- The value of a long-term commitment to a programmatic activity
- The provision of staff resources via a central coordination activity in the BMBF
- The increasing credibility of the HTS as its goals evolved, based on a stream of advice from legitimate experts in the research and industrial communities
- The importance of a cross-government approach, led by a single coordinating ministry
- The need for a clear link to government, securing both money and legitimacy
- The importance of developing a multi-level evaluation strategy from the start, and connecting this to programme governance

5.4 Lessons from the cases

These three cases suggest a number of lessons. They appear to us to be highly consistent with messages from the literatures studies in other parts of this report.

[8] The High-tech forum. Available at: http://www.hightech-forum.de/en/auftrag/

- Intervention design and governance
  - Broad political commitment over long periods of time is needed in order to implement and sustain a large, complex programme. Without this, it is vulnerable to the kind of ideologically-based discontinuity that generated the Top Sectors. The UK case is especially positive in this respect, because the political consensus resulted in legislation, though it is not clear that all large and complex programme require a specific legal basis
  - Cross-departmental or ministry coordination mechanisms enable the whole-of-government approach needed to address the goals of the three programmes studied, and those of interventions to address the societal challenges more generally
  - The programmes nonetheless need to be led – there should be a single organisation with the legitimacy to be responsible for overall coordination and that has adequate staff resources to organise and manage the programme. This is clearly present in the UK and Germany, but the lack of an over-arching strategy and structure for the Top Sectors promotes fragmentation
  - Correspondingly, there needs to be well-articulated multi-level governance, with different levels responsible for implementing different parts of the overall programme
  - The programmes benefit from high-level advice that derives from groups that have strong social legitimacy. It can be important to consult widely, but the strategy nonetheless needs legitimation from people with social authority
  - Perhaps the key part of legitimation must come from government, confirming society's ownership of the programme and therefore being able to provide resources
- Monitoring and evaluation
  - A multi-level evaluation structure is needed and should be connected to the governance system but independent of it
  - Reporting from both the programme leadership and the evaluators should be in the public domain
  - Evaluation should tackle the classical evaluation issues of relevance, efficiency and effectiveness. It is desirable for the evaluators and others to provide analysis intended to promote learning and the development of the programme but his does not replace the role of evaluation in accountability

6 Views from Swedish policymakers

This section is based on interviews with seven Swedish policymakers in the area of research and innovation. These interviews aimed to test the degree to which policymakers believe Sweden will increasingly launch larger and more complex interventions, the extent to which existing arrangements for evaluation are adequate and change needs in evaluation.

6.1 The use of large and complex interventions

Our interviewees saw limited use of large or complex programmes to date, exceptions being some long-standing interventions in the energy area and Vinnova's Strategic Innovation Programmes (SIPs). Generally, Swedish research and innovation policy suffers from poor horizontal coordination, with ministries and their agencies tending to remain within their own sectoral silos. The ability of policymakers and government to think and act in systems terms is insufficiently developed. Equally, in the areas that we considered, there are no strong examples of multi-level governance, except to a degree within individual silos.

A limited understanding within government ministries and agencies of the interdependence of policy areas, government agencies and a range of stakeholders at other levels is detrimental to the ability to see understand and observe system level impact. While several government agencies are starting to realise their own limitations and/or the limitations that their own practices bring to others, putting this realisation into practice is hard and slow.

It appears that the SIPs may be precursors of a new style in policy that – consistent with thinking in the sustainable transitions area – involves a wider range of stakeholders, intervention over extended periods, crossing the boundaries of the 'sector' responsibilities of individual ministries and involving a delegation of some of the programme design and implementation functions normally fulfilled by agencies to groups of beneficiaries, in the form of PPPs. This will challenge the ability of the Swedish policy system to design, manage, monitor and evaluate programmes.

6.2 The need for systemic evaluation

Even in the current situation, the policymakers feel that evaluations to an insufficient extent take a systemic perspective. Evaluations should better capture secondary (indirect) impact, i.e. beyond the primary object of a public intervention. Beyond this, further work is needed to bridge the gap between evaluations with limited scope (e.g. of a specific programme or instrument) and the way entire innovation systems or sub-systems perform. This is in part due to limited demand for such evaluations (both from the government and from its agencies), but it is also said to be due to lack of capabilities among evaluators. While the community doing research on research and innovation is said to be able to take a more systemic perspective, the methods needed at the same time to handle the micro and the more macro perspectives are not present in the same organisations and are thus rarely combined in the same evaluation.

Government analysis agencies (including the Swedish Agency for Growth Policy Analysis, the Swedish Agency for Public Management and the Swedish National Audit Office) are said to also shy away from taking a systemic perspective.

The desire for more systematic evaluation practices is not only based on the need for summative evaluation to address the systemic effects of individual interventions but also in genuine needs, both within government and within its agencies, for continuous and updated information on what they want to achieve, how they are to achieve it, how successful on-going interventions are, and completed ones were. To achieve this, they need systematic evaluation practices that go beyond the monitoring–evaluation–impact assessment perspectives of the past. Continuous in-house learning is the key driver, and this requires a substantial in-house involvement in evaluations. This in turn creates a tension between learning in-house and legitimacy (as offered by external evaluators). Different funding agencies

reach different conclusions in this context, with some arguing for insourcing of evaluations and others for outsourcing.

6.3 Improvement needs

While there is at times a desire from evaluation clients for more ambitious evaluation approaches, they rarely provide budgets that allow suppliers to experiment with new methods (or combinations thereof) and to include the high proportion of senior capacity that would be required to explore such new approaches and draw appropriate system-level conclusions. On the same note, government agencies are not rewarded for taking on difficult tasks (including procuring ambitious and risky evaluations).

There is a sense of frustration when the findings of major evaluations are not acted upon, particularly by the government. Evaluations that are (obviously) not used undermine evaluation as a policy instrument. There is also a sense of frustration at the lack of 'organisational learning' or organisational memory at all levels. Findings of evaluations often are not converted into conclusions that are used in policy, and conclusions drawn are quickly forgotten.

The change must come from the top. The government must start requiring its ministries to take a systemic perspective, ministries must instruct its agencies to collaborate and ultimately to procure evaluations that take a systemic perspective. This means that all levels must develop their capabilities as 'clients' and discontinue old practices. This is recognised as being very difficult, but the realisation of such needs is (ever so) slowly emerging.

6.4 Methodological issues

Some policymakers are keen to use positivist evaluation methodologies – ideally RCTs – because these are regarded as providing the strongest evidence about impact. It is recognised that such methods produce little evidence about the causation mechanisms that link interventions to impact and that they can be difficult to apply in many evaluative situations. There is nonetheless a desire to apply them.

Others reject positivist methods (at least when used on their own) because they need formative information from evaluation to understand the mechanisms of causation (for example, to support evolution of the intervention strategy over longer periods).

Few of our interviewees reflected on the difference between micro-level and systemic behaviour. One methodological bridge could involve developing better methods to capture spill-overs. Evaluations usually concentrate on effects among identifiable beneficiaries and users because spill-overs are very hard to capture. Connecting micro interventions to macro (or systemic) change remains difficult. Few evaluations try to do so, not least because this would expose evaluation commissioners to the risks associated with immature methods.

Despite the difficulties, policymakers are interested in attribution in the sense of knowing the 'net effect' of the intervention or knowing **how much** of the change that takes place apparently as a result of the intervention and what can be attributed to it rather than to other factors. No-one raised the idea of contribution analysis as a way to avoid this problem.

6.5 Implications

Faced with differing needs, it is perhaps not surprising that the policymakers we spoke to did not speak with one voice. In Sweden as elsewhere, one driver of difference is that ministries dominated by economists (typically finance and industry) like positivist, especially econometric, methods. At the same time there is a growing interest across the board in understanding the systemic role and effects of intervention – requiring a much broader set of methods.

One of the fundamental principles of complexity theory is non-reductionism, which states that the behaviour of complex systems cannot solely be explained by behaviour at the lowest levels but that other higher-level interactions need to be taken into account. Since we are here discussing interventions in complex systems, this implies that evaluations that start with low-level interventions are unlikely ever

to be able to explain systems behaviour without taking higher levels into account – and this is well beyond the frames of reference seen in commissioned evaluations.

The interviews suggest that while there is an increasing need to be able to govern and evaluate large and complex interventions, the silo'ed nature of the Swedish policy system makes it difficult to take the necessary systemic approach. Policymakers realise this and are starting to look for solutions. To some extent, they feel that relevant methods are to be found in different communities, making it hard to bring together the robust, micro-evaluation techniques available in the consulting community with the more systemic ones shared by some researchers.

While some parts of the system avoid the risks associated with more adventurous evaluation methods, there remains a need on the one hand to deploy robust evaluation techniques that are convincing to government and on the other hand to use approaches that address complexity and enable learning in the policy and implementation system. Some central authority – presumably the government – needs to encourage a more systemic and probably multi-level approach to evaluation that connects with the growing need for multi-level governance of large and complex interventions.

In doing this, there is a need to take a more realistic position about what various evaluation techniques can and cannot do, about ways in which they may be used to complement each other and about the extent that single evaluations focused on micro-level interventions can explain systemic change. It will be necessary to bring together different evaluation communities and evaluation methods in order to improve practice.

7 Lessons learnt

Here we aim to sketch an approach to evaluating CITPs in Sweden, based on the literature reviewed and our reading of the three case examples. The idea of sedimentary layers has appeared more than once in this review, and is a useful reminder that what we propose here relates only to that special category – most evaluation activities deal with smaller things and are best dealt with close to the interventions that need to be evaluated. So we aim to deposit a fresh layer of sediment here, rather than to dig a hole.

First we summarise our findings, based on the literature review, the three cases and our interviews with Swedish policymakers. Then we suggest a structure for governing and evaluating large and complex programmes specifically in the Swedish context.

7.1 Findings of the review

7.1.1 Why do we need to deal with CITPs?

Policies for research, innovation and industry have become broader and broader in their scope over the past decades, as we have learnt that individual initiative does not succeed on its own: its success depends upon the way the wider innovation system functions. Actors are thus interdependent. This understanding was already driving policy towards larger-scale interventions at the point where Western societies became concerned about the 'societal challenges' that appear to pose existential threats to society. The industrial growth that we have promoted for so long turns out now to be part of the problem. The societal challenges really do affect most or all of society and addressing them requires much broader engagement than was necessary for traditional, growth-orientated innovation and industry policy. The problem of coordinating different actors (not least within government) that was already becoming obvious in innovation policy is even more difficult in addressing the societal challenges.

This poses a further policy challenge. Whereas policies for industrialisation and growth can support innovation in various ways, they do not have to (and, experience suggests, they should not) try to specify particular outcomes. In addressing climate change, for example, there is no choice but to address ioxideemissions. That means there is a need to impose goals about which all or most of society agrees, and that cannot be done without wide involvement. A growing part of our policy will, at least for the time being, therefore need large, programmes to tackle problems that are both social and technical in character and that engage large parts of society. As the rather convoluted literature on transitions makes clear, we do not yet know well how to manage and govern such things, so there will need to be a period of experimentation.

7.1.2 How do we need to govern large and complex programmes?

The so-called societal challenges place new demands on the way we design, govern and evaluate government intervention in research and innovation.

In the post-War period, we largely relied on the idea that scientists know best about what research should be done and that – if given generous funding and left to their own devices – the results of their work would turn out to be socially useful. We still fund basic research in this way. From the 1960s, society started to make more demands for specific types of results, especially those useful to industrial development, so we developed ways to prioritise particular research themes in addition to funding 'free' or basic research. Over time we have learnt that coordinating different parts of the research and innovation system so as to produce successful innovation is complicated and requires coordination across government, science and industry. We are still a long way from finishing the task of devising the best governance and evaluation structures to do this.

The societal challenges exacerbate these difficulties. Now we have to deal with the 'quadruple helix' of government, science, industry and wider society and address the need for far-reaching changes across the responsibilities of different ministries and in the ways many different parts of society function. This means we now have to innovate new styles of governance and evaluation.

Each of the three generations of intervention aims to rectify different 'failures' in the research and innovation system, starting with the market failure that means private firms invest a socially sub-optimal amount in research, going on to address failures within the way our research and innovation systems work and culminating in the need to find ways to reach beyond market forces and intervene to address the societal challenges. Current policy needs to tackle all three types of failure.

Research and innovation policy instruments have evolved through the three governance generations from simple grant funding for individual researchers through bilateral research-industry collaborations to increasingly large-scale and complicated funding programmes involving many actors at once. At the extreme, large and complex programmes emerge to address major systemic shifts in national competitiveness and in order to tackle socio-technical transitions.

Big, complicated and complex programmes tend to need multi-level governance, coordinating the implementation of strategy but also providing the local strategic intelligence and flexibility to adapt design to local needs. Such governance seems more powerful and more likely to focus on societal goals if there a degree of central power or compulsion. Even then, however, there is a risk that the higher levels lose touch with the lower ones. But networked governance without strong central authority and links to government easily evolves to follow the goals of the network stakeholders rather than societal goals.

7.1.3 Evaluation

The purpose of public policy evaluation is to assess the relevance, efficiency and effectiveness of policy interventions and to give advice to policymakers about how to improve current and future interventions. This cannot be done without robust method, objectivity and ways to explain the mechanics of how interventions intervene.

Not all approaches produce evaluations that conform to these expectations. Broadly there are three alternatives, of which only the *realist* approach is fit for purpose. It is the only credible approach to evaluating programmes in general, as well as large and complex ones in particular.

Positivist evaluations tend to use econometric and other approaches (often inspired by clinical research in medicine) to try to establish statistical relationships between interventions and effects. These cannot explain mechanisms of causation and use simple and very narrow indicators of success. On their own they are rarely robust for use in policy. *Constructivist* evaluations focus on the perceptions and interests of stakeholders, providing information that can help them learn how to improve the intervention, but it has little interest in providing accountability – and especially in making sure that interventions continue to serve the needs of society as a whole rather than of the stakeholders. The *realist* approach focuses on providing objective, factual evidence about both mechanisms and effects. It develops and tests a 'theory' about how an individual intervention causes effects. It aims to be robust by using multiple methods (including positivist and constructivist ones, where appropriate) in parallel, providing information that meets both the need for accountability and the need for learning.

Theory-based methods for understanding causation mechanisms can be demand very high levels of evaluative effort if they explore in detail how every step in a causal chain works. There is some consensus in the evaluation community that very simple models are not very useful – especially for learning – so prescriptions for how to build more sophisticated theories and becoming more demanding. However, the large amount of effort needed to explore detailed theories of change implies that the evaluation designer needs to decide when this effort is justified and when a quicker approach is adequate.

Theory-based evaluation methods traditionally use linear logic models or theories of change. Complexity poses a challenge because it means there can be feedback loops leading to changes in the complex system. In that case, the causality becomes non-linear. Evaluators find themselves increasingly working in complex situations – not only in large complex programmes but also in much smaller interventions, for example in economic development or social policy. The complexity can be social – but it can also be technical. As a result, evaluators need to work with logic models that contain feedback or are circular. Evaluation needs to be done in a staged way that can cope with such changes in logic while still

maintaining their accountability to wider society and not becoming captive to the needs or views of the stakeholders involved.

Leading practitioners in evaluation in practice understand different things by 'complexity' and confront it in different ways. There is no clearly good practice that we can use here. Evaluators need to recognise that evaluating entire socio-technical systems is close to impossible. It is better to make use of what is already known about the performance of the intervention being studied and similar ones, focus on the parts of the intervention that seem most crucial to success, and compare similar interventions in different circumstances in order to make it possible to address complexity with limited evaluation resources.

A new interest in the evaluation literature is 'developmental evaluation'. This aims to tackle complexity but does so via the perceptions and the interests of the intervention managers and stakeholders, so it is 'constructivist'. It provides them with formative information to help them learn and make adjustments to the intervention. Development evaluation sets out to make itself accountable to the intervention stakeholders but not to society. It is therefore not alone appropriate for policy evaluation in the conventional sense.

If large-scale and complex interventions need multi-level governance, then they also need multi-level evaluation. It is tempting to consider setting up a national evaluation system covering a wide range of topics, so that large programmes could be treated within the set of parameters it defines. The fact that no country does this is perhaps a warning signal. Where national systems exist address a small part of the national system and focus on quality control and accreditation. They tackle stable systems with relatively homogenous components, such as schools, so they do not offer very much for evaluating interventions that are fine-tuned to specific circumstances and may change over time.

The main experience with multi-level evaluation comes from the EU. As with our discussion of multi-level governance, a major problem is that the strategic-level evaluations are out of touch with the lower-level ones, so they understand little about the mechanisms that cause change at 'grass roots' level. This seems to be partly because higher-level evaluations do not concern themselves with information from the lower levels and partly because different people and organisations plan and do the evaluations at high and low levels, often in ways that are not compatible. Some writers suggest evaluating individual, lower-level interventions and then in some way adding them up. Others propose a multi-level logic model, connecting the strategic, organisational and operational levels together. But none of these has been tried in practice. Nor has there been much work on meta-evaluation or systematic review to bring the different levels of the programme together. Here we confront a problem that is familiar from other fields as well as evaluation, including economics, biology and geography: the difficulty of connecting macro and micro levels of analysis. There is no recipe available off the shelf. There is a clear need for evaluators and programme designers to innovate.

7.1.4 Socio-technical transitions

Conceptually, there is a clear distinction between transition interventions and other large programmes. Transition interventions aim to make disruptive changes to large socio-technical systems in order to meet specific societal goals, such as reducing CO_2 emissions. Other CITPs do not have so specific goals – complex innovation programmes aim to promote successful innovations, for example, without necessarily specifying in advance what they will do.

However, from a governance and evaluation perspective, transition initiatives and complex innovation programmes differ only in so far as they are at different points on a spectrum of complexity. Each will in any case be unique and therefore require unique governance and evaluation arrangements, with a unique intervention logic and a unique set of indicators associated with the need to measure progress. From a policy perspective, therefore, the differences between transition initiatives and complex innovation programmes are small. It is perhaps more important to note that the similarities are large and that we therefore expect to be able to use similar tactics in designing governance and evaluation arrangements.

We need to understand about managing socio-technical transitions, because they will be needed in order to address at least some of the societal challenges. Economists have tried to explain how such transitions work in terms of past transitions in 'technological regimes' – in effect the systems of innovation that cluster around individual classes of technology, impeding or fostering the introduction of disruptive technologies. Historians of technology have adopted a broader approach to transitions – such as that between sailing ships and steam – that takes account of a wider range of social factors and actors. Those who now work on 'transitions management' emphasise the broader approach and the need to involve a wide range of stakeholders from many parts of society in achieving transitions, such as removing fossil fuels from the electricity production and distribution system.

Moving from observation of past transitions to management of future ones makes it especially important to identify which *functions* – such as knowledge development, market formation, creation of legitimacy – need to be performed in order to make the transition possible.

Most writing about transitions now involves a multi-level perspective. The central level is the 'socio-technical regime', which broadens the earlier idea of technological regimes to take better account of wider society. This is the battleground between alternative socio-technical systems. It is fed from below by a flow of innovations – some of which are absorbed as incremental changes and some of which have potential to trigger transitions. Above is the 'landscape', comprising politics, broader society and the broader context, such as infrastructures.

Intellectually, the transitions writers faced an important challenge in moving from historical analysis of transitions to specifying how to trigger them. Past transitions have generally happened rather spontaneously as individual innovations perturb the socio-technical regime. Addressing the societal challenges means that instead of transitions being triggered bottom-up, we need to find out how to drive them top-down on the basis of policy. This is why the sociotechnical transitions literature has been heavily criticised for failing to take account of politics, policy, agency and the ability of individuals to trigger changes – precisely the ingredients that are needed in moving from observing to acting.

Moving from description to prescription means we have to understand the barriers that need to be overcome in order to enable transitions. These include short-term political processes, fragmented institutions, lock-ins of various kinds (including societal attitudes and habits) and often the active resistance of incumbent producers and users, who feel they stand to lose from change. Managing transitions involves finding ways to overcome these barriers and will involve the 'creative destruction' of old structures as well as the creation of new ones. Much of the discussion about transitions management involves the creation of a multi-level network governance but in effect involves 'governance without government' and has only just started to consider who appoints and legitimises the governance or how it is connected to the interests of society as a whole rather than the stakeholders with an interest in the transition, or even who funds it.

A further problem with much of the transitions management literature has been that it is conceptual rather than empirical. In the last few years, however there have been a growing number of studies of climate transition management in cities. These suggest that practical transition management is strengthened by the legitimacy and regulatory power of government. Preferably, government should lead, though where government fails to act voluntary coalitions can nonetheless to some degree coordinate transitions. They confirm the need for 'arenas' or 'platforms' where transition management can be negotiated, discussed and coordinated as well as for 'reflexivity' in the sense of self-analysis and adaptation of strategy based on learning from experience and evaluation. It is important to involve the relevant stakeholders, especially those who can impede the transition. The misalignment between the shortness of election cycles and the long periods of time during which transitions must be managed means that discontinuities among the people in government, policymakers and policy can be disruptive factors. However, a common thread among the cities is disconnection from the national and international governance levels, so the scope of the transitions handled is regionally limited. City initiatives tend to be more effective if they succeed in connecting to the higher governance levels.

The transitions literature needs to go further to tackle the integration of transition management into wider (especially national) systems of government, policymaking and governance and to address the central issue of accountability. Evaluation in the transitions literature serves transition management, not society. This not sustainable. It is clear that the need for feedback and iterative strategy development is greater in transitions than in many kinds of government intervention and that traditional, linear logic models need to become more iterative. This means that evaluation methods have to evolve, not that evaluation should abandon its role in providing accountability to the government and society. A key part of that evolution is the need to develop better multi-level evaluation models that address the difficulties of communication between governance and evaluation at the different levels.

7.1.5 Evidence from case studies

Evidence from the three cases studied in this report – the UK Climate Change Act, 2008, the Netherlands Top Sector Policy and the German High-Tech Strategy – is largely consistent with what we see in the broader literature. Large, complex programmes can only be effective if there is longer-term political commitment. They need to be connected to a multi-level governance that stretches across the needed parts of government as well as a wide range of external stakeholders, including those with social authority. By inference, we have to reject the idea in parts of the transitions literature that it is possible to do governance without government. Programmes require a management organisation that has legitimacy and funding. An independent, multi-level evaluation is also needed, providing accountability as well as information for learning. Strategies need periodically to be reviewed and revised in a transparent way, taking account of evaluation and other evidence.

7.1.6 Views of Swedish policymakers

Our interviews with Swedish policymakers indicated that they want to see evaluations at the instrument level that are as robust as possible, while at the same time learning about the systemic effects of intervention. How to achieve this combination is not clear. Not all parts of government are willing and able to do the experiments and take the risks associated with developing a better systemic understanding and the research and evaluation community seems not to have resolved the methodological issues that would have to be tackled. The problem becomes more acute when large and complex programmes are considered, though so far there have been few of these. Some central authority – presumably the government – needs to encourage a more systemic and probably multi-level approach to evaluation that connects with the growing need for multi-level governance of large and complex interventions. In doing this, there is a need to take a realistic position about what various evaluation techniques can and cannot do, about ways in which they may be used to complement each other and about the extent that single evaluations focused on micro-level interventions can explain systemic change. It may be necessary to bring together different evaluation communities and evaluation methods in order to improve practice.

7.2 How should we evaluate large and complex programmes in Sweden?

Here, based on our findings from the review, we first specify requirements or criteria about the needed evaluation system that emerge from the review. Second, we dismiss certain possibilities, which do not appear to be useful. Third we sketch the way to handle the governance and evaluation of large and complex programmes.

The size and complexity of CITPs such as national innovation strategies or transitions strategies means we have to devise governance structures that (a) connect together central and local actions, for example among international authorities, nation states, regions and organisations and (b) ensure a level of 'horizontal' coordination among these hierarchies that ensures consistency across the programme. The literature shows that these are ambitious requirements, that the arrangements for meeting them need to be tailored to the specific programme (so they are likely to differ from case to case) and that they tend to be multi-level in character, so that there is a sensible division of labour. We also know that in such multi-level arrangements there is a need for distributed strategic intelligence, so that all levels have the capacity intelligently to play their part and to contribute to the whole. The literature shows that these governance arrangements are often 'incomplete' in the sense that particular initiatives do not even begin to tackle the whole of the problem addressed by the programme, but are nonetheless needed if we are to

build a wider pattern of change. One of the most difficult design decisions in intervention and in evaluation is where to set the boundary and the same applies to programme governance.

There are good reasons to decentralise evaluation. One involves the principle of 'subsidiarity': only centralising things that are better done centrally than at a lower level. In research and innovation policy, centralisation makes policy development and implementation remote from the delivery of policy and the context in which the policy operates, reducing its relevance. A second is the idea that a system that invokes subsidiarity needs 'distributed strategic intelligence', including evaluation, so that all the actors in the policy system operate on the basis of locally-relevant evidence.

However, the tendency for lower- and higher-level governance and evaluation to become disconnected means that multi-level governance must be designed specifically to avoid this. A central responsibility is needed for coordinating the evaluation of a large and complex programme so that the entire structure is supplied with strategic intelligence, without at the same time compromising the subsidiarity principle. Something more like an arena than a central boss is needed.

It is important to stress that such programmes are and will remain the exception rather than the rule. It will continue to be necessary to govern most research and innovation efforts using first- and second-generation arrangements. It would therefore be counter-productive to tear down the existing structure of organisations that largely reflects the needs of first- and second-generation governance and to replace them with a third-generation structure. The social effort of building third-generation governance structures will be considerable and each structure will have to be designed on the basis of the specific characteristics of its programme. These structures are inherently temporary (even if 'temporary' can mean quite a number of years) so it is important not to jeopardise the benefits of Sweden's well-developed system of ministries and agencies or its ability to handle first- and second-generation policies.

7.2.1 What criteria should a new approach be able to fulfil?

The review implies that the following criteria should be used in designing a new evaluation process.

- The overriding criterion is that the evaluation system be capable of producing policy evaluations that test the relevance, efficiency and effectiveness of the programmes and provides advice about how to modify current designs and improve future ones. This requires a realist approach
- The evaluation system needs ultimately to answer to society and therefore to answer to government either directly or through some agency
- It should create ad hoc structures relevant to the present programme but overlaid on (and making best use of) existing and more permanent structures and capabilities
- It needs to explore mechanisms of causation and therefore to involve theory-based evaluation techniques using mixed methods
- It should within reasonable bounds be cost-effective, if necessary focusing evaluation effort on critical uncertainties rather than attempting to evaluate entire systems
- The approach should be able to handle the specific characteristics of CITPs, which are not necessarily present in simpler interventions, such as
  - Complexity, emergence and systemic change
  - Inter-organisational cooperation and coordination
  - A wide range of stakeholders, beyond the 'usual suspects' in research and innovation policy
  - Multi-level governance and evaluation
  - The need periodically to cope with changes in the goals and intervention logic of the programme as these evolve over time while maintaining accountability
- It should include structures and processes that link evaluation (and governance) at high and low levels of the hierarchy, allowing them to learn from each other

- There simply is no method available off-the-shelf that adequately connects the evaluation of individual initiatives at the micro level with the macro or systems level. In so far as systemic aspects cannot be met by current evaluation actors using current techniques, therefore, the evaluation system should have discretionary income and authority to use it to promote methodological development

*7.2.2 What should a new evaluation approach **not** do?*

Based on the review, we can dismiss certain hypothetical possibilities at this stage.

- There is no evident reason to restructure the way most research and innovation evaluation is done in Sweden today – though effort is needed everywhere to improve the ability to handle systemic aspects of intervention
- In particular, creating a central evaluation agency would be counter-productive because it would remove the ability to generate relevant strategic intelligence by asking and answering locally-relevant evaluation questions, reducing the effectiveness of the agencies
- Nor is it practical to devise a standard set of metrics and evaluation issues and apply it to all programmes. That kind of approach works well in structured systems where the components to be evaluated are relatively homogenous, for example in the inspection of schools and the homologation of higher education institutions. CITPs are individually unique and cannot be handled in such a structure

7.2.3 How could these principles be operationalised?

Since evaluation is ultimately part of governance, some aspects of evaluation are inseparable from the wider governance of programmes. So here we need first to deal with governance. Where a large and complex programme can be handled within a single agency (or even a small number of agencies) there is every reason to use Sweden's existing structures. On the infrequent occasions when this is not the case, we envisage a system of governance that would involve

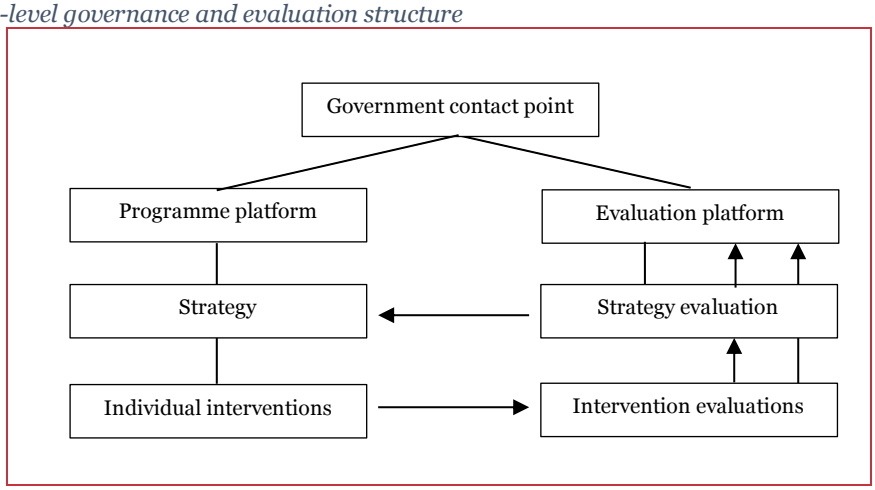
- The government designating a point in its structure to which a programme would ultimately answer. This is not easy in Sweden, since there is no government-wide governance for the research and innovation system (as Finland has had at some points in history in the Research and Innovation Council). The natural point of authority is the education minister, who has overall responsibility for innovation as well as research. But this would involve a dangerous narrowing of perspective on an intervention that by its nature should involve many if not all parts of government. In the absence of a whole-of-government coordination body, it may be necessary to allocate groups of relevant ministries on a case by case basis to each programme
- Creating a 'platform' to design and govern the programme, in which a wide range of agencies (or other agents such as cities) and other stakeholders are involved but which has staff resources and a leadership team of its own, able to lead the programme. The platform would coordinate individual interventions managed by existing agencies (or combinations thereof), using their own budgets, resulting in a hierarchical structure of interventions with a reporting structure leading back up to the platform
- As in the examples of the UK Climate Change Act and the German High-Tech Strategy, the platform would be responsible for revising the strategy and plan periodically – perhaps every four years, to fit into the government cycle and mesh with the Research and Innovation Bills – but the period should also be influenced by the needs of the individual intervention. Strategy revision should be informed by evaluation not only at the overall level but also at the level of sub-interventions, as well as by consultation with stakeholders and relevant sources of expert knowledge

The use of platforms to manage large, complex programmes is not without precedent in Sweden. A specific example is the IT-Delegation that ran the IT4 programme – the industrial R&D component of the national IT strategy in the 1980s. The delegation devised and implemented a strategy for academic-industrial collaboration in IT, coordinating state funds from STU (The Swedish National Board for

Technological Development, predecessor to VINNOVA), FMV (the defence procurement executive) and Televerket (Swedish Telecom).

Figure 16 illustrates the multi-level governance and evaluation structure we would suggest, based on the literatures reviewed. On the operative side, the programme platform is appointed by and reports to government. It devises an strategy and implements it through various agencies or other agents. (This could include public-private partnerships.)

Figure 16 Multi-level governance and evaluation structure



It is important that the evaluation function (which we call a platform, but it could be a committee or be assigned to one or more expert groups independent of government) is separate from the programme and does not report to the programme but to the government. This enables accountability as well as objectivity. It should devise and implement an evaluation strategy in discussion with government, the platform and the implementing agencies. It should then do (or more probably outsource, using relevant and independent expertise) evaluation at the strategy level on a cycle consistent with the (re)planning cycle of the programme.

Evaluations at the intervention level should be organised by the implementing agencies but must be reported to the evaluation platform and be partially aligned with needs defined buy that platform. This, combined with the use of agency on the operational side, should ensure not only the local relevance of the intervention evaluations but also their ability to feed into the higher-level strategic evaluations. The evaluation strategy should involve at least the following elements.

- An ex ante evaluation, in effect providing a design review for the programme
- An intervention logic articulated at a level of detail that allows the roles of various agents to be identified and that considers key assumptions upon which the success of the programme depends
- A hierarchy of evaluation questions linked to the logic, indicating which are to be answered at the overall level and which at the level of individual evaluations. This specification should leave sufficient space for the agencies to insert their own evaluation questions, relevant to their intervention and local circumstances. But there should also be enough commonality to allow a level of meta-evaluation or structured review
- A timetable for the periodic overall evaluation and the individual evaluations of interventions at the agency level
- Processes for reporting evaluation results up and down between the evaluation and governance levels shown in Figure 16
- Processes for dialogue with the government contact point, the programme platform and the agencies

Especially since this will be a new way of organising programmes and evaluations, it needs itself to be evaluated at the end of the first planning cycle.

8 Bibliography

Academy Health. (2017). *Evaluating Complex Health Interventions: A Guide to Rigorous Research Designs*. Washington DC: Academy Health.

Arnold, E. (2012). Understanding the long-term impacts of R&D funding: The EU framework programme. *Research Evaluation, 21*(5), 332-343.

Arnold, E., & Boekholt, P. (2003). *Research and Innovation Governance in Eight Countries A Meta-Analysis of Work Funded by EZ (Netherlands) and RCN (Norway)*. Brighton: Technopolis.

Arnold, E., & Carlberg, M. (2009). *Mapping Nordic R&D Cooperation Instruments*. Brighton: Technopolis.

Arnold, E., & Guy, K. (1986). *Parallel Convergence: National Strategies in Information Technology*. London: Frances Pinter.

Arnold, E., & Guy, K. (1992). *Evaluation of the IT4 Programme*. Brighton: Technopolis and SPRU.

Arnold, E., Åström, T., Boekholt, P., Brown, N., Good, B., Holmberg, R., . . . van der Veen, G. (2011). *Impacts of EU Framework Programmes in Sweden, VA 2008:11*. Stockholm: Vinnova.

Arnold, E., Kuhlmann, S., & van der Meulen, B. (2001). *A Singular Council: Evaluation of the Research Council of Norway*. Oslo: Royal Norwegian Ministry of Education, Research and Church Affairs.

Arnold, E., Malkin, D., Good, B., Clark, J., & Ruiz Yaniz, M. (2009). *Evaluating the National Innovation Strategy for Competitiveness*. Santiago de Chile: National Council for Innovation for Competitiveness.

Arrow, K. (1962). Economic welfare and the allocation of resources for innovation. In R. R. Nelson, *The Rate and Direction of Innovative Activity: Economic and Social Factors* (pp. 609-626). NBER Press.

Avelino, F., & Wittmayer, J. M. (2015). Shifting power relations in sustainability transitions: a multi-actor perspective. *Journal of Environmental Policy Planning, 18*(5), 628–649.

AWTI. (2014). *Status of the top sectors in 2014*. The Hague: AWTI.

Barnes, M., Sullivan, H., & Matke, E. (2004). *The Development of Collaborative Capacity in Health Action Zones: A Final Report from the National Evaluation*. University of Birmingham.

Baumann, B., Bührer, S., Greer, H., Hochreiter, H., Mayer, S., Simmonds, P., . . . Kuhlmann, S. L. (2004). *Good practices for the management of multi actors and multi measures programmes (MAPS) in RTDI policy*. Vienna: TIG Technologie Impulse Gesellschaft.

Befani, B., Ramalingam, B., & Stern, E. (2015). Towards Systemic Approaches to Evaluation and Impact. *IDS Bulletin, 46*(1), 1-6.

Bergek, A., Jacobsson, S., Carlsson, B., & Lindmark, S. (2006). *Analysing the Dynamics and Functionality of Sectoral Innovation Systems. RIDE/IMIT Working Paper No 84426-007*. Gothenburh: Chalmers University of `technology.

Berkhout, F., Smith, A., & Stirling, A. (2004). Socio-technical regimes and transition contexts. In B. Elzen, F. W. Geels, & K. Green, *System Innovation and the Transition to Sustainability* (pp. 48-75). Cheltenham: Edward Elgar.

Bernal, J. D. (1939). *The Social Function of Science*. London: Routledge.

Bhaskar, R. (1997). *A Realist Theory of Science*. London: Verso.

Blamey, A. A., MacMillan, F., Fitzsimons, C. F., Shaw, R., & Mutrie, N. (2012). Using programme theory to strengthen research protocol and intervention design within an RCT of a walking intervention. *Evaluation, 19*(1), 5-23.

Bos, M. W., Hofman, E., & Kuhlmann, S. (2016). An assessment method for system innovation and transition (AMSIT). *International Journal of Foresight and Innovation Policy,, 20*(4), 185-214.

Boswell, M. R., & Mason, S. G. (2018). Regional Climate Planning and Local Outcomes in California. In S. Hughes, E. K. Chu, & S. G. Mason, *Climate Change in Cities: Innovations in Multi-Level Governance* (pp. 59-78). New York: Springer.

Bourgeois, E., & Hughes, S. (2018). Metropolitan Governance of Waste for Climate Change Mitigation: Examining the Case of Montreal. In S. Hughes, E. K. Chu, & S. G. Mason, *Climate Change in Cities: Innovations in Multi-Level Governance* (pp. 77-96). New York: Springer.

Braun, D. (1993). Who governs intermediary agencies? Principal-agent relations in research policymaking. *Journal of Public Policy, 13*(2), 135-162.

Braun, D. (2003). Lasting tensions in research policy-making – a delegation problem. *Science and Public Policy, 30*(5), 309-322.

Brinkerhoff, D. W., & Brinkerhoff, J. M. (2011). Public-Private Partnerships: Perspectves on Purposes, Publicness and Good Governance. *Public Administration and Development, 31*, 2-14.

Britt, H. (2016). *Complexity-Aware Monitoring*. Washington DC: USAID.

Bush, V. (1945). *Science, the Endless Frontier: A Report to the President on a Program for Postwar Scientific Research*. Washington DC: NSF.

Busquin, P. (2000). *Towards a European Research Area, COM(2000) 6 final, 18.1.2000*. Brussels: European Commission.

Bussels, M., Happaerts, S., & Bruininckx, H. (2013). *Evaluating and Monitoring Transition Initiatives: essons from a field scan, Research Paper 5*. Leuven: HIVA, KU Leuven.

Caffrey, L., & Munro, E. (2017). A systems approach to policy evaluation. *Evaluation, 23*(4), 463-478.

Carlsson, B., & Stankiewicz, R. (1991). On the nature, function and composition of technological systems. *Journal of Evolutionary Economics, 1*(2), 93-118.

Chen, H. (1990). *Theory-driven Evaluation*. Beverley Hills CA: Sage.

Chen, H., & Rossi, P. (1989). Issues in the Theory-driven Perspective. *Evaluation and Program Planning, 12*(4), 299–306.

Cherp, A., Vinichenko, V., Jewell, J., Brutschin, E., & Sovacool, B. (2018). Integrating techno-economic, socio-technical and political perspectives on national energy transitions: A meta-theoretical framework. *Energy Research and Social Science, 37*, 175-190.

Clark, H., & Andersson, A. A. (2004). Theories of Change and Logic Models: Telling Them Apart. *American Evaluation Association*. Atlants GA.

Cole, H., Freeman, C., Jahoda, M., & Pavitt, K. (. (1973). *Thinking About the Future: A Critique of The Limits to Growth*. Brightoon: Sussex University Press.

Cyert, R. M., & March, J. G. (1963). *A Behavioral Theory of the Firm*. New Jersey: Prentice-Hall.

Dale, A., Burch, S., Robinson, J., & Strashok, C. (2018). Multilevel Governance of Sustainability Transitions in Canada: Policy Alignment, Innovation, and Evaluation. In S. Hughes, E. K. Chu, & S. G. Mason. New York: Springer.

Davies, R. (2005). Scale, Complexity and the Representation of Theories of Change Part II. *Evaluation, 11*(2), 133-149.

de Bresson, C. (1999). An Entrepreneur Cannot Innovate Alone; Networks of Entreprises Are Required. *DRUID Conference*. Aalborg.

Delahais, T., & Toulemonde, J. (2012). Applying contribution analysis: Lessons from five years of practice. *Evaluation, 18*(3), 281-293.

DFID. (2009). *Political Economy Analysis How to Note*. London: DFID.

Douthwaite, B., Mayne, J., McDougall, C., & Paz-Ybarnegaray, R. (2017). Evaluating complex interventions: A theory-driven realist-informed approach. *Evaluation, 23*(3), 294-311.

Edler, J., Ebersberger, B., & Lo, V. (2008). Improving policy understanding by means of secondary analyses of policy evaluation. *Researh Evaluation, 17*(3), 175-186.

Edler, J., Geourghiou, L., Blind, K., & Uyarra, E. (2012). Evaluating the demand side: New challenges for evaluation. *Research Evaluation, 21*, 33-47.

Edler, J., Kuhlmann, S., & Smits, R. (2003). *New Governance for Innovation. The need for Horizontal and Systemic Policy Coordination*. Karlsruhe: Fraunhofer ISI.

Edquist, C., & Johnson, B. (1997). Institutions and organisations in systems of innovation. In C. Edquist, *Systems of Innovation: Technologies, Institutions and Organizations*. London and Washington: Pinter and Cassell Academic.

Ehnert, F., Kern, F., Borgström, S., Gorissen, L., Maschmeyer, S., & Egermann, M. (2018). Urban sustainability transitions in a context of multi-level governance: A comparison of four European states. *Environmental Innovation and Societal Transitions, 26*, 101-116.

Elzinga, A. (1997). The science-society contract in historical transformation: with special reference to 'epistemic drift. *Social Science Information, 36*(3), pp. 411-455.

Emilianoff, C. (2014). Local Energy Transition and Multilevel Climate Governance: The Contrasted Experiences of Two Pioneer Cities (Hanover, Germany, and Växjö , Sweden). *Urban Studies, 51*(7), 1378–1393.

Ernø-Kjølheide, E., & Hansson, F. (2011). Measuring research performance during a changing relationship between science and society. *Research Evaluation, 20*(2), 131-143.

Forrester, J. W. (1968). *The Principles of Systems, 2nd ed.* Lawrence KS: Wright-Allen Press.

Forrester, J. W. (1971). *World Dynamics*. Cambridge: Wright-Allen Press.

Fosse Hansen. (2013). Systemic evaluation governance. New logics in the development of organisational fields. *Scandinavian Journal of Public Administration, 16*(3), 47-64.

Fosse Hansen, H., & Rieper, O. (2009). The Evidence Movement: The Development and Consequences of Methodologies in Review Practices. *Evaluation, 15*(2), 141–163.

Freeman, C. (1987). *Technology Policy and Economic Performance: Lessons from Japan*. London: Frances Pinter.

Freeman, C. (1996). *Long Wave Theory*. Cheltenham: Edward Elgar.

Gates, E. F. (2017). Learning from seasoned evaluators: Implications of systems approaches for evaluation practice. *Evaluation, 23*(2), 152-171.

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research Policy, 31*, 1257-1274.

Geels, F. W. (2010). Ontologies, socio-technical transitions (to sustainability), and the multi-level perspective. *Research Policy, 39*, 495–510.

Geels, F. W. (2014). Regime resistance against low-carbon transitions: introducing politics and power into the multi-level perspective. *Theory, Culture and Society, 31*(5), 21-40.

Geels, F. W. (2014). Regime Resistance against Low-Carbon Transitions: Introducing Politics and Power into the Multi-Level Perspective. *Theory, Culture and Society, 3*(5), 21-40.

Geels, F. W. (2014). Understanding barriers to transitions in the Multi-Level Perspective (MLP). *KAPSARC Workshop, Manchester University, 13-14 May 2014*. Manchester.

Geels, F. W. (2018). Disruption and low-carbon system transformation: Progress and new challenges in socio-technical transitions research and the Multi-Level Perspective. *Energy Research & Social Science, 37*, 224-231.

Geels, F. W., & Schot, J. (2007). Typology of sociotechnical transition pathways. *Research Policy, 36*, 399-417.

Genus, A., & Coles, A. M. (2008). Rethinking the multi-level perspective of technological transitions. *Research Policy, 37*(9), 1436-1445.

Genus, A., & Coles, A.-M. (2008). Rethinking the multi-level perspective of technological transitions. *Research Policy, 37*, 1436–1445.

Georghiou, L. (2008, April 24). Europe's research system must change. *Nature, 452*, 935-6.

Gerrits, L., & Verveij, S. (2015). Taking stock of complexity in evaluation: A discussion of three recent publications. *Evaluation, 21*(5), 481-491.

Gladwell, M. (2000). *The Tipping Point: How Little Things Can Make a Big Difference*. New York: Little Brown.

Gollata, A. J., & Newig, J. (2017). Policy implementation through multi-level governance: analysing practical implementation of EU air quality directives in Germany. *Joural of Eurpean Public Policy, 24*(9), 1308–1327.

Grin, J., Rotmans, J., & Schot, J. (2010). Introduction: from persistent problems to system innovations and transitions. In J. Grin, J. Rotmans, & J. Schot, *Transitions to Sustainable Development. New Directions in the Study of Long Term Transformative Change*. New York: Routledge.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. London: Sage.

Hölscher, K., Wittmayer, J. M., Avelino, F., & Giezen, M. (2017). Opening up the transition arena: An analysis of (dis)empowerment of civil society actors in transition management in cities. *Technological Forecasting & Social Change*, DOI10.1016/j.techfore.2017.05.004.

Hanberger, A. (2013). Framework for exploring the interplay of governance and evaluation. *Scandinavian Journal of Public Administration, 16*(3), 9-27.

Hawkins, A. J. (2016). Realist evaluation and randomised controlled trials for testing program theory in complex social systems. *Evaluation, 22*(3), 270-285.

Haxeltine, A., Wittmayer, J., & Avelino, F. (2013). *researchTransformative social innovations: A sustainability transition perspective on social innovation*. Rotterdam: Erasmus Universiteit.

Heimer, T., Fischer, T., Treperman, J., Mannfield, D., Reiß, T., Lindner, R., . . . Weber, M. (2011). *Begleitforschung der Hightech-Strategie – Analyse zu ausgewählten Aspekten*. Frankfurt: Technopolis.

Hekkert, M. P., Suurs, R. A., Negro, S. O., Kuhlmann, S., & Smits, R. (2007). Functions of innovation systems: A new approach for analysing technological change. *Technological Forecasting & Social Change, 74*, 413-432.

Hendriks, C. M., & Grin, J. (2007). Contextualising reflexive governance: the politics of Dutch transitions to sustainability. *Journal of Environmental Policy and Planning, 9*(4), 333–350.

Hertting, N., & Vedung, E. (2012). Systemic evaluation governance. New logics in the development of organisational fields. *Evaluation, 18*(1), 27–46.

Hess, D. J. (2014). Sustainability transitions: a political coalition perspective. *Research Policy, 53*(2), 278–283.

Hessels, L. K., van Lente, H., & Smits, R. (2009). In search of relevance: the changing contract between science and society. *Science and Public Policy, 36*(5), pp. 387-401.

Hodson, M., & Marvin, S. (2010). Can cities shape socio-technical transitions and how would we know if they were? *Research Policy, 39*, 477-485.

Homsy, G. C. (2018). Size, Sustainability, and Urban Climate Planning in a Multilevel Governance Framework. In S. Hughes, E. K. Chu, & S. G. Mason, *Climate Change in Cities: Innovations in Multi-Level Governance* (pp. 19-38). New York: Springer.

Hooghe, L. (1996). *Cohesion Policy and European Integration*. Oxford: Clarendon Press.

Hughes, S., Chu, E. K., & Mason, S. G. (2018). *Climate Change in Cities: Innovations in Multi-Level Governance*. New York: Springer.

Hummelbrunner, R. (2011). Systems thinking and evaluation. *Evaluation, 17*(4), 395-403.

Ingelstam, L. (2002). *System – att tänka över samhälle och teknik,*. Eskilstuna: Statens Energimyndighet.

Jänicke, M. (2016). The Multi-level System of Global Climate Governance – the Model and its Current State. *Environmental Policy and Governance, 28*, 108-121.

Jacobsson, S., & Bergek, A. (2004). Transforming the energy sector: the evolution of technological systems in renewable energy technology. *Industrial and Corporate Change, 13*(5), 93-118.

Jacobsson, S., & Jonsson, A. (2000). The diffusion of renewable energy technology: an analytical framework and key issues for research. *Energy Policy, 28*(9), 625-640.

Janssen, M., den Hertog, P., Korlaar, L., Groot-Beumer, T., Steur, J., Rienstra, Y., . . . Erven, B. (2017). *Evaluatie Topsectorenaanpak*. Utrecht: Dialogic.

Kaiser, R., & Prange, H. (2004). National Evaluation Capacity Development: Practical tips on how to strengthen National Evaluation Systems. *Journal of European Public Policy, 11*(2), 249-266.

Kelly, T. J. (2010). Five Simple Rules for Evaluating Complex Community Initiatives. *Community Investments, Spring 2010 Volume 22, Issue 1, 22*(1), 19-36.

Kemmerzell, J. (2018). Innovations in European Climate Governance and Their Impact on Local Climate Policy: An Analysis of German Major Cities. In S. Hughes, E. K. Chu, & S. G. Mason, *Climate Change in Cities: Innovations in Multi-Level Governance* (pp. 39-58). New York: Springer.

Kemp, R., Schot, J., & Hoogma, R. (1998). Regime shifts to sustainability through processes of niche formation: the approach of strategic niche management. *Technology Analysis and Strategic Management, 10*, 175-196.

Keskitalo, E. H., Juhola, S., Baron, N., Fyhn, H., & Klein, J. (2016). Implementing Local Climate Change Adaptation and Mitigation Actions: The Role of Various Policy Instruments in a Multi-Level Governance Context. *Climate, 4*(1), 7.

Kivimaa, P., & Kern, F. (2016). Creative destruction or mere niche support? Innovation policy mixesfor sustainability transitions. *Research Policy, 45*, 205-217.

Kline, S. J., & Rosenberg, N. (1986). An overview of innovation. In R. Landau, & N. Rosenberg, *The Positive Sum Strategy: Harnessing Technology for Economic Growth* (pp. 275-305). Washington DC: National Academies Press.

Kuhlmann, S. (2001). Future governance of innovation policy in Europe—three scenarios. *Research Policy, 30*, 953-976.

Kuhlmann, S., & Arnold, E. (2001). *RCN in the Norwegian Research and Innovation System, Background Report No 12 in the Evaluation of the Reseach Council of Norway*. Oslo: Royal Norwegian MInistry of Education, Research and Church Affairs.

Kuhlmann, S., Boekholt, P., Georghiou, L., Guy, K., Héraud, J.-A., Laredo, P., . . . Smits, R. (1999). *Improving Distributed Intelligence in Complex Innovation Systems*. Karlsruhe: Fraunhofer-ISI.

Laes, E., Gorissen, L., & Nevens, F. (2014). A Comparison of Energy Transition Governance in Germany, The Netherlands and the United Kingdom. *Sustainability, 6*, 1129-1152.

Langfeldt, L., Godø, H., Gornitzka, Å., & Kaloudis, A. (2012). Integration modes in EU research: Centrifugality versus coordination of national research policies. *Science and Public Policy, 39*, 88-98.

Laredo, P. (1995). Structural effects of EC RTD Programmes, 1995, 34(3), . *Scientometrics*, 473-487.

Lay, M., & Papadopoulos, I. (2007). An Exploration of Fourth Generation Evaluation in Practice. *Evaluation, 13*(4), 495-504.

Leeuw, F., & Furubo, J.-E. (2008). Evaluation Systems What Are They and Why Study Them? *valuation, 14*(2), 157-169.

Lehtonen, M. (2014). Evaluating megaprojects: From the 'iron triangle' to network mapping. *Evaluation, 20*(3), 278– 295.

Loorbach, D. (2010). Transition management for sustainable development: a prescriptive, complexity-based governance framework. *Governance, 23*(1), 161-183.

Loorbach, D. A. (2007). *Transition Management New mode of governance for sustainable development (PhD Thesis)*. Rotterdam: Erasmus Universiteit.

Loorbach, D., & Rotmans, J. (2010). The practice of transition management: examples and lessons from four distinct cases. *Futures, 42*, 237–246.

Loorbach, D., & van Raak, R. (2006). *Strategic Niche Management and Transition Management: different but complementary approaches*. Rotterdam: DRIFT, Erasmus University.

Luederitz, C. e. (2017). Learning through evaluation e A tentative evaluative scheme for sustainability transition experiments. *Journal of Cleaner Production, 169*, 61-76.

Lundvall, B. Å. (1992). *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. London: Frances Pinter.

Luukkonen, T., Arnold, E., & Martínez Riera, C. (2016). *Mutual Learing Exercise: Evaluation of Complex PPP Programmes in STI*. Brussels: DG Research and Inovation, European Commission.

Magro, E., & Wilson, J. R. (2013). Complex innovation policy systems: Towards an evaluation. *Research Policy, 42*, 1647-1656.

Malthus, R. (1798). *An Essay on the Principle of Population, as it affects the future improvement of society with remarks on the speculations of Mr. Godwin, M. Condorcet, and other writers*. Anonymous.

Manning, S., & Reinecke, J. (2016). A Modular Governance Architecture In-The-Making: How Transnational Standard-Setters Govern Sustainability Transitions. *Research Policy, 45*(3), 618-633.

Markard, J., & Truffer, B. (2007). Technological innovation systems and the multi-level perspective: towards an integrated framework. *Research Policy, 37*(4), 596-615.

Markard, J., Suter, M., & Ingold, K. (2016). Socio-technical transitions and policy change: advocacy coalitions in Swiss energy policy. *Environmental Innovation and Societal Transitions, 18*, 215–237.

Marks, G., Scharpf, F. W., Schmitter, P. C., & Streek, W. (1996). *Governance in the European Union*. London: Sage.

Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation, 18*(3), 270-280.

Meadowcroft, J. (2009). What about the politics? Sustainable development, transition management, and long term energy transitions. *Policy Science, 42*, 323–240.

Meadows, D., Meadows, D., Randers, J., & Behrens III, W. W. (1972). *The Limits to Growth*. New York: Potomac Associates.

Medical Research Council. (2006). *Developing and evaluating complex interventions: new guidance*. London: Medical Research Council.

Meyer-Krahmer, F., Nauwelaars, C., Galetta, D.-U., Santos, F., & Hunter, A. (2017). *Meta-Evaluation of Article 185 Initiatives*. Brussels: European Commission Directorate-General for Research and Innovation.

Miedzinski, M., Allinson, R., Arnold, E., Griniece, E., Cassingena Harper, J., Doranova, A., . . . Veen, G. v. (2013). *Assessing environmental impacts of Research and Innovation Policy*. Luxembourg: Publications Office of the European Union.

Mowery, D. C., & Rosenberg, N. (1979). The Influence Of Market Demand Upon Innovation: A Critical Review Of Some Recent Empirical Studies. *Research Policy, 8*(2), 102-153.

Nelson, R. R. (1959). The simple economics of basic scientific research. *Journal of Political Economy, 67*, 297-306.

Nelson, R. R. (1993). *National Innovation Systems*. New York: Oxford University Press.

Nelson, R. R., & Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, Mass and London: Belknap Press and Harvard University Press.

Normann, H. E. (2015). The role of politics in sustainable transitions: the rise and decline of offshore wind in Norway. *Environmental Innovation and Societal Transitions, 15*, 180–193.

OECD. (1999). *Managing National Innovation Systems*. Paris: OECD.

OECD. (2015). *System Inovation: Synthesis Report*. Paris: OECD.

OECD. (2016). *Evaluation Systems in Development Co-operation: 2016 Review*. Paris: OECD.

OECD Committee of Science Policy. (1968). *The Promotion and Organisation of Fundamental Research. Issues and Recommendations*. Paris: OECD.

Oliver, K., & Pearce, W. (2017). Three lessons from evidence-based medicine and policy: increase transparency, balance inputs and understand power. *Palgrave Communications, 3*(43), 1-7.

Patton, M. Q. (2006). Evaluation for the way we work. *Nonprofit Quaterly , 13*(1), 28–33.

Patton, M. Q. (2010). *Developmental Evaluation. Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guildford Press.

Patton, M. Q. (2015). *Qualitative research and evaluation methods (4th ed)*. Thousand Oaks CA: Sage.

Patton, M. Q. (2016). The developmental evaluation mindset: Eight guiding principles. In M. Q. Patton, K. McKegg, & N. Wehipeihana, *Developmental evaluation exemplars: Principles in practice* (pp. 289– 312). New York, NY: Guilford.

Pawson, R. (2002). Evidence-based Policy: The Promise of 'Realist Synthesis'. *Evaluation, 8*(3), 340–358.

Pawson, R. (2002). Nothing as Practical as a Good Theory. *Evaluation, 9*(4), 471-490.

Pawson, R. (2013). *The Science of Evaluation: A Realist Manifesto*. London: Sage.

Pawson, R. (2018). The Realist Foundations of Evidence-Based Medicine: A Review Essay. *Evaluation, 24*(1), 42-50.

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. London: Sage.

Pedersen, L. H., & Rieper, O. (2008). Is Realist Evaluation a Realistic Approach for Complex Reforms? *Evaluation, 14*(3), 271–293.

Pettigrew, A. M. (1990). Longitudina field research on change: theory and practice. *Organization Science, 1*(3), 267-292.

Ramos-Mejía, M., Franco-Garcia, M.-L., & Jauregui-Becker, J. M. (2017). Sustainability transitions in the developing world: Challenges of sociotechnical transformations unfolding in contexts of poverty. *Environmental Science and Policy*.

Rauschmayer, F., Bauler, T., & Schäpke, N. (2013). *Towards a governance of sustainability transitions: Giving place to individuals, UFZ Discussion Papers, No. 17/2013*. Leipzig: Helmholtz Zentrum für Umweltforschung – UFZ.

Raven, R., Kern, F., Verhees, B., & Smith, A. (n.d.). Niche construction and empowerment through socio-political work: a meta-analysis of six low-carbon technology cases. *Environtal Innovation and Societal Transitions, 18*, 164–180.

Rietschel, E. T., Arnold, E., Čenys, A., Dearing, A., Feller, I., Joussaume, S., . . . Stame, N. (2009). *Evaluation of the Sixth Framework Programmes for Research and Development 2002-2006*. Brussels: European Commission.

Rip, A., & Kemp, R. (1998). Technological change. In S. Rayner, & E. L. Malone, *Human Choice and Climate Change*. Columbus OH: Battelle.

Rogers, P. J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation, 14*(1), 29-48.

Rothwell, R. (1994). Towards the Fifth-generation Innovation Process. *International Marketing Review, 11*(1), 7-31.

Rothwell, R., Freeman, C., Horsley, A., Jervis, V. T., Robertson, A. B., & Townsend, J. (1974). SAPPHO updated – Project SAPPHO phase II. *3*, 258-291.

Sabatier, P. (1988). An advocacy coalition framework of policy change and the role of policy-oriented learning therein. *Policy Sciences, 21*, 129-168.

Sabel, C. F., & Zeitlin, J. (2008). Learning from Difference: The New Architecture of Experimentalist Governance in the European Union. *European Law Journal, 14*, 271-327.

Sarzynski, A. (2018). Multi-level Governance, Civic Capacity, and Overcoming the Climate Change "Adaptation Deficit" in Baltimore, Maryland. In S. Hughes, E. K. Chu, & S. G. Mason, *Climate Change in Cities: Innovations in Multi-Level Governance* (pp. 97-120). New York: Springer.

Schmookler, J. (1966). *Invention and Economic Growth*. Harvard University Press.

Sendzimir, J. P., Magnuszewski, P., Balogh, P., & Vári, A. (2006). Adaptive management to restore ecological and economic resilience in the Tisza river basin. In J. P. Voß, D. Bauknecht, & R. Kemp. Cheltenham: Edward Elgar.

Shove, E. (2003). Principals, agents and research programmes. *Science and Public Policy, 30*(5), 371-382.

Shove, E., & Walker, J. (2007). CAUTION! Transitions ahead: politics, practice and sustaiable transition management. *Environmental Planning A, 39*(4), 763-770.

Simon, H. A. (1945). *Administrative Behaviour*. New York: Free Press.

Smith, A., Stirling, A., & Berkhout, F. (2005). The governance of sustainable sosio-technical transitions. *Research Policy, 34*(10), 471-476.

Stame, N. (2004). Theory-based Evaluation and Types of Complexity. *Evaluation, 10*(1), 58-76.

Stame, N. (2006). Introduction: Streams of Evaluative Knowledge. In R. C. Rist, & N. Stame, *From Studies to Streams: Managing Evaluative Systems*. New Brunswick, NJ: Transaction Publishers.

Stame, N. (2008). The European Project, Federalism and Evaluation. *Evaluation, 14*(2), 117-140.

Stern, N. (2008). *The Economics of Climate Change*. London: HM Treasury.

Svensson, O., & Nikoleris, A. (2018). Structure reconsidered: Towards new foundations of explanatory transitions theory. *Research Policy, 47*, 462-473.

Tömmel, I., & Verdun, A. (2009). *Innovative Governance in the European Union: The Politics of Multilevel Policymaking*. Boulder CO: Lynne Rienner.

Theewes, J. (2012). *Dare to measure: Evaluation designs for industrial policy in The Netherlands*.

Thijs, G., & Segone, M. (2012). *National Evaluation Capacity Development: Practical tips on how to strengthen National Evaluation Systems*. Paris and Geneva: UNESCO and ILO.

Tõnurist, P., & Kattel, R. (2016). Can Research, Development, and Innovation Policies Cross Borders? The Case of Nordic–Baltic Region. *Science and Pubic Policy*.

Turnheim, B., Berkhout, F., Geels, F., Andries, H., McMeekin, A., Nykvist, B., & van Vuuren, D. P. (2015). Evaluating sustainability transitions pathways: Bridging analytical approaches to address governance challenges. *Global Environmental Change, 35*, 239-253.

Turnheim, B., Kivimaa, P., & Berkhout, F. (2018). *Innovating Climate Governance: Moving Beyond Experiments*. Cambridge: Cambridge University Press.

Ulli-Beer, S. (2013). Conceptual Grounds of Socio-Technical Transitions and Governance. In S. Ulli-Beer, *Dynamic Governance of Energy Technology Change Socio-Technical Transitions Towards Sustainability* (pp. 19-47). Heidelberg: Springer.

van der Meulen, B. (1998). Science Policies as principal-agent games. Institutionalisation and path dependency in the relation between government and science. *Research Policy, 27*, 397-414.

Vedung, E. (2010). Four Waves of Evaluation Diffusion. *Evaluation, 16*(3), 263-277.

Veugelers, R., & al, e. (2009). *Evaluation of the Finnish National Innovation System*. Helsinki: MEE and MEC.

Voß, J.-P., & Bornemann, B. (2011). The Politics of Reflexive Governance: Challenges for Designing Adaptive Management and Transition Management. *Ecology and Society, 16*(2), 9.

Vogel, I. (2012). *Review of the use of 'Theory of Change' in international development*. London: DFID.

von Bertalanffy, L. (1968). *General System Theory Foundations, Development, Applications*. New York: George Brazilier.

von Hippel, E. (1976). The Dominant Role of Users in the Scientific Instrument Innovation Process,. *Research Policy, 5*(3), 212-239.

Walton, M. (2016). Expert views on applying complexity theory in evaluation: Opportunities and barriers. *Evaluation, 22*(4), 410-423.

Weber, K. M., & Rohracher, H. (2012). Legitimizing research, technology and innovation policies for transformative change Combining insights from innovation systems and multi-level perspective in a comprehensive 'failures' framework. *Resarch Policy, 41*, 1037-1047.

Weiss, C. (1987). In D. Palumbo, *Where Politics and Evaluation Research Meet*. Newbury Park, CA: Sage.

Wholey, J. S. (2003). Using Policy Evaluation to Support Decision Making and Improve Program Performance',. *Japanese Journal of Evaluation Studies, 3*(2), 6-21.

Wiener, N. (1948). *Cybernetics: Or Control and Communication in the Animal and the Machine*. Paris and Cambridge, Mass: Herrmann & Cie, MIT Press.

Wittmayer, J. M. (2016). Insights and Lessons for the Governance of Urban Sustainability Transitions. In D. Loorbach, J. M. Wittmayer, H. Shiroyami, & J. Fujino, *Governance of Sustainability Transitions: European and Asian Experiences*. Tokyo: Springer.

Wittmayer, J. M., & Loorbach, D. (2016). Governing Transitions in Cities: Fostering Alternative Ideas, Practices, and Social Relations Through Transition Management. In D. Loorbach, J. M. Wittmayer, H. Shiroyami, & J. Fujino, *Governance of Sustainability Transitions: European and Asian Experiences*. Tokyo: Springer.

Wittmayer, J., Frantzesjaji, N., van Steenbergen, F., Roorda, C., & Henneman, P. (2012). *Introducing a profile for urban change agents: experiences from five European cities. In: Drift Working Paper. Erasmus University Rotterdam, Rotterdam, Creative Commons*. Rotterdam: Erasnmus University.

WK Kellogg Foundation. (2004). *Logic Model Development Guide*. Battle Creek MI: WK Kellogg Foundation.

Appendix A Climate Change Act

The Climate Change Act 2008 is the legislative foundation of the UK's approach to addressing climate change. The Act requires government response and action to reduce greenhouse gas emissions and prepare the UK for the impact of climate change, and provides a framework to deliver on these requirements. The Act sets the high-level target of reducing the CO2 emissions level by at least 80% of the 1990 levels by 2050 (mitigation) and support the UK's adaptation to the effects of climate change (adaptation).

A.1 Evaluation culture in the organisation managing the intervention

Historically, the UK government has a strong track record with evaluation. In line with this, Government's HM Treasury has developed guidance material for commissioning and conducting evaluations. Two documents, The Magenta Book and The Green Book form the backbone of the evaluation processes across all government departments[9].

In 2014, BEIS (then known as the Department for Business, Innovation and Skills) set out its Evaluation Strategy in which they set out their commitment to maintaining and developing a robust evidence base across the department's activities and investments (BIS, 2016). This Strategy is updated by Evaluation Plans published every two years, including updates of their monitoring and evaluation coverage of the department's supported programmes and agencies.

A.2 Policy background and origins of the intervention

Prior to 2008, climate change policy had been steadily gaining momentum both within the UK and internationally. From the early 2000s, there had been discussion around setting definitive targets for the reduction of carbon emissions. Early assessments such as the report *Energy- The Changing Environment* (published 2000 by the Royal Commission on Environmental Pollution) or the 2003 *Energy White Paper* made a strong case for cutting emissions and provided an indication of the cost of such an intervention.

Climate change and environmental policy was on the agenda for both of the main political parties. In 2005, Prime Minister Tony Blair placed it high on the international agenda at the G8 summit in Gleneagles and commissioned the influential report, *The Economics of Climate Change: The Stern Review,* published in 2006. This report set out the economic costs of unmitigated climate change and the need for intervention. The leader of the opposition party, David Cameron, strongly supported the Change Act, as did the other opposition parties.

The political will was bolstered by a range of proactive stakeholders, including a number of instrumental non-governmental organisations who maintained momentum and mustered public support. This included widespread national campaigns such as The Big Ask delivered by Friends of the Earth that encouraged over 130,000 people to ask their Parliamentary representative to support the Bill.

In 2008, the Climate Change Act was passed with a significant majority and at the time was the first in the world to impose legally binding emission reduction targets. Further, it was one of the first in the world to create a framework for a cross-government approach to addressing climate change. While some stakeholders held that the Act was not ambitious enough, it was widely accepted and supported. As of 2017, the Climate Action Network ranked the UK third in the world for action on climate change (Climate Change Network, 2017).

[9] The Magenta Book[9] is the Government's guidance on designing evaluation, with a particular focus on providing detailed guidance on process and impact evaluation methodologies while The Green Book[9] provides guidance on the economic appraisal and evaluation of policies, projects and programmes.

A.3 Strategy and plan for the intervention, intervention logic and its evolution over time

In order to achieve its goals, the Act sets out a framework consisting of two branches of action.

- Emissions reduction, in the form of legally binding, incremental Carbon Budgets that place an economy-wide cap on green-house gas emissions over a five-year period up until 2050
- Adaptation planning, through a five-year cycle beginning with a Climate Change Risk Assessment (CCRA) and a responsive National Adaption Plan (NAP) working to reduce vulnerability to the impact of climate change. The NAP is a register of all actions taken by the government to increase awareness of and resilience to climate change that is renewed every five years

The UK government is legally required to meet these targets and is accountable to the UK Parliament, and is open to judicial review should it fail to comply.

As climate change policy touches many different systems and parts of the economy, the Act requires input from a number of government departments. The main department responsible for collating and coordinating this input with regard to the Carbon Budgets is the Departments for Business, Energy and Industrial Strategy (BEIS). Domestic adaptation policy including the NAP is the responsibility of the Department for Environment and Rural Affairs (Defra). The government also established a liaison group which functions as an official level working group representing the national authorities who may undertake actions under the Climate Change Act. As many aspects relating to climate change are regionally devolved, Wales, Northern Ireland and Scotland each have their own Climate Change Strategy and under the Act are committed to supporting the implementation of UK-wide policies[10].

The Act also stipulates the establishment of the independent Climate Change Committee (CCC) to ensure the targets set are evidence based, and independently assessed. The CCC, functioning as custodian of the of UK climate policy as per the long-term objectives set in the Act provides advice regarding targeting setting and is responsible for producing annual progress reports for Parliament on the progress made towards both mitigation and adaptation.

Prior to the budget setting, the CCC provides a recommended budget. The Government will then accept and legislate this budget, or not. This review process is complemented by BEIS's own internal analysis of what target the Carbon Budgets should be set at prior to accepting the CCC's recommendations. These studies scope and generate full economic assessments of different policy options (DECC, 2016). Similarly, BEIS, and other relevant Departments will commission a wide range of research to assess the emission reduction impacts of sector specific policies e.g. on each policy aiming to reduce energy use in homes, regulation on cars, funding for renewables and innovation in green tech etc.

As it stands, the five Carbon Budgets currently legislated are in line with the CCC's recommended budgets and currently run up until 2032. The budgets are set 12 years in advance to allow policy-makes, business and individuals enough time to prepare. Once a budget has been set, the onus is placed upon the Government to ensure the appropriate policies are implemented to meet the targets set. In 2011, the Government published The Carbon Plan, which set out the scenarios and long-term plan for the government to meet the first four budgets up until 2027. More recently, the Clean Growth Strategy published in October 2017 contains the Government's plan to meet the legislated fourth and fifth carbon budgets, covering UK emissions in the periods 2023-2027 and 2028-2032.

The adaptation aspect of the Act is covered by the development of The National Adaption Programme (NAP), first published in 2013. The first NAP was developed in response to the Climate Change Risk Assessment 2012 (CCRA), a catalogue of the risks potential risks to the UK as well as a progress of wider stakeholder consultation and research. This Programme is the Government's strategy for addressing the main risks and opportunities for preparing for the impact of climate change. While the Risk Assessment

[10] Scotland passed the Climate Change (Scotland) Act in 2009, Wales passed the Environment (Wales) Act in 2016 and though Northern Ireland have not yet introduced specific climate change legislation, Northern Ireland Executive have included emissions targets in its Programme for Government (2011-2015).

conducted is UK wide, the NAP applies only to England. Devolved administrations are able to request adaptation assessments of their national strategies from the ASC.

A.4 Governance of the intervention and the evaluation

The Climate Change Act is widely viewed as an innovative piece of climate legislation, effective in creating an 'overarching' or whole-economy governance structure to guide progressive action on climate change. The governance of the actions are pursued across government departments in parallel, as demonstrated by the concurrent Clean Growth Strategy (BEIS, 2017) and Defra's 25-Environment Plan (Defra, 2018). This two-pronged approach reflects the distinction that has been made between two parts of the Act, emissions reduction and adaptation, though it is not yet possible to judge whether a more thoroughly integrated approach would have been more effective (Fankhauser et al, 2018).

As the implementation is managed by government departments, the governance of the climate change policies is subject to political fluctuation and transition. With changes in government, so too come new Secretary of States and departmental ministers, and in turn the policies and modes of implementation. While the Climate Change Act is set in place to attempt to maintain a degree of consistency, more research is necessary to understand how the Act has shaped the day-to-day policy making for these government departments (Fankhauser et al. 2018).

According to the Act, Parliament is responsible for holding the Government accountable to its obligations. This includes the Government's responsibility to produce a carbon budget, but also its responsibility to respond to the regular reports produced by the CCC. To date, this process has for the most part been upheld with only minor delays. As the carbon budgets legislated by government have always been in line with the recommended budget set by the CCC. As such, it is not evident what the process would be in this situation. While the CCC would undoubtable provide comment or feedback on the situation, once again the onus would be placed upon Parliament to formally respond.

A.5 The evaluation

The Committee on Climate Change (CCC) functions as an independent advisory body to ensure the targets set are evidenced based and independently assessed. Intrinsic to the act is the mandatory progress monitoring, with regular reports to Parliament and the public to ensure transparency and the continued presence of the Act at the forefront of social consciousness. The Adaptation Sub-Committee (ASC) sits within the CCC and is responsible for providing advice and reporting on the progress with regards to the National Adaptation Programme.

Given the different structure and implementation process of each pathway, the evaluation and progress reporting of the carbon budget and the adaption plan are delivered separately, each with their own evaluation framework and monitoring tools.

In addition to the activities of the CCC, BEIS, and other relevant Departments will commission a wide range of research to assess the emission reduction impacts of sector specific policies e.g. on each policy aiming to reduce energy use in homes, regulation on cars, funding for renewables and innovation in green tech etc. As such, while the actions of the CCC are mostly closely related to providing evaluative feedback on the implementation of the Act, it is important to acknowledge that the assessment provided is not the only source of evidence utilised by the government.

Notably, these reports do not seem to offer feedback or reflection upon the Act as a whole, but instead currently offer monitoring and tracking of existing progress and outcomes.

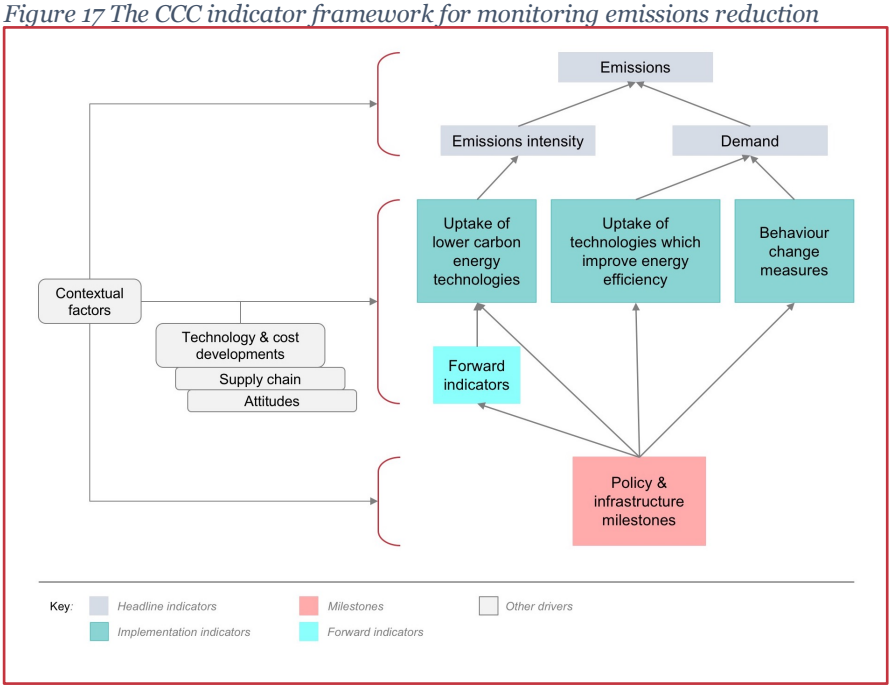
A.5.1 Monitoring and evaluation strategy, plan and framework and associated indicators

Emission Reduction

The monitoring and reporting of progress of emissions reduction is conducted by the CCC. The framework for the monitoring was defined in the first progress report to Parliament (CCC, 2008).

While one approach to measuring emissions reduction would be to measure the actual emissions levels and deem it successful should it fall within budget, this approach fails to capture the progress on the implementation of measures required to meet long-term targets and the long-lead times required for some actions to have impact. Monitoring mitigation covers both the emissions and the progression towards implementation itself. The framework developed (see Figure 15) includes emissions, drivers of emissions, forward indicators for these drivers, policy milestones and contextual factors.

Figure 17 The CCC indicator framework for monitoring emissions reduction



Source: Committee on Climate Change (2008)

The indicators utilised to measure the progression towards the Carbon Budgets in 2017 are divided into broad themes relative to their sector: power; buildings; industry; transport emissions; agriculture, land-use and forestry; and water; and F-gases. Indicators are also measured and analysed according to the economy wide progress and that of the devolved administrations.

These trajectories are consisting of a subset of pathways that are then sub-divided into specific avenues of policy measures, technical adjustments or social changes. In turn, each avenue has a handful of indicators used to measure the progression towards the goal across the different levels of the indicator framework.

The headline indicator 'Emissions' includes a sectoral breakdown of economy wide emissions, and is fed by indicators measuring emissions intensity and demand, reflecting the supply and demand side factors that dive emissions. These headline indicators are complemented by implementation indicators that track the progress towards implementing the measures required for sustainable emissions reduction. These indicators reflect the different pathways across the eight different focus areas.

Forward indicators are used where appropriate and provide a measure of whether the trajectories are on track to deliver future measures. This is pertinent where long-term planning and processes are required to form the foundation of implementation and take place long before the measure will result in emissions reductions. All of these indicators are preceded by policy and infrastructure milestones, that provide the planned measures with the appropriate enabling framework.

Adaptation

The conceptual framework for the monitoring for the National Adaptation Plan was initially proposed by the ASC in 2010 and it has continued to be developed and adjusted (ASC, 2010; ASC, 2015). The current framework seeks to assess whether the UK is becoming more or less vulnerable to climate change risks, the uptake of adaption actions and whether climate change risks are being taken into account within the decision-making process. This framework seeks to measure the six different themes presented within the NAP and each theme is divided into specific, measurable factors that are deemed to be the most important (ASC, 2015)[11]. These factors are called 'adaptation priorities' and are evaluated with consideration made to

- The presence of a plan: assessment of the existence of explicit policies or plans addressing the priority area by encouraging adaptation or removing barriers
- Actions taking place: assessment of whether the activities laid out in the NAP are taking place or are on track for future delivery. This also includes those taking place outside of the NAP
- Progress made: assessment of the overall progress towards adaptation, the impact of any existing policies or actions. This also takes into account the current and future risks, and the time taken to deliver measurable changes in vulnerability

Based on this framework, ASC developed an assessment toolkit which serves as the foundation for the evaluations published by the ASC every two years. This assessment toolkit comprises two components, each with its own constituent indicators.

1. Monitoring changes in climate risks using indicators: quantitative data collection, requiring robust, long term and spatially distributed datasets of indicators of risk, adaption actions and climate impact.

2. Evaluating preparedness for future climate: Analysing the decision-making processes to determine if the action is sufficient and correctly targeted to determine if adaption is sufficient to address the desired outcomes. This should focus on the levels of up-take of policies relevant for the current situation and the uptake of and reflection upon the policy measures with long-term or systemic consequences which may depend upon future climate.

Given the complexity of monitoring adaptation, the process of indicator development utilised a systems approach. Seeking to understand the range of impacts, drivers and actions in parallel and grasp how they were connected, where actions would be necessary and the best approaches (Miller et al., 2011). The process stepped away from a linear view of attributing indicators to particular changes in climate change vulnerability or specific adaptation actions or policies.

This process of indicator identification and development was initially outsourced to Ricardo AEA (later known as AEA Technology). The process utilised systems maps to identify potential impacts, drivers and actions in five particular sectors, focusing on a discrete set of impacts for further indicator development (AEA Technology, 2011). This focussed list of impacts were identified through literature review, workshops and with the members of ASC, and expert consultation on sectoral and methodological aspects. This focussed list was then further narrowed down to a set of indicators, based on an evidence-based assessment, and reflected the indicators that were both possible and desirable to measure. The assessment was based on three criteria: the significance of the impact in the current situation, the sensitivity of the impact to climate change, and the expected changes in impact.

[11] Given the complex nature of climate change, this framework is applied to a subset of priority areas: land use planning; national infrastructure; building design; natural resources, and emergency planning.

In parallel to the systems mapping, a review of availability, relevance and quality of datasets was undertaken. The availability of datasets was a significant limitation upon the indicator selection process, and data were collected for only 40% of the indicators proposed by AEA (AEA Technology, 2011).

The ASC has continued to develop, update and expand the list of indicators utilised within the assessment process, with a refresh every two years. Reviewing is necessary as the availability and relevance of the data available changes over time. The current list of indicators utilised by the ASC includes around 200 different indicators.

A.5.2 Evaluation tools and methods

The indicators underpinning the progress reports produced by both the ASC and the CCC are largely quantitative and are derived from large datasets collected by a range of other stakeholders, such as government agencies and departments, charities, and research and consultancy firms. As such, the development and analysis of these indicators is dependent upon strong coordination activities between the relevant departments and agencies and the CCC, coordinated by the Government's liaison team. Both committees have provided detailed breakdowns of the processing and presentation requirements for these data.

Building upon these datasets, the CCC's committees and secretariat undertake a range of analytical activities. Further data sources include policy and legislative documents that serve to populate the indicators in each sector regarding policy measures implemented. This in turn contributes to the analysis of the cost-effectiveness of the implementation at that point in time, the degree to which climate change considerations are integrated into the decision-making processes and the degree to which sufficient adaption activities are currently underway.

Complementary to this analysis is the constant and intensive dialogue with stakeholders and government departments (e.g. Department for Transport), providing feedback and comments on the analysis and guiding the analytical process itself. As such, this process provides an element of qualitative data and opportunity for sense checking the results (Fankhauser et al, 2018).

A.5.3 Evaluation actors, management and contracting

The CCC is a Government non-departmental public body and is sponsored and funded by the UK Government, the Scottish Government, the Welsh Assembly Government and the Northern Ireland Executive. As such, it is responsible to the UK Parliament, the Secretary of State for Business, Energy and Industrial Strategy, the Secretary of State for the Environment, Food and Rural Affairs and the Environment and Climate Change Ministers from the devolved administrations of Northern Ireland, Scotland and Wales. The CCC operates at arm's length from its associated government departments and is not equipped with formal decision-making power.

The CCC's central committee comprises an independent Chairman and seven independent members. The ASC comprises an independent Chairman (who also sits on the central committee) and five independent members. The central committee itself meets once a month, and is responsible for determining the work programme, providing advice and recommendations. The CCC as an entity employs a Chief Executive responsible for the management of the CCC and ASC and a secretariat of around 30 members of staff to provide analytical and corporate support. This secretariat provides the CCC with strong in-house capacity with expertise across a range of the technical aspects of climate change such as transport, energy, economics and climate science. The Adaptation Sub-Committee also utilises six of the members of secretariat staff. The Secretariat is responsible for taking forward the decisions of the committee and undertaking a large proportion of the work, including gathering and analysing the data and evidence that feed into the reports produced by the CCC.

The CCC and the ASC have commissioned a number of external research consultations to inform the advice it provided. These consultations both complement and feed into the progress reports provided to parliament and regard specific technology developments, climate impacts, indicator development, or sector specific analysis.

The CCC is subject to a Cabinet Office mandated review every three years assess its function, delivery mechanisms and governance arrangements. The latest of these Triennial Reviews[12] conducted in 2013 found that the CCC provided necessary, impartial expertise, fulfilled an important function and that there remained the need for a high level of independence (DECC, 2013). Notably however, as the CCC is still financially dependent upon the government, an independent and consistent source of funding would better align with international best practice (Fankhauser et al, 2018).

A.6 Assessment of the success of the intervention

A.6.1 Evaluation and monitoring outputs

The most recent report for Parliament generated by the ASC provided evidence that the risks posed by climate change have continued to increase across a number of priority areas since 2015 and that the policies and actions put forward by the National Adaptation Programme would be insufficient to counter them (ASC, 2017). While progress was made against some of the objectives and actions within the programme were largely being delivered, progress against other objectives has been hindered, increasing vulnerability.

While the first three carbon budgets were met, it has been noted by the CCC that this achievement may be partially attributable to the widespread reduction in the use of coal for power generation (CCC, 2017). By contrast, emissions from transport were higher in 2016 than any other year since 2009 (CCC, 2017). The most recent progress report delivered to parliament regarding the carbon budgets urgently called for new policies to reach its targets, drawing attention to the governments inaction to meet their legal requirements. While it seems this call was addressed in the Government's most recent, relevant strategy, the Green Growth Strategy, the CCC's independent assessment found that while strong commitment is present, the proposed policies and actions require greater support to avoid the risk of under-delivery (CCC, 2018). Further, this assessment noted that the Government is not on track to meet the targets of the fourth and fifth carbon budgets. Overall, the perception of the intervention from the evaluation perspective is that while the most recent strategy proposes some positive policy directions, there is not clear link between the policies proposed and the objectives set within the carbon budgets (CCC, 2018).

The Triennial Reviews conducted by the CCC's Review Team have not found any major shortcomings in the structure or function of the CCC or the ASC (DECC, 2013). Further, while the power of the CCC is limited, the advice they provide is directly reflected in legislation and policies implemented across the UK. As such, the impact of entity is not insignificant.

The Government has not yet commissioned an evaluation of the Act in its entirety. However, independent assessment of the Act provides some indication of the evolution and impact since its fruition (Fankhauser et al. 2018). While the implementation of the Climate Change Act has been susceptible to periods of political upheaval, particularly in the aftermath of the UK's decision to leave the European Union, the Act has been successful in implementing some wider societal changes. First and foremost, the Climate Change Act has facilitated a fundamental shift in the ways in which political debate on climate change is conducted (Fankhauser et al. 2018). Through the creation of routine target setting, monitoring, reporting and debate, the Act has fostered a standardised process for addressing a large societal challenge. Further, the Monitoring and Evaluation framework have established an agreed framework and empirical evidence base, against which future policy instruments may be measured.

A.6.2 Views of the intervention and evaluation managers

The over-arching perception of the Act was positive, viewed as a progressive and powerful tool for supporting continued action towards climate chance mitigation and adaptation. Concerns about the governance and implementation of the intervention largely centred around mechanisms for ensuring accountability.

[12] As of 2015, the Government renamed these reviews the Tailored Reviews and introduced a new set of guidelines for such reviews. For more information on this review process, see https://www.gov.uk/government/publications/tailored-reviews-of-public-bodies-guidance

The recent reports produced by the CCC highlight the government's risk of under-delivering, concerns lingered around the processes for ensuring adherence. While Parliament has accepted these responses, there is hope that the Parliament would be more active in holding the government accountable should the policies prove insufficient to meet the targets. It was also acknowledged that the degree to which Parliament will address Government failure to act upon the recommendations made by the CCC will be tested in the coming years.

While the CCC is a non-elected public body and is not equipped with the power to hold the government accountable to its obligations, it was also thought that the Committee advisory function was the most appropriate role for it to play. Questions surrounding the capacity for such a body to be able to enforce accountability also raised the issue of legitimacy to hold such power.

A.7 Lessons learnt

A.7.1 For the intervention

The Act has both benefited from and fostered political consensus, in turn significantly strengthening the longevity and importance of the issue of climate change. This consensus is partially attributed to the high levels of investment from all political parties at its establishment, and partially attributed to the strength of the legislative architecture it created. As such, one of the lessons to take the invention forward is maintaining the same degree of consensus, potentially through greater buy in from all political parties (Fankhauser et al. 2018).

Despite generally high-levels of adherence to the legislative requirements, the Act would benefit from stronger safeguards to hold the government accountable. For example, it seems there is a need to implement a statutory response time in which the government is required to publish the Carbon Plans so as to prevent undue delays. Questions have also been raised regarding the degree to which dividing the Act into two parallel pathways of adaptation and mitigation has negatively impacted the coordination between the two complementary activities. Where each is governed by different departments and strategies, sectors or policy areas where the line between them is blurred may in turn suffer from confusion, conflicting policies or simply neglect (Fankhauser et al. 2018).

There are various developments in the wider policy landscape that will require the adaptation of the Climate Change Act and targeted adjustments to the legal framework. This includes ensuring the governance structures in place are compatible with wider contextual factors such as the Paris Agreement and policies associated with Brexit and prevent political backsliding.

One of the potential changes that could made to ensure greater independence of the CCC would be to switch the funding source to an appropriate non-government entity such as the National Audit Office[13].

A.7.2 For future intervention design

A comprehensive legal framework is an essential tool for coordinating actions that require such a broad spectrum of policy interventions across a wide range of sectors. The comprehensive nature of the framework should pertain not only to the range of long- and short-term targets, but also the distribution of responsibilities across the policy landscape. This is also complemented by the clarity of the of the direction of travel, bolstered by strong levels of consensus over the value of this policy direction. In order to ensure the government's response time is reasonable, a time limitation could also be placed.

The strong guardrails for policy imposed by the Climate Change Act have been successful in maintaining the UK's commitments to climate action despite political upheaval (Fankhauser et al. 2018). The strength of the legislative commitment and maintaining the high levels of government accountability have ensured that the Act is able to consistently place pressure for change and encourage action.

[13] The National Audit Office is responsible for scrutinising Parliamentary public spending. This public audit perspective helps Parliament hold Government accountable and improve public services. For more information, see https://www.nao.org.uk/about-us/

Whilst partially attributed to the legal framework, the presence and activity of the CCC has been important. In offering advice, monitoring and recommendations, the CCC serves to maintain clarity of purpose and offer strategic guidance for policies. An independent body such as this has served to ensure consistent policy delivery and evidence-based decision making. Further considerations should be to ensure the independence of the CCC, particularly with regards the funding supporting the organisation.

A.7.3 For future evaluations

Integral to both future interventions and their evaluation is the approach taken to their establishment. An important consideration is the degree to which policies are established in line with an outcomes-based approach that in turn allows them to be evaluated effectively. This process thus requires policy makers and evaluators to work in collaboration to create an integrated approach. In line with this, is the emerging need for the CCC increasingly to evaluate not only mitigation focussed indicators but the policies themselves. In doing so, evaluation process will be better equipped to offer feedback on the policies themselves. Further, the requirement for forward looking indicators is becoming increasingly apparent. Such indicators allow both evaluators and policy makers to be 'one step ahead' and in turn adjust their policies accordingly. Such forward looking indicators however are necessarily dependent upon awareness of the intended outcomes.

One of the key aspects of the evaluation is the strength of the technical and human capacity to understand, use and manipulate the data available. As the indicators selected for the measurement of both the adaptation and the emission reduction policies are often based on more than one dataset, as such, high level of competence and understanding of the data to assess and combine dataset with difference parameters is required. Future evaluations of this kind will be required to mobilise similarly high levels of capacity.

A.8 References

A.8.1 Literature

Adaptation Sub-Committee (2010) How well prepared is the UK for climate change? Available at: http://www.theccc.org.uk/reports/adaptation/1st-progress-report-2010

Adaptation Sub-Committee (2011a) Adapting to climate change in the UK: measuring progress. Available at: http://www.theccc.org.uk/reports/adaptation/2nd-progressreport-2011

Adaptation Sub-Committee (2012) Climate change – is the UK preparing for flooding and water scarcity? Available at: https://www.theccc.org.uk/wp-content/uploads/2012/07/CCC_ASC_2012_interactive_2.pdf

Adaptation Sub-Committee (2015) Progress in preparing for Climate Change: 2015 Report to Parliament. Available at: https://www.theccc.org.uk/wp-content/uploads/2015/06/6.736_CCC_ASC_Adaptation-Progress-Report_2015_FINAL_WEB_070715_RFS.pdf

Adaptation Sub-Committee (2017) Progress in preparing for climate change – 2017 Report to Parliament. Available at: https://www.theccc.org.uk/wp-content/uploads/2017/06/2017-Report-to-Parliament-Progress-in-preparing-for-climate-change.pdf

AEA Technology (2011) Provision of research to identify indicators for the Adaptation Sub Committee. Commissioned by the Adaptation Sub-Committee. Available at http://www.theccc.org.uk/adaptation

Climate Action Network (2017) Climate Change Performance Index Results 2017 http://www.caneurope.org/docman/climate-energy-targets/3015-climate-change-performance-index-2017/file

Committee on Climate Change (2009) Meeting Carbon Budgets – the need for a step change. Available at: https://www.theccc.org.uk/archive/aws2/21667%20CCC%20Report%20AW%20WEB.pdf

Committee on Climate Change (2017) Meeting Carbon Budgets – Closing the policy gap. Available at: https://www.theccc.org.uk/wp-content/uploads/2017/06/2017-Report-to-Parliament-Meeting-Carbon-Budgets-Closing-the-policy-gap.pdf

Committee on Climate Change (2018) An independent assessment of the UK's Clean Growth Strategy: From ambition to action. Available at: https://www.theccc.org.uk/wp-content/uploads/2018/01/CCC-Independent-Assessment-of-UKs-Clean-Growth-Strategy-2018.pdf

Department of Energy & Climate Change (DECC) (2013) Triennial Review of the Committee on Climate Change Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/270886/committee_climate_change_triennial_review_2013.pdf

Department for Business, Innovation & Skills (2016) Evaluation Plan 2016. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/495162/BIS-16-98-evaluation-plan-2016.pdf

Fankhauser, S., Averchenkove, A. and Finnegan, J. (2018) 10 years of the UK Climate Change Act. Available at: http://www.lse.ac.uk/GranthamInstitute/wp-content/uploads/2018/03/10-Years-of-the-UK-Climate-Change-Act_Fankhauser-et-al.pdf

Miller, K., M. Harley and N. Kent (2012) Climate Change Adaptation-Related Indicators, Sniffer, Edinburgh.

OECD (2013) National Level Monitoring and Evaluation of Climate Change Adaptation in the United Kingdom. Available at: http://www.oecd.org/env/cc/National%20level%20ME%20in%20the%20United%20Kingdom.pdf

A.8.2 Interviews

Name	Role	Organisation
Kathryn Brown	Head of Adaptation	Climate Change Committee Secretariat
Adrian Gault	Chief Economist	Climate Change Committee Secretariat

Appendix B Dutch Top Sectors

In 2011, the Dutch Ministry of Economic Affairs (EZ) launched its Enterprise Policy (Ministry of Economic Affairs, 2011). In addition to some traditional financial instruments (such as subsidies and grant funding), the Policy also included their new approach to supporting innovation systems, the Top Sectors approach. The primary objective of the novel approach is to strengthen innovation systems in the country, focusing on nine sectors of investment, and foster greater alignment of government, academia and industrial institutions. This is primarily achieved through Public-Private Partnerships (PPPs). These nine sectors[14], the Top Sectors, are the areas of investment which both the business and research communities in the Netherlands excel. These are characterised by high labour productivity, exports and R&D investments. The Top Sectors approach constitutes only a small part of the overall enterprise policy supported by EZ and receives a relatively limited portion of the overall budget. As such, it is important to consider the Top Sector approach as a complementary part of the wider landscape of innovation support.

The new enterprise policy marked a shift away from direct subsidies and towards other instruments such as subsidies (delivered through the TKIs) and support for PPPs, with the intention of fostering more proactive engagement with innovative activities and placing the incentive with the entrepreneur. The Top Sectors are complemented by wider generic deregulation and the abolition of €500 million worth of business subsidies to provide funds for these other support mechanisms.

B.1 Evaluation culture in the organisation managing the intervention

The Netherlands has a long history of evaluation in the sphere of public research, now a widely accepted and constructive management tool (OECD, 2014). In line with this, the EZ has an open approach to the evaluation of its policies and follows the government level approach as stipulated in the report *Dare to measure: Evaluation designs for industrial policy in The Netherlands* (Theeuwes Committee, 2012). This report outlines systematically developed evaluation design protocols, specifically for measuring economic impact, and the methods required for implementation. This approach is applied to the evaluation of many of the financial instruments applied by the Ministry. However, this approach is best suited for simple instruments and cases where the evaluation does not focus on causation mechanisms. The Top Sector approach is not a typical policy instrument but instead includes a number of actors, instruments and goals working in tandem. Hence a new evaluation methodology was developed by the consultancy firm Dialogic for the evaluation of the Top Sectors approach conducted in 2017.

B.2 Policy background and origins of the intervention

The Top Sectors approach succeeded the Innovation Programme approach, which ran from 2006 to 2010. That was a predominantly top-down, industry-specific strategy. The Top Sectors took a bottom up approach and began with an open call for consortia. During 2011, researcher centres and businesses self-organised into 'Top Teams' which later managed the Top Sectors themselves.

The top sector approach was introduced

- To replace subsidies and grants with fiscal policy
- To encourage PPPs through regular financing
- To reduce the fragmentation of innovation policy
- To increase government involvement across a number of different ministries

The introduction of the approach marked a shift in the overall innovation support system in two ways. First, the new focus on both a sector-based and integrated approach, which is culturally grounded in the

[14] The nine top sectors are: Horticulture and propagation materials, Agri-food, Water, Life sciences and health, Chemicals, High tech, Energy, Logistics, Creative industries.

Dutch polder model, a consensus-based economic and social policy approach widely implement in Dutch policy making during the 1980s and 90s (OECD, 2014). Second, the approach involved a new framework for funding and organising PPP. In particular, the emphasis of these PPP also shifted towards one more focused on the demand-driven activities and focus. Hence the government's role in the system has also shifted, away from its role of subsidies based management towards network management (AWTI, 2015).

B.3 Strategy and plan for the intervention, intervention logic and its evolution over time

The strategy for the Top sectors was developed in close consultation with the private sectors and academic institutions (Ministry of Economic Affairs, 2011). The aim was to produce a cohesive policy approach that would serve a range of sectors. By engaging the private sector, the intention was to orientate the structure of the Top Sectors towards demand-driven research and innovation, in turn bolstering productivity in the Netherlands. The approach comprises two programmes: support for the Top Sectors and the MKB Innovatiestimuleringsregeling Topsectoren (MIT) programme, consisting of support instruments to encourage SME affiliation with the Top Sectors.

The Top Sectors provide an approach for coordination and cooperation between the business community, universities, research centres and the government, allowing them to work together on both their own sectors and coordinate with others. Each Top Sector is a self-organising entity, coordinating and organising a range of different activities within their specific sectors, coordinated by a Top Team consisting of representatives from each part of the quadruple helix. The core activities of the Top Sector are programming, building Public-Private Partnerships (PPP) and dissemination, delivered through consortia known as the Top Consortia for Knowledge and Innovation (TKIs). Each Top Sector, may host one or more smaller TKI consortia. For example, the Energy Top Sectors hosts TKIs for Urban Energy, Wind at sea and Bio-based Economy, each working to address the needs of the specific technical areas.

The Top Sectors are connected by three common focus areas – knowledge and innovation, internationalisation, and human capital – while also fostering a supportive regulatory environment. The Top Sector should enable cross-fertilization and collaboration between different Top Sectors ns, including the Knowledge and Innovation Agenda (KIA), the International export and knowledge agenda and the Human Capital Agenda (HCA). These agendas lay out the out the strategic direction for the top sector and the consortia within and potential actions or specific focus within these directions. The approach also offers tailor made solutions to eliminate bottle-necks impeding the collective innovation plans[15].

The Top Team is responsible for deciding the research subjects and allocating funding to the TKIs. This funding is then distributed to support the consortia's activities. Participating research institutions contribute in kind to the PPP while industry partners are required to pay in cash for the participation. EZ contributes 30 cents per € invested by businesses.

The activities of the TKI within the Top Sector includes cooperative, basic/pre-competitive research projects and a range of innovation activities. To bring innovative SMEs more readily into top sectors, the EZ also provides funds for their participation in the Top Sectors in the MIT scheme. This scheme involves a number of projects, including advisory services, feasibility projects, R&D cooperation, knowledge vouchers and innovation network activities.

Though largely true to the original structure, the Top Sector approach has evolved. Between 2010 and 2016, the policies were aimed at supporting and developing the sectors, strengthening their internal connections. Since 2016, this focus has shifted to include a more mission orientated focus. The latest reports produced by the Government reflecting on the Top Sector approach and the outcomes of the most recent evaluation, and call for a requirement of the Top Sectors to also seek to address these

[15] Notably there are also three cross-over knowledge domains (IT, Nano-technology and Biotechnology) that were added into the approach later. While these cross sectoral domains are supposedly to function as Top Sectors, they have not yet been developed to reach their potential partly due to the absence of a clear mandate and governance structure. There is however growing interest to be able to mobilise these domains more effectively in the future.

challenges and key technologies. The Top Teams and knowledge institutions were asked to formulate a new Knowledge and Innovation Agenda (KIA) in 2017, which are linked up with the National Science Agenda, priority departmental agendas and Horizon 2020 (Ministry of Economic Affairs, 2017).

B.4 Governance of the intervention and the evaluation

At the highest level of research and innovation policy, the Government publishes national Knowledge and Innovation Agendas that address the societal challenges and key enabling technologies. These are based on extensive consultation and are used in the Top Sector policy to bring together the sectoral agendas with government department agendas. The Knowledge and Innovation contracts form the financial basis for the Knowledge and Innovation Agendas, and reserves funds for the cross-sectoral themes.

The governance and ownership of the policy instruments that feed into the top sectors sometimes, though do not always, sit within EZ[16]. Perhaps more characteristically though, the management of instruments is coordinated and managed by the participating academic institutions or companies involved. Indeed, the role the government have assigned themselves for delivering the approach is not clear: it has shifted away from its role as manager and instead placed the responsibility on academic institutions and private companies[17]. For example, the direction and project funding decisions are made entirely within the PPP with no input from the Government.

The Top Sectors themselves are led by the Top Teams, consisting of a range of stakeholders. The Top Team is responsible for the management while a dedicated executive office is responsible for the day-to-day activities, provides support to the Top Sectors and provides a central point of contact. The Top Team is also supported by an Advisory board (or direction-team), consisting of prominent stakeholders from private, public and non-government organisations. The Top Teams also have science and industry officers who meet regularly to discuss progress and initiatives. They coordinate at the project level as well as at the institutional level, so the coordination is rather detailed.

Each of the TKI consortia within the Top Sectors consists of business and academic institutions, coordinated by a programme office, a supervisory board and an executive committee. The programme office coordinates the programme of projects delivered by the consortia, including the design, setting up and implementation. The programme office is also responsible for knowledge sharing and dissemination on behalf of the consortia. The executive committee and supervisory boards are respectively responsible for the strategic leadership of the consortia and ensuring compliance to auditing and regulatory processes. While the government may serve to facilitate coordination of the network, the top sector's own activities also serve as a level of grassroots governance.

The Advisory and Evaluations Teams are associated within TKIs and provide advice to the Top Team on progress, budgeting, and provide feedback to the TKI. The recommendations provided by this team also feed into the decisions taken by the Top Team regarding to project portfolio of the Top Sectors.

B.5 The evaluation

The monitoring of the Top Sector approach is incorporated into the Progress Reports produced by the EZ for Parliament. The Progress Reports on Enterprise Policy provide detailed overview of the current progress of the enterprise policy, and include the Enterprise Policy Monitor as an annex. The Monitor is updated annually and is largely held by the Central Bureau for Statistics (CBS).These serve to provide reflection on the progress made, and support a form of accountability for action. In addition to these reports, the Ministry is required to evaluate all of the instruments it supports, including those under the banner of the Top Sector approach. As such, the EZ have contracted individual evaluations of both

[16] For example, the Department for Health is able to implement instruments, largely related to the activities and needs of the Health Top Sector.

[17] Notably, the degree to which government departments are involved in the governance of the Top Sector varies, with some playing a stronger role in the strategic direction of the Top Sector while others provide significantly less commitment.

instruments, the PPP payment scheme for the TKI and the MIT (Dialogic, 2016; Technopolis Group, 2017).

Some of the Top Sectors have been subject to their own, individual evaluations such as the of the Top Sector Energy evaluation conducted by CE Delft and Trinomics in 2014 (CE Delft & Trinomics, 2014). This example evaluation aimed to measure and understand to degree to which the Top Sector as a whole is working to meet its own strategic targets, and in some places to define a baseline against which future reviews may be measured. This includes reviewing the programme of activities undertaken by the Top Sector, and the means through which they are being delivered. These reviews are sector specific, with monitoring frameworks closely tied to the specific objectives of the Top Sector.

In 2014, the Dutch Advisory Council for Science, Technology and Innovation (AWTI) undertook a review of the approach,. It was acknowledged that it was still too early to attempt to measure any tangible outcomes, so the review largely concerned the delivery model and a review of the policies at a meta-level (AWTI, 2014).

The full evaluation of the Top Sectors approach was conducted by Dialogic, an independent research and consultancy firm in 2017 (Dialogic, 2017). This evaluation primarily sought to gain insight into the bottlenecks, the suitability and the cost-benefit of the whole approach, the process and impact pathways to impact, and its effectiveness and efficiency. Given the complexity of the approach and in the absence of a suitable counter-factual, the evaluation required a novel approach. In collaboration with Harvard Kennedy School for Public Policy, Dialogic developed a new evaluation methodology for the EZ (Janssen, 2016). The framework developed was based on modern transformative industrial policy that integrated existing frameworks of evaluating policy impact and policy design (Janssen, 2016). Following a pilot of the approach, the full evaluation was conducted by Dialogic (Dialogic, 2017).

B.5.1 Monitoring and evaluation strategy, plan and framework

The Enterprise Policy Monitor provides a standardised monitoring process. Coordinated by CBS for the Government, this largely involves extensive monitoring of economic development. However, these indicators were not found to be indicative of the impact of the Top Sectors and as such only offer a small part of the data required for a full evaluation (Dialogic, 2017). In particular, they offered no indication of the added value of the combination of policies employed in the Top Sector approach meaningfully contributed to the achievement of the objectives.

The evaluation framework developed by Dialogic comprises two lower-level frameworks: one for studying policy impact and one for studying policy design (Janssen, 2016; Dialogic, 2017). These two frameworks provide a framework of qualitative indictors and questions to explore in a standardised fashion. This complex model provides a number of functions with which to apply the set of indicators.

To measure the impact of the policy, the first step is to identify the bottlenecks in the system or what problem the policy sought to address. From this the impact achieved is then subdivided into the direct results of the incentives and additionality (i.e. does it really make a difference?). The novelty of the instrument is also questioned, seeking to delineate whether the instrument is new or simply incorporated into the new strategy. The source of the policy is an important aspect to help understand whether the policy was generated through a top-down or bottom-up approach. When analysed together, these questions serve to give an impression of the effectiveness of the approach in strengthening the innovation system.

B.5.2 Evaluation tools and methods

The evaluation conducted by Dialogic (2017) was predominantly based on qualitative data, due in part to the limited availability of data regarding the implementation and outcomes of specific instruments. For the analysis of the approach within the nine sectors, across the three common domains, as well as the overarching analysis, the evaluators made use of the existing documentation which included partial evaluations, annual reports, mission statements, implementation agendas, policy notes, progress reports and advisory notes.

The extensive document review was complemented by just under 100 semi-structured interviews with both internal and external stakeholders. The interviewees included not only representatives from the Top Teams of the Top Sectors, but also participating companies, academic institutions and government representatives.

Notably, the evaluators also conducted a quantitative analysis following the publication of the formal evaluation (Janssen, 2018). This study tested twelve of the TKI for their capacity to facilitate collaborative R&D projects in the Netherlands, utilising data regarding organisations that financially participated in collaborative projects. As this was not included within the formal evaluation process, the degree to which the findings are incorporated into the policy development process is less clear, however the study does demonstrate a method for measuring one of the specific objectives of the Top Sectors and in doing so may be utilised in future evaluations. This study also provides the Top Sector approach with a quantitative benchmark against which future progress could be measured.

The approach taken regarding the evaluation of the Top Sector Energy (CE Delft and Trinomics, 2014) involved a two-pronged approach. On the one hand, the evaluation sought to measure the impact of the Top Sector upon emissions and on the other hand, understanding the technological impact of the top sector on the structure of the energy system in the Netherlands. While the former focused on collating emissions data and attempting to calculate the baseline emissions and quantify the change instigated by the Top Sector, the latter focused on qualitative data gathered through interviews. This demonstrates the ways in which the evaluations concerning the individual Top Sectors may not be coherently be transposed across to other Top Sectors or compared effectively with others.

B.5.3 Indicators employed in monitoring and evaluation

The indicators employed in the monitoring of the Top Sectors are compiled by the Centre for Policy Statistics and the CBS. Measured since 2010, these (macro-) economic indicators serve to provide a picture of the economic importance and strength of the Top Sectors over time. The monitoring covers more than 30 indictors in total across themes of macroeconomics, business, employment, R&D, innovation, education and green growth (CBS, 2017). The monitoring data is then processed by the EZ and Entrepreneurial Netherlands (RVO.nl) and used to produce the annual Progress Reports.

Notably, while this data provides an indication of the delivery of activities (such as numbers of firms participating in the Top Sectors, distribution of R&D funding, patents, exports etcetera) and currently offer and indication of the baseline, it is not yet possible to attribute any of the changes to the instruments or actions under the Top Sectors approach. Furthermore, these indicators do not offer any insight into the role these organisations play in facilitating socio-technical transitions (Janssen, 2016). As such, the indicators utilised in the 2017 evaluation have been sourced through primary qualitative research conducted by the Dialogic.

While monitoring the financial contributions of participating organisations is necessary, there is very little monitoring of the level of participation or benefits to institutions or businesses present within the Top Sector community that do not financially contribute. As such, understanding the levels of participation or impact upon the wider community is difficult to ascertain.

B.5.4 Evaluation actors, management and contracting

The external consultancy firms and agencies contracted to conduct evaluations of the Top Sector approach and its constituent instruments report directly to EZ. These evaluations are coordinated and overseen by steering committees, with representatives from the relevant government departments.

Dialogic's evaluation of the Top Sectors (2017) was contracted off the back of a pilot study they had delivered previously in association with the development of the evaluation framework. As such, while there may have been an open call for the evaluation, as the incumbent with knowledge and expertise of the subject matter they were almost assured the contract. The initial pilot study was initiated through direct contact.

B.6 Assessment of the success of the intervention, based on

The Top Sectors approach as a whole was found to be effective and efficient (Dialogic, 2017). The Top Sectors have provided a central platform for the sectors to share knowledge, collaborate and coordinate their activities. In doing so, the approach has facilitated greater collectively and a central strategic direction for the sectors, some of which are prone to fragmentation, as the approach has generated a mass and collective momentum for implementing changes to the innovation system. When combined with the orientation away from away from a subsidy-based innovation landscape towards one with greater emphasis on collaborative endeavours, it follows the approach has had at least some impact upon the landscape. The approach has worked to increase the engagement of new players, especially in sectors dominated by a large few. However, the extent of the impact has been varied in number and scope depending on the sector.

The approach has also been successful at facilitating greater cooperation across different regions within the Netherlands (Dialogic, 2017). Participating in the Top Sector means that geographical distance is no longer a barrier to forming new collaborative relationships. However, while the top sectors have facilitated greater PPP, this mostly concerns companies operating in the same ecosystem as those represented in the top sector already. Furthermore, while the increase in demand orientated research has helped to re-affirm the position of research institutions within the eco-system, the emphasis on ground-breaking innovation has been slow to emerge[18]. Notably however, it has also been acknowledged that the approach is in relative infancy. Furthermore, systemic changes such as that supported by the approach take many years to support direct attributable impact upon innovation. As such, while this impact may not have fully manifested yet, these impacts are expected in the future.

While the approach has successfully increased demand-orientated research conducted at research institutes, there is a need for a larger degree of orientation provided by government. The sectoral approach has helped identify the common pinch points and bottle necks within the landscape, shared across sectors and in doing so provide a clear point of action for the government to address within their general enterprise policy. However, the degree to which the Top Sectors are working to address societal challenges and contribute towards the long-term goals of the country is unclear.

The Top Sector approach has also been marked by lower levels of engagement from Government. This is partly attributed to the lower levels of transparency and formal accountability within the network, unclear distribution of responsibility and the low level of buy -n from across Government departments. Again, this feature is partly attributed to the bottom-up approach. However, the benefits of clearer frameworks of responsibility are evident, particularly with the emerging interest to address societal challenges.

The general perception of the Top Sector participants is positive, with acknowledgement that it has been successful in a number of ways.

Indeed, as the Top Sectors have received long-term financial support they have been able to consolidate their communities and have become firmly established. This firm foundation will be able to support greater impact in the future. Furthermore, this foundation will also provide a platform for more diversification of research, amplifying the potential for more progressive innovative work in the future. While the evaluation found the research to be less ground breaking, it was also acknowledged that this was understandable given the budget cuts, as in such cases, the sectors have sought to support their own networks. Further, some Top Sectors were perceived as now supporting very innovative working, further building upon this firm foundation. In light of this, it was also acknowledged that the maturation process for PPP was long but worthwhile. Time was required to learn how to manage the partners and allow to cooperation efforts stabilise.

[18] Notably, this finding largely depends on the definition of ground breaking. In this case, the research conducted by the Top Sectors was noted as being progressive however was not seen to be as disruptive as research or innovation stemming from interdisciplinary or cross-sector research.

The nature and scale of the PPP model within the Top Sector approach is double-edged. On the one hand, the level of private sector buy-in leverages power within the sector and provides a platform for industrial direction setting. On the other hand, the Top Sectors hinder progress towards radical innovation, limited partly by the size of the sector, and partly by the presence of large incumbent interests within the Top Sector community.

Internal accountability seems to vary across all the Top Sectors and was found to be the weakest element within the evaluation conducted by Dialogic (2017). Though the annual reports generated by EZ for Parliament offer an update regarding the progress made towards by the top sectors and the common domains, these reports offer less in the way of formal accountability and reflective policy evaluation.

B.7 Lessons learnt

B.7.1 For the intervention

The effectiveness of the intervention would be strengthened by providing more guidance for the ground-breaking innovation visions of the Top Sectors and facilitating market formation through innovation procurement policy (Dialogic, 2017). Neither of these were formally stated as part of the initial objectives of the approach. There is scope for these objectives to be incorporated into the approach upon its next iteration.

Dialogic also recommended that the Top Sector approach should have strong links to societal challenges (2017). Given the strength of business interests in driving the direction of the actions, the activities cannot be assured to address these social interests. Notably however, one of the successes of the intervention has been not only its engagement of large business in PPP and but its stronger emphasis on the demand driven actions of the consortia, as such, care will need to be taken to ensure the Top Sectors maintain the balance and continue to be relevant for their industrial partners.

B.7.2 For future intervention design

The absence of any clear delineation of responsibility within the approach contributed to the lack of transparency and could compromise the sustainability of the intervention and future evaluations. While the Top Sectors have relatively successfully self-governed, striking the balance between research and innovation to meet industrial needs and actions towards addressing social challenges is challenging and requires at least some government involvement. In order for government to be able to influence the pathways and directions of research in such a context, it follows that its presence 'at the table' would need to be stronger.

As evidenced by the challenges faced in harmonising the monitoring data collected with the evaluation framework later developed, greater attention should be paid to evaluation and monitoring needs at the beginning of the of policy development process.

In relation to the evaluation process conducted by Dialogic, much debate centred around defining the goals of the intervention. Whilst some stakeholders agreed to measure against the initial ambitions of the approach, others questioned whether it would be more appropriate to address the newer missions orientated goals. This issue highlights the importance of defining clear milestones for evaluation within the timeline of the project, in doing so drawing clear lines between the objective setting process and the evaluation activities.

B.7.3 For future evaluations

One of the key lessons from the evaluation conducted by Dialogic (2017) was the development and testing of the evaluation framework. The framework and process developed will not only serve as a useful tool for the evaluation of the Top Sectors in the future but will also inform evaluations of other large and complex programmes. As this methodological framework was very experimental, it was also acknowledged that this could be developed further. For example, the quantitative analysis conducted by Dialogic outside of the formal evaluation (Janssen, 2018) could be incorporated to provide another data point to feed into the analysis.

The evaluation processes of both CE Delft (2014) and Dialogic (2017) highlighted the variable quality and consistency of the monitoring activities across the different Top Sectors. This posed problems with regard to defining the distinction between the control group and those participating. Unable to delineate the interests, benefits and wider participation rates of those involved also poses problems for the definition of the appropriate KPIs. The Dialogic evaluation benefited from being able to define a beneficiary group of those financially contributing to the PPP. As the private financial contributions are clearly stated, the quantitative analysis conducted independently was able to generate a form of quantitative analysis. It was also acknowledged however that this analysis may have failed to capture the wider benefits to the sector.

B.8 References

B.8.1 Literature

Advisory council for Science, Technology and Innovation (AWTI) (2014) Status of the top sectors in 2014 (English summary). Available at: https://english.awti.nl/publications/documents/publications/2014/10/2/status-of-the-top-sectors-in-2014

CBS (2017) Monitor topsectoren 2017, Methodebeschrijving en tabellenset. Available at: https://www.cbs.nl/nl-nl/visualisaties/monitor-topsectoren-2017

CE Delft and Trinomics (2014) Review Topsector Energie. Available at: http://trinomics.eu/evaluation-of-the-dutch-topsector-energie/

Dialogic (2016) Tussenevaluatie TKI-toeslagregeling. Available at: https://www.dialogic.nl/file/2016/12/2016.034-1626.pdf

Dialogic (2017) Evaluatie Topsectorenaanpak. Available at: https://www.dialogic.nl/file/2017/08/evaluatie-topsectorenaanpak-deel-1-hoofdrapport.pdf

Janssen, M. (2016) What Bangs for Your Bucks? Assessing the Design and Impact of Transformative Policy. Available at: https://growthlab.cid.harvard.edu/files/growthlab/files/janssen_rfwp_69.pdf

Janssen, M., Bogers, M. and Wanzenböck, I. (2018) Systemic innovation intermediaries and their moderating influence on R&D collaboration tendencies. [Working Paper] Available at: https://conference.druid.dk/acc_papers/jnqol3zg2bns103zotlkvh4lnz7l40.pdf

Ministry of Economic Affairs (2011) To the Top: Towards a new enterprise policy. Available at: https://www.government.nl/documents/parliamentary-documents/2011/02/04/to-the-top-towards-a-new-enterprise-policy

Ministry of Economic Affairs (2016) Progress through renewal, 2016 Enterprise Policy report. Available at: https://www.topsectoren.nl/publicaties/publicaties/rapport-grondstoffenefficientie-lei-wur/november-2016/30-11-16/rapportage-bedrijvenbeleid-engels

Ministry of Economic Affairs and Climate (2017) Progress report 2017 - Navigating under full sail. Available at: https://www.topsectoren.nl/publicaties/publicaties/rapporten-2017/oktober/09-10-17/vgr-2017---engelse-versie

OECD (2014) OECD Reviews of Innovation Policy: Netherlands 2014. OECD Publishing. http://dx.doi.org/10.1787/9789264213159-en

Technopolis Group (2017) Evaluatie MKB Innovatiestimuleringsregeling Topsectoren (MIT), 2013 – 2016. Available at: https://www.rijksoverheid.nl/documenten/publicaties/2017/03/17/eindrapport-evaluatie-van-de-mit-regeling-mkb-innovatiestimulering-topsectoren

Theeuwes Committee (2012) Dare to measure: Evaluation designs for industrial policy in The Netherlands Final report of the Impact Evaluation Expert Working Group. Available at: https://www.innovationpolicyplatform.org/system/files/Dare%20to%20measure%20Evaluation%20designs%20for%20industrial%20policy%20in%20NL.pdf

B.8.2 Interviews

Name	Role	Organisation
Dr. Matthijs Janssen	Senior researcher/consultant	Dialogic
Dr Stephan Slingerland	Senior consultant	Trinomics
Prof. Dr. Nico van Meeteren	Executive Director	Top Sector Life Science
Dr Henry van der Wiel	Economic advisor	Ministry of Economic Affairs and Climate Policy

Appendix C High-Tech Strategy - Germany

The High-Tech Strategy (HTS) is an inter-departmental strategy launched in August 2006 to promote research and innovation. Associated with a government investment to date of €50bn (Foreign & Commonwealth Office, 2018), the strategy supports a wide range of programmes and instruments with the aim of securing economic, cultural and social prosperity (BMBF, 2010).

The HTS began with a focus on enabling technologies in support of industry, then shifted increasingly towards the societal challenges in Phase II, influenced in part by the publication of the Lund Declaration[19] and the emerging importance of the societal challenges in EU R&D policy. Finally, in Phase III, there has been a further shift from a 'technology-led' towards an 'innovation-led' strategy, geared not only at generating enabling technologies but also at creating solutions for societal challenges.

The HTS currently involves eleven federal ministries: Education and Research; Economic Affairs and Energy; the Interior, Building and Community; the Environment, Nature Conservation and Nuclear Safety; Transport and Digital Infrastructure; Family Affairs, Senior Citizens, Women and Youth; Labour and Social Affairs; Food and Agriculture; Health; Economic Cooperation and Development; as well as the Federal Foreign Office.

C.1 Evaluation culture in the German Federal Government

The German Federal Ministries have a long tradition in technology programme evaluation, dating back at least to the evaluations of Indirect-specific technology programmes by the former Federal Ministry of Research and Technology (BMFT) in the 1990s. BMFT was merged with the Ministry of Education in 1994 to form the Federal Ministry of Education and Research (BMBF). However, the use of evaluation was not mandated and many programmes were not evaluated (Kuhlmann, 2003). Indeed, systematic evaluation of policy measures seems to still be lacking (EFI, 2008; EFI, 2013), even though BMBF maintains a central evaluation function, because evaluation still is in practice invited but not obligatory. Important programs are duly evaluated but many others are not.

C.2 Policy background and origins of the High-Tech Strategy

Historically, the **Science Council** (Wissenschaftsrat, established 1957) and the **German Research Foundation** (Deutsche Forschungsgemeinschaft, established 1951) are responsible for policy co-ordination and for the channelling of research resources into specific areas. The Science Council is also responsible for evaluating research institutes, whereas the German Research Foundation is a grant-awarding body formed by members of the industry, academia and research institutes (Allen, 2009).

Historically, BMFT and BMBF policies have focused on 'push' policies for research and the generation of new technologies. An underlying aspect of the HTS was to connect this 'push' better with the 'pull' of societal needs. This has been realised in the three-stage evolution of its focus, described above. It was the first policy measure developed to involve a large number of relevant stakeholders in the research and innovation sector: Federal Government, Federal Länder, Foundations, Industry and the Science Council. This is also the first time that a national strategy was developed to span most of the ministries in Germany (Allen, 2009), requiring an integrative approach to co-ordinate the various collaborations and to implement the HTS (Federal Ministry for Economic Affairs and Energy, 2014).

C.3 Strategy and plan for the High-Tech Strategy, intervention logic and its evolution over time

The HTS is thus an overarching approach to coordinating policy and funding by triggering inter-ministry co-operation. Its influence goes beyond research and innovation funding to affect regulation and

[19] The Lund Declaration Addendum (2009) Available at: https://www.vr.se/download/18.7dac901212646d84fd38000337/1340207547972/Lund_Addendum.pdf

legislation for research and innovation. Its inter-ministry character that also links it to the political level has made its evolution through three phases possible.

- Phase I (2006-2009) focused on key technologies and lead markets. The aim was to create the conditions in Germany that will enable researchers and organizations to gain leading positions in markets that are both technologically advanced and likely to grow in importance in the future(Allen, 2009)
- Phase II (2010-2013) prioritised the use of key technologies to address the major societal challenges
- Phase III (2014-2017) of the High-Tech Strategy was called Innovations for Germany. This was developed as a broad innovation strategy with new topics and new instruments enabling innovation being added (BMBF, 2014). The progress made within this phase, and the re-establishment of the priority areas was agreed upon in the 2017 report 'Progress through Research and Innovation: Report on the Implementation of the High-tech Strategy' (BMBF, 2017). This second iteration of 'Innovations for Germany' takes the strategy up to 2020

In the early years, the strategy was critiqued for lacking originality and dismissed as a repackaging of existing efforts. However, the strategy has evolved and increased its efforts to link R&D to societal impact. The focus on the societal challenges increases the potential for cooperation within the EU, which is an explicit aim in the current Strategy. It has been associated with greater government financial commitment and, in the most recent Phase (Innovations for Germany), the inclusion and tracking of cross-cutting 'missions'.

- Priority challenges with regard to value creation and quality of life
- Networking and transfer
- The pace of innovation in industry
- Innovation-friendly framework
- Transparency and participation.

Innovation for Germany does not only include technological innovation, but also social innovation, emphasising the social purpose of investments in R &D (BMBF, 2014).

Many of the funding measures implemented in Phase III are based on systemic approaches. For example, the HTS promotes networking and transfer, as well as the innovation activities of SMEs and the creation of start-ups (BMBF, 2014). Further, the Federal Government aimed to create innovation-friendly framework conditions, such as improving the availability and process to access of innovation financing and adjusting copyright law to complement the need of education and research sectors (BMBF, 2014; EFI, 2015).

The High-Tech Strategy functions as an umbrella strategy stretching over a great deal of research and innovation intervention. Whilst many programmes are officially associated with the HTS, others simply define themselves as 'working towards the High-Tech strategy. This looser form of association raises questions about where the boundary of any overall evaluation should eventually lie.

While the initial strategy put forward in 2006 included reference to regular reviews and reports of the implementation of the strategy. From 2018, the Federal Government's Report on Research and Innovation would document the progress made under the Strategy (BMFT, 2016). Published annually, this report offers an insight into the specific activities under each of the 'missions'.

C.4 Governance of the High-Tech Strategy

The HTS is managed by the Federal Ministry of Education and Research (BMBF), which BMBF financially supports almost all of the programmes under the HTS umbrella, though notably programmes also receive financial input from other relevant government ministries. In line with the cross-government approach of the Strategy, some of the instruments and programmes implemented under the strategy are also delivered by other Federal Ministries.

Initially, the implementation of the HTS was to be organised by all relevant forces for innovation in industry, science and government (BMBF, 2006). Two platforms were established to support the strategy and facilitate coordination with non-government organisations: the Industry-Science Research Alliance on the Technology Prospects of Markets of the Future; and the Council for Innovation and Growth.

The Industry-Science Research Alliance (Forschungsunion Wirtschaft – Wissenschaft) comprised representatives from the industrial and science sectors and worked to provide support for both the implementation and continued development of the strategy. Active between 2006 and 2013, the alliance provided advice within the framework of Phase I and Phase II of the HTS. This remit covers providing advice across different levels of implementation, from both the strategic direction of specific innovation fields to the strategic development of the cross-cutting measures. The Alliance was replaced in Phase III by the High-Tech Forum, also functioning as an advisory body. The Forum meets three times a year and is made up of civil-society representatives and stakeholders from academia and business. It advises on the Strategy's development of the new thematic areas and on the implementation of suitable related programmes, integrating its work with the contributions of the other existing committees.[20]

The High-Tech Forum reports to the Federal Government and it seems that the inputs it provides significantly influence the implementation and further development of the HTS (BMBF, 2017). However as noted by the EFI (2015), the delayed establishment of the Forum limits its active work to only two years (within the legislative term). As such, the opportunity for the Forum to provide meaningful input during the early stages of the Strategy's development was missed. In line with this, the absence of a clear and binding process for integrating such external advice into the strategy is noted.

The Council for Innovation and Growth functioned as the advisory council to the Chancellor and provided advice regarding overarching problems and issues within the innovation system. This Council however was only active from 2006 to 2008, when its role providing advice for Government was taken over by The Commission of Experts for Research and Innovation (Expertenkommission Forschung und Innovation – EFI). EFI is an independent expert council of six people, supported by a team of analysts, and provides scientific policy advice to the Federal Government on education, research and innovation policies and it contributes with analyses and international comparisons of the innovation environment (European Commission, 2012). The EFI also publishes annual reports containing their key findings. These reports review encompass the national research and innovation system, and as such include reference to the HTS. The degree to which the reports address the HTS specifically varies from year to year. EFI does not necessary seek to provide an all-encompassing evaluation but instead highlights key areas for development.

C.5 The evaluation

The HTS lacked an overall evaluation strategy from the outset and has not since established one. The individual ministries responsible for programmes within the HTS are also responsible for evaluating them. There is no mechanism for coordinating these evaluations and so far no attempt has been made to do a meta-evaluation or otherwise to bring their results together.

In 2011, a study of the framework conditions of HTS Phase I was commissioned by the Federal Ministry of Research and Education (Technopolis Group et al, 2011). This work was not deemed by the evaluators as a fully-fledged impact analysis, but rather as a monitoring exercise accompanying the High-Tech Strategy. The evaluation project was perceived as a supporting mechanism for the federal ministries to assess the design and the structure of the HTS up to 2011. 'Are the challenges framed in the right way?' was one example of question characterising the evaluation, which studies the innovation system for each thematic areas, looking to identify drivers and barriers. The study also approached cross-cutting issues across the challenges such as the effect of the HTS in SMEs, in venture capital, in public procurement and in skills. The study took a micro perspective, studying each of these thematic areas in focus, and developed an analytical framework centred around the availability and accessibility of funding, ease and

[20] The High-tech forum. Available at: http://www.hightech-forum.de/en/auftrag/

incentives of company formation, access to markets, procurement processes, standardisation, access and availability of skilled workers, and the interconnectedness of these different elements within the field. The methods used in this study were primarily qualitative, reviewing the existing literature and conducting interviews.

The Phase III Strategy includes reference to commissioning evaluations to ensure impact and cost-effectiveness (BMBF, 2014) as part of a plan from the Federal Government to report on implementation and further development of the new strategy at regular intervals. Notably however, reference is made only to evaluations of funding instruments and does not include mention of evaluating the over-arching strategy.

In 2015, the EFI recommended the development of an overall evaluation strategy, covering the indicators, success criteria and evaluation methods in addition to a method for integrating the evaluation outcomes into policy (EFI, 2015). The review even went so far as to stipulate the inclusion of such a framework within the process for designing further research and innovation policy measures.

EFI has historically included reviews of the HTS in its annual reports on the research, innovation and technical performance of Germany. These reviews, though forming only a small part of the overall reports produced, offer critical assessments of the strategy in light of the overarching needs of the innovation system and recommendations for future development and improvements. The extent to which they influence and inform the development of existing and new strategies is not clear. More recently, the Federal Cabinet approved a report on the High-tech Strategy and its progression, detailing the implementation process and responding to the EFI's 2017 annual report (BMBF, 2017). However, as mentioned previously there is no obligation for these recommendations to be considered in the strategy development process.

C.6 Assessment of the success of the High-Tech Strategy

In the absence of an all-encompassing evaluation, assessing the success of the HTS is challenging. The reports generated by the EFI largely indicate that the HTS is a welcome approach, particularly in light of the increasing complexity and number of instruments present within the innovation support ecosystem (EFI, 2017).

The transitions through each of the Strategy's phases have also been acknowledged as important and necessary developments. Government has recognised the changing needs and interests of relevant stakeholders and adjusted the strategy accordingly. This approach, though sometimes suffering from the absence of clear long-term goals, also allows for a great deal of flexibility and responsive evolution. On the hand however, the most recent phase is marked by its mission orientated focus. The reorientation of the approach towards societal needs and the incorporation of social innovation is an important shift in perspective. EFI have as of yet been unable to draw conclusions regarding the success of this approach due to delays in implementation (EFI, 2017).

The HTS has been applauded for its success in strengthening inter-department cooperation, a trait noted as being an integral part of fostering a productive research and innovation policy (EFI, 2017). In light of this however, the strategy would benefit further still from institutionalising these inter-departmental collaborations and presenting a coherent overall evaluation.

C.7 Lessons learnt

C.7.1 For the intervention

One of the key lessons learnt from the High-Tech Strategy is the value of reliability and continuity of research and innovation funding. In line with this, the value of maintaining the Strategy in at least a similar format has been recognised (BMBF, 2017; EFI, 2017). However, further recommendations arise in regard to the timescales for implementation. Indeed, between 2013 and 2017, the implementation and thus impact of the HTS was limited by the legislative delay. Hence, EFI sought to highlight the value of acting quickly to ensure the maximum impact of the strategy could be achieved.

In addition to providing advice regarding the specific foci of the Strategy, the EFI also set out lessons for the governance of the High-Tech Strategy. In particular, the value of cross-department engagement and communication was highlighted by the EFI (2017) which called for the formal establishment of a Federal Committee of State Secretaries for the High-tech strategy. This too would work to facilitate a coherent external image for the Strategy and its programmes.

Though the Strategy has been successful in facilitating inter-departmental cooperation to a degree, the EFI recommends a permanent body to formalise these relationships (EFI, 2107). The establishment of a Federal Committee of State Secretaries for the HTS would serve to formalise the conduits for effective communication, networking and cooperation and in turn foster an external image for the strategy's governance.

The involvement of industry in the governance means that there has been a powerful lobby speaking in favour of the HTS.

C.7.2 For future intervention design

The high levels of international attention devoted to the High-Tech Strategy demonstrates interest in the holistic approach to innovation policy. It has been important effectively to engage the wide range of relevant stakeholders within the strategy development process. Without that, the resulting strategies risk both their transparency and a failure to adapt to new challenges (EFI, 2015; BMBF, 2017). Future interventions will have to make significant efforts to engage a wide range of stakeholders in the early stages of development to ensure relevance for the whole society.

C.7.3 For future evaluations

While the High-Tech Strategy has benefited from the advice provided by various platforms and independent bodies, these inputs do not constitute a formal evaluative process, the absence of which has not yet been addressed. One of the lessons learnt for the development of future large-scale interventions such as the High-Tech Strategy would be the inclusion of a review or evaluation process. In the view of the EFI, the benefit and efficiency of such a strategy needs to be understood in clear terms through a pre-set methodology, developed during the early days of the intervention itself (EFI, 2015). While efforts may have been made to rectify this in the latest iteration (Innovation for Germany) the approach is not yet clear and seems to lack the scope of a full-scale evaluation of the intervention.

Greater attention needs to be paid to clarify the goals of the each of missions, and the hierarchy within them. Coupled with setting milestones at the beginning of the legislative period will also facilitate future evaluations (EFI, 2017). These recommendations are tethered to the absence of an overarching evaluation framework and defined approach. As such, future evaluations should be preceded by the development of such a framework, and a formal process of goal setting.

C.8 References

C.8.1 Literature

Allen, M. M. C. (2009). The national innovation system in Germany. In: V.K. Narayanan and G.C. O'Connor, editor(s). Encyclopaedia of Technology and Innovation. Oxford: Blackwell, p. 375-389. Available at: https://www.research.manchester.ac.uk/portal/files/32298747/FULL_TEXT.PDF

European Innovation Scoreboard (2017) Available at: https://www.rvo.nl/sites/default/files/2017/06/European_Innovation_Scoreboard_2017.pdf

Commission of Experts for Research and Innovation (EFI) (2015) Research, innovation and technological performance in Germany – EFI Report 2015, EFI, Berlin.

Commission of Experts for Research and Innovation (EFI) (2017) Research, innovation and technological performance in Germany – EFI Report 2017, EFI, Berlin. Available at: https://www.e-fi.de/fileadmin/Gutachten_2017/EFI_Report_2017.pdf

Current development and challenges, The New High-Tech Strategy – Innovations for Germany. Available at: https://www.e-fi.de/fileadmin/Inhaltskapitel_EN_2015/2015_A3.pdf

European Commission (2012) Innovation Landscapes – A study on innovation approaches in three selected EU member States. Available online: https://ec.europa.eu/research/industrial_technologies/pdf/innovation-landscapes_en.pdf

Federal Ministry of Education and Research (BMBF) (2006) The High-Tech Strategy for Germany. Available online: https://www.fona.de/pdf/publikationen/bmbf_the_high_tech_strategy_for_germany.pdf

Federal Ministry of Education and Research (BMBF) (2010) Ideas. Innovation. Prosperity: High-Tech Strategy 2020 for Germany

Federal Ministry of Education and Research (BMBF) (2014) New High-Tech Strategy: Innovations in Germany. Available at: https://www.bmwi.de/Redaktion/EN/Artikel/Technology/high-tech-strategy-for-germany.html

Federal Ministry of Education and Research (BMBF) (2016) Bundesbericht Forschung und Innovation 2016. Available at: https://www.bmbf.de/pub/Bufi_2016_Hauptband.pdf

Federal Ministry of Education and Research (BMBF) (2017) Fortschritt durch Forschung und Innovation. Available online: https://www.bmbf.de/pub/Fortschritt_durch_Forschung_und_Innovation.pdf

Kuhlmann S. (2003) Evaluation of research and innovation policies: a discussion of trends with examples from Germany, in Int. J. Technology Management, Vol. 26, Nos. 2/3/4.

Technopolis Group, Fraunhofer ISI, Fraunhofer Fokus (2011) Begleitforschung der Hightech- Strategie - Analyse zu ausgewählten Aspekten.

The Lund Declaration Addendum (2009) Available at: https://www.vr.se/download/18.7dac901212646d84fd38000337/1340207547972/Lund_Addendum.pdf

Foreign & Commonwealth Office (2018) UK SIN country snapshot – Germany. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/680376/171024_Germany_snapshot__002_.pdf

C.8.2 Interviews/Consultations

Name	Role	Organisation
Dr. Pascal Hetze	Head of programmes for innovation policy	Stifterverband für die deutsche Wissenschaft
Mr Engelbert Beyer	Head of Directorate of Innovation Strategies	Federal Ministry of Research and Education
Professor Frieder Meyer-Kramer	State secretary at the time of the launch of High Tech Strategy	Retired

technopolis |group| United Kingdom
3 Pavilion Buildings
Brighton BN1 1EE
United Kingdom
T +44 1273 204320
E info@technopolis-group.com
www.technopolis-group.com