

A formative approach to the evaluation of Transformative Innovation Policies

Jordi Molas-Gallart^{1,*}, Alejandra Boni¹, Sandro Giachi² and Johan Schot³

¹INGENIO (CSIC-UPV), Camí de Vera s/n, Universitat Politècnica de València, 46022 València, Spain, ²Independent Researcher, Cádiz, Spain and ³Utrecht Centre for Global Challenges, Department of History and Art History, Utrecht University, Janskerkhof 2-3A, 3512 BK Utrecht, The Netherlands

*Corresponding author. Email: jormoga@ingenio.upv.es.

Abstract

Transformative Innovation Policies (TIPs) assert that addressing the key challenges currently facing our societies requires profound changes in current socio-technical systems. To leverage such ‘socio-technical transitions’ calls for a different, broad mix of research and innovation policies, with particular attention being paid to policy experiments. As TIPs diffuse and gain legitimacy they pose a substantial evaluation challenge: how can we evaluate these policy experiments with a narrow geographical and temporal scope, when the final objective is ambitiously systemic? How can we know whether a specific set of policy experiments is contributing to systemic transformation? Drawing on TIPs principles as developed by and applied in the activities of the Transformative Innovation Policy Consortium and on the concept of transformative outcomes, this article develops an approach to the evaluation of TIPs that is operational and adaptable to different contexts.

Key words: formative evaluation; sustainability transitions; transformative outcomes; flexible theories of change; transformative innovation policies.

1. Introduction

Transformative Innovation Policies (TIPs) are based on the notion that addressing the key challenges currently facing our societies requires profound changes in current socio-technical systems (Weber and Rohrer 2012; Schot and Steinmueller 2018). To leverage such ‘socio-technical transitions’ calls for a broad mix of research and innovation policies, with particular attention being paid to societal experimentation. These are demonstration projects focused on societal and/or ecological challenges and involving many actors, including social, grassroots, and civil society innovators. They address long-term policy objectives often accompanied by long-term targets and plans to achieve them, and can be supported by strategic visioning and foresight processes. They may deploy policy mixes building on traditional policy instruments, such as R&D subsidies, tax incentives, programmes for building R&D and innovation platforms, and policies for stimulating entrepreneurship. To become part of the TIP mix, policies need to focus on enabling a transformation (Rogge, Pfluger and Geels 2020). Following Schot,

Kivimaa and Torrens (2019), we argue that all TIPs should be executed as Experimental Policy Engagements (EPEs). This notion is introduced to signal that transitions are complex and long-term processes that can be modulated through TIPs but not controlled. In other words, TIPs engage with ongoing transitions that are influenced by many other actors and factors.

As TIPs diffuse and gain legitimacy, they pose a substantial evaluation challenge: the evaluation of such EPEs or sets of EPEs with a narrow geographical and temporal scope, when the final objective is ambitiously systemic. Participants in EPEs need to learn whether the activity has set them up on the way to systemic transformation. This article, therefore, seeks to develop an evaluation approach that is suitable to the evaluation of TIPs.

The problem we face can be seen as a specific instance of the common ‘attribution’ challenge posed by the impact assessment of policies that occur a long way upstream from their intended final objectives, as for instance societal challenge-driven research policies, or local interventions aiming at socio-economic development (Smutylo 2001). In these situations, the results of an intervention

can be no more than a contribution to (and not a determining cause of) the systemic changes being pursued.

The article draws on the work of the Transformative Innovation Policy Consortium (TIPC) (<http://www.tipconsortium.net/>).¹ The research process was part of the co-creation journey between a number of Science, Technology, and Innovation (STI) agencies and researchers working together in the TIPC. Demands for a new evaluation method were expressed in several meetings and initial versions of our approach were presented and discussed in four TIPC workshops and training/learning activities in 2018 and 2019 and other bilateral interactions involving officers from science and innovation agencies in six different countries (South Africa, Colombia, Mexico, Norway, Finland, and Sweden). Next to these general interactions, a case study was conducted with the Swedish Innovation Agency, Vinnova, in order to assess and learn about the added value of our evaluation proposal for their practice (December 2018–April 2019). This case study should be seen as a collective thought experiment, which is part of a broader co-creation journey. Together with the Vinnova project team members, we explored what would have happened if our evaluation method would have been applied to the evaluation of an initiative that had been designed with transformative goals in mind. This article contains an evaluation proposal grounded in this co-creation journey and in previous relevant evaluation and experimentation literature. It is noticeable that, although there is a large literature on the impact assessment of research and innovation policies, it is not focused on assessing transformative approaches to innovation aiming at systemic transitions. Therefore, we draw on a relatively new set of proposed evaluation frameworks and approaches developed for sustainable innovation and sustainable transition policies.

The outline of the article is as follows. In Section 2, we characterize TIP against the backdrop of other STI policies. This section can be seen as an articulation of the demand side leading to a better sense of what needs to be incorporated in the design of any TIP evaluation. In Section 3, we discuss various approaches closely related to the evaluation of TIPs in order to assemble our building blocks for a new evaluation approach, which we present in Section 4. This formative approach is consistent with TIP principles and aims to be operational and adaptable to the different TIPC contexts. In Section 5, we illustrate the implications and potential of applying the proposed approach with reference to a case of a TIP that, despite its transformative goals, was assessed using existing evaluation techniques. This case is used because it was part of the research process leading to the formulation of our formative evaluation proposal.

2. Characterizing TIPs

The TIPC co-creation journey started with a critical consideration of the shortcomings of current approaches to research and innovation policies. The model discussed by [Schot and Steinmueller \(2018\)](#) [see also [Weber and Rohracher \(2012\)](#) and [Daniels et al. \(2020\)](#)] proposes the existence of three innovation policy frames. In the ‘first frame’, policy is based on a linear understanding of innovation: innovation emerges from a process that starts with the generation of new knowledge through basic and applied research, the further development of such knowledge into new technologies, which when applied generate welfare and growth. Within this frame, policy (and evaluation) objectives can be defined and operationalized by focusing on the quality, nature, and mix of R&D

inputs and how they shape the excellence, innovativeness, and viability of the knowledge system.

The innovation systems literature has provided a ‘second frame’ for innovation policy, stressing that the progression from new knowledge, to new technologies, innovation and growth is far from automatic and does not necessarily move in a single direction. New technological development can spur, for instance, basic and fundamental research and the extent to which new knowledge and technologies will lead to innovation and growth is contingent upon a variety of institutional factors and the linkages among different participants in an innovation system. The focus on systemic failures has provided a different rationale for innovation policies, moving beyond R&D investment levels to the institutional conditions and inter-organizational links and learning that can promote innovation. Another policy objective within this frame can be to encourage actors to become more entrepreneurial, including the promotion of commercialization activities among knowledge producers. Yet, the type of innovation that is thus promoted, its direction and characteristics are of lesser concern. Within this frame, policy (and evaluation) objectives can be defined and operationalized by focusing on the scope, scale, and quality of interactions among various actors in the innovation system, the level of commercialization, and the availability of skilled actors needed to participate in the interactions.

The third innovation policy frame focuses attention on addressing societal and environmental challenges through socio-technical system change (which is different from knowledge production and product and process innovation). From this perspective, the directionality of innovation, and the connection between the ecological, social, and the technological arenas become key concerns. Directionality means that innovation policy will not just stimulate specific technological options, but will look into the social and environmental drivers and consequences of each option, then aim for a deliberation on desirable policy directions and eventually foster some desired directions for innovation, while blocking undesirable ones. Of course, this is an iterative process, and not all consequences and directions can be known upfront, so a flexible approach is required. To address directionality, TIPs need to incorporate deep learning and reflexivity, which in this context we take it to imply the questioning and reframing of underlying assumptions about desirable directions. Deep learning or second-order learning typically emerges if the diversity of opinions and beliefs among stakeholders are acknowledged and embraced. Because focusing on disruptive change can result in disagreements among the stakeholders, TIPs require broad consultation processes to discuss different rationales and perspectives in order to broaden the scope of inputs into policy definition, uncover innovative ideas, and minimize legitimization problems later on.

This frame calls for reorienting frame 1 and 2 policies towards transformation, for example, by focusing R&D investments on the Sustainable Development Goals, and by stimulating grassroots innovators. Building, in particular, on the sustainability transition literature ([Grin, Rotmans and Schot 2010](#); [Smith, Voß and Grin 2010](#); [Markard, Raven and Truffer 2012](#); [Loorbach, Frantzeskaki and Avelino 2017](#)), and the transition management and strategic niche management strands, it promotes societal experiments as a promising policy instrument that can be used to explore and facilitate the development of possible transition pathways as well as to coordinate with a wide range of sectoral policies for energy, mobility, health-care, water, food, etc.

During the co-creation journey the conclusion was reached among TIPC members and researchers that STI policies aiming for socio-technical system transitions should be executed as EPEs (Torrens, Johnstone and Schot 2018; Schot, Kivimaa and Torrens 2019). These engagements aim at making unfolding transformation processes more transformative and become experimental because they are time-bounded attempts to influence the transformation in a reflexive and learning-oriented manner. EPEs can support three core transition processes: building and nurturing of niches (or alternative practices), expanding and mainstreaming niches into the wider world (or system diffusion), and the opening up and unlocking of regimes (Grin, Rotmans and Schot 2010; Markard, Raven and Truffer 2012; Kivimaa and Kern 2016; Schot, Kivimaa and Torrens 2019).

The identification and logic of these three processes rests on the multi-level perspective (MLP) on socio-technical transitions as defined by Rip and Kemp (1998); Geels (2002, 2010), and Geels and Schot (2007). The starting point of this framework is that transitions are a change (transition) in socio-technical systems. These systems are stable and dominant configurations of practices, relations, discourses, culture, legislation, etc. providing ways of realizing a particular societal function (Smith, Vofß and Grin 2010). Furthermore, these system elements are put in place, maintained and destroyed by a wide range of actors whose behaviour is configured by rules they also construct in concrete actions. Here, we draw on the sociological duality of structure principle introduced by Giddens (1984). Actors are not passive rule-followers but knowledgeable agents who actively use rules to interpret the world, make decisions and act. These rules contain behavioural instructions, beliefs, and values concerning all system dimensions. Together they form a socio-technical regime. In the end, a system transformation is not only about changing the system, but also about constructing a new regime (rule-set) using the innovative capacities of all relevant actors. It is for this reason that learning and reflexivity have such an important role to play. Actors need to use their agency, question the rules they use in their daily practices, unmake them, and become active rule-makers.

The main contribution of MLP is the idea that a system transition can be understood as the result of interactions across three levels: landscape, regime, and niche (Rip and Kemp 1998; Geels 2002; Grin, Rotmans and Schot 2010). Niches are protective spaces where different ideas, models, configurations, and ways of doing try to survive and develop. Niches present configurations whose characteristics are different from those of the regime: they may work with different principles; may use different technologies; present different relations among stakeholders; or may privilege different sources of knowledge and alternative cultures. Systems and regimes are usually stable, but are permanently exposed to pressures derived from external, powerful, and long-term economic, social, cultural, or environmental trends, which constitute the 'landscape'. Instead, niches usually evolve quickly as they are spaces of experimentation and change (Kemp, Schot and Hoogma 1998). Niches are home to transformative ideas and practices, but their potential is constrained or enabled through the more powerful structures of the regime (Bos and Grin 2008). Systems transitions may take place when the regime is destabilized because of the heavy pressure of the landscape, so windows of opportunity may be open for niches—if they are mature enough—to influence or even completely replace the regime (Geels 2002). Since transitions involve political struggles among niche and regime actors, conflicts are to be expected, but can be resolved.

This characterization of TIP and the underlying MLP on system change brings up a number of demands for evaluation practice. It should pay attention to directionality in relation to societal and ecological challenges, to deep learning and reflexivity, and should be based on inclusive and participatory processes. TIPs aim at system change, so evaluation should focus on identifying processes contributing to such change accepting that conflicts among actors need to be embraced in the evaluation process.

3. Evaluating TIPs for sustainability: some foundations

TIP evaluation involves assessing the changes associated with or leading to socio-technical transitions. This is a very challenging task. As argued above, we encounter the problem of relating ambitious medium and long-term systemic goals with geographically and time-bounded EPEs. Furthermore, TIPs are explicitly navigating innovation into specific directions, and considering some types of innovation as undesirable. In determining directionality TIPs propose inclusive approaches to policy definition and implementation and we maintain that such inclusiveness needs to be also expressed through evaluation practices. We need, therefore, to use evaluation approaches that allow for the participation of policy stakeholders while providing evidence of the extent to which a policy is contributing to a systemic change in the desired direction.

These challenges are not unique to TIPs, but are common to the assessment of all policies that support sustainability transitions and address environmental problems. There is a substantial body of research and practice that has sought to assess the results, efficiency, and effectiveness of these activities. We can distinguish three main kinds of evaluative work. Some studies undertake policy assessment from an academic perspective; that is, they try to determine policy success and failure factors and try to improve our understanding of how policies work and yield results, but without being directly part of the policy process. In contrast to this work, we find evaluative research that is almost undistinguishable from the policy initiatives they assess: this occurs in the case of many policy experiments. Policy experimentation is very common among sustainability policies, and although there are different types of experimentation, 'learning is an essential justification' for them (Kivimaa et al. 2017). When the main objective of an (experimental) policy intervention is to learn about the effectiveness and efficiency of the policy tools being experimented upon, it is designed in order to be assessed and such assessment is part of the experimentation. Finally, policy evaluation practice constitutes a body of policy assessments intended to feedback into the policy cycle in such a way that it becomes part of such cycle. Policy evaluations have a wide variety of goals, can be implemented in different stages of the policy cycle and use varied methodologies, but they are always targeted to a specific policy or portfolio of policies and constitute a tool for their definition and management.

A very substantial part of the evaluative work addressing sustainability transitions is linked to policy experimentation and has seldom appeared in mainstream evaluation journals. A review of these approaches can be found in Luederitz et al. (2017). In this work, a team of 28 researchers built an analytical framework for the evaluation of 'sustainability transition experiments' (Luederitz et al. 2017: 61 passim) based on an extensive review of an evaluative work mainly related to transition experiments. Extracting different

assessment perspectives from different studies they constructed a list of 25 ‘features’ (Luederitz et al. 2017: 64 *passim*) with associated evaluative questions to be applied to the assessment of transition experiments to make them more effective and efficient. Extracting the list from a compilation of the literature allows the authors to avoid associating their approach to any ‘single theoretical interpretation of transition experiments’, but rather to ‘provide a broad array of features that are of importance across different framings of sustainability transition experiments’ (Luederitz et al. 2017: 72). The absence of an underlying theoretical foundation, and therefore of a specific Theory of Change (ToC) providing the logic of how the inputs invested in a policy are expected to lead to a set of outputs and relevant outcomes, allows this design to be used for the comparison of different experiments regardless of their underlying theoretical foundations and ToCs. The goal of the authors is to provide a tool for comparison across experiments ‘to facilitate and accelerate learning across different experiments’ (Luederitz et al. 2017: 72).

Our approach is different in that we aim to develop an approach within an experiment or policy. In this way, we are firmly established within the third type of evaluative approaches described above: the development of a policy evaluation tool applicable to a specific intervention and providing input into the other policy cycle tasks (policy definition and implementation). To this end we will establish a set of outcomes against which to deploy a monitoring strategy, based on a ToC providing the logic of intervention, which is in turn substantiated by a theoretical argument. Like Luederitz et al. we also focus on learning, but our use of the formative approach described below rests on an understanding of the specific logic of an intervention and the contextual conditions underpinning it. For all these differences our approach reflects many of the ‘features’ Luederitz et al. identify. For instance, our approach stresses that the collaborative practices characterizing TIPS are extended to evaluation and in so doing we strengthen the collaborative ‘feature’ of transition experiments and we intend to ‘support individual and organizational learning’ through a process of reflexive monitoring and evaluation.

Reflexive processes ‘enable the challenging and change of presumptions, current practices, and the underlying institutions, either in the design of a project or in its management’ (van Mierlo, Arkesteijn and Leeuwis 2010: 145). A reflexive approach to evaluation will encourage learning across actors seeking to contribute to sustainable development by working on system innovation (van Mierlo, Arkesteijn and Leeuwis 2010; van Mierlo et al. 2010; Arkesteijn, van Mierlo and Leeuwis 2015). ‘Reflexive monitoring and evaluation’ has emerged as a specific approach that distinguishes itself from more common ‘result-oriented’ evaluations by considering learning how to contribute to system innovation the central goal of evaluation. ‘Result-oriented approaches’ focus on accountability and steering, and on a set of predefined objectives, while ‘reflexive monitoring and evaluation’ put ‘the prevailing values and institutional settings up for discussion’ (van Mierlo et al. 2010: 36). Therefore, in this perspective learning goes hand in hand with a constant process of questioning dominant values and institutions and is not connected to policy steering. Policy steering is not at the centre of the evaluation objectives because of the way in which the approach understands the implications of complexity. System change is understood to be complex and, consequently, the changes generated by the policies aiming at such change are thought to be unpredictable. Referring to Rogers (2008), they argue that ‘Outcomes of project interventions cannot be predefined but are

emergent because system innovations are highly complex without clear causal strands and linear paths [...]. System innovations consist of many different social and technical components that cannot be usefully identified in advance and are partly invisible and/or intangible’ (van Mierlo, Arkesteijn and Leeuwis 2010).

Like the ‘Reflexive Monitoring and Evaluation’ approach, we intend to use the evaluation process to add a reflexive layer to the policy definition and implementation process, and consider that challenging and changing dominant assumption, practices, and associated institutions are core aspects of transformative innovation. Yet, we interpret the implications of complexity differently. We agree that the behaviour of a complex system cannot be predicted with exactitude, yet this does not imply that all possible results are equally probable. The assessment of the different probabilities of a set of possible results is common to many analysis of complex systems. From our perspective, the history and theory of socio-technical changes provides important clues as to the most likely components and characteristics of such change and provides guidance so that policy evaluation can be used as a tool to navigate systemic change into desirable directions. The approach we propose is based on a ToC. A ToC ‘sets out why it is believed that the intervention’s activities will lead to a contribution to the intended results’ (Mayne 2011). It defines the expected relations between the resources invested in an intervention and their effects, and the assumptions under which we expect such effects.

We are not alone in using a description of the situation to be transformed, the desired goals, and the steps linking them, as a tool for the evaluation of policies aiming at complex socio-technical change. Taanman (2014) bases his approach to the monitoring of sustainable transition programmes on the generation of ‘transition scenarios’ describing ‘how the current situation is expected to be transformed in the desired situation. In this respect, a transition scenario is similar to a policy theory, ToC, programme theory or plan’ (Taanman 2014; Section 5.1). Taanman states his objective as the enactment of ‘fundamental change in the dominant culture, structure, and routines of a regime’. The scenarios ‘can be used to frame which changes in culture, structure, and practices and which sustainability criteria are relevant to monitor’. Finally, given the uncertainty of systemic change processes, Taanman recognizes that ‘Transition scenarios can change over time’ (Taanman 2014; Section 5.1). On similar lines we will propose below the use of flexible Theories of Change. We will draw on the techniques developed by Dutch consultancy HIVOS, stressing the use of action research tools for the definition and redefinition of Theories of Change (van Es, Guijt and Vogel 2015).

Taanman’s approach is comprehensive and complex, considering the application of different monitoring modes at different stages and policy levels. Complex systems transitions cannot be achieved through a single policy or experiment but will require the combination of different policy tools. Evaluators are therefore confronted with, both the systemic character of the policy aims, and the multi-level nature of the potential policy interventions. Taking a systemic view requires a shift in interest from the project or programme levels to the level of the whole system that the policy initiatives are trying to effect (Caffrey and Munro 2017). The evaluators’ main focus may thus change from the analysis of a specific intervention, to the study of the effects of portfolios of interventions and the systemic impacts of policy mixes involving an integrated evaluation of the different policy instruments and their interactions (Magro and Wilson 2013). Turnheim et al. (2015) propose to tackle the gap

between specific actions and the systematic transformation they are aimed at through ‘an integration strategy based on alignment, bridging, and iteration’ of learning-based evaluations of local initiatives with socio-technical analysis at regime level, and quantitative system modelling at the landscape levels. Yet, this is a complex approach encompassing several analytical layers to align a set of complementary interventions. These approaches assume a set of different policies pursuing the same rationale and which can be assessed with coherent criteria. Yet, policies aiming at sustainable transitions are often defined as local experiments, which are only sometimes brought together under a programme of interventions. Although broad sets of policy mixes with a common rationale are still rare, it is still important to identify the levels of the policy activity or set of activities we are evaluating.

Taanman (2014) identifies three levels in interventions seeking sustainability transitions: projects, programmes, and transition field. The three levels form a hierarchy: ‘the higher levels provide the context for lower levels and lower levels influence the higher level context’ (Taanman 2014; Section 3.2). His monitoring approach may address one of the levels or, more often, the interactions between two levels (for instance, the relationship between the transition dynamics in the ‘gas system’ and a programme of activities) or how the developments at all three levels are related to each other.

We find the distinction between different levels of policy action necessary, but we see each higher level as a providing an additional layer of policy activity, connecting lower level interventions, rather than just a context that influences and is influenced by lower strata. By defining the levels in terms of policy interventions we can focus the evaluative analysis on the results of the policy, rather than engaging in comprehensive assessments of systemic changes. Yet, even without engaging in an extensive systemic analysis, assessing the results of policy interventions remains a very challenging task because the changes sought will typically occur a long way downstream from the intervention and be the result of a complex interaction of factors which may or may not be directly related to the interventions.

A common approach to evaluate the effects of this type of programmes is to focus on their outcomes. Evaluators have developed different approaches focusing policy evaluation on the policy outcomes rather than their long-term impacts (Earl, Carden and Smutylo 2001; Wilson-Grau and Britt 2013; Belcher, Davel and Claus 2020). Outcomes are defined ‘as changes in the behaviour, relationships, activities, or actions of the people, groups, and organizations with whom the programme works directly. These outcomes can be logically linked to a program’s activities, although they are not necessarily caused by them’ (Earl, Carden and Smutylo 2001: 1).² We are going to adopt this definition and focus our evaluation approach on outcomes. Our contribution is that, instead of adopting a completely open position in which all emerging outcomes are treated equally without any attempt at orienting the study in any specific direction (which is the common approach in many outcome-based evaluation methodologies) we will orient and structure our analysis using 12 categories of transformative outcomes (TOs) defined below. These categories offer guidance about the kinds of transformative change that we need to trigger and thus help profile a policy in terms of its transformative potential. Furthermore, our approach will aim to establish whether and how the activities enacted by the policy have contributed to the selected outcomes. Following Mayne (2011), this analysis of policy contributions uses the ToC to infer causation. The ToC lays out what Pawson and Tilley (1997)

refer to as the ‘Context-Mechanism-Outcome’ processes that explain how interventions generate or fail to generate results, thus adopting a generative understanding of causation.

4. A formative evaluation approach for TIP

4.1 A flexible ToC focused on TOs

We propose that the formative evaluation of TIPs focuses on the analysis of TOs that can be expected to accrue while the intervention is still ongoing. These outcomes can be traced back to the immediate results (outputs) generated by the intervention and they contribute to the transition process that the policy seeks to enact. Such links constitute a ‘ToC’ understood as an account of what is expected to happen; that is how policy inputs lead to activities that contribute to relevant changes (TOs) which in turn will contribute to systemic change.

To build the ToC and define TOs we work with existing transitions theory, in particular the MLP as outlined above. An elaboration of these outcomes has been published separately (Ghosh et al. 2020); here we provide a summary in the next section. Our approach is similar to theory-oriented approaches in that it revolves around a description of an expected process of change that ‘consider programmes in their context, which includes actors’ environments ... and public service culture and behaviour’ (Stame 2004: 63). But unlike most theory-oriented approaches in which the evaluator builds the theory behind the programme interpreting the understanding of what may happen offered by actors involved in the intervention (Stame 2004), we actively use transitions theory to co-produce with the policy actors a ToC that focuses on transformative changes.

4.2 Six guiding principles

We propose a formative approach to TIP evaluation that is coherent with TIP principles and builds on the existing evaluation practices described above. Our approach emphasizes participation, focuses on the analysis of TOs, and pursues the improvement of TIPs definition and implementation. On this basis, we formulate six principles to guide the evaluation of TIPs. These principles have been discussed in TIPC meetings and were accepted by the members as reasonable starting points.

1) *Adopt a formative approach to evaluation.* By a formative approach we mean a style of evaluation which is conducted with the participation of stakeholders with the main purpose of improving the definition and implementation of the interventions being evaluated. Under this perspective, evaluation should be understood as a reflexive practice aiming at helping policy actors to navigate their TIPs and contributing to their capacities to do so. In such a practice, failure should be seen as a learning opportunity on the context, conditions, and activities conducive to transformative processes. In addition, evaluation can help refining transformative innovation theory by providing information about how to make EPEs work effectively. Our approach to the application of formative evaluation to innovation policies draws on a stream of evaluation work dating back to, at least, the mid-80s and the evaluation of the UK Alvey programme (1984–90). This evaluation of a British programme to support R&D in the information technology sector, developed a real-time evaluation approach and provides an early example of formative STI policy evaluation (Molas-Gallart and Davies 2006). The Alvey evaluators argued that that real-time evaluation had several advantages

over ex-post approaches, particularly the fact that it provided actionable feedback to those working in an intervention (Hobday 1988). The UK Alvey Programme evaluation became a referent in the early 90s. The use of evaluation approaches that were explicitly characterized as real-time *and* formative can be traced to another evaluation of a programme in the IT sector (Eschenbach, Hafkesbrink and Lütz 1995). Formative approaches were soon after developed as part of a new mode of evaluation in which evaluators would get directly involved in learning exercises with all programme stakeholders, playing the role of facilitators rather than that of external experts, and leading to a more flexible and experimental approach to innovation policy formulation (Kuhlmann 1999).

2) *Integrate evaluation with policy design and implementation.* Following from our understanding of formative evaluation, we see evaluation as part of the policy process and, therefore, as a task that should share in the overall characteristics we aim this process to have. Specific policies, their implementation, and evaluation should be coherent with the stated research and innovation policy objectives (directionality, societal goals, and system impact). Evaluation thus becomes a strategic part of the design and implementation process of TIPs.

3) *The evaluation process should be inclusive and participatory.* The inclusivity characterizing TIPs should also be present in the evaluation process. Traditional evaluations are often led by external evaluation experts who implement and plan them. In contrast, participants in TIPs should also join in their evaluation, with external evaluation experts mainly acting as facilitators paying, for instance, attention to the power dynamics that may lead to some voices being heard more than others. Therefore, evaluation should facilitate participation and open debate, channelling power conflicts, and differences in interest and perceptions. The groups and communities participating in the evaluation process will be varied and have different access to resources and different interests. Managers and grassroots participants, for instance, may have different perspectives on the definition of the problems to be addressed, and be unequal in terms of the power they hold. An evaluation design should be attentive to such differences.

4) *Use a mix of methods and techniques.* Rather than being driven by formalized standard protocols, evaluation practice needs to be adaptable and flexible, selecting different methods and techniques according to the policy context and its transformative nature. Quantitative techniques can provide synthetic assessments that allow for comparison across different units of assessment, and can provide, under specific conditions, robust assessments of the net impact of an intervention. Yet, the assessment of transformative impact is difficult to be achieved with 'standard' indicators. This difficulty is in part attributable to the nature of social values, which are often linked to incommensurable dimensions and perceived differently depending on cultural background and personal preferences. In these situations, qualitative methods can provide a better approximation to impact assessment by providing a fined-grained, contextualized description of TOs through detailed narratives. Finally, participatory techniques can help increase participation and the inclusiveness of the evaluation process.

5) *Use a nested approach to assess multi-level TIPs.* TIPs can operate at different levels. Niche projects are local initiatives attempting to generate or support a specific niche. Programmes may bring together several niche projects and will seek to develop links and relationships among them that will facilitate scaling up. Finally, several programmes can combine with other policies in policy mixes

that aim to realize socio-technical system change. Impact understood as transformation of a socio-technical system cannot accrue from a single niche-level experimental policy. Each small-scale experiment can contribute to socio-technical change, and such contribution can be enhanced by combining them with those of other experiments grouped in policy programmes. Each policy intervention can therefore be evaluated on its own, but our expectations of what the policy can achieve will differ according to its level. Such expectations should, however, be coherent across levels. The outcomes that are pursued at project level form part of and contribute towards outcomes at programme level and these, in turn, towards the outcomes and impacts at the policy mix level. The outcomes achieved at each level are, therefore, nested within and will contribute towards those of the higher level.

6) *Use a flexible ToC.* Many evaluations use Theories of Change to structure their work. The ToC approach to policy evaluation has a long history. The term was coined by Carol Weiss who proposed that programme evaluation be built on potential causal models of the programmes and defined ToCs as 'the chain of assumptions explaining how activities lead step by step to the expected outcomes' (Weiss 1998: 2). Building on Weiss, Connell and Kubisch (1998) define a ToC approach as a systematic and cumulative study of the links between activities, outcomes, and contexts of the initiative being evaluated. We do not understand these links as simply reflecting a cause-and-effect relationship, since contexts, activities, and outcomes are co-evolving. A ToC is typically defined by policy stakeholders and starts by identifying the main changes that an intervention is aiming to achieve. Policy goals are therefore defined as changes to a baseline situation. Next, participants work backwards from such intended changes to identify the processes that will lead to them, and how these processes will be triggered by the intervention. In this way stakeholders, with the help of evaluation experts, produce an expected process linking the activities³ triggered by an intervention with its results. Our ToCs will be *flexible*, implying that they should not be understood as a fixed causal chain between inputs, activities, TOs, and impacts. Rather, they can be revisited and redefined as a result of the formative evaluation process. The ToCs will be used to foster learning (first and, specially, second order⁴) and reflexivity among participants and to help assess if the policy is contributing to move towards its objectives. Following ToC conventions, we will distinguish five elements, which we will align with the MLP in transitions theory.⁵

- Context: the background 'socio-technical landscape' influencing socio-technical regime change, but which is not directly addressed by the intervention.
- Inputs: the resources available to actors to enact change, including the inputs provided by the policy intervention.
- Activities: the interventions which together constitute the experiment. These activities are linked to:
- TOs in three areas drawn from MLP: (1) building and nurturing of niches; (2) expanding and mainstreaming niches; and (3) opening up and unlocking regimes. All these outcomes are identifiable in individuals, groups, and organizations involved in the experiment (see Table 1 below).
- Impact: the emergence of a new, sustainable socio-technical system(s) that will deliver on the ultimate policy goals in terms of reduction of inequality, CO₂ reduction, air pollution, etc.

Table 1. Twelve types of transformative outcomes, adapted from Schot, Kivimaa and Torrens (2019), Ghosh et al. (2020)

Niche building	
Shielding	Offering protection for niche experiments and normalizing these protection measures. Protection can be offered through subsidies but also market benefits, such as a VAT exemption, or cultural protection by trying to change the meaning or perceptions of a specific solution through a media campaign
Learning	First order (optimizing existing behaviour) and second order (changes in frames and assumptions) in or across several system dimensions (science, technology, innovation; markets; culture and symbolic meanings; industrial strategy)
Networking	Participation in the niche of a wide range of diverse (in terms of niche and regime actors, and in terms of regime dimensions) stakeholders Building and strengthening ties among actors in a niche Creation of a community of practice ensuring resource mobilization Emergence of intermediaries in facilitating the above
Navigating expectations	Creating space for voicing new and alternative expectations and bridging the diversity of expectations building a shared vision
Niche expansion and embedding	
Upscaling—increasing user adoption	Spread of the adoption of new practices and rules, bandwagon effect
Replication	Replication of niche conditions in different contexts Adaptation of a niche in a different locality
Circulation	Circulation of ideas, people, tacit knowledge, rules across niches, and system dimensions Emergence of system intermediaries
Institutionalization (formal and informal rules)	Developing standard definitions, narratives, regulations, and preferred types of behaviours, beliefs, and values Establishment of certification schemes, protocols... Development of a mature market niche
Opening up and unlocking regimes	
Destabilizing and de-aligning regimes	Disrupting policy frameworks and governance arrangements taking advantage of tensions between regime dimensions Phasing out of policies and implementation of other policies disrupting the dominant socio-technical system
Unlearning and deep learning of regime actors	Second-order learning among regime actors—change existing values and beliefs Unlearning routines based on existing skills and capabilities Emergence of new policy assumptions
Empowering niche–regime interactions	Creation of formal and informal linkages between niche and regime actors Emergence of intermediators facilitating such linkages
Changing perceptions of landscape pressures	Regime actors develop new interpretations of the nature and consequences of trends (such as climate change, loss of biodiversity, pollution, rising inequality, digitalization, urbanization) and shocks

The focus on TOs is a key element in our method and is linked to our theoretical understanding of TIPs (Schot, Kivimaa and Torrens 2019). We argue that there are three main transformative processes in the transition from a local niche where a new sustainable socio-technical environment emerges to the change in socio-technical regime: (1) building or constructing the niches; (2) accelerating their growth and expansion and embedding them in the regime, and (3) opening up the existing regimes, destabilizing their practices, and unlocking path dependencies. The three groups of TOs mentioned above mirror these three processes. Schot et al. (2019) have identified and defined in detail 12 different types of TOs, 4 in each of these 3 groups. A summary of this typology can be found below in Table 1. When co-constructing the ToCs with experiment participants we will identify how the expected outcomes can be mapped against these 12 types. It is important to note that we are not proposing that experiments should cover in a comprehensive manner all of the outcome types. In most cases this would be unfeasible. What the typology offers is a guide that enables users to become aware of how their activities are positioned

against the range of processes required to achieve socio-technical transformation.

5. The implications of our approach

In this section, we provide a brief illustration of the differential implications of applying our approach to a transformative innovation programme that had already been implemented and assessed. The case is a long-term project on circular economy in the emerging Forest Chemistry sector, funded by the Swedish Innovation Agency (Vinnova) Challenge-Driven Innovation (CDI) programme. The case was selected together with Vinnova, with the objective of exploring ex-post the value-added of our approach. The question was whether the project would have developed in another way using a formative evaluation approach with TOs. The case study can thus be seen as a thought experiment conducted together by Vinnova and the TIPC research team. The TIPC team consisted of the authors of this article and the Vinnova team of seven people, strategic advisors, internal

analysts, and evaluators. The TIPC team studied a number of internal⁶ and three scientific documents (Allmér 2017; Fuenfschilling, Bauer and Clemente 2017; OECD 2016). Subsequently, a workshop was organized in Stockholm (29 January 2019) in which we introduced our approach, carried out a joint case study review, and drew a set of conclusions.

We first present the characteristics of the CDI programme in the context of Sweden's need to transform industry and the business sector in a sustainable way and present the specific project with its transformative goals. We then compare our approach to the archetypical evaluation approach that was used for the project and highlight how our formative evaluation framework could enhance socio-technical transformation.

5.1 The Swedish CDI programme

The origin of the CDI programme lies in the 2009 Lund Declaration and the need to move the Swedish STI system towards flexible approaches able to tackle current societal challenges. Vinnova launched the programme in early 2011 with the main purpose of converting societal and environmental challenges into opportunities for economic growth. CDI focused on four related areas: Information Society 3.0, Sustainable Attractive Cities, Future Healthcare, and Competitive Production. 'These were all areas in which Sweden had both a strategic interest and a good innovation track record' and where 'advances in the development of adequate solutions to many societal problems will need to be made' (OECD 2016).

The impact goals of the CDI programme aimed at generating solutions for both sustainable growth and the internationalization of Swedish technology. In operational terms, the programme sought to improve coordination and mobilization amongst business actors, and promote cross-sector collaboration and user- and demand-driven innovation initiatives. Each project had to illustrate its impact logic accordingly, addressing specific challenges in line with the programme's objectives. The CDI programme funded projects following a stage-gate process based on three stages: initiation, collaboration, and follow-up. The projects were thus assessed before being allowed funding for scaling up.

The CDI programme is an example of a 'frame 1' (Schot and Steinmueller 2018) R&D programme with 'frame 2' features because of its attention to the interaction among actors, and several characteristics of a 'frame 3' (TIP) approach because of its orientation towards addressing societal challenges (Fuenfschilling, Bauer and Clemente 2017; Smallman 2018: 250). In our work, we did not focus on the entire CDI program but on one specific exemplary case study: the Forest Chemistry project. This project had progressed successfully through the various stages of the CDI programme and therefore provided a suitable example for our thought experiment.

5.2 The Forest Chemistry project: the innovation journey

The 'Forest Chemistry' project took place between 2011 and 2017, and developed through the three stages that the CDI programme envisaged. This resulted in granting three different subprojects: the Forest Chemistry (*Skogskemi*) projects I and II, and the Forest Methanol (*Skogmetanol*) project. The aim of the Forest Chemistry projects was to develop new 'green chemicals' production technologies using the residues generated by the forestry industry. The main technological objective was the development of a system by which

the methanol generated by sulphate pulp mills dedicated to paper production would be used for purifying NOx emissions in the local chemical industry. Underpinning this project was a transformative vision of contributing to the building of a circular economy. Not only did this involve the development of new technologies, but also the construction of new industrial links between paper pulp mills and the chemical industry. To reach these objectives called for the deployment of several policy interventions including government subsidies to make the price of the methanol supplied to the chemical industry attractive. A variety of parties participated in the projects including sulphate pulp mills, chemical companies, and a 'Support Platform' formed by Vinnova policymakers, researchers, and representatives of the forest and chemical industries.

The project started with a preliminary study of the components present in forest raw material and its potential use in the chemical industry. Based on its results, RISE Processum (a research institute-owned bio-refinery developer) pursued increased co-operation between forest and chemical industries to identify chemicals, processes, and value chains with large potential. Therefore, in the first stage (November 2011–March 2012) the project focused on knowledge generation, identifying three value chains with the greatest technological and market potential for Sweden: methanol, butanol, and olefins. In stage 2 (August 2012–November 2014), pre-feed, system analyses, and technical evaluations were performed leading to a focus on the development of technology for cleaning the methanol from stripper gases. A new company joined the consortium in this phase: a sulphate pulp mill that had a high emission of NOx gases, and therefore a need for an improved cleaning process. In the third stage (May 2015–June 2017), the project reached maturation. In 2015, the consortium included five actors operating in different stages of the value chain for cleaning stripper gases: research organizations, equipment suppliers, sulphate pulp mill, and an organization representing end customers. In 2016, the consortium achieved the demonstration of a flexible pilot equipment designed by the equipment supplier and installed on the sulphate mill. A 1,000 h of continuous operation test showed that the pilot equipment obtained results in line with the lab environment and that the purification process worked (NOx emissions were significantly reduced, almost enough to use the methanol as a green input chemical). However, in 2017 the utility equipment manufacturer decided to sell the technology. It had not been implemented up to the day we discussed it in the workshop.

According to Vinnova representatives, the project suffered from coordination difficulties due to lack of communication among sub-projects and partners. Partners from different sectors did not reach a common understanding on the long-term tasks and goals of the project. Therefore, although short-term commitment to the project was successful, the project could not ensure the long-term engagement of participants.

5.3 The evaluation approach

In the workshop, we considered how things could have been done differently with a formative evaluation approach focused on TOs. We concluded that, despite its challenge-driven and transformative ambitions, the Forest Chemistry project followed a linear and traditional approach from design to implementation, with limited opportunities to diversify project options and redefine arrangements in order to stretch the project towards a more transformative focus. From a TIP perspective, this project could have been staged as a

process of niche construction with four key transformative processes: shielding, learning, networking, and navigating of expectations (see Table 1). It was considered that within the scope and time frame of the project it was not feasible to address TO addressing niche expansion and regime destabilization processes.

Although various expectations played a key role in the first stages of the projects (Allmér 2017), in stage 2 participants closed down alternative options for technology development. This decision was solely based on techno-economic viability criteria, while other criteria linked to the social and environmental challenges were downplayed. This strategy is, however, consistent with the norms and rules that characterize the current socio-technical system. The concern about economic viability and the expectation that technical solutions can provide low risk and profitable solutions for the firms involved can explain the early closing-down of alternative technological development paths. To do otherwise would call for additional resources and make the decision process more complex. Similarly, discussing alternatives in a participatory and reflexive way is costly both in terms of time and resources. Yet, the initial transformative expectations and visions required a broader approach moving beyond the concentration on developing and implementing a technological solution. Our TOs draw attention to the need to build a broad and deep network that would not only sustain resource mobilization but also allow actors representing social and ecological interests to voice their concerns and expectations. Similarly, second-order learning and reflexivity were largely absent, actors did not question their own assumptions, but focused on realizing one particular option. But such activities could only be developed in a protected niche that could shield participants from the immediate rules and economic logic of the incumbent socio-technical system. In short, although the Forest Chemistry project should have aimed at constructing a niche for green chemicals inducing a circular economy, the project focused too early on successful implementation with one particular technology. This choice was reinforced by the way in which evaluation was implemented. A formative approach revolving around TOs would have introduced reflexive practices. Such practices would have increased awareness of how the early focus on technology solutions led the programme towards operating within the existing institutional frameworks. By keeping a focus on TOs, policy direction could have been maintained even when it cut against the grain of established practice. Instead, the CDI program stage-gate mechanism led to a summative approach accompanied by tight timeframes, which many participants criticized because it could not be dovetailed with the long processes of dealing with the regulatory and technical aspects of developing applications and it hindered the broader involvement of civil society actors (Fuenfschilling, Bauer and Clemente 2017). As a result, the easiest and more straightforward options tended to be chosen instead of pursuing riskier and potentially transformative alternatives.

It was agreed during the workshop that our evaluation approach may have helped to address these gaps. It supports, from the very beginning, the management of expectations and visions of the emerging network. Such expectations and visions can be expressed through a flexible ToC, designed and adapted by means of participatory techniques. This ToC can then be used for developing indicators for TOs; but it must be noted that these indicators are the result of a reflexive process involving a wide range of participants, who then use them to discuss and guide their choices. Note that this process is very different from the requirement to find easily quantifiable and difficult to ‘game’ indicators, which can allow a comparative

measure (usually against a benchmark) of project achievements. These latter indicators are needed to make ‘stop-go’ decisions at the stage gate, and can be typically found in measures of technical or financial performance. Yet, transformative policies are guided by more complex socio-technical achievements which can seldom be rendered by easily measurable and comparable indicators. In a formative approach, the indicators linked to the ToC will be used by the project participants to inform assessments of the degree to which they are making progress into the desired trajectory of change (instead of being used by external actors as an objective measure on which to base funding decisions). Therefore, they do not need to be comparable across projects. In other words, usable indicators may not provide decontextualized measures and may even be qualitative.

On the basis of the discussions during the workshop, Vinnova concluded that a formative evaluation could make a positive contribution to their transformative ambitions. Accordingly, they have started a pilot with TIPC on formative evaluative evaluation with a newly started initiative for building a sustainable health and food system (<http://www.tipconsortium.net/experiment/swedens-innovation-policy-experiment/>).

6. Conclusions

We have argued that an evaluation approach for TIPs needs to be formative, aiming to improve the definition and implementation of the interventions under evaluation and involving the policy participants. This requires evaluation to be conducted in real-time, as a form of constructive monitoring. The reflexive process provided by formative evaluation and the focus on TOs can drive policies towards achieving their transformative goals. This is a challenging task as it faces the obstruction of the promoters of the existing socio-technical systems, bringing institutional inertia and entrenched incumbent interests. As illustrated by the CDI project, in the absence of a continuous reflexive process, initiatives aiming at transformative change can easily refocus on technical goals and lose sight of the transformative challenges ahead. Transformative change needs to be supported by a clear strategic drive which is deliberately pursued through the policy implementation processes; such is the goal of the change pursued here.

To be able to assess in real-time the degree to which the interventions are progressing towards the achievement of long-term systemic goals, the evaluation approach needs to be focused on TOs. We propose to use 12 TOs as a heuristic to reflect on the transformative potential of an intervention as it develops. To encourage reflexivity, the ToCs need to be co-created and flexible, and should be revisited as part of the formative, real-time evaluation processes.

Although ToCs are common in policy evaluation in other domains (for instance in development), they were seldom used in the evaluation of innovation policies. Our interaction with STI policy-makers suggested the importance of anchoring evaluation on a generic ToC that would help build a common rationale and theory-based justification for TIPs: a stylized view of the transformative change processes derived from transitions theory (Grin, Rotmans and Schot 2010; Markard, Raven and Truffer 2012). The resulting approach is innovative and provides an answer to the problem of assessing the downstream contributions and impact of current policy interventions, in a way that is coherent with the TIP approach.

In conclusion, we argue that, instead of acting as a perfunctory check at specified points of the project, TIP evaluation at its most effective should be a key element of the policy definition and implementation process, across different policy levels. As shown with the example of the

Forest Chemistry project, the level of effort required for this type of evaluation is of a different order of magnitude from the evaluative analysis that supports the archetypical approach to summative evaluation. Yet, the role and function of these types of evaluation are very different: in our formative approach evaluation is part and parcel of a different way of defining and implementing policy, through which the different stakeholders in a policy monitor and reassess policy results as they happen. It is a form of real-time monitoring⁷ embedded in the policy process.

Developing a new approach for transformative evaluation is a reflexive, participatory process that is interwoven with all stages of the policy process. As the policy evolves and adapts, so will the evaluation. Ultimately, TIPs will need a new evaluative strategy that must be co-created through the same actors who conduct the EPEs.

Conflict of interest statement. None declared.

Notes

1. The Transformative Innovation Policy Consortium (TIPC) consists of a set of research and policy partners, including innovation and research agencies from Finland, Sweden, Norway, South Africa, and Colombia. The Consortium's key objective is to examine and expand on current policy approaches to assist in solving urgent social, environmental, and economic issues. TIPC aims to shape and deliver a new transformative innovation policy framework. During the process, all participants are positioned as active co-researchers and policy co-designers.
2. Different outcome-focused evaluation approaches define 'outcomes' in different ways. For instance, Belcher, Davel and Claus (2020) consider them only in terms of behaviour change, 'conceptualized (and emphasized) as changes in knowledge, attitudes, skills, and/or relationships (KASR)'. Note that distinct from outputs: the direct results of an activity supported by the intervention. Such outputs are directly caused by the intervention, but they stop short of representing a change in behaviour, relationships, and actions.
3. Such activities are the direct consequence of the inputs that an intervention brings to bear. ToC practitioners often use the term 'structure' to refer to the wider set of resources that can be called upon to achieve project goals (Cohen and Franco 1992). Structural elements include the budget and the human resources available to support a programme, its planning, method and operational principles, and the norms under which the intervention operates.
4. The possibility of reviewing and changing the ToC is a natural consequence of our emphasis on second-order learning as one of the objectives of the evaluation process. Second-order learning implies the revision of original assumptions, values, and objectives and, therefore, a possible redrafting of the original ToC.
5. The distinction between outputs, outcomes, and impacts is very common in evaluation, although the specific definitions can vary somewhat. For instance, Boekholt et al. (2014) define outputs as the direct results generated by the intervention, outcomes as the immediate benefits for the beneficiaries of the intervention, and use the term 'impact' to refer to the wider effects on society.
6. The internal documents studied were:
Andersson, E. (ed.) (2014) Skogskemi—Gasification Platform. Sub project report to the Skogskemi project. The Skogskemi Project, Örnsköldsvik, Sweden: SP Processum AB. Available at: <https://www.processum.se/images/dokument/Ovrigt/>

Gasification_platform.pdf.

Joelsson, J. M., Engström, C., and Heuts, L. (2015). From green forest to green commodity chemicals: evaluating the potential for large scale production in Sweden for three value chains (Final report from Skogskemi: a project with the programme Challenge-Driven Innovation). Vinnova report: VINNOVA. Available at: <https://www.vinnova.se/publikationer/from-green-forest-to-green-commodity-chemicals/>.

Joelsson, J. M., and Mossberg, J. (eds) (2014). Skogskemi—Systems Analysis. Sub project report to the Skogskemi project. The Skogskemi Project, Örnsköldsvik, Sweden: SP Processum AB. Available at: https://www.processum.se/images/dokument/Ovrigt/Systems_analysis.pdf.

Johansson, E., and Pettersson, L. E. (eds) (2014). Skogskemi—Olefins Value Chain. Sub project report to the Skogskemi project. The Skogskemi Project, Örnsköldsvik, Sweden: SP Processum AB. Available at: <http://www.processum.se/images/dokument/Ovrigt/Olefins.pdf>

Jönsson, L. J. (ed.) (2014) Skogskemi—Sugar Platform. Sub project report to the Skogskemi project. The Skogskemi Project, Örnsköldsvik, Sweden: SP Processum AB. Available at: https://www.processum.se/images/dokument/Ovrigt/Sugar_platform.pdf.

Vinnova (2017) Challenge-Driven Innovation: Societal Challenges as Opportunities for Growth. Available at: https://www.vinnova.se/contentassets/d3fa44cef3c2408eb3d2a2b3e6da3f77/vi_17_06t.pdf.

7. Another example of real-time 'impact assessment' can be found in Joly, Matt and Robinson (2019).

Acknowledgements

The authors wish to thank the members of the Transformative Innovation Policy Consortium research and support team and the staff members of Vinnova involved in the case study discussed in this article for the insights they have shared on numerous occasions. We are also very grateful to two anonymous referees for their detailed and helpful comments. Support for the completion of this article was also received through the Spanish National Research Council (CSIC) Project 201910E044 and Vinnova (Reference Number: 2018-04975).

References

- Allmér, C. (2017) *Expect the Unexpected. Differences in Vision and Expectation in the Developing Bioeconomy*. Master's Programme in Innovation and Spatial Dynamics, Lund University [Online]. <<http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=8917414&fileId=8917415>> accessed 30 Dec 2019.
- Arkesteijn, M., van Mierlo, B., and Leeuwis, C. (2015) 'The Need for Reflexive Evaluation Approaches in Development Cooperation', *Evaluation*, 21: 99–115.
- Boekholt, P., Arnold, E., Giarracca, F., and Ploeg, M. (2014) Evaluation Reference Model. For TAFT IE's Taskforce Benchmarking Impact, Effectiveness and Efficiency of Innovation Instruments. Brighton: Technopolis Group. <https://www.technopolisgroup.com/wp-content/uploads/2020/03/TAFTIE-Evaluation-Reference-Model.pdf>
- Belcher, B. M., Davel, R., and Claus, R. (2020) 'A Refined Method for Theory-Based Evaluation of the Societal Impacts of Research', *MethodsX*, 7: 100788.
- Bos, B., and Grin, J. (2008) "Doing" Reflexive Modernization in Pig Husbandry: The Hard Work of Changing the Course of a River', *Science, Technology, & Human Values*, 33: 480–507.

- Caffrey, L., and Munro, E. (2017) 'A Systems Approach to Policy Evaluation', *Evaluation*, 23: 463–78.
- Cohen, S., and Franco, R. (1992) *Rationalising Social Policy: Evaluation and Viability*. Santiago: CEPAL.
- Connell, J. P., and Kubisch, A. C. (1998) *Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects and Problems*. in Karen Fulbright-Anderson, Anne C. Kubisch, and James P. Connell (eds.), *New Approaches to Evaluating Community Initiatives Vol. II: Theory, Measurement, and Analysis*. Washington D.C.: The Aspen Institute.
- Daniels, C., Schot, J. W., Chataway, J., Ramirez, M., Steinmueller, E., and Kanger, L. (2020) 'Transformative Innovation Policy: Insights from Colombia, Finland, Norway, South Africa and Sweden', in Cele M. B. G., Luescher T. M., and Fadji A. W. (eds) *Innovation Policy at the Intersection. Global Debates and Local Experiences*. Cape Town: HSRC Press, 9–23.
- Earl, S., Carden, F., and Smutylo, T. (2001) *Outcome Mapping. Building Learning and Reflection into Development Programs*. Ottawa: International Development Research Centre.
- Eschenbach, R., Hafkesbrink, J. and Lütz, S. (1995) 'Evaluation of the Promotion of Pre-Competitive Research and Development Cooperation in the Case of Microsystems Technology (MST)', in Becher G. and Kuhlmann S. (eds) *Evaluation of Technology Policy Processes in Germany*, pp. 165–196. Netherlands: Springer-Science-Business Media, B.V.
- Fuenfschilling, L., Bauer, F., and Clemente, J. (2017) *Transformative Innovation Learning History: Sweden. Challenge-Driven Innovation: A New Avenue for Transformative Policy at Vinnova*. Transformative Innovation Policy Consortium. <https://www.tipconsortium.net/wp-content/uploads/2019/05/TIPC_Sweden_InnovationHistory_CDI_Final-4.pdf> accessed 28 Jun 2020.
- Geels, F. W. (2002) 'Technological Transitions as Evolutionary Reconfiguration Processes: A Multi-Level Perspective and a Case-Study', *Research Policy*, 31: 1257–74.
- Geels, F. W. (2010) 'Ontologies, Socio-Technical Transitions (to Sustainability), and the Multi-Level Perspective', *Research Policy*, 39: 495–510.
- Geels, F. W., and Schot, J. (2007) 'Typology of Sociotechnical Transitions Pathways', *Research Policy*, 36: 399–417.
- Ghosh, B., Kivimaa, P., Ramirez, M., Schot, J., and Torrens, J. (2020) *Transformative Outcomes: Assessing and Reorienting Experimentation with Transformative Innovation Policy*. Transformative outcomes, TIPC Working Paper, TIPCWP 2020-02. Online access: <http://www.tipconsortium.net/wp-content/uploads/2020/07/Transformation-outcomes-TIPC-working-paper.pdf>
- Giddens, A. (1984) *The Constitution of Society. Outline of the Theory of Structuration*. Berkeley: University of California.
- Grin, J., Rotmans, J., and Schot, J. (2010) *Transitions to Sustainable Development: New Directions in the Study of Long Term Transformative Change*. New York and Abingdon: Routledge, 1–10.
- Hobday, M. (1988) 'Evaluating Collaborative R&D Programmes in Information Technology: The Case of the U.K. Alvey Programme', *Technovation*, 8: 271–98.
- Joly, P.-B., Matt, M., and Robinson, D. K. R. (2019) 'Research Impact Assessment: From Ex Post to Real-Time Assessment', *fteval Journal for Research and Technology Policy Evaluation*.
- Kemp, R., Schot, J., and Hoogma, R. (1998) 'Regime Shifts to Sustainability through Processes of Niche Formation: The Approach of Strategic Niche Management', *Technology Analysis & Strategic Management*, 10: 175–98.
- Kivimaa, P., Hildén, M., Huitema, D., Jordan, A., and Newig, J. (2017) 'Experiments in Climate Governance—A Systematic Review of Research on Energy and Built Environment Transitions', *Journal of Cleaner Production*, 169: 17–29.
- Kivimaa, P., and Kern, F. (2016) 'Creative Destruction or Mere Niche Support? Innovation Policy Mixes for Sustainability Transitions', *Research Policy*, 45: 205–17.
- Kuhlmann, S. (1999) 'Distributed intelligence: combining evaluation, foresight and technology assessment', *IPTS Report*, No. 40 (Special Issue: Evaluation and Research Activities), pp. 16–22.
- Loorbach, D., Frantzeskaki, N., and Avelino, F. (2017) 'Sustainability Transitions Research: Transforming Science and Practice for Societal Change', *Annual Review of Environment and Resources*, 42: 599–626.
- Luederitz, C., Schäpke, N., Wiek, A., Lang, D. J., Bergmann, M., Bos, J. J., Burch, S., Davies, A., Evans, J., König, A., Farrelly, M. A., Forrest, N., Frantzeskaki, N., Gibson, R. B., Kay, B., Loorbach, D., McCormick, K., Parodi, O., Rauschmayer, F., Schneidewind, U., Stauffacher, M., Stelzer, F., Trencher, G., Venjakob, J., Vergragt, P. J., von Wehrden, H., and Westley, F. R. (2017) 'Learning through Evaluation—A Tentative Evaluative Scheme for Sustainability Transition Experiments', *Journal of Cleaner Production*, 169: 61–76.
- Magro, E., and Wilson, J. (2013) 'Complex Innovation Policy Systems: Towards an Evaluation Mix', *Research Policy*, 42: 1647–56.
- Markard, J., Raven, R., and Truffer, B. (2012) 'Sustainability Transitions: An Emerging Field of Research and Its Prospects', *Research Policy*, 41: 955–67.
- Mayne, J. (2011) 'Contribution Analysis: Addressing Cause and Effect', in Forss K., Marra M., and Schwartz R. (eds) *Evaluating the Complex*, pp. 53–96. Piscataway, NJ: Transaction Publishers.
- Molas-Gallart, J., and Davies, A. (2006) 'Program Theories and Evaluation Practice: The Experience of Science, Technology and Innovation Policies in Europe', *American Journal of Evaluation*, 27: 64–82.
- OECD (2016) *OECD Reviews of Innovation Policy: Sweden 2016*. Paris: OECD Publishing. <https://www.tipconsortium.net/wp-content/uploads/2019/05/TIPC_Sweden_InnovationHistory_CDI_Final-4.pdf> accessed 28 Jun 2020.
- Pawson, R., and Tilley, N. (1997) *Realistic Evaluation*. London: SAGE Publications.
- Rip, A., and Kemp, R. (1998) 'Technological Change', in Rayner S. and Malone E. L. (eds) *Human Choice and Climate Change. Vol. II, Resources and Technology*, pp. 327–399. Columbus, OH: Battelle Press.
- Rogers, P. J. (2008) 'Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions', *Evaluation*, 14: 29–48.
- Rogge, K. S., Pfluger, B., and Geels, F. W. (2020) 'Transformative Policy Mixes in Socio-Technical Scenarios: The Case of the Low-Carbon Transition of the German Electricity System (2010–2050)', *Technological Forecasting and Social Change*, 151: 119259.
- Schot, J., Kivimaa, P., and Torrens, J. (2019) *Transforming Experimentation. Experimental Policy Engagement and Their Transformative Outcomes*. Transformative Innovation Policy Consortium. <http://www.tipconsortium.net/wp-content/uploads/2019/07/Transforming-Experimentation.pdf>.
- Schot, J., and Steinmueller, W. E. (2018) 'Three Frames for Innovation Policy: R&D, Systems of Innovation and Transformative Change', *Research Policy*, 47: 1554–67.
- Smallman, M. (2018) 'Citizen Science and Responsible Research and Innovation', in Hecker S., Haklay M., Bowser A., Makuch Z., Vogel J., and Bonn A. (eds) *Citizen Science: Innovation in Open Science, Society and Policy*. London: UCL Press, 241–253.
- Smith, A., Voß, J.-P., and Grin, J. (2010) 'Innovation Studies and Sustainability Transitions: The Allure of the Multi-Level Perspective and Its Challenges', *Research Policy*, 39: 435–48.
- Smutylo, T. (2001) 'Crouching Impact, Hidden Attribution: Overcoming Threats to Learning in Development Programs', *Block Island Workshop on Across Portfolio Learning*, 22–24 May.
- Stame, N. (2004) 'Theory-Based Evaluation and Types of Complexity', *Evaluation*, 10: 58–76.
- Taanman, M. (2014) 'Looking for Transitions. Monitoring Approach for Sustainable Transition Programmes', Doctor, Erasmus Universiteit Rotterdam, Rotterdam.
- Torrens, J., Johnstone, P., and Schot, J. W. (2018) 'Unpacking the Formation of Favourable Environments for Urban Transformation. The Case of the Bristol Energy Scene', *Sustainability*, 10: 879.
- Turnheim, B., Berkhout, F., Geels, F., Hof, A., McMeekin, A., Nykvist, B., and van Vuuren, D. (2015) 'Evaluating Sustainability Transitions Pathways: Bridging Analytical Approaches to Address Governance Challenges', *Global Environmental Change*, 35: 239–53.
- van Es, M., Guijt, I., and Vogel, I. (2015) *Theory of Change Thinking in Practice. A Stepwise Approach*. The Hague: HIVOS.

- van Mierlo, B., Arkesteijn, M., and Leeuwis, C. (2010) 'Enhancing the Reflexivity of System Innovation Projects with System Analyses', *American Journal of Evaluation*, 31: 143–61.
- van Mierlo, B., Regeer, B., Vam Amstel, M., Arkesteijn, M., Beekman, V., Bunders, J., de Cock Buning, T., Hoes, A.-C., and Leeuwis, C. (2010) *Reflexive Monitoring in Action. A Guide for Monitoring System Innovation Projects*. Boxtel, The Netherlands: Boxpress.
- Weber, K. M., and 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'', *Research Policy*, 41: 1037–47.
- Weiss, C. H. (1998) *Evaluation. Methods for Studying Programmes and Policies*. London: Prentice Hall.
- Wilson-Grau, R., and Britt, H. (2013) *Outcome Harvesting*. Cairo: Ford Foundation Middle East and North Africa Office.