


Evaluating transformative innovation policy in a formative way: Insights from Vinnova's food mission experiment

Alejandra Boni ^{1,*}, Diana Velasco^{2,*}, Jordi Molas-Gallart², Johan Schot³

¹INGENIO (CSIC-UPV), Universitat Politècnica de València, Camí de Vera, 46022 València, Spain

²Consejo Superior de Investigaciones Científicas, INGENIO (CSIC-UPV), Universitat Politècnica de València, Camí de Vera, s/n, ES-46022 València, Spain

³Centre for Global Challenges, Utrecht University, Janskerkhof 2-3a, 3512 BK Utrecht, The Netherlands

*Corresponding authors. Emails: aboni@ingenio.upv.es (A.B.) and diana.velasco@csic.es (D.V.)

Abstract

This article presents the insights from an evaluation of a transformative policy experiment, in the food domain, conducted at the Swedish Agency of Innovation (Vinnova). To be consistent with the principles and objectives of these policies, it was necessary to implement a formative evaluation approach developed in the Transformative Innovation Policy Consortium (TIPC), that is supportive of experimental policies designed for highly complex and uncertain environments. The article presents the main characteristics of this approach and its differences and similarities with other proposals to assess transformative innovation policies. Next, the article describes how the evaluation was conducted in Vinnova during a 2-year engagement. Subsequently, the main insights gleaned from the process are highlighted in terms of (1) understanding the expectations of actors; (2) what has been learned by doing, from a technical and organizational point of view; and (3) the constraints that emerged during the engagement. It is then argued that the TIPC- formative evaluation approach challenges the role of evaluators, stretching them, and requires different attitudes from those that are common in evaluation practice.

Keywords: transformative innovation; formative evaluation; policy evaluation; Vinnova; TIPC; evaluators' role.

1. Introduction

Disruption and shifts in current socio-ecological and technical systems due to massive landscape pressures such as climate change, economic crises and pandemics seem to be accelerating in recent decades. These macro tendencies have raised society's awareness, provoking changes in perceptions and behaviours that are activating different social (and generative) fields (Scharmer 2018). Adaptive responses to local and global challenges have been emerging and building resilience across interconnected systems where diversity and redundancy, connectivity, and polycentric governance in systems act as good principles to promote a just transition to a viable world (Biggs, Schlüter and Schoon 2015).

In this context, innovation policy fields are increasingly turning to policies focused on ambitious societal goals (Kuhlmann and Rip 2018). Design, implementation and evaluation policy strategies are taking a systemic approach by identifying and understanding a broad range of factors that influence the possibilities for societally desirable innovations to become successful (Hekkert et al. 2020; Janssen, Bergek and Wesseling 2022; Bergek and Haddad 2022). While the urgency and legitimacy of such policies mainly stem from societal developments, the 'transformative' innovation policy frame (Weber and Rohracher 2012; Schot and Steinmuller 2018), drawing inspiration from the literature on sustainability transitions, is the latest proposal. We emphasize, in agreement with Janssen, Bergek and Wesseling (2022), that this transformative innovation policy brings with it responsibilities regarding delivering concrete advice on how to evaluate systemic innovation and transition programmes.

We argue that the evaluation of systemic innovation and transitions programmes requires new approaches to evaluation and, consequently, new competencies for evaluators. Ofir and Rugg (2021: 48) pointed out that there is a sense of urgency for the evaluation field 'to revisit, redesign and reconfigure evaluation theories and practices to support the large-scale, transformative changes our societies and ecosystems need'. An evaluation that supports and accelerates system transformation requires a different approach (Patton 2021). Since policy environments are complex and very diverse, and policy objectives are ambitious and radically innovative, there is a high degree of uncertainty about the ways in which an intervention will develop and the effects it will have. In such scenarios policy becomes experimental, requiring evaluation practices that support the development of the experiment rather than judging its final results against a set of well-defined criteria derived from precise expectations. Conventional results-based management, linear theories of change, and measuring and attributing predetermined impacts are no longer (if they ever were) suited to a world defined by uncertainties and interdependencies (Ofir and Rugg 2021).

Similar concerns have been raised by the transition community. Janssen (2019) highlights the difficulties of tracing causal mechanisms in complex systems given that policies usually generate unanticipated system dynamics. Luederitz et al. (2017) and van Mierlo, Arkesteijn and Leeuwis (2010) have remarked on the importance of (reflexive) learning in monitoring and evaluation practices as a way of capturing the emergent dynamics of systemic challenges; while Haddad et al. (2019) have emphasized that evaluators not only assess programmes against a set of relevant impacts and system-level

transformative outcomes, but they also analyse interactions between instruments and stakeholders.

The Transformative Innovation Policy Consortium (TIPC) has developed a formative approach to the evaluation of Transformative Innovation Policies (TIPs) (Boni, Molas-Gallart and Giachi 2019; Molas-Gallart et al. 2021) that involves assessing the changes associated with or leading to socio-technical transitions (Rip and Kemp 1998; Geels 2002; Geels and Schot 2007). Evaluation becomes a real-time reflexive practice that contributes to the design and implementation of TIPs, whilst at the same time also contributing to their capacities to do so. Inclusivity via participatory methods is a core feature of this type of evaluation. Policymakers and the actors involved in the design and implementation are in control of the evaluation process, with external evaluation experts mainly acting as facilitators. The evaluation aims to facilitate participation and open debate, acknowledging and channelling power conflicts and differences in interests and perceptions.

By exploring a particular experimental policy engagement developed within the TIPC, based on an evaluation of the process of designing a mission-oriented innovation policy, we present key lessons that nurture the specific formative approach to evaluation for TIP along with its constraints. As stated above, our approach to evaluation is a real-time reflexive practice, similar in purpose to what has been developed by the ASIRPA team (Matt et al. 2023). Consequently, although our evaluation influenced the design of a mission-oriented policy, its contribution went beyond the design phase. As we describe in Section 3, we developed a Monitoring and Learning Plan aimed at testing the implementation of the food mission. Additionally, we had different meetings with Vinnova staff to foster reflection on the main learnings derived from the formative evaluation. This experience generated insights into the roles and attitudes that evaluators should develop to carry out this type of evaluation.

The structure of this paper is as follows: in Section 2, we provide an overview of the TIPC formative evaluation approach and compare it with other proposals for assessing TIPs; in Section 3, we describe the case study, including the different phases of the evaluation, the methods used, and the outputs achieved; in Section 4, we present our findings and insights, focusing on expectations, technical and organizational learnings, constraints that emerged during the engagement, and the implications of our approach for evaluators' attitudes; and finally, in Section 5, we provide some final conclusions.

2. TIPC's formative approach to the evaluation of Transformative Innovation Policies

As mentioned above, TIPC has developed a formative approach to the evaluation of TIPs taking as a heuristic the Multi-Level Perspective (MLP) (Rip and Kemp 1998; Geels 2002; Geels and Schot 2007). The evaluation process is based on reflexive monitoring practices (van Mierlo et al. 2010) for fostering socio-technical systems transformation. In this sense, it is a theory-based approach to evaluation that links general evaluation concepts and approaches with the sustainability transitions conceptualization of outcomes described in current innovation and transitions studies literature (Haddad and Bergék 2023). In the TIPC approach, the MLP is used as a generic transformative Theory of Change (ToC) which connects

with specific ToCs tailored to individual interventions. The evaluation is formative since it provides participants with tools to widen their perception of the role and scope of innovation policies and their contribution towards systems change by defining and working towards processes and outcomes directed towards transformation (Janssen, Bergék and Wesseling 2022).

Moreover, the TIPC approach uses participatory methods where different stakeholders define themselves the assessment strategy for transformational policies. This creates increasing demands on evaluators to mobilize and empower relevant stakeholders from different fields, sectors, and policy levels; coordinate and align potentially divergent perspectives on problems and possible solutions; and manage conflicts of interests (Weber and Rohrer 2012). Table 1 indicates the main characteristics of TIPC formative evaluation.

2.1 Multi-level perspective and the 12 transformative outcomes

A transition is a change in socio-technical systems, which are stable and dominant configurations of markets, user preferences, industries and industrial strategies, policies, cultural practices and technologies providing specific ways of serving a particular basic need or societal function (Smith, Voß and Grin 2010). Socio-technical systems are constructed and maintained by actors who are guided by a set of formal and informal rules, which together form a regime. These systems change when shifts occur at different levels, as theorized by the MLP, specifically when dynamic and complex interactions emerge between actors advancing new solutions and ideas in niches, deviating and pushing the dominant rules. Although regimes seek to remain stable, they get also pressured by long-term exogenous trends and shocks, or landscape, breaking the dominant configuration of the system and giving space to the niches to replace the dominant regime (Rip and Kemp 1998; Geels 2002; Geels et al. 2016). This is a simplification of complex and long-term change processes, but it is helpful in order to orientate transformation pathways. Significantly, the MLP provides analytic lenses to determine systems transitions when three conditions are met: (1) a regime is destabilized, (2) niches provide strong alternatives at scale, and (3) landscape trends and shocks are perceived by regime and niche actors as a window of opportunity for a transition.

Building upon the MLP, TIPC developed and experimented with a formative evaluation methodology focused on 12 Transformative Outcomes (TOs) that together address the three conditions referred to above. The TOs are divided into three macro-processes: (1) building and nurturing niches, (2) expanding and mainstreaming niches, and (3) unlocking and opening up regimes (see Table 2).

The TOs refer to a goal—for example, a better network—but also to the related process (networking) developed over time. As a result of the formative evaluation process, these outcomes can be measured at a particular moment in time, but with a process-oriented approach to help understand how the current activities are contributing (or not) to the transformation purposes.

The TOs guide the definition of signs of change in a socio-technical system, providing distinctive conceptual and practical insights to measure systems transformation. In the TIPC proposal, the TOs are not used to measure impact; they are signs of progression along the change process (Wise and

Table 1. Key principles of the formative evaluation approach (based in [Molas-Gallart et al. 2021](#): 435–436)

Adopt a formative approach to evaluation	An evaluation 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.
Integrate evaluation with policy design and implementation	Evaluation as part of the transformative policy and, therefore, coherent with the stated research and innovation policy objectives (directionality, societal goals and system impact). Evaluation as a strategic part of the design and implementation process of TIPs.
The evaluation process should be inclusive and participatory	The inclusivity characterizing TIPs should also be present in the evaluation process. Participants in TIPs should join in their evaluation, with external evaluation experts mainly acting as facilitators; for instance, paying 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 interests and perceptions.
Use a mix of methods and techniques	Rather than being driven by formalized standard protocols, evaluation needs to be adaptable and flexible, selecting different methods and techniques according to the policy context and its transformative nature.
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 between them that will facilitate scaling up. Finally, several programmes can combine with other policies in policy mixes.
Use a flexible Theory of Change (ToC)	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 activities triggered by an intervention with its results. Our ToCs will be flexible, implying that they should not be understood as a fixed causal chain; rather, they can be revisited and redefined as a result of the formative evaluation process. The ToCs will be used to foster learning and reflexivity among participants and to help assess if the policy is contributing towards advancing its objectives.

[Arnold 2022](#)). At the same time, the TOs are crucial to influence the directionality of the intervention through a reflexive learning approach. The intervention logic needs to be continually revisited and adapted based on an iterative process of looking back (reflexivity) and looking forward ([Wise and Arnold 2022](#)). This understanding of system transformation is different from what is proposed by [Janssen \(2019\)](#) and [Bergek and Haddad \(2022\)](#). They propose to use technological innovation systems literature to assess impact of how a transformative policy affects functions favouring the creation and diffusion of new economic activities. Examples of functions that can be analysed are development and diffusion of knowledge, entrepreneurial experimentation, formation of markets, guidance of actors' search processes, mobilization of resources, legitimation of technologies and actors, and the development of positive external economies. By analysing these functions, analysts or evaluators can identify weaknesses in the system and assess the impact of policies on each process.

The TIPC formative evaluation approach can also be defined as an initiative-based learning approach to analyse transitions ([Turnheim et al. 2015](#)). This approach engages with concrete interventions, *involving diverse social actors with the aim of fostering innovation and upscaling innovative sustainability solutions* ([Turnheim et al. 2015](#): 244). As we describe in Section 3, the experiment conducted between TIPC

researchers and Vinnova can be considered as an initiative-based learning and the presentation of our findings is guided by the four elements proposed by [Turnheim et al. \(2015\)](#): understanding expectations; learning by doing; environment and resource constrains; and responding to emergent processes.

However, our approach to evaluation has limitations and is not able to identify long-term impact. In fact, in contrast to other approaches ([Luederitz et al. 2017](#); [Williams and Robinson 2020](#); [Haddad and Bergek 2023](#)), TIPC's formative evaluation shifts its focus from measuring impacts to support system change processes. The evaluation process deliberately embeds directionality towards sustainability, identifying and pursuing outcomes that are considered to be conducive to systemic change. This proactive stance enables the design of strategic interventions that drive transformative change, fostering a continuous learning process involving key stakeholders. Evaluation thus becomes proactive supporting implementation and intervention redesign, instead of taking a retrospective, reactive stance. ([Molas-Gallart et al. 2021](#)). The accomplishment of TOs provides directionality and strategic orientation towards transformation ([Wise and Arnold 2022](#)).

2.2 Flexible ToC

In the experimental policy engagements developed in the first phase of TIPC (2018–22), policy initiatives at different levels

Table 2. Twelve types of transformative outcomes, adapted from Schot et al. (2019) and Ghosh et al. (2021)

Niche building and nurturing	
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	The first order (optimizing existing behaviour) and second order (changes in frames and assumptions) in or across several system dimensions (science, technology, innovation; markets; culture & 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 The 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 mainstreaming	
Upscaling	Increased adoption of new practices and rules, diffusion of technologies, the 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, and rules across niches and system dimensions The emergence of system and niche intermediaries.
Institutionalization (formal and informal rules)	Developing standard definitions, narratives, regulations, and preferred types of behaviors, 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 The emergence of new policy assumptions
Empowering niche-regime interactions	Creation of formal and informal linkages between niche and regime actors The emergence of intermediaries 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

developed change pathways, guided by the TOs, through a specific ToC built collaboratively by the policy stakeholders involved in the process. Our approach shares commonalities with the adaptive and systemic approaches to evaluation (Patton 2011, 2020) but differs in its grounded theoretical base (the MLP and the 12 TOs), informing how system change is expected to happen. Unlike most theory-oriented approaches, in which the evaluator builds the programme theory interpreting the expectations of the actors involved in the intervention (Stame 2004), we actively use TIP theory to co-produce, with the policy actors, a ToC that focuses on TOs (Molas-Gallart et al. 2021).

Policy interventions focus on unlocking transformation by selecting groups of TOs, rather than implying that a single intervention should comprehensively cover all of the outcome types. Ideally, a nested approach to transformation is developed through a whole set of interventions, starting at a national or regional policy level providing a framing ToC that is then developed by programmes and specific interventions that have their own ToCs.

Our approach shares commonalities and differences with that presented by Haddad and Bergek (2023). Both share the

importance of grounding the evaluation in transition theory; moreover, both approaches are directed at defining targeted socio-technical configurations, actors and institutions and use a ToC as a way of representing the intervention. A main difference is the use of the context, mechanism and outcomes analysis proposed by realist evaluation. This step involves assessing ‘behavioural additionality’ which refers to the extent a policy has influenced or contributed to intermediate transformative processes at the system level. The purpose of this step is to understand how the policy has impacted the system and to identify any potential transformative processes that have resulted from the policy’s implementation. By assessing behavioural additionality, evaluators can gain insight into how the policy is working and whether it is contributing to broader system-level changes (Haddad and Bergek 2023).

In contrast, TIPC’s approach develops a flexible ToC using as a base some of the techniques developed by the Dutch consultancy HIVOS, which stress the use of action-research tools for the definition and redefinition of ToCs (van Es, Guijt and Vogel 2015). In this approach, ToCs are typically developed collaboratively by policy stakeholders and evaluation experts who facilitate the process. It starts by identifying the changes that an

intervention aims to achieve. These changes are often framed as policy goals. From there, stakeholders work backwards to identify the processes that are likely to lead to these changes, and how the intervention will trigger these processes. This involves developing an expected process linking the inputs, activities and TOs. The process helps to depict possible evolution pathways with multiple causalities and multiple expected effects. That is why it is crucial to explicitly identify the assumptions made by the participants throughout the process. In this line, ToCs are not fixed causal chains, but flexible frameworks that can be revisited and redefined as needed throughout the formative evaluation process. This flexibility is based on the feedback produced by the ongoing learning and reflexivity among participants and can help to assess whether the policy is moving towards its objectives. Overall, ToCs are an important tool for fostering collaboration, learning and reflexivity among policy stakeholders and evaluation experts, and can help to ensure that interventions are effective in achieving their intended outcomes (Molas-Gallart et al. 2021).

Finally, both the TIPC formative approach to evaluation and the ideas presented by Patton (2023) highlight the changing role of evaluators in the context of system transformation. Evaluators are no longer just external observers of change, but active participants in the change process, requiring them to have a deeper understanding of the context and a range of competencies to effectively engage with stakeholders and contribute to the development and implementation of interventions. The TIPC framework emphasizes the importance of co-production and collaboration in the evaluation process, with evaluators being flexible, reflexive and willing to engage with stakeholders. Similarly, Patton highlights the need to move beyond traditional evaluation methods and the willingness to take risks, experiment and innovate in the approach to evaluation. In summary, both the TIPC formative approach to evaluation and the ideas presented by Patton recognize the importance of the changing role of evaluators in system transformation and emphasize the need for evaluators to develop the competencies and attitudes required to effectively engage with stakeholders and contribute to the development and implementation of interventions.

However, although the TIPC-formative evaluation and Patton's approach to evaluation (2011) share the same concern on the evaluator's role, there is a main difference between the two approaches. TIPC uses a specific Theory of Change inspired by transition theory, while Patton's developmental evaluation is not supported by any specific theory. In TIPC's approach, the ToC is a central component, while Patton's proposal focusses on dynamics, permeable boundaries, interdependencies and emergent interconnections which are difficult to capture in a ToC. One common feature is the relevance of learning, although Patton's proposal goes further and contends that learning is central for accountability.

3. The case of food policy at Sweden's Innovation Agency (Vinnova)

In 2019, Vinnova started an experimental process to design mission-oriented policies for mobility and food challenges guided by the Sustainable Development Goals (SDGs). In December 2019, the agency engaged with TIPC researchers to apply the TIPC approach to evaluation (Boni, Molas-Gallart and Giachi 2019) with two aims: (1) explore the practical use of the TOs and the construction of flexible ToCs to stay

focused on transformation, inducing second-order learning, reflexivity and anticipation along the process; and (2) enhance the transformative potential of Vinnova's approach to mission-oriented innovation policy. These goals are coherent with the aims of TIPC work in-country: to create cross-learning processes for the TIPC members; provide tools and pathways to address global societal challenges in situ; and develop and test methodologies, theories and tools as a continuous co-creating process tailored to the country's needs.

Vinnova's leadership recognized that this type of work implied a different approach to policymaking, by experimenting with strategic design methods that integrate upfront complexity and uncertainty in the policy instruments design (Hill 2022), and also required a different approach to evaluation to avoid transformational failure (Weber and Rohracher 2012). The whole engagement was focused on the process developed by the Vinnova team part of the engagement and the TIPC researchers. As part of the research process, TIPC researchers also engaged with other policymakers from other government organizations such as the food and procurement agency, a design agency supporting the process, and stakeholders that were part of the mission design. In this sense, what was evaluated was the transformative capacity of Vinnova's food mission-oriented innovation policy, by applying the TIP framework to define systems change, pathways to advance towards that change and continuous monitoring to assess directionality and strategic reorientation (Wise and Arnold 2022).

Furthermore, the engagement sparked a deeper reflection on challenge-led innovation policy expressed through missions. Mission-oriented innovation policy (MOIP) involves setting well-defined goals or missions to be accomplished within a specific timeframe (Mazzucato 2018; Janssen et al. 2021). The TIP framework challenges targeted mission-oriented policies focussed on large science and technology support programmes (Diercks, Larsen and Steward 2019) by focusing on systems change beyond industrial, economic or technological advancements. Designing, implementing, and evaluating transformative missions necessitates an experimental approach that nurtures, strengthens, and scales niches while at the same time destabilizes systems regimes. It requires an open-ended and long-term perspective, recognizing that socio-technical transitions are complex, dynamic, and fuzzy processes (Geels 2022). To effectively navigate this complexity, the evaluation approach embraces uncertainty by incorporating a reflexive layer that facilitates the reframing of desirable directions. This critical approach was consistently applied by both teams throughout a four-phase process.

3.1 Phase 1: Developing a preliminary ToC and selecting the TOs (February–July 2020)

For 6 months, the teams from Vinnova (an analyst, the head of the strategic design area, two programme managers) and TIPC (four researchers and one communications manager) held five collaborative sessions to develop a mutual understanding on evaluation, define the object of the evaluation and set the bases on which to develop a ToC. At the time, the Vinnova team was advancing on the definition of two missions, one focused on mobility and the other on food. Since the mobility mission was more advanced in its process, and after several discussions, the Vinnova and TIPC teams decided to focus on the food mission. The teams saw an opportunity to start the evaluation process from the mission-design phase,

reaping benefits from learning and reflexivity throughout the process.

The first two workshops provided the bases to understand the methodologies, principles, and experiment scope. The next three workshops were dedicated to building a common language, further exploring how TIPC's evaluation could strengthen the mission design and transformational purposes, and to monitoring the learning process of the participants. The initial aim to build a full ToC for the food mission was discussed given the tensions between a fully explorative process of mission definition and deployment—where each phase is developed from the results of the previous one—and the prescriptive nature of ToCs, even if they are flexible and revisited during the process.

The differences in the approaches required the team to be open and explore together the best of both approaches to contribute to the ultimate aim, which was to develop a transformative mission-oriented innovation policy for the food socio-technical system. Consequently, the team made a deliberate choice to prioritize the identification of desired outcomes and associated assumptions, rather than delving into detailed definitions of outputs and activities. This decision was guided by two key considerations: Firstly, the mission was intentionally designed to have an open-ended nature, allowing for the emergence of diverse action pathways aligned with the desired outcomes. By focusing on outcomes rather than predefined outputs, the team aimed to foster flexibility and adaptability throughout the mission implementation. Secondly, the high-level nature of the policy framework called for a ToC that provided a broad guiding framework rather than prescriptive action pathways. This approach acknowledged that specific interventions and detailed action pathways would be developed through specific initiatives in connection with the mission's overarching objective.

In the third workshop, thirty outcomes were initially defined by the Vinnova team related to changes in local food production and distribution, local shops, shopkeeper skills, neighbourhoods and others. For each outcome, the main actors involved were also defined.

The fourth and fifth workshops resulted in a shorter list of six outcomes mainly related to opening and unlocking regimes (see Table 3). The Vinnova team decided to focus on this macro process in the first stage of the mission design and implementation, given their emphasis on creating a protected space to foster emerging collaborations among regime and niche actors towards the mission aim. Therefore, the outcomes were focused on developing a new way of working with a diversity of voices, showcasing major practice changes in incumbents, and creating bridges between government agencies, large and small companies, and consumers and social groups. Through further reflections and discussions with the research team, it became evident that directing resources towards nurturing and empowering niche initiatives was crucial in achieving the initial mission outcomes. This insight reinforced the understanding that the TOs should not be interpreted as having a strict sequential or hierarchical order. Rather, they serve as guiding principles that channel energy and resources towards different intervention points aimed at fostering systemic change. It is important to acknowledge that achieving all 12 transformative outcomes in a single intervention is not realistic, given the long-term and complex nature of sustainability transitions. However, recognizing the interdependencies among the outcomes is vital to unlock

transformation. Thus, the focus should be on pursuing multiple interventions that collectively contribute to advancing the mission's objectives.

Regarding the practical use of the TOs, one of the Vinnova team members gave the following account:

I think the 12 transformative outcomes are a very fruitful tool for distinguishing and monitoring the systemic aspects of the mission-oriented approach. The challenge lies in connecting the actual operational principles of the mission-oriented work with these outcomes. (Written communication as part of the preparation phases for the workshop)

Beside the joint workshops, a closer and more fluid relationship was built with one of the Vinnova team members, who acted as an intermediary, helping in translating the theory-based formative evaluation methodology to the Vinnova context and the strategic design principles to the TIPC researchers. Having such an intermediary proved to be crucial in guiding the process and defining the best approach to advance in the engagement. The team meetings were co-designed and co-developed with the intermediary, and this was conducive to building trust between both teams and navigating expectations more effectively. The final part of each of the three workshops was dedicated to reflecting on the encounter, on how the participants felt, what worked and what did not. As a parallel strategy to reflect upon the engagement, the Vinnova team member who served as an intermediary and two TIPC team members used a diary in which they recorded their reflections on the learning process from the perspective of knowledge, attitudes and interactions (van Mierlo et al. 2010). The diary, called the *dynamic reflexive binnacle*,¹ was an effective tool for recording the process and reflecting on what type of learning was being developed.

3.2 Phase 2: Developing the final ToC (August–December 2020)

During the second half of 2020, the context for the formative evaluation work changed. Vinnova launched a new organizational configuration which included the creation of a food area within the agency. The changes resulted in Vinnova investing less time and fewer team members in the activity. The team was reduced to two people (the person who had acted as an intermediary and one of the project managers, who became head of the newly created food area), while other members of the original team were reallocated to other tasks. The enduring involvement of the new head of area signalled that the engagement continued to be seen as important, despite the lessened resources. The teams held four workshops during this period, along with one-to-one encounters and preparation of a regional (for European Nordic region) TIPC learning event. Work continued on reducing and refining the six previous outcomes (Table 3), focusing on the mid and long term. As a product of the deliberation process, the Vinnova team decided to focus on three outcomes that could be monitored in the short and medium term. The outcomes reflect a focus on destabilizing the food regime and the connection with niche actors as a strategy to further develop the transformation process. Furthermore, they should not be understood as having a strict sequential order or varying levels of complexity. It is essential to develop strategies concurrently

Table 3. Shorter list of outcomes of the preliminary ToC

Macro-process	Transformative outcome	Outcome
Opening up and unlocking regimes	Unlearning and deep learning in regimes	<ol style="list-style-type: none"> 1. Develop a new way of working within the state, new methods, practices, less linear and reactive, more circular, dialogue-driven and open, allowing for new voices other than the ‘usual suspects’. 2. Among incumbent policy and industry actors of the Swedish food policy system, there is still an overarching goal of economic growth—where sustainability comes second. By allowing these actors to recognize this actively and explore what it means to put sustainability first, the mission work can contribute to the unlearning processes of, not least, policymakers. 3. Changing behaviour with show and tell, through ‘champions’, to make a convincing case for behavioural change in the business-crucial parts of the organization. By questioning the assumption that there is a contradictory relationship between sustainability and economic growth, the regime actors can change their beliefs. By providing an alluring arena for thought exchange, we can get regime actors to converse around risk-sharing with niche actors from other parts of the system.
	Empowering niche-regime interactions	<ol style="list-style-type: none"> 4. Regime actors: providing the internal niche-actors with a platform, including the more business-strategic parts of the actors 5. Vinnova is seen as a more neutral party than most other agencies within the food policy segment (thanks to our historical (non-)role)—through the type of dialogue/workshop processes that the mission work has set up, a new kind of communications arena has occurred, which both incumbent and niche actors are willing to engage in.
	Changing perceptions of landscape pressures	<ol style="list-style-type: none"> 6. When a state agency (with a rather good reputation in the policy and industry system of Sweden) such as Vinnova dares to recognize the need for transformative change and a new way of approaching SDGs, incumbent and niche actors dare to think that a new order is possible—meaning both a real threat and a possibility. By standing up for other actors, and actively advocating a more radical approach towards sustainability, Vinnova can contribute to normalizing a shift from growth to sustainability.

across all three outcomes. These are the final outcomes depicted in the ToC (Figure 1):

- O1. New understanding of the food system in terms of agents and configuration (strengthening regime-niche interactions, unlearning and deep learning in regimes, and learning in niches).
- O2. Policy and business actors within the food system change their perceptions and behaviours with regard to the relationship between the three dimensions of sustainability (unlearning and deep learning in regimes, changing perceptions of landscape pressures).

- O3. Government agencies work together to produce alternatives for new sustainable Retail systems (dealigning and destabilizing regimes).

3.3 Phase 3: Monitoring, evaluation and learning plan. Broadening learning (January 2021–December 2021)

During 2021, Vinnova’s re-organization took full effect, affecting the resources available to implement further activities of the food mission and therefore its evaluation. The work focused on defining a monitoring, evaluation and learning

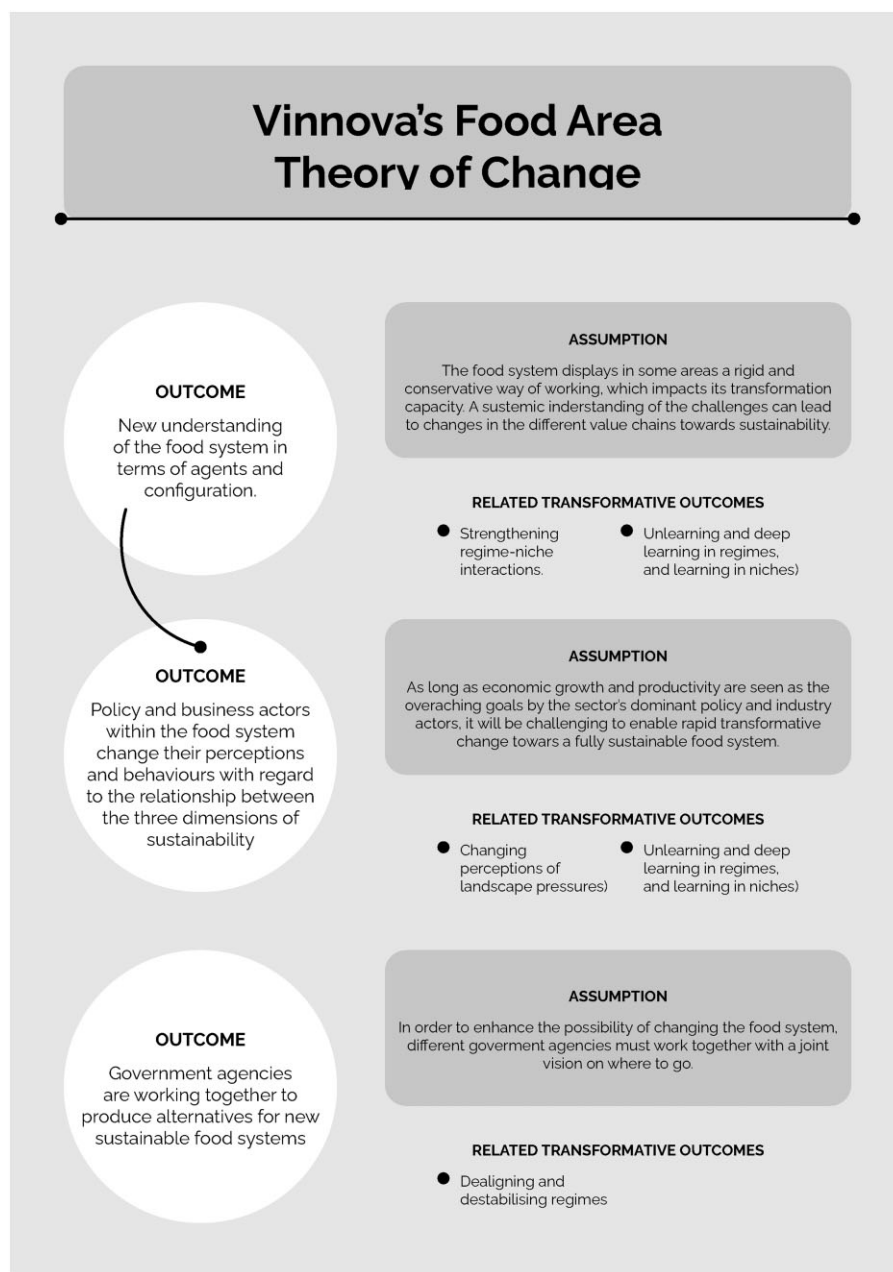


Figure 1. Vinnova's food area Theory of Change.

(MEL) plan based on the outcomes defined. New members joined from the Vinnova side, including a division director and a project manager, who was involved in the first phase of the process. The team defined qualitative and quantitative indicators to monitor the process for each of the outcomes and the linked assumptions. The TIPC researchers used indicative evaluation questions, questions related to the TOs, definition of indicators and the methods for data collection (a detailed table with all this information is presented in [Supplementary Annex S1](#)). The teams then decided to test the MEL plan in the area of the food mission that was most developed, related to the transformation of the public school meals served nationwide. This process included the analysis of internal data such as reports, digital boards, slides, webpages and spreadsheets with participants of different engagements. This analysis was complemented by five interviews with

policymakers from the national food and procurement agencies, an entrepreneur, other policymakers within Vinnova and two practitioners who were part of a design agency hired by Vinnova to support the participative workshops with different actors in the food national ecosystem. The results of this analysis were presented in a workshop with the Vinnova team involved in the process and project managers and analysts from other areas as a strategy to embed the learning process within the agency.

The process concluded with implications to develop further phases that include other outcomes related with niches within the mission ToC. Moreover, in this process, the group reflected on how transitions are processes featuring controversy and consensus at different levels and stages. Therefore, the following steps might also involve controversy rather than prioritizing consensus among participants. The workshop

also provided feedback for the management structure, capacities and capabilities within Vinnova, by demonstrating the necessity of adapting routines to agile design processes and formative evaluation tools that need to be embedded within the policy strategy cycle (Howlett and Ramesh 1995). It was also clear that the food area and its experimental methodologies required further shielding from the higher managerial structure.

Furthermore, the work expanded beyond the food area, thus reconnecting with the original objective of disseminating learning to the whole organization. Results from this phase highlighted the importance of gathering, as in the first phase, broader areas within Vinnova and inviting key allies from other organizations before the process has ended.

3.4 Phase 4: Closing down on main learnings, opening up on new endeavours (January–March 2022)

The experimental policy engagement between both teams was finalized in March 2022 with two in-person workshops in Vinnova. The workshops were co-designed with the Vinnova team, including the head of the systems transformation and a senior analyst. The first workshop aimed to expand critical learnings with other areas within Vinnova and strategic partners for the strategic innovation programmes (SIP) 2.0, such as the Swedish Energy Agency, Technopolis, the Government Offices, the Swedish Research Council for Sustainable Development (Formas) and the Swedish Agency for Growth Policy Analysis. The second workshop was a dedicated space between the food area and the TIPC researchers aimed at summarizing the main learnings from the engagement, presenting the current state of affairs for the area and reflecting together on possible ways to move forward.

For the first workshop, TIPC researchers produced a Learning History (LH) to expand learning beyond the food mission experience. LHs are used mainly in organizational learning traditions (Roth 1996; Roth and Kleiner 1998; Parent, Roch and Béliveau 2007) to present specific experiences and their underlying challenges and takeaways from a general relatable story that acts as a vehicle to ignite reflection in different contexts. The main component of the LH is the direct voices of the protagonists—in this case, the Vinnova team involved in the engagement—on specific topics emerging from the engagement. The LH used in the workshop concerned the value of TIPC's approach to policy design and presented three challenges to reflect upon: (1) mainstreaming learning coming from specific policy testing is not a straightforward process; (2) mainstreaming lessons and tools require simplification for others to adapt and adopt to their needs—however, co-learning, co-design and formative evaluation require more profound understanding and involvement from the policy-making teams; and (3) destabilizing dominant practices and routines in policy also means destabilizing one's organization. The session enabled a collective and active reflection on the obstacles and enablers for system transformation using the inputs from the formative evaluation developed by the Vinnova-TIPC team.²

The second workshop included all the food area members, including the two Vinnova members active in the engagement. A visualization of their ToC and its guiding TOs helped the team to reflect upon the need to place more emphasis on

building, nurturing and strengthening niches. The workshop deepened strategies in terms of policy instruments (i.e. calls, direct funding, the search for intermediary organizations) to expand the current ToC and achieve the aims of the food missions. After the workshop, new calls on mobilization and learning initiatives for future innovation towards niche creation were launched.³ This work is ongoing and the results are yet to be seen.

For each of the four phases, different methods were used to support the engagement process. It was clear from the beginning that keeping an open mind to different approaches and honouring radical experimentation was the only way to move forward. Table 4 presents a summary of the different participants, outputs and methods used in the formative evaluation.

4. Experiencing the evaluation. Learnings from the process

In this section, we describe our main takeaways in implementing the TIPC-formative evaluation approach, which is based on the four elements highlighted by Turnheim et al. (2015) to describe initiative-based learning interventions in transition: understanding expectations, learning by doing, environment and resource constraints, and responding to emergent processes. Additionally, we provide insights on the role and attitudes of evaluators, which we consider crucial in the evaluation of TIP.

All the quotes appearing in the following five subsections have been anonymized. They are taken from the in-depth interviews conducted in each of the phases, the Learning History and the dynamic reflexive binnacle (see Table 4).

4.1 Understanding expectations

'We have found that clearly expressing each other's assumptions and expectations in the intervention at the start of the process is a factor of success'. (Vinnova team member).

Starting an engagement is always challenging, more so when it comes to evaluation. The process of building the right group energy that empowers and leads to action requires time and trust. However, engagements are bound by time constraints and output deliveries, which can make the team focus on the set plan rather than the final purpose. We did not spend enough time at the beginning of the engagement getting to know each other and understanding each other's (TIPC's and Vinnova's) perspectives and methodologies. The TIPC team built a methodology with set steps and felt pressed to deliver tangible outputs such as the ToC. As we explained before, the methodologies were at odds with the approach to setting transformation pathways, but the team only understood this after several working sessions and one-to-one reflection sessions with the person who would take on the role of mediator. Once misunderstandings and frustrations were expressed and addressed, both teams had the right mindset and energy to work together. Gaining trust by clearly expressing expectations at the start of the process would have saved time and energy and is crucial to the evaluation process. The following quote from a Vinnova team member expresses the relevance of dedicating time to share conceptions and agree on goals and expectations.

Table 4. Participants, methods and outputs of the Vinnova-TIPC formative evaluation process

Phase	Who was involved	Methods	Outputs
<i>Developing a preliminary ToC and selecting the TOs</i>	Four Vinnova team members Four TIPC researchers	Training exercises as learning platforms (highly interactive, based on experience); slide-decks, collaborative online boards as sources for reflection and interaction; rapid feedback interviews; in-depth interviews with Vinnova team members; one-to-one reflection meetings	slide-decks reflexive dynamic binnacle Preliminary ToC
<i>Developing the final ToC</i>	Two Vinnova team members Four TIPC researchers A communications officer and a designer from TIPC	One-to-one reflection meetings; collaborative online boards; documentary analysis as input for workshops design; online mentoring/coaching; in-depth interviews with Vinnova and TIPC members;	Reflexive dynamic binnacle Final ToC and infographic of the ToC
Broadening learning	Four Vinnova team members Four TIPC researchers Key stakeholders from other government agencies, design agency and a food entrepreneur	Historical data analysis; documentary analysis; in-depth interviews with policymakers from other Swedish Agencies and Vinnova team in charge of the school food mission; reflective practice session about data analysis results with different areas in Vinnova	Learning history; Monitoring, evaluation and learning plan Report with the evaluation analysis of the activities developed during 2021 with implications for further phases
Closing down on main learnings, opening up on new endeavours	Six Vinnova team members Four TIPC researchers Key stakeholders from other government agencies Broader officers in Vinnova from other areas	In-depth interviews with Vinnova members; Training exercises as learning platforms; rapid feedback interviews; training exercises as learning platforms; reflective practice session with the food area group	Learning histories Final Reports with implications for the evaluation of the food mission-oriented policy strategy

When I think of setting up a similar project in the future or in another context, I think one of my main takeaways is that it would be necessary to spend more time talking at the beginning of the project. Within the group, you need to allow for everyone to clearly speak out about expectations and assumptions. What are the assumptions about each other, and what do we as researchers and practitioners, as well as individuals, expect when it comes to working with each other. (Vinnova team member).

Both teams finally built a safe and respectful space to talk about the process, which was fundamental to enabling co-creating processes, as was expressed by a Vinnova team member:

Having the opportunity, in a safe space, to have a serious conversation about what we are doing while we are doing it and to build up trust. To come to that point has taken some time, with some excursions outside of the map that we intended to be within, but I think I have learned a lot. (Vinnova team member).

4.2 Learning by doing

4.2.1 A 'bricolage of techniques'

Nevertheless, understanding each other's points of view when the participants have different sets of capabilities, professional and cultural backgrounds can be stressful, challenging and time-consuming. It requires the team to remain flexible regarding the expected outputs from each interaction and to acknowledge that deviations from the original plan can be a

source of meaningful learning and should be accepted and even embraced. To this end, we used various sets of techniques. The TIPC team organized face-to-face and online workshops, used online interviews and developed two tools (the dynamic reflexive binnacle and the learning history) to capture the learnings accumulated through the process. We acted as *bricoleurs*: 'Bricolage is built on a foundation of eclecticism, an open-minded approach that eschews rigidly following recipes about how things ought to be done but instead considers multiple methods, designs, and inquiry possibilities, often combining diverse approaches in creative and situationally appropriate, insightful, and useful ways'. (Patton 2020: 101). As in the Blue Marble evaluation approach proposed by Patton (2020), in our formative evaluation approach, a *bricolage* of techniques proved to be essential given the diversity of situations, applications and the dynamic nature of the different policy interventions with which we interacted.

4.2.2 Rooting the evaluation approach in theory

Our formative evaluation approach has a strong theoretical background based on transitions theory. We have realized how this distinctive feature makes it meaningful for participants, allowing them to reflect on their intervention and, eventually, reorient it.

In the Vinnova case, the theoretical inputs were difficult to introduce. When the TIPC team started the engagement, Vinnova was designing its intervention following an experimental strategic-design approach, as we mentioned before. According to this approach, a way of delivering ambitious change is a continuous design of a prototype that can be

tested and refined over time (Young 2010), without pre-defining specific desired outcomes (Gaziulusoy and Erdoğan Öztekin 2019). Instead, we started our engagement attempting to build a ToC and then deriving the TOs from it. This resulted in a clash between the two different perspectives, which brought tensions and difficulties. As one of Vinnova's participants recalled:

From the broader perspective of the mission-oriented work and its design-oriented approach, it has turned out to be complex and somewhat 'unfitting' to map out a ToC in the 'classical way' [input-output-activities-actors-outcomes] at this point in time. The mission-oriented work is not based on a pre-defined map; rather, the work is to build up along the way, with an overall direction [the missions] as the guiding principle. (Vinnova participant).

To address these initial differences, the TIPC team focused on the objectives and practices laid out in the previous two sections: developing trust and mutual understanding through the generation of safe spaces, and the use of a variety of techniques in a flexible way. The evaluators abandoned the idea of mapping out a ToC and focused, instead, on identifying the four most relevant outcomes that the food policy team wanted to achieve (see Figure 1).

Despite these difficulties in using the theory, it was acknowledged by the Vinnova team that the ToC and the TOs could be valuable in challenging and mobilizing 'regime' actors.

Then you discovered other things and deeper dimensions in them than you would have done otherwise, so [it was] very valuable. Yeah, extremely valuable. I think in many cases, necessary. If you really want to mobilise, change the ambitions of the actors that could actually influence the game. (Vinnova's leadership).

It was also acknowledged that the use of theory during the evaluation can provide a transformative narrative, although it can be supported with stronger data.

When I sit in different discussions where I had been over about how we should construct new ways of governing initiatives with the potential to transform systems, that is a perspective that I need, but I'm not yet really able to verbalise it as well as I should be, to use it as arguments and we need strong arguments if we're going to get some of these perspectives [...] So, I would say the theory around transformative outcomes is super interesting. We need to test it more; we need more concrete cases, basically, that I can point to. And maybe this is my own limitation, that I haven't enough data basically on this. (Vinnova's leadership).

4.2.3 The role of intermediaries

The challenges described in the previous subsections were successfully addressed thanks to members in both teams taking the roles of intermediaries. Such a role was not planned but became essential for the process, as one of the evaluators recalls:

When you try to bring something as new as the transformative outcomes, which is somewhat theoretical, to a group of people who work within different contexts, you need a translator, someone who translates our words. I think we are fulfilling this role as well, but it is obvious to me that you need someone from the agency, from the organisation. (TIPC team member).

Online interactions make it more difficult to grasp an experiential understanding of the context, and thus the role of an intermediary becomes even more crucial. The person who took on this role helped the team of evaluators better understand the agency's culture, and its emergent design approach to missions, making the theoretical framework and methodology more applicable and context-sensitive. In addition, she provided updated information on relevant organizational changes, participants' duties, time constraints and inner organizational tensions. Reflection interactions between the intermediary and the TIPC team became essential for establishing trust with the intermediary and, through her, with other Vinnova participants.

As one TIPC member pointed out:

Her intermediation has opened up possibilities to understand each other's expectations and to unblock different stages in the process when frustration was growing among all the team members [...] I have had one-to-one conversations with her, and she has helped me to better understand the mission's group dynamics and reflection processes. By doing that, I have gained trust, and I have enlarged my readiness to reflect on the processes. Then, I consider that having a person in her role is fundamental for each of the projects we envisage in the coming times. (TIPC member).

4.2.4 Constraints

However, time and resource constraints, along with changes in the organizational structure on the Vinnova side, created stress in the engagement and limited the initial scope of the evaluation, which included monitoring indicators for each of the outcomes defined by the team. The Vinnova teams and leadership also reported that although the process was highly relevant for the food area team, the investment of time from the Vinnova members and its context-dependency makes it difficult to apply to other contexts, as one of the heads of department mentioned:

I think it has the potential to have a lot of value, the project itself was started as an experiment and it was started as a niche, if you want [...] What we realise now, two years in, is that it is quite challenging to translate these experiences to our mainstream way of working and, for a while, we were really struggling with understanding how to scale this. Because, as we've discussed many times, it is very time-consuming to add a layer of evaluation through the actual project (Vinnova's leadership).

This statement outlines one of the limitations of the formative evaluation approach: it requires resources dedicated to the engagement. From the TIPC side, the 2-year contract agreement guaranteed the involvement needed. However, reflecting on the engagement from Vinnova's side, we argue that the experimental character of this evaluation was an

element that hindered the adequate provision of time. Embedding, in a more specific manner, the formative evaluation engagement as part of the workload of Vinnova's workers may have been a way of overcoming the barrier of the lack of time. However, this would have required a clear commitment from the organization to the experimental nature of the formative evaluation, which is somewhat difficult in such a large organization.

Another struggle experienced through the engagement was the scaling-up of results. As has been presented in Section 3.4, the formative evaluation helped the food mission team to act directly in their policy design and shape the expected outputs inspired by the TOs. However, we were unable to properly diffuse these valuable results. Vinnova is a large organization, in which multiple processes are in play concurrently; the formative evaluation was one of many such processes, one very different from other traditional policy approaches. Without a doubt, its experimental niche nature will require time and different organizational capacities to upscale the approach to other programmes and areas.

4.3 The evaluators' role and attitudes

As we have described, the engagement was a source of learning for both teams starting with the reconfiguration of assumptions on the role of researchers and practitioners in highly experimental settings where co-creation and co-production of knowledge are core to the process. A Vinnova team member described in some detail the problems we faced:

One problematic, and commonly occurring, aspect of this is the way we as practitioners tend to relate to researchers as consultants [...] researchers are viewed as experts who are supposed to help the practitioners to solve specific issues. Practitioners, who tend to be busy and focused on practical solutions, thus enter the exchange with a 'what's-in-it-for-us' mentality, and expect the researchers to deliver tools and advice in a similar way as consultants would. In the case of TIPC, I think this generally occurring tendency was a bit further enhanced by the ambitious (and, indeed, partly selling) tone and professional format of communications material of the consortium. Actually, in the same way, the high expectations of Vinnova (in general and within TIPC) might partly derive from the agency's capacity and strategy regarding communications and self-proclaimed profile. Another, quite contradictory, aspect of the imbalance in expectations, lies in the simultaneously existing assumption that researchers do not know the reality of the practitioners as well as the practitioners do themselves. (Vinnova team member).

Practitioners questioned the TIPC researchers' role as evaluators and brought them to a 'discomfort zone'. To deal with this situation, the TIPC team had to acquire some of the attitudes that Patton (2021: 123) highlights as relevant for transforming the field of evaluation.

Questioning prevailing assumptions: The experimental ethos and the reflexive practice of TIPC's formative evaluation approach start with the evaluators themselves. For the TIPC team, the assumption of the need for a 'classical' ToC to conduct the formative evaluation was questioned. Once it was clear that such an approach was hindering rather than helping the process, an adaptation to focus on the

identification and monitoring of desired outcomes was a way to move forward.

Empathy and humility: Building trust and mutual understanding, as well as recognizing the essential role of intermediaries, requires putting aside the 'expert' ego and engaging with the practitioners with a high degree of empathy.

Openness to new opportunities, ideas and ways of thinking and valuing of multiple perspectives: To act in context and provide suitable support, the TIPC team used a combination (*bricolage*) of techniques. Both teams left their comfort zone and experimented to combine different practical and theoretical approaches.

Comfort with ambiguity and unfamiliar situations: The COVID-19 pandemic forced the teams to work remotely. This was an 'unfamiliar' situation that forced a reshaping of the initial plan. The digital setting, coupled with the nature of the engagement, stretched the TIPC team to be facilitators beyond the familiar researchers' role, which involved enhancing different competencies and dealing with ambiguity. The last factor was the lack of familiarity of both teams with working with new colleagues coming together from different realms of practice. Taken together, both teams were often outside their comfort zones, and yet the combination of flexibility, openness and humility and the individual and collective reflection they engaged in after each interaction helped them adapt and work comfortably.

Adaptability and the ability to be cognitively nimble: This attitude is closely related to the previous ones. Due to the experimental character of the engagements, the TIPC team had to be adaptive and agile. The *bricolage* of methods and the flexibility in using the ToC are examples of these attitudes. Challenges like the difficulties in applying a theory-based approach required adaptations that, while preserving the theoretical base, led to diverse ways of setting up the evaluation activities. The evaluators adapted their approach and played facilitating and mediating roles to make the interactions more dialogical without losing theoretical depth.

5. Conclusions

The experimental nature of TIPC's formative evaluation develops on the flexibility and agile approach to context-based interventions. As we showed in the Vinnova engagement, both teams had to explore diverse entry points in their sense-making undertaking of the innovation policy design. The evaluation was conceived as part of the process instead of an initial or final stage, which anticipated and brought to light assumptions, disagreements and possible transformational failures. The interactions became richer once trust was enhanced and captured emerging learning was used as active guidance for the subsequent stages, making the process transparent, shareable and relevant to others. Reflexivity as a base allowed the teams to navigate ambiguity and uncertainty through intermediation and effective methods and techniques.

The use of theory to root the evaluation helped the Vinnova team to focus on their intended transformative outcomes. The TOs were core to the reflection process, in order to view the current possibilities and maximize the possible impact of the intervention. Having a set of outcomes based on opening and destabilizing regimes and any outcome on niche building and nurturing helped the team to reflect on their assumptions and weaknesses in the mission design. The last engagement phase was particularly enlightening for the design

of the subsequent policy instruments, as explained above. The theory helped the Vinnova team to have a broader systemic view of the main socio-technical systems, question the level of inclusivity in their engagements, and have a clear directionality to focus their efforts.

However, we have also highlighted the limitations in making the TIPC-formative evaluation approach inside the organization sustainable and scalable. The lack of time on the part of Vinnova's participants was overcome by adapting methodologies, reflecting on the process and having a member of Vinnova acting as an intermediary. This is a relevant insight that should be carefully considered in future engagements. Furthermore, a more detailed agreement from the two sides on the human resources needed and time allocation could help to fit the formative evaluation process into organizational dynamics. The same can be said for the scaling up of the insights of the evaluation. A formative approach to evaluation should carefully consider how learning is embedded into the organization. The engagement described in this paper was successful in capturing learning among the participants; however, more efforts should be placed on embedding what has been learned into the organization.

Finally, a significant source of reflection and learning has been to interrogate ourselves about our role as evaluators. We argue that such a formative evaluation approach challenges the role of evaluators, stretching them and requiring different attitudes from those that are common in evaluation practice. Evaluators questioned prevailing assumptions and exercised empathy and humility easily; however, opening up to new opportunities, ideas and ways of thinking and valuing multiple perspectives required further efforts to adapt methodologies without losing the core principles of the theoretical framework. The teams also found it difficult to be comfortable with ambiguity and unfamiliar situations. Some discomfort was present in both engagements, but participants learned how to deal with it and become more relaxed in situations where disagreements and differences in practices and culture emerged. Adaptability has been essential throughout the engagement, but, given the theory-led approach, there were limits in terms of being *cognitively nimble*. However, the evaluators combined different theoretical foundations when applying their evaluation approach, adapting it to the needs and concepts held by the participating practitioners. Although Patton (2021) does not mention reflexivity as a key attitude, it was very relevant in the case context. A reflexive practice was crucial to overcoming challenges and enabling learning.

Supplementary data

Supplementary data are available at *Research Evaluation Journal* online.

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.

Funding

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).

Conflict of interest statement. None declared.

Notes

1. The dynamic reflexive binnacle was a tool designed by the TIPC team during the engagement to keep track of changes in attitudes, interactions and knowledge (based on Beers, van Mierlo and Hoes, 2016) of the engagement participants. Its aim was to encourage and keep a record of the learning process, containing reflections on the challenges and successes throughout the interventions. It recorded changes in individual and collective routines as a result of their interaction. It formed part of the monitoring and evaluation process.
2. Learning histories were used to reflect on the engagement process and its results. One of the learning histories can be found in <https://www.tipconsortium.net/publication/learning-histories-vinnova-balancing-theory-and-practice/>.
3. See: <https://www.vinnova.se/en/calls-for-proposals/mobiliserings-och-larinsatser-for-2022-01228/a-new-recipe-for-the-food-system-mobilization/>.

References

- Beers, P. J., van Mierlo, B., and Hoes, A.-C. (2016) 'Toward an Integrative Perspective on Social Learning in System Innovation Initiatives', *Ecology and Society*, 21: 33.
- Bergek, A., and Haddad, C. R. (2022) 'Evaluating Transformative Innovation Policy Outcomes as Unfolding Processes of Change in Socio-Technical Configurations', in G. Vélez-Cuartas, and O. Y. Romero-Goyeneche (eds) *Transformative Metrics: Contributions to the Studies for Monitoring and Evaluating How Science, Technology, and Innovation Can Address Social and Environmental Challenges*, pp. 63–105. Universidad de Antioquia.
- Biggs, R., Schlüter, M., and Schoon, M. L. (2015) *Principles for Building Resilience: Sustaining Ecosystem Services in Social-Ecological Systems*. Cambridge University Press.
- Boni, A., Molas-Gallart, J., and Giachi, S. (2019) 'Towards a Framework for Transformative Innovation Policy Evaluation'. *TIPC Research Report*. Science Policy Research Unit (SPRU), University of Sussex Business School.
- Diercks, G., Larsen, H., and Steward, F. (2019) 'Transformative Innovation Policy: Addressing Variety in an Emerging Policy Paradigm', *Research Policy*, 48: 880–94.
- Gaziulusoy, I., and Erdoğan Öztekin, E. (2019) 'Design for Sustainability Transitions: Origins, Attitudes and Future Directions', *Sustainability*, 11: 3601.
- 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. (2022) 'Causality and Explanation in Socio-Technical Transitions Research: Mobilising Epistemological Insights from the Wider Social Sciences', *Research Policy*, 51: 104537.
- Geels, F. W., and Schot, J. (2007) 'Typology of Socio-Technical Transitions Pathways', *Research Policy*, 36: 399–417.
- Geels, F. W., Kern, F., Fuchs, G., Hinderer, N., Kungl, G., Mylan, J., Neukirch, M., and Wassermann, S. (2016) 'The Enactment of Socio-Technical Transition Pathways: A Reformulated Typology and a Comparative Multi-Level Analysis of the German and UK Low-Carbon Electricity Transitions (1990–2014)', *Research Policy*, 45: 896–913.

- Ghosh, B., Kivimaa, P., Ramirez, M., Schot, J., and Torrens, J. (2021) 'Transformative Outcomes: Assessing and Reorienting Experimentation with Transformative Innovation Policy', *Science and Public Policy*, 48: 739–56.
- Haddad, R. C., Nakic, V., Bergek, A. and Hellsmark, H. (2019) 'The Policymaking Process of Transformative Innovation Policy: A Systematic Review', in *Proceedings of the 4th International Conference on Public Policy (ICPP4)*.
- Haddad, C. R., and Bergek, A. (2023) 'Towards an Integrated Framework for Evaluating Transformative Innovation Policy', *Research Policy*, 52: 104676.
- Hekkert, M. P., Janssen, M. J., Wesseling, J. H., and Negro, S. O. (2020) 'Mission-Oriented Innovation Systems', *Environmental Innovation and Societal Transitions*, 34: 76–9.
- Hill, D. (2022) *Dark Matter and Trojan Horses: A Strategic Design Vocabulary*. Strelka.
- Howlett, M., and Ramesh, M. (1995) *Studying Public Policy: Policy Cycles and Policy Subsystems*. Oxford: Oxford University Press.
- Janssen, M. J. (2019) 'What Bangs for Your Buck? Assessing the Design and Impact of Dutch Transformative Policy', *Technological Forecasting and Social Change*, 138: 78–94.
- Janssen, M. J., Torrens, J., Wesseling, J. H., and Wanzenböck, I. (2021) 'The Promises and Premises of Mission-Oriented Innovation Policy—A Reflection and Ways Forward', *Science and Public Policy*, 48: 438–44.
- Janssen, M. J., Bergek, A., and Wesseling, J. H. (2022) 'Evaluating Systemic Innovation and Transition Programmes: Towards a Culture of Learning', *PLOS Sustainability and Transformation*, 1: e0000008.
- Kuhlmann, S., and Rip, A. (2018) 'Next-Generation Innovation Policy and Grand Challenges', *Science and Public Policy*, 45: 448–54.
- Luederitz, C., Schöpke, 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.
- Matt, M., Robinson, D. K., R., Joly, P. B., Van Dis, R., and Colinet, L. (2023) 'ASIRPA Real-Time in the Making or How to Empower Researchers to Steer Research towards Desired Societal Goals', *Research Evaluation*, 32: 412–25.
- Mazzucato, M. (2018) *Mission-Oriented Research & Innovation in the European Union: A Problem-Solving Approach to Fuel Innovation-Led Growth*. Brussels: Publications Office of the European Union.
- Molas-Gallart, J., Boni, A., Giachi, S., and Schot, J. (2021) 'A Formative Approach to the Evaluation of Transformative Innovation Policies', *Research Evaluation*, 30: 431–42.
- Ofir, Z., and Rugg, D. (2021) 'American Journal of Evaluation Section on International Developments in Evaluation: Transforming Evaluation for Times of Global Transformation', *American Journal of Evaluation*, 42: 47–52.
- Patton, M. Q. (2011) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. Guilford.
- Patton, M. Q. (2020) *Blue Marble Evaluation: Premises and Principles*. Guilford.
- Patton, M. Q. (2021) 'Evaluation Criteria for Evaluating Transformation: Implications for the Coronavirus Pandemic and the Global Climate Emergency', *American Journal of Evaluation*, 42: 53–89.
- Patton, M. Q. (2023) 'Evaluating Transformation Means Transforming Evaluation'. In Beth Edmondson (ed.) *Sustainability Transformations, Social Transitions and Environmental Accountabilities*, pp. 15–38. Cham: Springer International Publishing.
- Parent, R., Roch, J. M., and Béliveau, J. (2007) 'Learning Histories: Spanning the Great Divide', *Management Research News*, 30: 271–82.
- Rip, A., and Kemp, R. (1998) 'Technological Change', in S. Rayner, and E. L. Malone (eds) *Human Choice and Climate Change. Vol. II, Resources and Technology*, pp. 327–99. Columbus, OH: Batelle Press.
- Roth, G. (1996) *Learning Histories: Using Documentation to Assess and Facilitate Organizational Learning* (SWP 3968). <<http://34.201.211.163/bitstream/handle/1721.1/2668/SWP-3968-37617965.pdf?sequence=1&isAllowed=y>> accessed 10 June 2023.
- Roth, G., and Kleiner, A. (1998) 'Developing Organizational Memory through Learning Histories', *Organizational Dynamics*, 27: 43–60.
- Scharmer, O. (2018) *The Essentials of Theory U: Core Principles and Applications*. Berrett-Koehler Publishers.
- 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> accessed 10 June 2023.
- 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.
- 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.
- Stame, N. (2004) 'Theory-Based Evaluation and Types of Complexity', *Evaluation*, 10: 58–76.
- 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*. 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.
- 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', *Res Policy*, 41: 1037–47.
- Williams, S., and Robinson, J. (2020) 'Measuring Sustainability: An Evaluation Framework for Sustainability Transition Experiments', *Environmental Science & Policy*, 103: 58–66.
- Wise, E., and Arnold, E. (2022) *Evaluating Transformation—What Can we Learn from the Literature?* (No. 2022/10). Lund University, CIRCLE-Centre for Innovation Research.
- Young, G. (2010) 'Design Thinking and Sustainability', *Zumio Meaningful Innovation*, 61: 1–27.