REALISING SOCIETAL CHALLENGES: TOWARDS CHALLENGE-LED MONITORING AND EVALUATION

VINCENT BAARSLAG, AMBER GEURTS & FRANS VAN DER ZEE
DOI: 10.22163/FTEVAL.2024.632

ABSTRACT

Challenge-led innovation policies place societal challenges and transitions at the focal centre and goal of innovation policy. This new genre of innovation policy not only requires new approaches in agenda-setting, programming, implementation and management, but also requires a renewed view and practice of **monitoring and evaluation** in order to be able to steer innovation policies towards societal goals. In this paper, we focus on the necessity and usefulness of a different view and way of monitoring and evaluating challenge-led R&I policy and its implementation – which we term **challenge-led monitoring and evaluation**. To define this new approach, we conducted a literature review. Our analysis identifies the bottlenecks as well as potential routes to arrive at an appropriate monitoring and evaluation framework for challenge-led innovation policies. Next, our analysis highlights how governance and institutionalised evaluation culture and practice are presented as (part of) solutions to all four identified bottlenecks but usually considered an afterthought requiring 'experimentation'. However, overcoming other identified bottlenecks in challenge-led monitoring and evaluation is contingent upon altering (a) governance and organisational structures as well as (b) institutionalised assumptions and practices. Therefore, we argue that in order to further develop challenge-led monitoring and evaluation, the roles of governance and organisational structures as well as institutionalised assumptions and practices should be problematised and prioritised as bottlenecks.

1. INTRODUCTION

Innovation policy in the Netherlands and Europe has been developing rapidly in recent years. Rather than only targeting economic growth and productivity, we see an increasing recognition – both in policy practice as well as in science - for a new genre of innovation policy that aims to focus research and innovation more strongly on addressing persistent societal challenges (Kuhlman & Rip, 2018; Schot & Steinmueller, 2018). Consequently, previous innovation policy frameworks, which focused on fixing market failures by repairing private firms' underinvestment in research and development (frame 1) or on fixing system failures in national innovation systems and strengthen innovation networks (frame 2), are being complemented with a new generation of innovation policies to address and find solutions for societal challenges (frame 3, Schot & Steinmueller, 2018). Based on different emerging streams of literatures, these frame 3 innovation policies are framed as 'mission-oriented innovation policy (Mazzucato, 2018; Hekkert et al., 2020) or 'transformative innovation policy' (Diercks et al., 2019; Haddad et al., 2022). In this paper, we place these different perspectives under a common umbrella that we call 'challenge-led innovation policy (CIP)' (Rathenau Instituut, 2020; 2021).

This new genre of challenge-led innovation policy not only requires new approaches in agenda-setting, programming, implementation and management (Janssen et al., 2021). In light of this new type of policy, we argue that **monitoring and evaluation** also requires an innovative approach, which is different on a number of dimensions from the usual, 'traditional' monitoring and evaluation of research and innovation policies. **Challenge-led monitoring and evaluation** is meant to effectively use the potential and increase the effectiveness of utilising research and innovation in solving societal challenges and realising system transitions.

Monitoring and evaluation is important in addressing societal challenges and transitions (see e.g. Janssen, 2019; Luederitz et al., 2017; Turnheim et al., 2015; Weber et al., 2014; Wittmann et al., 2022). Monitoring and evaluation is a key element in the realisation of a balanced and reliable evidence base that enables politics and public administrations to conduct an informed debate and to make targeted assessments about the deployment, progress, and outcomes of poli-

cies for societal challenges and transitions. Monitoring and evaluation provide the evidence to enable political and administrative accountability and hence fulfil a vital democratic function. At the same time, monitoring and evaluation also enable politics and public administration to maintain or, where necessary, shift policy direction by learning and offering room for improvement and adjustment. Learning and adjustment can mean intensifying, accelerating, but also phasing out or even stopping particular research and innovation (R&I) efforts, thus being able to even better steer mission and transition processes towards the envisioned societal goal. Monitoring and evaluation can thus provide knowledge bases and policy intelligence to inform system-level transformative policies for R&I now and in the future.

It also becomes increasingly clear that monitoring and evaluation of challenge-led R&I policies should differ in important respects from the usual, more traditional evaluation of R&I policies as it sets different design requirements in terms of the evaluation perspective, framework and approaches to be used. Furthermore, as the limitations of traditional monitoring and evaluation frameworks are becoming increasingly evident, calls for reflection, reconceptualization and experimentation with altogether different monitoring and evaluation practices increases (Arnold et al., 2018; Haddad et al., 2022; Molas-Gallart et al., 2021; Wittmann et al., 2022). However, the complexity and wickedness of societal challenges impose various conceptual challenges for such a challenge-led monitoring and evaluation, including '**the multidimensionality and interaction of effects, the different analytical levels, the long time horizon associated with mission goals and the empirical diversity of missions**' (Wittmann et al., 2022, p.31).

While there have been numerous attempts to initiate and substantiate the discussion of what it takes to develop a monitoring and evaluation framework that can be considered 'challenge-led', there has been no convincing synthesis, yet, of what this innovative view and role means for monitoring and evaluation, let alone how it could be further shaped and implemented. With the increasing importance of challenge-led R&I in policy and society, whereby the bottlenecks, potential pitfalls and possible solutions of challenge-led monitoring and evaluation have until now been underexposed, such a synthesis would fulfil a societal need and serve a public interest. Renewal of monitoring and evaluation practices, but also a government that learns and adjusts more consistently and transparently on the basis of (continuous) monitoring and evaluation is necessary in view of the pressing societal urgency of solving various societal challenges. Expectations regarding the role of R&I in solving societal challenges,

and the associated use of public funds accentuate this necessity. In this article, we therefore focus on the emergence of a different perspective on the monitoring and evaluation of challenge-led R&I policy. The research question central to this article thus reads: **"How does the emergence of challenge-led research and innovation policies impact requirements for monitoring and evaluation, and how can challenge-led monitoring and evaluation strengthen policy and governance of research and innovation aimed at societal challenges?"**

We address our main research question using a systematic literature review. In the next sections, we describe our methodology for the literature review and present our main findings. We conclude our analysis with a discussion of the subsequent implications and steps to be taken as from here.

2. METHODOLOGY

In order to address our central research question we conducted an in-depth literature review (Elsbach & van Knippenberg, 2020; Torraco, 2016) focusing on generating insights as well as identifying active debates regarding why and how challenge-led monitoring and evaluation is different from 'traditional' monitoring and evaluation and what is needed to implement challenge-led monitoring and evaluation practices. This review process consists of several steps, including the search for relevant articles, their selection , and their analysis. To complement our literature review, we have also conducted semi-structured interviews with civil servants and evaluation experts at different policy settings and levels, involved with different societal challenges. These interviews helped us to identify the main bottlenecks and crystallise our findings further.

SEARCH AND SELECTION

To arrive at a set of relevant articles, we conducted a Web of Science article search from 2004 -2022. Our search included search terms in which 'innovation policy' or 'innovation programme' was combined with 'evaluation', 'assessment' or 'monitoring' concepts. This approach means that any governance or government capacities notions around monitoring and evaluation were excluded from the search process. This decision was made to focus on what challenge-led monitoring means, rather than whether government has the capacity to do so. In addition, we included search terms regarding 'sustainability', 'transition', 'transformation', 'missions', 'SDG' or 'societal challenge' in combination with 'evaluation', 'assessment' or 'monitoring' concepts. We used several

variations of these keywords to increase the number of potentially relevant search results. Finally, to augment the articles identified during our literature review of academic peer reviewed papers and to correct possible omission bias we also searched Google for grey literature and other articles that might be useful. This resulted in a total of 304 articles and book chapters.

The main selection process involved two rounds. In the first round, we primarily judged articles based on the title, the abstract and keywords. This resulted in a key literature list of 51 articles. In the second round, we inspected the full texts of articles in order to check whether and how the main constructs we are interested in have been mentioned in the body of the text. In case we were unsure about an article, two of our authors would debate the relevance of the article. As a result of this selection process, several articles were excluded for the final analysis step. After the second step, we ended up with 30 relevant articles, out of which 15 articles discussed core issues to our research questions, while 15 other articles discussed closely related issues. A summary of our search and selection process is shown in Figure 1.

Figure 1: Flowchart article selection
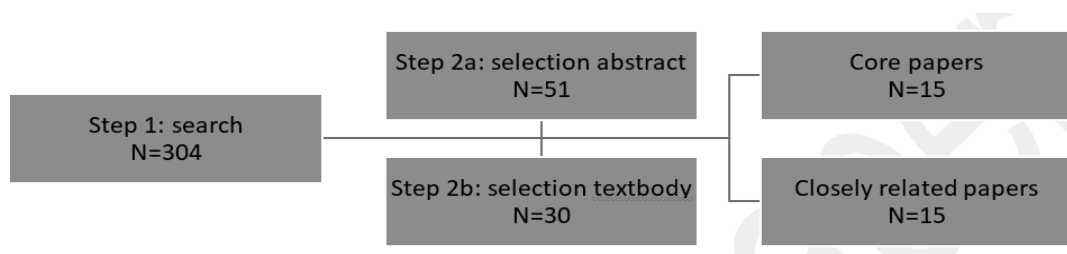
3. ANALYSIS

The analysis of our selected sample of core and closely related papers was conducted in two rounds. First, we analysed the core texts to identify bottlenecks and solutions that are discussed, described or highlighted in the papers. In this step, we largely relied on the text and wording used in the papers and solely identified a difference between 1) the discussion of a bottleneck or 2) the discussion of a solution. In the second step, we analysed the identified bottlenecks to define aggregate dimensions. In this step we identified four dimensions of bottlenecks, namely 'operationalisation', 'attribution', 'accountability' and 'learning'. Next, we also analysed the identified solutions and found six dimensions of proposed solutions, namely 'operationalisation', 'attribution', 'ac-

countability', 'learning', 'governance and organisation', and 'institutional rigidity'. We refined our analysis by revisiting the literature and our analyses frequently, and corroborating our findings using semi-structured interviews with experts. In the case we were unsure about a dimension, the authors would debate the dimensions and its characteristics.

4. FINDINGS

Our analysis shows that with the emergence of the new genre of challenge-led innovation policy, the anticipated scope of monitoring and evaluation has also changed. As such, challenge-led monitoring and evaluation focuses not only on the result (the innovation outcome), but also on the question of how that outcome relates to the desired impact (i.e. addressing the societal challenge) and whether ongoing (intermediate) innovation results are still in line with and steering towards that desired impact. That is, societal challenges and the context in which they are shaped are often complex, with goals that cover a longer term, whereby the solution directions can sometimes also shift and change over time. The complexity and systemic character of societal challenges, stemming from their wicked nature, have implications for monitoring and evaluation. These implications are discussed next.

4.1. IDENTIFYING BOTTLENECKS OF CHALLENGE-LED EVALUATION

By analysing the relevant literature, four important bottlenecks for challenge-led monitoring evaluation can be demarcated. These bottlenecks concern 1) accountability, 2) learning, 3) operationalisation, and 4) attribution. Though we acknowledge that these bottlenecks can also be observed in some of the 'traditional' R&I policy evaluations, our analysis shows that the wicked character of societal challenges that challenge-led innovation policy aims at, exacerbates these bottlenecks (Duckett et al., 2016; Wanzenböck et al., 2020). That is to say, the four demarcated bottlenecks are not unique to challenge-led monitoring and evaluation, but manifest themselves differently and more extensively when compared to a 'traditional' innovation policy monitoring and evaluation context.

4.2. CATEGORISING BOTTLENECKS: UNDERSTANDING THE CHANGE DIMENSION

When further analysing these bottlenecks it was found that they can be placed into two different categories: (1) bottlenecks related to a fundamental shift in the role of R&I policy evaluation within a challenge-led innovation policy (CIP) context, and (2) bottlenecks related to the operationalisation and attribution of transformative impact in a CIP context. The two bottlenecks related to **accountability** and **learning** originate from a fundamental shift in the role of monitoring and evaluation within a CIP context: the socio-economic and environmental legitimations of CIPs, as well as the increased emphasis on reflexivity and adaptivity alter the evaluation perspective, imposing additional requirements on monitoring and evaluation of CIPs in terms of providing accountability and facilitating learning. In contrast, identified bottlenecks related to **operationalisation** and **attribution** in large part preceded the emergence of CIPs and are merely magnified by the wicked context that characterises CIPs. Figure 2 illustrates the identified bottlenecks and their categorization. Next, the aforementioned bottlenecks and their categorisation are further explained.

Figure 2: Four bottlenecks of challenge-led monitoring and evaluation

Cause:	Bottleneck:	Why?:
Fundamental shift in the role of evaluation within a CIP context	Accountability	IP now (partially) accountable for adequately addressing societal challenges through supported transition pathways
	Learning	Increased need for learning due to reflexivity and adaptivity being emphasised in CIP frameworks
Wicked nature of CIP context	Operationalization	Measuring additionality increasingly difficult due to wickedness of challenges, long timespan
	Attribution	Attributing results increasingly difficult due to complexity of policy mixes, need for outcome-based evaluation

Note: IP is short for Innovation Policy, CIP is short for Challenge-led Innovation Policy

4.2.1   A FUNDAMENTAL SHIFT IN THE ROLE OF EVALUATION WITHIN A CIP CONTEXT

The change in policy goals and legitimization towards addressing societal challenges constitutes a fundamental shift in the role of innovation policy (Schot & Steinmueller, 2018). In contrast to innovation policy frameworks for economic growth and competitiveness, CIPs no longer primarily aim to enhance economic growth and increase productivity through stimulating technological progress and enhancing national systems of innovation (Bush, 1945; Lundvall, 1992). With

the shift towards CIP frameworks, innovation policies are directly assigned a role in addressing and finding solutions for societal challenges. Accordingly, we find that the intervention logic of CIPs is oftentimes inverted with respect to 'traditional' innovation policy frameworks. Using Theories of Change (ToCs), CIPs often take desired impacts rather than specific interventions as their point of reference (Weiss, 1995). This inverted intervention logic has consequences when considering monitoring and evaluation for accountability as well as for learning. Based on our analysis, we argue that this reframing comprises a fundamental shift in the role of monitoring and evaluation within a challenge-oriented context, calling for broader scoping when providing accountability and stimulating learning for CIPs.

First, **accountability**. Establishing democratic and political accountability is one of the primary functions of policy monitoring and evaluation as it contributes to citizens' trust in government and provides evidence regarding policy performance that can be used to justify implemented policy as well as inform future policy making decisions (OECD, 2020). With the shift towards CIPs, justification of R&I policy now also comes from ameliorating societal challenges (Schot & Steinmueller, 2018). This means that CIPs contribute to and are (at least partially) to be held accountable for generating solutions to societal challenges, despite the complex character and the long timescales at which (solutions to) societal challenges unfold (Janssen, 2019; Robinson & Mazzucato, 2019). As such, the concept of accountability is fundamentally different with respect to 'traditional' accountability for innovation policy: in addition to accountability for the expenditure of public resources on innovation, accountability now also encompasses adequately addressing societal challenges through supported transition-pathways. This includes taking into consideration the socio-economic and ecological implications of supported transition pathways and a review of the uptake of solutions (Wanzenböck, 2020). As framings of both problems and fitting solutions are oftentimes highly contested, accountability is meant to support stakeholders' need for trust in government and deliver evidence regarding policy performance longitudinally – answering the question whether government is still doing the right thing given the societal goal (rather than the question did we do things right). Finally, because societal challenges often lack clear ownership structures, their governance tends to be more distributed, bottom-up - and less unidirectional and top-down - thus inhibiting or at least making accountability more challenging (Arnold et al., 2018; Hertting & Vedung, 2012).

Second, **learning**. Monitoring and evaluation can assist in the assessment of value in regard to the aims, goals or objectives of a programme or policy. Aside from gaining insight into achievements, such an evaluation also enables reflection and the possibility to assist in the identification of (future) changes needed (Kuhlmann & Rip, 2018; Mazzucato, 2018; Haddad, 2022). That is, the wickedness and long term horizon of societal challenges raise the need for reflexivity to look at the progress made and how. However, it also raises the need for adaptivity to be built into the monitoring and evaluation framework in order to mitigate transformational failures (Amanatidou et al., 2014; Weber & Rohracher, 2012). Various CIP-frameworks emphasise this need for reflexivity and adaptivity in some shape or form by highlighting the need to integrate learning throughout the policy cycle and/or the multiple organisational/govern-ance levels (Kattel & Mazzucato, 2018; Kuhlmann et al., 2018; Loorbach, 2010; Lindner et al., 2016; Molas-Gallart, 2021). This need for reflexivity and adaptivity stems from the uncertainty in developments of the societal challenge on the one hand and possible solutions on the other hand. Accordingly, effectively facilitating diverse learning needs of various stakeholders across governance levels of CIP is an active field of study (Amanatidou et al., 2014; Aranguren et al., 2017; Borras, 2011; Haddad et al., 2022; Janssen, 2022; Luederitz; 2017, Magro & Wilson, 2019). Aranguren et al. (2017), for instance, noted the importance of demand for evaluation when it comes to transformative capacity. In a similar vein, Magro & Wilson (2019) highlighted legitimacy of the governance of evaluation as a prerequisite for dealing with contestation. Facilitating policy learning in practice remains, however, difficult as it is highly complex, can conflict with providing accountability, and often does not enjoy priority from all involved stakeholders (Amantidou et al., 2014; Magro & Wilson, 2019).

4.2.2. OPERATIONALISATION AND ATTRIBUTION OF TRANSFORMATIVE IMPACT IN A CIP CONTEXT

Next to the fundamental shift in the role of evaluation within a CIP context, our analysis shows that the wicked context of societal challenges and CIPs also accentuates the limitations of current monitoring and evaluation practices, giving prominence to bottlenecks related to the **operationalisation** of transformative impact and the **attributing** of results to specific policy interventions and policy targets (Arnold et al., 2018; Duckett et al., 2016; Grillitsch et al., 2018; Magro & Wilson, 2019; Janssen, 2019; Luederitz, 2017; Rittel & Webber, 1973). That is, within CIP the (a) growing number of policy targets and instruments, that are (b) interacting with one another and (c) adapted in an uncontrolled way makes building an evidence-base for policy making and policy learning increasingly

difficult (Adam, 2018). As such, CIPs exacerbate pre-existing bottlenecks related to operationalisation and attribution (Amanatidou et al., 2014; Aranguren et al., 2016; Haddad, 2022).

First, **operationalisation**. Operationalisation focuses on collecting and interpreting data that gives insight in performance, including expected and realised results. The shift towards CIPs makes operationalisation increasingly difficult, given the wickedness of the societal challenges, the potential of change over time, and the increasing blurring of policy domains (Kivimaa, 2022; Lindner et al., 2021; McLaren & Kattel, 2022). Measuring the additionality of innovation policies applied to complex systems is already a difficult matter (Arnold, 2004; Janssen, 2019; Magro & Wilson, 2013). However, measuring additionality in terms of transformative impact is even more complex due to the difficulty of 1) attributing transformative impact to policy measures and 2) the long timescales at which impact tends to manifest (Amanatidou et al., 2014; Haddad, 2022; Janssen, 2019; Molas-Gallart, 2021). In their evaluation framework centred around transformative outcomes, Molas-Gallart et al. (2021), for instance, emphasize the difficulty of establishing causal links between innovation policy and transformative outcomes. In a similar vein, Janssen (2019) explains that **"as transformative policy involves adaptation of elements of the specific environment firms are active in, it leads to systemic change which opens up possibilities also for firms not directly involved. The result is the lack of a good counterfactual"** (p. 82). Hence, the need for outcome/impact-based monitoring and evaluation, combined with elevated contextual complexity associated with CIPs complicates the operationalisation of challenge-led monitoring and evaluation.

Second**, attribution**. The need for outcome-based evaluations, stemming from the inverted intervention logic associated with CIPs, makes attributing results to individual policy elements increasingly difficult. Additionally, because (the solutions to) societal challenges span across various policy domains, governance levels and regional borders, CIPs are often part of a complex policy mix consisting of policy instruments, strategies, characteristics and processes spanning across multiple dimensions (Flanagan, 2011; Rogge & Reichardt, 2016). As a consequence, CIPs can rarely be evaluated outside of the policy context it is part of, magnifying the so-called attribution problem (Adam, 2018; Bovaird, 2014; Kern, 2019; Schuch, 2017). Belcher & Hughes (2021) explained that **"if the treatment itself is multi-pronged, evolving, and/or under-specified, there will be uncertainty as to which variation of the emergent research-informed innovation is responsible and for which specific effects"** (p. 160). Accordingly,

Haddad et al. (2022) identified "**attributing the effects of policy and perform-ing ex-ante evaluation**" as one of the main challenges for policy practitioners working with CIPs. The inherent limitations with regards to attribution in a complex policy mix, in conjunction with the need for outcome-based evalua-tions, has led scholars and practitioners engaged in the evaluation of CIPs to increasingly forgo establishing the effects of a single policy instrument, focus-ing on establishing contribution instead (Janssen, 2019; Kivimaa & Kern, 2016; Molas-Gallart, 2021).

Our analysis of the literature thus indicates that the monitoring and evaluation of challenge-led innovation policies requires consideration of the bottlenecks from a challenge-led perspective. While we acknowledge that not all chal-lenge-led policies require a completely novel approach to evaluation (i.e. there might not be the need to renew everything all at once, nor to renew everything every time), the identification of these bottlenecks ensures that policy-makers can identify issues that are likely to occur based on the bottleneck. Further-more, our distinction of two categories of bottlenecks shows that the challenges towards making monitoring and evaluation challenge-led is different depend-ing on the category of changes.

4.3   LITERATURE ON ADDRESSING BOTTLENECKS OF CHALLENGE-LED EVALUATION

Our analysis of relevant literature also indicates suggestions to address the bottlenecks that we discussed in the previous section. Figure 3 summarises the main issues per bottleneck, as well as the possible solutions that have been proposed or developed in the literature. Note, however, that the litera-ture screened is in itself not elaborate or clear regarding what these solutions could look like in practice, or what would ensure that these solutions will make a difference in the monitoring and evaluation practice. Next, we will discuss the potential routes to arrive at an appropriate monitoring and evaluation frame-work for challenge-led innovation policies as identified in the literature.

Figure 3: Issues and solutions of addressing bottlenecks of challenge-led evaluation

Bottleneck:	Issue:	Possible solution:
Accountability	Complexity, uncertainty & contestation	Theory-based evaluation, multi-level governance and evaluation system, separate formative & summative evaluation
Learning	Insufficient facilitation of reflexivity and adativity	Formative evaluation across governance levels and domains
Operationalisation	Ambiguity	Qualitative and normative approach, Flexible ToC
Attribution	Interaction effects, Ambiguity	Contribution to outcome, Contribution to impact
Governance & organisation	Lacking reflexive & adaptive governance, politicisation & continuity of the approach	Experimentation in governance and organizational structures
Institutional rigidity	Institutionalised evaluation practice not designed for CIP	Experimentation in evaluation approaches

Note: ToC denotes theory of change.

4.3.1 POTENTIAL ROUTES TO ADDRESS BOTTLENECKS RELATED TO A FUNDAMENTAL SHIFT IN THE ROLE OF R&I POLICY EVALUATION WITHIN A CIP CONTEXT

The **accountability** bottleneck concerns the need to establish accountability for adequately addressing societal challenges through supported transition pathways. CIPs are often nested within a complex challenge-led policy mix and are characterised by a high degree of uncertainty and contestation both in terms of challenges and solutions (Arnold et al., 2018; Belcher & Hughes, 2021; Wanzenböck et al., 2020). Taking into consideration these characteristics, scholars and evaluation practitioners have suggested providing accountability using: theory-of-change-based evaluation, a transparent and reflexive multi-level governance and evaluation system with a variety of stakeholders, and/or inclusive or participative formative evaluation combined with separate external summative evaluation (Amanatidou et al., 2014; Arnold, 2018; Belcher & Hughes, 2021; Magro & Wilson, 2019; Miyaguchi, 2022; Molas-Gallart et al., 2021). Thus, challenge-led monitoring and evaluation is defined by its transparency of what can be evaluated along the theory-of-change, and its inclusiveness of stakeholders in the monitoring and evaluation process over time.

The **learning** bottleneck deals with the elevated need for reflexivity and adaptivity that characterises CIP frameworks and imposes additional requirements

regarding the extent to which challenge-led monitoring and evaluation should facilitate learning. In response, numerous frameworks that promote learning through formative evaluation across governance levels and policy domains have been presented (Amanatidou et al., 2014; Aranguren et al.„ 2017; Janssen, 2019; Luederitz, 2017; Molas-Gallart, 2021; Räkköläinen and Saxén, 2022; Wittmann et al., 2022). Hence, challenge-led monitoring and evaluation's emphasis on learning and adaptivity emphasizes that addressing societal challenges is a process, which thus requires steering towards the societal goal.

Interestingly, the focus of and attention for accountability and learning is, in some sense, also paradoxical as it arises from an inherent tension that exists between accountability and learning. That is, as accountability is demanded but increasingly complex from a challenge-led perspective, actors might be less willing to share information that could inform learning. In addition, as evaluation with the goal of learning also concerns expenditure (time, extra staff/human capital), doing so should pay off but should also be accounted for. And, finally as accountability and learning from a challenge-led perspective demands an inclusive process, it will become increasingly difficult to actually ensure accountability (e.g. a butcher inspecting and testing its own meat) (Amanatidou et al., 2014; Arnold et al., 2018; Haddad et al., 2022; Magro & Wilson, 2019; van der Steen et al., 2018).

4.3.2 POTENTIAL ROUTES TO ADDRESS BOTTLENECKS RELATED TO THE OPERATIONALISATION AND ATTRIBUTION OF TRANSFORMATIVE IMPACT IN A CIP CONTEXT

Such paradoxical tensions appear less so considering the bottlenecks related to the operationalisation and attribution of transformative impact in a CIP context. The **operationalisation** bottleneck has to deal with ambiguity that stems from the wickedness of societal challenges, their longitudinal horizon and their blurring of policy domains. Consequently, it has been proposed that theory-based qualitative, normative and flexible approaches should be used to operationalise societal challenges for evaluation, in which complex, context-specific considerations can be taken into account more (Amanatidou et al., 2014; Aranguren, 2017; Arnold, 2018; Janssen, 2019; Luederitz, 2017; Molas-Gallart, 2021).

Finally, the main issue that the **attribution** bottleneck is confronted with are interaction effects that stem from the complexity of challenge-led policy mixes. Thus, as a solution, scholars and evaluation practitioners have suggested mov-

ing away from aiming to establish attribution, towards establishing contribution (Janssen, 2019; Kivimaa & Kern, 2016; Wittmann et al., 2022). As such, room is provided to consider the extent to which a solution contributes to addressing a societal goal (or not), and whether the policy is still doing the right thing(s).

4.3.3 ADDITIONAL POTENTIAL ROUTES TO ADDRESS BOTTLENECKS

Finally, our analysis also indicates two additional bottlenecks that appear insufficiently problematised and addressed in the literature and policy practice. These bottlenecks appear as they are presented as (part of) solutions to all four identified bottlenecks, but they are presented as an afterthought that needs 'experimentation'. These bottlenecks are: 1) 'governance and organisation' that is lacking in reflexivity and adaptivity and increasingly politicised at the expense of the continuity of the approach, and 2) 'institutional rigidity' sustaining an institutionalised evaluation culture and practice that is not designed to fulfil the evaluation needs within CIP frameworks. These bottlenecks highlight potentially significant issues and solutions concerning further development of challenge-led monitoring and evaluation. That is, ameliorating other identified bottlenecks in challenge-led monitoring and evaluation is contingent upon altering (a) governance and organisational structures as well as (b) institutionalised assumptions and practices. Therefore, we argue that in order to further develop challenge-led monitoring and evaluation, the roles of governance and organisational structures as well as institutionalised assumptions and practices should be problematised and prioritised as bottlenecks.

5. DISCUSSION – CONCLUSIONS AND RECOMMENDATIONS

Our literature review has shown that existing literature provides ample input on the bottlenecks, issues and possible solutions to develop challenge-led monitoring and evaluation (see Figure 2 and 3). By delineating the bottlenecks and their potential solutions, we aim to provide policy-makers and evaluators strategic guidance to understand what it takes to develop a challenge-led monitoring and evaluation 'fit-for-purpose', and which issues might hamper possible ways forward.

An important implication of our review is that governance and institutionalised monitoring and evaluation culture and practice are important bottlenecks

to consider. While existing literature does not considered these a bottleneck per se, the literature does mention that both governance and institutionalised culture and practice both require 'experimentation' (see Figure 3). To develop challenge-led monitoring and evaluation further, we therefore argue that we should de-emphasise the well-known bottlenecks and focus instead on the governance and organisation of monitoring and evaluation for societal challenges, and question the institutionalised monitoring and evaluation culture and practices. That is, we argue that altering governance and organisational structures as well as institutionalised assumptions and practices should be prioritised because ameliorating other identified bottlenecks in challenge-led evaluation is contingent upon doing so. In other words, while the main issues of the bottlenecks towards a challenge-led monitoring and evaluation can be indicated (see Figure 2), the potential routes forward are largely dependent on the institutionalisation of challenge-led monitoring and evaluation practices that are incorporated into the governance and organisation. After all, a challenge-led monitoring and evaluation is not automatically ingrained, and therefore a different design, and a change of mind-set and practices within the existing monitoring and evaluation practice is required. We thus argue that knowing that challenge-led monitoring and evaluation is necessary does not automatically lead to a different monitoring and evaluation practice. Current practices, with their habits and routines, standards, and socio-culturally accepted norms regarding monitoring and evaluation, whether imposed or not, play a role and are difficult to set aside. What is needed, therefore, is a culture change and the institutionalisation of a new, monitoring and evaluation norm or standard.

As a result, challenge-led monitoring and evaluation touches on a more institutional question of recalibrating how we deal with monitoring and evaluation in policy assessments and society at large, and what the role of policymakers, evaluators, and financiers and intermediaries is in this context. We thus argue that this not only requires further reflection and elaboration of the meaning and consequences of monitoring and evaluating bottlenecks for societal challenges, but above all a change in the governance and organisation of monitoring and evaluation as well as the culture regarding the where, who, when and how monitoring and evaluation should be carried out. That is, as the bottlenecks of challenge-led evaluation ensure the emergence of novel practices, actors, or owners that are not institutionalised, these practices, actors and owners compete with highly institutionalised ones. Consequently, these competitions are the institutional sources of contestations that policymakers, evaluators, financiers and intermediaries that want to develop a challenge-led evaluation are likely to be confronted with.

5.2 CONCLUSION

An emerging stream of research can be identified that sets out to identify and address the bottlenecks of developing monitoring and evaluation approaches for challenge-led innovation policies. Although these efforts are still fragmented, they are starting to contribute to the establishment of a common understanding of how the emergence of challenge-led innovation policies impacts the requirements for monitoring and evaluation, and how challenge-led monitoring and evaluation could strengthen policy and governance of research and innovation aimed at societal challenges. This understanding helps in the development of challenge-led monitoring and evaluation approaches 'fit-for-purpose'. However, to establish consensus, new norms among actors, and novel practices regarding the monitoring and evaluation of challenge-led innovation policies, dealing with identified governance and institutional bottlenecks requires more attention.

REFERENCES

Adam, C., Steinebach, Y., & Knill, C. (2018). Neglected challenges to evidence-based policy-making: The problem of policy accumulation. **Policy Sciences**, **51**(3), 269-290.

Amanatidou, E., Cunningham, P., Gök, A., & Garefi, I. (2014). Using evaluation research as a means for policy analysis in a 'new'mission-oriented policy context. **Minerva**, **52**(4), 419-438.

Aranguren, M. J., Magro, E., & Wilson, J. R. (2016). Regional competitiveness policy evaluation as a transformative process: From theory to practice. **Environment and Planning C: Politics and Space,** 35(4), 703-720.
Arnold, E. (2004). Evaluating research and innovation policy: a systems world needs systems evaluations. **Research evaluation**, 13(1), 3-17.

Arnold, E., Åström, T., Glass, C., & de Scalzi, M. (2018). **How should we evaluate complex programmes for innovation and socio-technical transitions?** Retrieved from technopolis |group| United Kingdom website: http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-250284

Belcher, B.M. & Hughes, K. (2021). Understanding and evaluating the impact of integrated problem-oriented research programmes: Concepts and consider-

ations. **Research Evaluation**, 30(2), 154–168, https://doi.org/10.1093/reseval/rvaa024

Berger, P. L. & Luckmann, T. (1966). **The Social Construction of Reality: A Treatise in the Sociology of Knowledge**. Garden City, NY: Anchor Books.

Borrás, S. (2011). Policy learning and organizational capacities in innovation policies. **Science and Public Policy**, 38(9), 725-734.

Bovaird, T. (2014). Attributing outcomes to social policy interventions–'gold standard'or 'fool's gold'in public policy and management?. **Social Policy & Administration**, 48(1), 1-23.

Bush, V., (1980). Science--the endless frontier : a report to the President on a program for postwar scientific research. [Washington, D.C.] :**National Science Foundation.**

Diercks, G., Larsen, H. & Steward, F. (2019). Transformative innovation policy: Addressing variety in an emerging policy paradigm. **Research Policy** 48, 880–894.

Duckett, D. Feliciano, D. Martin-Ortega, J. Munoz-Rojas, J. (2016). Tackling wicked environmental problems: the discourse and its influence on praxis in Scotland. **Landsc. Urban Plan.**, 154, pp. 44-56

Elsbach, K.D. & van Knippenberg, D. (2020). Creating high impact literature reviews: An argument for integrative reviews. **Journal of Management Studies**.

Flanagan, K., Uyarra, E., & Laranja, M. (2011). Reconceptualising the 'policy mix'for innovation. **Research policy**, 40(5), 702-713.

Grillitsch, M., Hansen, T., Coenen, L., Miörner, J., & Moodysson, J. (2019). Innovation policy for system-wide transformation: The case of strategic innovation programmes (SIPs) in Sweden. **Research Policy**, 48(4), 1048-1061.
Haddad, C. R., Nakić, V., Bergek, A., & Hellsmark, H. (2022). Transformative innovation policy: A systematic review. **Environmental Innovation and Societal Transitions**, 43, 14-40.

Hajer, M. (1995). **The politics of environmental discourse: ecological modernization and the policy process.** Clarendon Press.

Hekkert, M. P., Janssen, M. J., Wesseling, J. & Negro, S. O. (2020). Mission-oriented innovation systems. **Environmental Innovation and Societal Transitions**, 34, 76-79.

Hertting, N., & Vedung, E. (2012). Purposes and criteria in network governance evaluation: How far does standard evaluation vocabulary takes us?. **Evaluation**, 18(1), 27-46.

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. & Wanzenböck, I. (2021). The promises and premises of mission-oriented innovation policy—A reflection and ways forward. **Science and Public Policy**, 48(3), 438–444.

Kattel, R., & Mazzucato, M. (2018). Mission-oriented innovation policy and dynamic capabilities in the public sector. **Industrial and Corporate Change**, 27(5), 787–801. https://doi.org/10.1093/icc/dty032

Kern, F., Rogge, K. S., & Howlett, M. (2019). Policy mixes for sustainability transitions: New approaches and insights through bridging innovation and policy studies. **Research Policy**, 48(10), 103832.

Kivimaa, P. (2022). Transforming innovation policy in the context of global security. **Environmental Innovation and Societal Transitions**, 43, pp. 55-61

Kivimaa, P., Kangas, H. L., & Lazarevic, D. (2017). Client-oriented evaluation of 'creative destruction'in policy mixes: Finnish policies on building energy efficiency transition. **Energy Research & Social Science**, 33, 115-127.

Kivimaa, P., & Kern, F. (2016). Creative destruction or mere niche support? Innovation policy mixes for sustainability transitions. **Research policy**, 45(1), 205-217.

Kuhlmann, S. & Rip, A. 2018. Next-Generation Innovation Policy and Grand Challenges. **Science and Public Policy**, 45(4): 448–454.

Lindner, Ralf et al. (2021). Mission-oriented innovation policy: From ambition to successful implementation, Perspectives - Policy Brief, No. 02 / 2021, FraunhoferInstitut für System- und Innovationsforschung ISI, Karlsruhe.

Loorbach, D. (2010). Transition Management for Sustainable Development: A Prescriptive, Complexity-Based Governance Framework. **Governance**, 23(1), 161–183. https://doi.org/https://doi.org/10.1111/j.1468–0491.2009.01471.x

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. & Westley, F. R. (2017). Learning through evaluation – A tentative evaluative scheme for sustainability transition experiments. **Journal of Cleaner Production**, 169, 61- 76.

Lundvall, B. A. (1992). National systems of innovation: towards a theory of innovation and interactive learning.

Molas-Gallart, J., Boni, A., Giachi, S., & Schot, J. (2021). A formative approach to the evaluation of Transformative Innovation Policies. **Research Evaluation**, 30(4), 431-442.

Magro, E., & Wilson, J. R. (2013). Complex innovation policy systems: Towards an evaluation mix. **Research policy**, 42(9), 1647-1656.

Magro, E., & Wilson, J. R. (2019). Policy-mix evaluation: Governance challenges from new place-based innovation policies. **Research policy**, 48(10), 103612.

Mazzucato, M. (2018). Mission-oriented innovation policies: challenges and opportunities. **Industrial and Corporate Change,** 27, 803–815.

McLaren, J. and Kattel, R. (2022). Policy capacities for transformative innovation policy: A case study of UK Research and Innovation. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2020-04). Available at: https://www.ucl.ac.uk/bartlett/public–purpose/wp2022–04.

Miyaguchi, T. (2022). Importance and Utilization of Theory-Based Evaluations in the Context of Sustainable Development and Social-Ecological Systems. In: Uitto, J.I., Batra, G. (eds) **Transformational Change for People and the Planet. Sustainable Development Goals Series.** Springer, Cham. https://doi.org/10.1007/978–3–030–78853–7_15

OECD (2020), **Improving Governance with Policy Evaluation: Lessons From Country Experiences**, OECD Public Governance Reviews, OECD Publishing, Paris, https://doi.org/10.1787/89b1577d-en.

Räkköläinen, M., Saxén, A. (2022). Pathway to the Transformative Policy of Agenda 2030: Evaluation of Finland's Sustainable Development Policy. In: Uitto, J.I., Batra, G. (eds) **Transformational Change for People and the Planet. Sustainable Development Goals Series**. Springer, Cham. https://doi.org/10.1007/978-3-030-78853-7_16

Rathenau Institute (2020). **Maak werk van opgavegericht innovatiebeleid: Bericht aan het parlement.** Den Haag: Rathenau Institute.

Rathenau Institute (2021). **EU-missies voor maatschappelijke opgaven – Inspiratie uit Horizon Europe voor opgavegericht innovatiebeleid**. Den Haag: Rathenau Institute.

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. **Policy sciences**, 4(2), 155-169.

Robinson, D. K. R., & Mazzucato, M. (2019). The evolution of mission-oriented policies: Exploring changing market creating policies in the US and European space sector. **Research Policy**, 48(4), 936–948. https://doi.org/10.1016/j.respol.2018.10.005

Rogge, K. S., & Reichardt, K. (2016). Policy mixes for sustainability transitions: An extended concept and framework for analysis. **Research Policy**, 45(8), 1620-1635.

Schot, J. & Steinmueller, W. E. (2018). Three frames for innovation policy: R&D, systems of innovation and transformative change. **Research Policy**, 47(9), 1554-1567.

Schuch, K., Campbell, D., Carayannis, E. G. and Edler, J. (2017): Evaluation of Research, Development, and Innovation. In: **Encyclopedia of Creativity, Invention, Innovation and Entrepreneurship**; p. 1-8; share and cite: doi:10.1007/978-1-4614-6616-1_200025-1

Scott, R. (1995). **Institutions and organizations**. London, Sage Publications.

van der Steen, M., Faber, A., Frankowski, A., & Norbruis, F. (2018). Opgavegericht evalueren. **Beleidsevaluatie voor systeemverandering**. Den Haag: NSOB.

Termeer, C. J., & Dewulf, A. (2019). A small wins framework to overcome the evaluation paradox of governing wicked problems. **Policy and Society**, 38(2), 298-314.

Torraco, R. J. (2016) Writing integrative literature reviews: Using the past and present to explore the future. **Human Resource Development Review**, 15(4).

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

Wanzenböck, I,. Wesseling, J.H., Frenken, K., Hekkert, M.P., Weber, K.M., A framework for mission-oriented innovation policy: Alternative pathways through the problem–solution space, **Science and Public Policy**, Volume 47, Issue 4, August 2020, Pages 474–489, https://doi.org/10.1093/scipol/scaa027

Weber, K. M., & Rohracher, H. (2012). Legitimizing research, technology and innovation policies for transformative change: Combining insights from innovation systems and multi-level perspective in a comprehensive 'failures' framework. **Research policy**, 41(6), 1037-1047.

Weber, M.; Polt, M. (2014): Assessing mission-orientated R&D programs: combining foresight and evaluation. **Fteval - Journal for Research and Technology Policy Evaluation**, (39), 5–10.

Weiss, C. H. (1995). Nothing as Practical as Good Theory : Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families. **New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts**, 1, 65--92.

Wittmann, F., Hufnagl, M., Roth, F., Lindner, R., & Kroll, H. (2022). Towards a framework for impact assessment for mission-oriented innovation policies. A formative Toolbox approach. **Fteval – Journal for Research and Technology Policy Evaluation**, 53, 31-42.

AUTHORS

CORRESPONDING AUTHOR:
VINCENT BAARSLAG
Rathenau Instituut, Anna van Saksenlaan 51, The Hague,
The Netherlands
Email: v.baarslag@rathenau.nl, vincent.baarslag@gmail.com
Orcid: https://orcid.org/0000-0003-2030-053X

AMBER GEURTS
Institution & Address: TNO, TNO Vector, Anna van Buerenplein 1, The Hague,
The Netherlands
Email: amber.geurts@tno.nl
Orcid: https://orcid.org/0000-0003-0478-0713

FRANS VAN DER ZEE
Institution & Address: Rathenau Instituut, Anna van Saksenlaan 51, The
Hague,
The Netherlands
Email: f.vanderzee@rathenau.nl