Chapter 5. Exploring the tensions between accountability and learning: the need for alignment

Chapter 2 emphasised alignment as an important aspect of the governance of niche experiments with a Triple P purpose because the multistakeholder context in which niche experiments take place gives rise to different requirements. These requirements may hamper the development of the niche experiment but cannot be ignored if the Triple P business is to survive without niche protection. Alignment needs to be created between people, planet and profit values, but also between the niche experiment and its various institutional contexts, including the funder of the niche experiments. This chapter focuses on this latter type of alignment, exploring the tensions between the more accountability-oriented needs of funders and the learning needs of managers of niche experiments. By looking more closely at the differences in needs and expectations, I formulate several strategies that managers may use to align the emergent learning approach often taken in the governance of niche experiments with the accountability needs of funders.

5.1 Introduction

While today both accountability and learning are considered important motives for programme evaluation, the literature shows that it is not self-evident that evaluation focuses on both motives at the same time. Different scholars state that accountability and learning are incompatible, and therefore a trade-off exists between accountability and learning as reasons for and results of evaluation (Feinstein, 2012; Guijt, 2010). Tensions thus seem to exist between accountability and learning (e.g. Van Der Meer & Edelenbos, 2006; Chouinard, 2013; Newcomer & Olejniczak, 2013). These apparent tensions between accountability and learning pose challenges to evaluators. Although evaluators are increasingly requested to facilitate and support learning within programmes, the call for accountability remains. In this chapter we will look at these tensions in more detail in the context of niche experiments with a Triple P purpose (integrating people, planet and profit values): projects aiming at changing current practices into a more sustainable direction. We will start with a theoretical exploration of accountability in the context of evaluation, distinguishing three types of
accountability: upwards, downwards and internal accountability. Subsequently, the needs of both funders and project managers of niche experiments with a Triple P purpose with regard to upwards, downwards and internal accountability are scrutinised, resulting in an adjusted accountability framework for the evaluation of niche experiments. Moreover, three alignment strategies are induced from our study, which may support bringing in line the different needs for evaluation. We conclude with a reflection on evaluation as a way to develop the (dynamic) capabilities that are necessary in today’s complex and changing society.

5.2 Niche experiments and accountability

System innovation programmes make it possible for Triple P projects\(^\text{19}\) to operate in a protected space, a ‘niche’ in which selection pressures prevailing in current regimes, such as regulatory requirements and demands for competitiveness and profitability, are reduced during the most vulnerable start-up period (Geels & Schot, 2007; Schot & Geels, 2008, see Figure 5.1b). Without this protection the projects would risk losing their radical, Triple P nature, and soon turn into business as usual (see Figure 5.1a). At the same time, these projects do find themselves in several institutional contexts that cannot be ignored: all those involved in the development of the Triple P project have to deal with requirements from their respective institutions; participating entrepreneurs have to weigh the expected return on investment, participating scientists are required to publish in peer-reviewed journals, participating policy-makers need to show output in terms of realised policy objectives and participating NGOs need to show to their supporters to what extent their interest is incorporated. Moreover, system innovation programmes support niche experiments through different means, often including financial support, which brings in a financial accountability dimension (see Figure 5.1b). Thus, whilst niche experiments, by their very nature, are protected from surrounding systems, questions of accountability to external parties do arise.

\(^\text{19}\) In this article niche experiment and project are used interchangeably (corresponding to use in literature and wordings chosen by interviewees), referring to initiatives aiming at sustainable development.
Exploring the tensions between accountability and learning: the need for alignment

According to the Webster dictionary, accountability means "an obligation or willingness to accept responsibility or to account for one's actions"\(^{20}\). This counts for public officials, as they are accountable to their electorate, for businesses, as they are accountable to their shareholders, but also for individuals, taking responsibility for one's own actions in all domains of lives. In the scientific discussion around the relationship between NGOs and their funders, accountability is commonly defined as “the means by which individuals and organisations report to a recognised authority and are held responsible for their actions” (Edwards & Hulme 1996, p967). In this context accountability is mostly understood in terms of reporting on funds in relation to activities and outcomes of these activities; organisations are expected to report to what extent the designated money has been spend on the designated purposes (Najam, 1996). Organisations they need to report to are funding organisations, such as donors, investors and governments that set reporting requirements to the initiative or enterprise they fund. This is also true for niche experiments, which are generally funded through large governmental innovation programmes, such as the Dutch interdepartmental programmes ICES/KIS I-III to strengthen the country’s research and development capacity or through departmental innovation programmes in particular domains (e.g. agriculture, or mobility).

Evaluation as a mechanism for accountability

There are several mechanisms for accountability, such as annual project reports and financial audits, but evaluation in particular is a widely used mechanism for accountability (Ebrahim, 2005; Ovens et al., 2012). The focus of evaluation for accountability is on assessing the effectiveness and efficiency of a programme or project (Feinstein, 2012). Effectiveness answers the question whether programmes or projects are delivering the expected results (Lehtonen, 2005) while efficiency refers to the question whether the right means for achieving the objectives were employed and often entails a cost-benefit analysis (Feinstein, 2012). The role of the associated performance measurement or programme evaluation is generally to gain insight into the outputs and outcomes (Jann & Wegrich, 2007), based on the systematic collection and analysis of information about the activities and results (Van Der Knaap, 2001) to provide evidence of impact and effectiveness of the programme (De Lancer Julnes, 2006). Often logic models (Chen, 1990), logical frameworks (e.g. Baccarini, 1999), programme theories (Patton, 2008) or theories of change (Connell et al., 1995; Fulbright-Anderson et al., 1998) are used to specify the rational underpinning of a particular programme, by ex-ante explicating expected relations between inputs, activities and desired outcomes which are used ex-post to assess whether and to what extent programme goals and objectives have been achieved.

While it is legitimate for a funder to require insight into how their funds have been spent, whether it is an investor in a new business or a donor of a non-profit organisations (Ebrahim, 2005), the focus of evaluation on accountability is also perceived to have considerable unfavourable effects (see e.g. Benjamin, 2008; Ebrahim, 2005; Joshi & Houtzager, 2012; Lehtonen, 2014; O’Dwyer & Unerman, 2007; Perrin, 2002; Richmond et al., 2003). As organisations often rely on positive evaluations for future funding, satisfying the donors needs for information becomes not only a time-consuming activity, especially if this is not considered relevant for internal decision making (Ebrahim, 2005), it also feeds 'strategic misrepresentation' (Lehtonen, 2014). Another adverse effect of the emphasis in evaluation on accountability to funders is that discrete, measurable, and proven approaches using planned project management are favoured over innovative, uncertain and more risky approaches that require

---

21 In evaluation literature one often speaks of programmes rather than projects. In the practice of niche experiments, it is often hard to distinguish between projects and programmes. We have used all terms interchangeably throughout the article.
emergent, governance approaches (Ebrahim, 2005; Lehtonen, 2014). Dominant forms of programme evaluation (e.g. performance measurement) presuppose a relatively stable programme, of which the activities, goals and intended effects can be univocally described (Regeer et al., 2009). However, a growing body of literature shows that many contemporary projects, programmes and initiatives, ranging from those that aim for social change as a response to persistent problems (e.g. sustainable development, poverty alleviation), to so-called megaprojects (e.g. nuclear waste facilities, large infrastructural projects), and Triple P business initiatives, have longer term objectives, that are more intangible (Ebrahim, 2005) and moreover, that are redefined during the course of the process (de Wildt-Liesveld et al., 2013) which, in turn, may fundamentally change the nature, scope and rationale of the project (Lehtonen, 2014). As Ebrahim (2005, p61) points out:

for an organisation “... that aims to feed schoolchildren a daily warm breakfast, there may be no problem with regular reporting on the number of children fed. But for an organization that aims to address broader public policies concerning urban poverty, such measures may provide limited useful information on how to tackle long-term systemic change.”

Similarly, in the case of megaprojects, many evaluation approaches “erroneously assume that the complexity and irreducible uncertainties inherent in megaprojects can indeed be controlled through careful ex ante planning of appropriate governance measures” (Lehtonen, 2014, p283, referring to Sanderson, 2012).

Naturally, this has consequences for the role evaluation can play in these processes. Whilst goal-oriented evaluation is an important mechanism for financial accountability, it does not sufficiently take into account the emergent nature of complex projects and their multifaceted environment. Over recent years, there has been a gradual shift from a government-centred policy paradigm, warranting goal-oriented programme evaluation, towards a more inclusive policy making model, known as governance (Rhodes, 1997) or deliberative democracy (Fischer, 2003). This shift has been accompanied by the emergence of a new evaluation paradigm with the focus on learning. Within this learning-oriented evaluation paradigm, rather than accounting for funds spent, evaluation aims to support “the process of continuous reflection on visions, strategies, actions and contexts that enable continual readjustments” (Guijt, 2010, p281). Increasingly, emphasis is placed on incorporating evaluation in the intervention process (see e.g. Friedman, 2001; Patton, 2008) by performing evaluation ex-durante rather than ex-post evaluation. Also, since Guba & Lincoln’s eminent work on
fourth generation evaluation (1989), the call for involving end-users in designing evaluation frameworks and gathering and interpreting data, has only gained prominence. Thus, in order for evaluation methodologies to be utilization focused (Patton 2008), they should be participatory (e.g. Cousins & Earl, 1992), responsive (Abma & Stake, 2001), systemic (e.g. Cabrera et al., 2008; Williams & Imam, 2007), and reflexive (e.g. Arkesteijn et al., 2015). In Table 5.1, with the risk of oversimplifying matters, the main differences between goal-oriented evaluation, directed towards financial accountability, and learning-oriented evaluation as they are often portrayed in literature, are exposed.

<table>
<thead>
<tr>
<th>Evaluation for accountability</th>
<th>Evaluation for learning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goal of evaluation</strong></td>
<td>Assessing the outcomes of a programme, focusing on effectiveness and efficiency.</td>
</tr>
<tr>
<td><strong>Outcomes/ result of evaluation</strong></td>
<td>Feedback to funders on how money was spent</td>
</tr>
<tr>
<td><strong>Evaluation phase</strong></td>
<td>Ex post evaluation of goals</td>
</tr>
<tr>
<td><strong>Evaluation methodology</strong></td>
<td>Applying an established methodology</td>
</tr>
<tr>
<td><strong>Involvement of relevant actors</strong></td>
<td>Little/no stakeholder participation</td>
</tr>
</tbody>
</table>

While both accountability and learning are considered important motives for programme evaluation, the literature shows that it is not self-evident that programme evaluation focuses on both motives at the same time. As we have seen, accountability and learning are often seen as incompatible. But are accountability and learning really two opposing features of evaluation? Let us return to the concept of accountability and examine potential relations to learning more meticulously.

**Downwards accountability and internal accountability**

The common conceptualisation of accountability as we have discussed so far has also been referred to as upwards accountability (Ebrahim, 2005; Edwards & Hulme, 1996),
functional accountability (O’Dwyer & Unerman, 2007) or financial accountability (Richmond et al., 2003). It concerns relationships with funders and often reflects a hierarchical relationship (hence ‘upwards’ accountability) between ‘principal’ (in this case the funder) and ‘agent’ (the project), where the principal uses evaluation “as a tool to ensure desired results (directing and motivating action)” (Benjamin, 2008, p326). However, in the corporate world, development field, and the public sector alike, it is recognised that organisations are accountable to multiple actors, notably to those that are affected by the programmes and projects they execute. For instance, in the case of health care, accountability is defined as “a multi-layered and multifaceted component of the ethical relationship of health care providers and institutions to patients and consumers” (Cassel & McParland, 2002, 250). And in the development field, NGOs are considered to be accountable to “groups to whom NGOs provide services” although it may also include communities or regions indirectly affected by NGO programs (Najam, 1996, p345, as cited in Ebrahim, 2005, p60). In this field, this type of accountability has been termed “downwards” accountability (Edwards & Hulme, 1996, Ebrahim, 2005), however, it is closely linked to the concept of ‘social accountability’ that has been introduced in the public sector as the “ongoing and collective effort [of actors in civil society] to hold public officials to account for the provision of public goods which are existing state obligations” (Houtzager & Joshi, 2008, p3). The concept of downwards accountability opens up many possibilities for more learning oriented evaluation approaches, in particular participatory approaches that involve multiple stakeholders in the endeavour.

Finally, next to upwards accountability and downwards accountability, organisations, programmes and projects are said to be accountable to themselves, to their own mission, which is referred to as ‘internal accountability’ (Ebrahim, 2005, p60). In this case, evaluation is not employed to assess progress towards externally defined objectives, but towards internal goals and missions. Moreover, in situations of changing external environments and unexpected effects, evaluations have the potential to assist in making timely adjustments, of objectives as well as programme activities, and facilitate day-to-day decision-making.

Reconciling accountability with learning through evaluation

Although in the context of literature on evaluation, evaluation for accountability and evaluation for learning are often presented as oppositional (as in Table 5.1), a closer examination of the concept of accountability (following Ebrahim, 2005, see also
Andreanus & Costa, 2014; Siddiquee & Faroqi, 2009) allows for a more nuanced view, with different roles of evaluation for different types of accountability (see Table 5.2).

Table 5.2. Reconciling accountability and learning through evaluation

<table>
<thead>
<tr>
<th>Accountability</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Upwards accountability</strong></td>
<td><strong>Accountable to:</strong></td>
</tr>
<tr>
<td></td>
<td>• Funders of initiative or programme (investors, government programmes, donors)</td>
</tr>
<tr>
<td></td>
<td>Goal-oriented programme evaluation (e.g. performance measurement)</td>
</tr>
<tr>
<td><strong>Downwards accountability</strong></td>
<td><strong>• Those affected by the programme</strong> (beneficiaries, service-users, citizens etc.)</td>
</tr>
<tr>
<td></td>
<td>Participatory evaluation, responsive evaluation</td>
</tr>
<tr>
<td><strong>Internal accountability</strong></td>
<td><strong>• Mission of the initiative, programme or organisation</strong></td>
</tr>
<tr>
<td></td>
<td>Utilization focussed evaluation, reflexive evaluation</td>
</tr>
</tbody>
</table>

The first type of evaluation, evaluation for upwards accountability clearly serves the purpose of assessing performance, or effectiveness, as well as efficiency of organisations, programmes or projects by gauging resource use in relation to services provided and impacts achieved. Learning may occur by considering the outcomes of the evaluation, in terms of lessons learned, to improve subsequent projects or activities. However, it has been argued that the potential contribution of evaluation for accountability to subsequent decision making about future actions is limited (e.g. Patton, 2008), unless outcomes are so poor that future funding is threatened (Ebrahim, 2005). An important reason is that evaluation is commonly considered to be a practice performed by external evaluators and as such is separated from the learning process of project managers or organisations. Therefore, the third type of evaluation emphasises the potential of evaluation to contribute to internal learning processes and decision-making by working together closely with project teams or staff during the process of intervention, and is closely related to process of organisational learning and change. The aim is to conduct the evaluation in such a way that results can be used in practice (utilisation-focused, Patton, 2008) and that it leads to adjustments along the way. Adjustment can include practical improvements, strategic adjustments and rethinking the core driving values, corresponding to single, double and triple loop learning respectively (Guijt, 2010), making the evaluation reflexive. The second type of evaluation corresponds to downwards accountability and starts from a multi-constituency perspective (D’Aunno, 1992), involving relevant stakeholders at different stages of the evaluation process, to assess not only the outcomes or impacts of programmes, but also to "reassess the desirability of those very outcomes" (Ebrahim
2005, p70). Participatory evaluation practices have in the past two decades proliferated under banners ranging from multi-stakeholder evaluation, participatory action research, 360-degree evaluation, and empowerment evaluation.

The above framework (Table 5.2) brings together accountability and learning through evaluation by expanding the notion of accountability with downwards and internal accountability. It is especially the latter two that warrant the use of learning oriented evaluation methodologies. However, the question remains whether evaluation for upwards accountability and evaluation for learning are reconcilable. To gain insights in the supposed accountability versus learning dichotomy regarding evaluation, we have examined the views of both funders and project managers of programmes and projects aiming at sustainable development in a broad sense (addressing amongst others sustainable agriculture, mobility, and regional development), which we refer to as niche experiments with a Triple P purpose. This chapter compares the needs and requirements of project managers and funders regarding the evaluation of these multifaceted initiatives. Based on the experiences, we formulate strategies that can be put in place by managers to comply with the requirements of funders in a way that will not jeopardise the flexibility required to manage these Triple P initiatives.

5.3 Methodology

This study explores the tensions between accountability and learning in the context of niche experiments with a Triple P purpose. This exploration is based on a qualitative data set, comprising 30 semi-structured interviews conducted between 2008 and 2012.

The 30 semi-structured interviews comprised 17 interviews with project managers and 13 interviews with funders. The interviewees referred in this chapter to as project managers, were all involved in the management of niche experiments in the context of the sustainable development of agriculture. Some managers were involved as initiating entrepreneurs (n=2), and others as consultants (n=3). The largest number of interviewees (n=12), was involved as intermediary project manager, appointed by system innovation programmes. All interviewed funders were involved in system innovation programmes aiming at sustainable development, either from the government (national level (n=8), provincial level (n=2)), or as funders of niche experiments conducted within system innovation programmes (n=3).
Respondents were interviewed at their work place and each interview lasted about one hour. All interviews were audiotaped after informed consent and transcribed verbatim. Summaries of the interviews were sent back to the interviewees for member check. The interviews were analysed by two independent researchers on the needs and expectations regarding upwards, downwards and internal accountability, resulting in an adjusted accountability framework for niche experiments with a Triple P purpose.

In the next two sections we will examine the diverse needs, expectations and experiences regarding evaluation expressed by both funders and project managers. As we consider evaluation to be an important accountability mechanism, the expressed needs will give insight into the spectrum of ‘forms of accountability’ as experienced (or enforced) by the interviewees. Evaluation needs, and their underlying reasons, thus reflect the broader context in which interviewees operate, and the multiple demands they need to comply with in order to reach their purpose. We will see that these complex contexts do not only form constraints but they also provide possibilities to reconsider the tension between accountability and learning.

5.4 Funders’ and project managers’ evaluation needs in context

5.4.1 Funders’ evaluation needs in context

When policy makers were asked for their main reason for administering evaluations of niche experiments, they generally expressed two interrelated needs: the need to determine the effectiveness and the need to determine the efficiency of a programme of project, both related to upwards accountability.

One of the policy makers describes ‘effectiveness’ as “the extent to which the programme achieved the things we wanted it to achieve and thus contributed to the predefined policy goals”. Whether a programme is effective in contributing to the overall policy of the Ministry is assessed through answering questions such as: What are the results of the programme? What are the effects of the programme on the problem that it was designed to solve? Is the problem solved or have the activities aggravated the problem? Or as one policy maker put it: “Can you make it plausible that the programme contributed to the policy goals?” This question is important to answer as “it doesn’t matter how well you carry out your programme: the Minster has to be able to prove that it is safer or more sustainable after the execution of the programme.” To determine if a programme contributed to the policy goals, evaluation should focus
on assessing whether the policy programme achieved the predefined objectives and outcomes.

**Table 5.3. Evaluation needs of funders in relation to spectrum of accountability**

<table>
<thead>
<tr>
<th>Accountability</th>
<th>Evaluation needs funders</th>
<th>Reasons for this need</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upwards accountability</td>
<td>• Need insight into effectiveness and efficiency of programme</td>
<td>• Political accountability</td>
</tr>
<tr>
<td></td>
<td>• To determine whether a programme should be adjusted or terminated</td>
<td></td>
</tr>
<tr>
<td>Downwards accountability</td>
<td>• -</td>
<td>• -</td>
</tr>
<tr>
<td>Internal accountability</td>
<td>• Need insight into outcomes at project level</td>
<td>• To adjust policy or innovation programme</td>
</tr>
<tr>
<td>regime level</td>
<td></td>
<td>• To define new programmes</td>
</tr>
<tr>
<td>Internal accountability niche</td>
<td>• Need insight into the process at project level</td>
<td>• To support continuous reflexive process of policy making</td>
</tr>
<tr>
<td>level</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

‘Efficiency’ was seen as another important objective of policy evaluation. According to the policy makers, “efficiency considers the question of whether the investment of two million Euros has been well spent.” Efficiency thus refers to the spending of resources in relation to results obtained. Was money spent efficiently or could the programme have spent less money while achieving the same results? For the interviewees it is important to answer these questions and to be transparent about the money spent because it is public money from taxes. Naturally, in the political context, financial accountability is closely related to political accountability, as politicians are accountable to their electorate, via their representatives in parliament. Insight into the effectiveness and efficiency of policy programmes is required to answer questions from the House of Representatives (Kamervragen in Dutch). These questions can focus on diverse aspects, such as the coherence of a policy programme, the amount of money spent and results. As one policy maker noted: “A question that could be posed by the House of Representatives is: Are you aware of the fact that you spent over two million Euros on the programme ‘Growing with Future’? Can you inform me on the effects of this invested capital?” Explaining the role of policy makers in answering this question, he continued with; “When a parliamentary question regarding our programme is posed, I have to convince the Minister in five minutes that the two million Euro investments was spent well.” The Minister then uses this information during a meeting of the House of Representatives to convince representatives that the programme has been effective in realising its goals. The needs of funders regarding evaluation for upwards accountability can be found in Table 5.3. We found that none of the funders mentioned
needs regarding downwards accountability. Evaluation needs regarding internal accountability will be discussed below.

The information on effectiveness and efficiency of a programme can also be used by the Dutch government to adjust or terminate a programme when, according to one policy maker, “the programme does not deliver the predefined objectives.” Although the main reason for evaluating effectiveness and efficiency is to ensure the programme is accountable (to the funder, to parliament), the process of adjusting or even terminating a policy programme also involves a process of learning from the outcomes of the policy evaluation, which brings us to the topic of internal accountability.

As the literature discussed above primarily focusses on questions of accountability from the perspective of programmes or projects, internal accountability is also understood at the level of those receiving funding rather than those providing funding. Rather than only complying with the requirements of a funding agency, donor or grant programme, receiving teams also need to comply with their own mission, longer-term vision and ideas about ethical conduct. This is referred to as internal accountability. Interviews with funders show that internal accountability also plays at the level of the funder: they too want to learn from the conducted projects to assess and possibly redefine objectives or adjust grant programmes.

First, the interviewees from national government said they learned from policy evaluations in order ‘to revise current policies’. For example, policy goals are assessed and tested, and adjusted when necessary. As one of the policy makers said: “The evaluation can be used as input to revise the current policy. Some points of the policy document may not be achieved within the programme. The question then is why they are not achieved: did we use the right resources? The answers to these questions may result in an adjustment of the policy objective.”

Outcomes of policy evaluation are also used to guide the content of new policies, providing thoughts and ideas for upcoming policies and programmes. As one of the policy makers mentioned: “On account of the data of an evaluation, you may develop the policy for the upcoming years. For this, you have to determine the learning points or the points that are still under developed. These points you can amplify and implement in new policies.” Thus, learning from a programme, in particular related to areas not yet achieved, may form the foundation of new policies. In addition, programmes may also reveal new insights or new questions to be put on the policy agenda.
In the case of funders from innovation programmes outside of governmental departments, internal accountability is of particular interest. More than departmental policy makers, they emphasise the need for a continuous feedback loop between project activities and results and the overall objective of the grant programme.

Although the main focus of the needs and expectations regarding evaluation of policy makers is on upwards accountability and related to this, internal accountability at regime level, some of the interviewees see learning at niche level as an integral part of the evaluation process. One of the policy makers for example, talked about three circles within policy evaluation. Explaining the circles he states: “There is a circle that focuses on the outcomes of the programme. For example, do farmers within the programme increase the use of a specific type of crop protection? The second circle includes the process within the programme: how is the process developing? Are the right actors involved? Does the programme interact with its wider environment? And then there is a third circle that involves the questions: How can you support the programme? What can the programme learn? These questions may result in an improved programme.” In our terminology, evaluation in the first circle contributes to upwards accountability and evaluation in the third circle contributes to internal accountability at regime level. The second circle refers to the types of evaluation we would find under internal accountability at niche level; it deals with questions that are helpful in making timely adjustments, to better reach the intended purpose (or even adjust the purpose), during the course of the project.

The policy makers who mentioned this type of internal learning state that it is important to be reflexive during the execution of a project. One interviewee for example mentioned that it is essential to ask questions during project execution, such as: “Are we doing the right things? How realistic are the objectives we formulated? Do we all agree on the objectives? Can we achieve our objectives with the resources and activities we are employing?” Rather than perceiving policy making as a linear process of defining objectives, devising policies, implementation and evaluation, some of the interviewees suggested that policy making is a continuous reflexive process. Especially in the case of system innovation programmes, interviewees consider that this reflexive process is important and can be aided by new forms of evaluation.

In correspondence with our findings, in Table 5.4 internal accountability is split into two types of internal accountability: internal accountability at regime level and internal accountability at the niche level.
Table 5.4. Evaluation needs of project managers in relation to spectrum of accountability

<table>
<thead>
<tr>
<th>Accountability</th>
<th>Evaluation needs managers</th>
<th>Reasons for this need</th>
</tr>
</thead>
</table>
| Upwards accountability  | • Ways to phrase project objectives that are flexible enough for the project and concrete enough for funders  
                         | • Ways to make intangible processes and immaterial results tangible                      | • To comply to requirements funder without jeopardizing the necessary flexibility of projects |
| Horizontal accountability| • Determine who are relevant stakeholders in the surroundings of the project             | • To realise alignment with the various institutional contexts in the surroundings of the niche experiment |
|                         | • Assess possible hampering factors posed by different institutional contexts             |                                                                                       |
| Internal accountability  | • Insights necessary to make decisions in the day-today work of managing niche experiments (notably regarding multi-stakeholder context) | • To achieve the purpose of the project  
                         | niche level                                                                                      | • To achieve horizontal accountability                                                             |
|                         |                                                                                           |                                                                                       |
| Internal accountability  | • Knowledge of institutional contexts and possibilities for creative solutions            | • To influence hampering conditions at regime level (in order to achieve horizontal accountability) |
| regime level            |                                                                                           |                                                                                       |

5.4.2 Project managers' evaluation needs in context

Project managers of niche experiments supporting Triple P initiatives talk about upwards accountability primarily as a burden, taking a lot of time with no clear benefit. According to one project manager: "The funding really put a mark on the programme, both financially and content wise. The conditions they set were limiting, for instance in terms of what partners to involve and under what conditions, but moreover a lot of management attention was needed; half of the time we were busy with bureaucracy rather than innovation." And that is not the only problem managers of niche experiments face with regard to upwards accountability. They also face the tension between the requirement of showing concrete results on the one hand and the necessary space for niche experiments on the other. "The government is accountable with regard to the money spent and needs to show results. Accountability and results are part of our contemporary [governance] culture. But for transitions we need space and direction, space for projects to emerge and direction of where roughly we want to go. Space and direction are far more important than accountability and results."
Transitions are to do with trust, with giving space to make things emerge, but that relates badly to money." This tension gives rise to challenges for project managers in terms of acquiring funds: "Even though I do not have a strict plan to be executed, I do have to write one - the funder requires this. They really would not give money if we would say: 'it depends on the entrepreneurs what we will be doing'. So, you do have to say beforehand, we will have so many meetings about this and that, etc." It implies that new types of performance measures are required for upwards accountability; measures that better reflect the emergent process of niche experiments. One of the managers of a system innovation programme, that as a rule participate actively in projects, rather than steer from a distance, says: "If we are able to make intangible processes tangible and immaterial results concrete, the funder would be very happy with that." See the first row of Table 5.4 for a summary of project managers' evaluation needs in relation to upwards accountability.

Like the funders, also project managers do not mention evaluation needs in relation to downwards accountability. They do, however, talk a lot about accountability towards parties other than funders (upwards accountability) and themselves (internal accountability). This is the case because niche experiments by definition are multi-stakeholder projects, bringing together entrepreneurs with actors from knowledge institutes, governmental and non-governmental organisations, and other businesses. In Table 5.4 we have therefore replaced downwards accountability for horizontal accountability. The multi-stakeholder nature of niche experiments brings in new accountability dimensions: "What I see a lot in these types of change processes is that you always need to involve more actors. This makes that you are very dependent on structures that you cannot really influence. Structures that are very robust and steady [...] and that bring in many institutional interests." For instance, scientists that participate may on the one hand contribute by bringing in relevant knowledge and taking up new research questions, on the other hand they are accountable to their institutions. " Universities and science need to account for how much they publish. In principle, science can really contribute [to niche experiments], but this is at the expense of fundamental research. So there is a tension there."

Other examples of horizontal accountability are the need to comply with laws and regulations: "For instance, in one of the projects, an air-cleaning system was used in the stable that was based on best practices, a demonstrable sound system. However, it was innovative and not yet accepted by legislation." Another example concerns an urban planning project with the ambition to develop houses without external sewage system, which again was not allowed under current legislation.
Thus, as an extension of downwards accountability, which concerns the responsibility to take into consideration all those affected by the project and involve them in all phases of the project planning, monitoring and evaluation, horizontal accountability concerns the responsibility to take into consideration all relevant parties in the environment of the project. Evaluation can contribute by assessing the diverse institutional contexts and the actors that need to be involved in the niche experiment to deal with possible obstacles. As one of the interviewees remarks about the role of the evaluator: "In one of the projects, different parties are involved; entrepreneurs, governments and knowledge institutes. They intend to realise the project together. But what we see is that citizens living in the area, and the nongovernmental organisations have no access to this network. That is a problem in the network. In this case, the evaluator is a person that indicates that certain parties are missing."

Besides horizontal accountability, project managers also express the need for internal accountability. Project managers considered that reflection on actions taken, and decisions made, may result in adjustments that improve the effectiveness of the project. Time to reflect, however, seems to be limited within innovative contexts. For example, one manager indicated: "I am extremely busy and I have to make many decisions. Therefore I tend to act rather intuitively." Another project manager expressed: "Often I am ruled (directed) by the hectic daily business. Therefore, I don't have or don't make time available to reflect on current practices." Managers also said they benefited from the 'outsider perspective' effectuated by an evaluator. One of the project managers illustrated this as: "I need feedback from outsiders; I am not capable to pull my own head out of the swamp." Another said: "We as a project are focused on realising a specific target. To prevent becoming practical too soon within the project a helicopter view from a distance is very helpful." A third project manager articulated: "Sometimes I wonder: 'are we doing the right things? Should we direct our energy on other things?' It is helpful to discuss these issues with someone." Thus, ex-durante evaluation can contribute to making timely adjustments along the way to remain internally accountable.

As we saw above dealing with the multi-stakeholder nature of niche experiments is a challenge to many project managers. Internal accountability at niche level is largely about learning how to deal with the multiple contexts of the project in order to reach its purpose; that is, in order to be accountable to the mission of the endeavour. As one of the interviewees points out: "There are various institutional, regulatory and financial obstacles in the context of a project. However, we need to put the term 'obstacle' into perspective, because it could also be seen as the task or challenge of the project to deal
with these hampering factors. And to find solutions within the context of a given institutional environment. We see it as the goal of project to see how you can change this; it is a project objective rather than an obstacle.” It means not accepting hampering factors as unchangeable conditions, but actually taking on the challenge by circumventing or confronting these factors. Referring to the above example of the houses without external sewage system, a project manager says: “If it appears that a house without sewage is not possible on the basis of legislation, an important part of your ambition drops. You can say, 'I accept that', but you can also say 'if we really find this important, because we want to keep the nutrients in the area or so, we must do something with this". By bringing in actors that may be able to make changes in their home organisations (in this case government), a reflection process is started at regime level. 'You need to put these legal people together and start trajectories to change legislation, in politics, but also in their own organisation. Otherwise you keep talking about beautiful ambitions, but you forget that in practice some structures hamper achieving these ambitions." We refer to this process as internal accountability at regime level.

Table 5.4 summarises the needs of project managers regarding evaluation and shows that the needs of project managers are different in various respects from the needs of funders, but they share that they operate in a complex context, which requires them to engage with different actors, reflect on their realities and adjust strategies to accommodate the multiple requirements whilst keeping direction. This insight opens up opportunities for the conceptualisation of alignment strategies between niche experiments and their surroundings, including funders and hence overcome the seeming incompatibility between learning (often associated with internal accountability) and accountability (often implying upwards accountability). In order to discuss these we will first consider the consequences of our results so far for the accountability framework that we introduced in the beginning (Table 5.2).

5.5 An adjusted accountability framework for niche experiments

Following Ebrahim (2005) we argued that the concept of accountability should be broadened from upwards accountability to also include downwards accountability and internal accountability (see Figure 5.2), thereby opening up many possibilities for more learning-oriented evaluation approaches. Our findings allow for a further conceptualisation of accountability in the context of evaluation; we propose two adaptations. Whilst downwards accountability reflects the need to take into consideration the fact that programmes are accountable to multiple actors - a need that
is recognised by the public, corporate and non-profit sectors alike - its focus is limited to those actors that are affected by the programme (also referred to as social accountability).

**Figure 5.2.** The original framework for accountability (based on Table 5.2)

Following our findings, the first adaptation we propose is to refer to downwards accountability as 'horizontal accountability'. This also reflects the move in governance from vertical principal-agent relationships to horizontal network governance (Benjamin, 2008; Lehtonen, 2014) and the multiple constituencies involved. Similarly, in the business world it is increasingly argued that “to build and sustain corporate success, companies must synchronize business objectives and market positions with political and regulatory activism and social and environmental engagement”22. Today, companies do not only need a 'licence to sell' (i.e. by producing goods that match

---

22 www.aligningforadvantage.com, 04-03-2015
market demands), they also need a ‘licence to produce’ (i.e. by complying to rules and regulations regarding safety, labour conditions etc.) and even a ‘licence to operate’ (i.e. the informal approval of products and operations by civil society groups and the public at large, non-compliance resulting in public protests or boycotts) (Casimir & Dutilh, 2003, see also Regeer, Mager, & Oorsouw, 2011). Synchronising business activities with the demands, interests and ideas of other stakeholders is done through processes of “value co-creation” when it concerns customers (Prahalad & Ramaswamy, 2004), but also through interactions with the entire network of suppliers, partners, other business, employees, governments and non-governmental organisations (Ramaswamy & Gouillart, 2010) on so called ‘next-practice platforms’ (Nidumolu et al., 2009). Similarly, in the case of Triple P business initiatives, where multiple actors are involved in the governing of projects, horizontal accountability implies that project teams are accountable, besides to funders, to multiple actors including consumers, suppliers, citizens, national and international regulators, civil society and non-governmental organisations. This is reflected in Figure 5.3 where outward arrows indicate that niche experiments are accountable to multiple actors, including both funders and beneficiaries. This implies that not only downwards, but also upwards accountability is conflated into the concept of horizontal accountability.

The second adaptation we propose is to distinguish between internal accountability at niche level and regime level. Internal accountability refers to the fact that any programme or organisation should be held responsible to its own mission, values and ideals. This is particularly true for niche experiments, where keeping the ambition high and not reverting to suboptimal solutions under pressures from the environment, is considered the primary challenge (e.g. Hoes et al., 2012; van Mierlo et al., 2010). It also implies taking into account the multifaceted environment of projects. This means that boundaries between what is internal and what is external to the project or programme become ‘vague, fluid and subject to constant change’ (Benjamin & Greene, 2009, referred to by Lehtonen, 2014).
Learning and reflection are said to be indispensable aspects of the governance of niche experiments (e.g. Loorbach, 2010; Raven, van den Bosch, & Weterings, 2007), due to the many uncertainties, their experimental nature (it is not known beforehand exactly what the result will be) and the multiple stakeholders that each bring in requirements from their respective institutions. The internal goal, to establish a Triple P business, therefore also has an indivisible external component. In the adapted framework, internal accountability does not only refer to the niche level but also to the regime level (see Figure 5.3) because work within the niche experiment may also "critically scrutinize and even attempt to adapt their structural context and the self-evident assumptions embedded therein." (Grin, 2010, p274). Internal accountability at regime level may mean that a government department adjusts their objectives and measures as a result of outcomes of programme evaluation, but it may also mean that universities adapt their publication requirements for career development, or that people and planet values become an integral part of accountability frameworks of companies.

The adapted accountability framework for niche experiments thus includes two shifts: the first one involves the conflation of upwards and downwards accountability into horizontal accountability and the second one involves the distinction of internal accountability at niche level and at regime level. Creating alignment with the multiple contexts (including the funders) that are relevant to the Triple P initiative (i.e. horizontal accountability) is key in order to increase its resilience outside of the niche protection. If accountability to funders has become one of the multiple facets of
horizontal accountability, and if internal accountability at niche and regime level is already part of the core business of governing niche experiments, can we perhaps learn from alignment strategies regarding the multi-actor nature of niche experiments for the benefit of evaluation?

5.6 The need for dynamic capabilities to create alignment

For learning-oriented evaluation methodologies to contribute to the diverse accountability requirements, they would need to support project managers by making visible the uncertainties resulting from a changing environment and turning what can be seen as 'threats' of the contexts into the core task of governing niche experiments. As we already saw above, rather than considering unconducive conditions as environmental factors outside of the scope of the project, in the context of niche experiments dealing with them, and changing them, is considered as 'the task or challenge of the project [...] the goal of the project'. And this task carries many uncertainties with it. It can be argued that managers of niche experiments face functional task uncertainty: they know what to do, but not how to do it (Whitley, 1984). As one of the interviewees expresses it: "In principle, everything is already there, the only question is how to organise it. I think that the how-question is the most important question that each project manager faces. How can I do this?" He continues that project managers cannot stick to their normal routines: "It is more than their daily work. They really need to find the pitfalls and obstacles that have hampered the transition since years."

Evaluation can contribute to this process of internal learning by introducing routines that help to develop dynamic capabilities. Dynamic capabilities are those capabilities required to deal with dynamic and unpredictable environments (see e.g. Teece, Pisano, & Shuen, 1997; Eisenhardt & Martin, 2000; Zahra, Sapienza, & Davidsson, 2006). The following citation regarding the role of evaluation in the context of development organisations, clearly resonates with much of the writing in the field of strategy research (e.g. Eisenhardt & Martin, 2000) as well as with the experiences of project managers of niche experiments:

"Development organizations that only monitor very selective aspects of their work and environment (such as meeting financial targets) risk deceiving themselves into thinking that their environment is stable. On the other hand, organizations that attempt to anticipate future uncertainties (e.g., by conducting strategic reviews or scenario planning that involve various levels of staff in
strategic discussion and that may require alliances with other organizations such as research institutions, funders, and even competitors) may be better positioned to recognize and respond to environmental change." (Ebrahim, 2005, p76)

In recent years various evaluation methodologies and tools have been developed that can be argued to support the process of developing dynamic capabilities (e.g. Van Mierlo et al., 2010, De Wildt-Liesveld et al., 2015). The evaluator can contribute by organising reflection within a project team on questions of how to engage multiple actors, how to keep ambitions high, how to keep the process open, in short how to realise a Triple P initiative in contexts in which selection pressures, such as regulatory requirements and demands for competitiveness and profitability, prevail. Learning-oriented evaluation can contribute to the development of alignment strategies, aided by the available dynamic capabilities. Alignment refers to the ‘bringing in line’ of two different worlds (Wenger, 1998), in this case the worlds of the niche experiments and the surrounding contexts, implying that development-oriented niche experiments needs to be adaptive with respect to the results-oriented regime (de Wildt-Liesveld, Regeer, & Bunders, forthcoming). In other words, alignment strategies aim to create overlap between different worlds or discursive spheres (see Hendriks & Grin, 2007), ensuring that the innovation developed within the niche is comprehensible and logical in the eyes of different stakeholders (Hoes & Regeer, 2015). We have identified two alignment strategies from our data, which we will present in the next section; boundary people and boundary objects.

5.6.1 Alignment strategies: boundary people and boundary objects

The multiple stakeholders that are indispensably engaged in niche experiments are not only involved to bring in their experiences, views and knowledge for the benefits of developing an innovative Triple P initiative, they also form the bridge to the multiple institutional settings. They can be considered 'boundary people': people that can bridge, or create alignment between different worlds. This requires careful selection of participating actors: "Accountability to your many constituencies is achieved by good casting. You do not want to select people that participate only to defend their organisation. Rather we invite people that do work from a particular interest, for instance animal welfare, but at the same time are able to change hats." Personal background may play a role in the ability to change hats: "Of course, I am a civil servant and when I arrive at a farmers, the farmer will treat me as a civil servant. However, as soon as I mention that we have a farm at home, they start laughing and
the tone totally changes. Suddenly I am one of them. And then I get so much more done."

Is there any evidence that the alignment strategy of ‘boundary people’ can also be used in the relationship between funder and niche experiment? Often funders are at distance from the projects and do not have inside knowledge about what goes on at field level. One of the policy makers that is responsible for different grant programmes explains: "We do receive interim reports of these projects, but we don’t really hear about the challenges these projects face. I coordinate different grant programmes, with multiple projects in each of them, and apart from this I do many other things, so I would not be able to say what exactly happens in the projects."

Others examples show that the funder can be a boundary person too. By becoming actively engaged within the programme, by for example taking part in the programme team and attending meetings, funders gain more inside knowledge of the projects which makes it easier for them to report to their superiors or constituency. One of the policy makers stated: “Sometimes I drive to Lelystad [city in the Netherlands] where the project is located, to talk with the project participants and to see in real life what the project is about. Being there enables the sharing of content related activities and results.” In the case of recent large-scale innovation programmes, with a separate programme bureau outside of governmental departments, working closely together with projects has become more common. "After we have decided to approve a project, the project leader is responsible for the execution of the project and I act as a coach for the project leader, and think along with the project leader. I also participate in project team meetings as one the funders and bring in the interest of our programme. And I make connections between the project and other partners in our network, for instance knowledge institutes." This role of the funder as active participant is also recognised by project managers. One of them says: "I remember the first project meeting that I attended. I had just started my new job with the knowledge brokering agency in charge of the project. One person of the funding agency was there too, but I didn’t realise that. He really was a member of the project team from the start. I have never associated him with the funder. From the beginning we were a team and the project really benefitted from this later on.”

A second alignment strategy that can be induced from the interviews is to create 'boundary objects'; objects that make sense in the eyes of different stakeholders. Not only the Triple P business idea itself can be considered a boundary object, as it needs to maintain coherence across different contexts, other boundary objects can be created to
support the process of realising a Triple P initiative. An example of a boundary object is the ‘status aparte’ (a legal status) that was granted to one of the niche experiments, based on the idea that the government should support innovations that comply with the values on which legislation is based, even if they do not comply with the existing legislation. The ‘status aparte’ is a boundary object, because it has legal status in the context of legislation and it provides the necessary innovation space in the context of the niche experiment. The earlier mentioned attempt to write a project plan that leaves room for emerging development, but at the same time is concrete regarding the number of meetings and required budget is another example of a boundary object. Also developing new types of indicators, that on the one hand do justice to the processes involved, thereby making the task of evaluating contributory to rather than distracting from the core business, and on the other hand comply with the requirements of funders are examples of boundary objects.

Star and Griesemer (1989) define boundary object as an object or strategy that is recognised as relevant and valuable by different actors participating in the programme and their respective home base organisations. A boundary object thus develops and maintains coherence across intersecting social worlds (Star & Griesemer, 1989), in other words, creates alignment (Wenger, 1998). The individuals who mediate between incommensurable worlds are referred to as boundary people (e.g. Edelenbosch, Kupper, Krabbendam, & Broerse, 2015) or boundary actors (Keshet, Ben-Arye, & Schiff, 2013). Also within evaluation literature, the importance of strategies that focus on crossing boundaries between funder and project has been stressed:

"[I]n addressing complex social problems, ends often are clarified and redefined through practice (Harmon, 1995; March & Olsen, 1985). These scholars suggest that accountability efforts should encourage dialogue and greater understanding between principal and agent (Fry, 1995; Harmon, 1995).” (Benjamin, 2008, p 326).

In sum, 'boundary people' and 'boundary objects' have been identified as important strategies that interviewees have used to create alignment between the niche experiment and its multiple contexts; they have aided horizontal accountability, supported by internal accountability at both niche and regime level.
5.7 Concluding remarks

This chapter dealt with issues of evaluation for accountability and learning in the context of niche experiments that support the realisation of Triple P initiatives. These niche experiments are temporarily funded by government programmes, which brings in an accountability dimension. Increasingly learning-oriented evaluations are employed for programme evaluation, but they do not always seem to meet the requirements of funders. Evaluators can become caught in a web of tensions between projects and funders. Particularly in the case of evaluating niche experiments, a gradual change of expectations, from in-depth reporting on learning experiences, failures, mistakes etc. at the start of the project (i.e. needs regarding internal accountability), to presenting discrete results and clear recommendations towards the end (i.e. needs regarding upwards accountability), is often experienced. Project managers and evaluators are thus in need of strategies to bridge the gap, or better to create alignment, between requirements of funders on the one hand and needs of the project on the other. There is a seeming incompatibility between evaluation for learning (often associated with internal accountability) and evaluation for accountability (often implying upwards accountability). In this chapter we have proposed an adapted accountability framework for evaluating these types of projects, which reconciles learning with accountability. Horizontal accountability is central in this adapted accountability framework; it refers to the fact that a niche experiment with a Triple P purpose is accountable to multiple actors including consumers, suppliers, citizens, national and international regulators, civil society and non-governmental organisations, as well as funders.

We believe that the insights and framework developed in this chapter are also applicable to evaluation beyond the context of niche experiments. Programme evaluation, involving the analysis of (causal) links between project activities and their outcomes, is increasingly considered to be challenging (e.g. Perrin, 2002; Williams & Imam, 2007), particularly when it concerns projects or programmes that deal with complex issues, involve multiple actors and find themselves in unstable environments. They are faced with interconnected ecological, social, political and economic systems, as well as with a plurality of values and perspectives brought in by the different actors involved, and as such are "influenced dramatically by a range of ambiguous and uncertain external and internal forces" (van Marrewijk, Clegg, Pitsis, & Veenswijk, 2008, p599). Rather than perceiving changing project objectives and unintended and unexpected effects as signs of inadequate project design, it is increasingly argued that the uncertainties emerging from case-specific contexts should be seen as a "source of
learning, reflexivity and adaptive governance" (Lehtonen, 2014, p279, see also Arkesteijn, van Mierlo, & Leeuwis, 2015). The adapted accountability framework accommodates exactly these arising needs.

**Evaluation strengthens goal-orientation**

Evaluation is means to an end, not an end in itself. Rather than treating accountability and learning as two different things, we have argued that actively addressing the multiple requirements for accountability, through learning-oriented evaluation approaches, strengthens the goal-orientation of projects. We would like to illustrate this by considering, what the implications of the adapted framework are for Triple P initiatives. The forces, such as regulatory requirements and demands for competitiveness on the market, that work from the outside in (see arrows moving in, Figure 5.1), are no longer seen as threats but as needs for accountability (see arrows moving out, Figure 5.4a), along with those other needs for accountability regarding beneficiaries, funders, scientists, NGOs, but also suppliers, consumers, other businesses, etc. If a niche experiment is successful in employing alignment strategies and developing dynamic capabilities, not only within the project team, but also, as an extension, in the associated institutional contexts, the chances are enlarged that the Triple P initiative survives as a Triple P business without protection (Figure 5.4b). At the same time, accountability towards funders (formerly called upwards accountability) can also be accounted for, through the same learning-oriented evaluation methodologies and alignment strategies employed to support the processes of horizontal accountability.

**Fig 5.4a.** Triple P initiative supported by niche experiment: alignment is created

**Fig 5.4b.** Triple P initiative is now a Triple P business that survives without protection
To conclude, in our information- or network society, project and programmes become increasingly complex, requiring the involvement of multiple stakeholders and alignment with multiple contexts. The local needs to be aligned with the global, environment with profit, human rights with health care. The development of dynamic capabilities is a prerequisite for forward-looking public services, non-governmental organisations, knowledge institutions and corporations that are willing to embrace the complexities of our contemporary world and anticipate future uncertainties. We hope to have shown that the field of evaluation is in a position to contribute to this important endeavour in a significant and meaningful way.