



Universität für Bodenkultur Wien
Department für Wirtschafts- und
Sozialwissenschaften

Assessment Practices in the Policy and Politics Cycles: A Contribution to Reflexive Governance for Sustainable Development?

James Meadowcroft & Reinhard Steurer

InFER

Institute of
Forest, Environmental, and Natural Resource Policy

Diskussionspapier / Discussion Paper 2-2018

2018

Diese Reihe ist ein Publikationsorgan von **InFER**, dem Institut für Wald-, Umwelt- und Ressourcenpolitik der Universität für Bodenkultur Wien. Der Inhalt der Diskussionspapiere unterliegt keinem Begutachtungsverfahren, allein die Autorinnen und Autoren zeichnen verantwortlich. Anregungen und Kritik seitens der Leserinnen und Leser sind ausdrücklich erwünscht.

This series is edited by **InFER**, the Institute of Forest, Environmental, and Natural Resource Policy at the University of Natural Resources and Applied Life Sciences, Vienna (BOKU). The discussion papers are not subject to review procedures. Therefore, responsibility for the content lies solely with the author(s). Comments and critiques by readers are highly appreciated.

ISSN 2072-764X

Also published as

James Meadowcroft & Reinhard Steurer (2018): Assessment Practices in the Policy and Politics Cycles: A Contribution to Reflexive Governance for Sustainable Development? In: Journal of Environmental Policy & Planning, 20:6; 752-774

Bestelladresse / orders to:

Institut für Wald-, Umwelt- und Ressourcenpolitik
Universität für Bodenkultur Wien
Feistmantelstr. 4
A – 1180 Wien
Tel: + 43–1–47 654–73200
e-mail: infer@boku.ac.at

In dieser Reihe erschienene Diskussionspapiere können von folgender Homepage als PDF-Files geladen werden: <http://www.boku.ac.at/wiso/infer/publikationen/diskussionspapiere>

The papers published in this series can be downloaded in PDF-format from:
<http://www.boku.ac.at/wiso/infer/publikationen/diskussionspapiere>

Eigenverlag des Instituts für Wald-, Umwelt- und Ressourcenpolitik, Universität für Bodenkultur Wien
Published by the Institute of Forest, Environmental, and Natural Resource Policy, University of Natural Resources and Life Sciences, Vienna (BOKU)

ABSTRACT

This article examines systematic assessment practices linked to sustainable development policies. We consider five types of assessment—monitoring, policy evaluation, formal audit, peer review, and specialist reporting—and explore their fate in the policy and electoral politics cycles. In contrast to traditional views of the policy cycle, we note that systematic assessments provide complementary feedback around the entire policy cycle. However, despite this omnipresence, their policy relevance is usually severely limited, *inter alia* because the policy cycle captures only parts of the political reality. A major concern for politicians (but not necessarily for policy or governance scholars) that goes far beyond the formulation and implementation of policies is the broader cycle of electoral politics that determines the state's political personnel as well as government priorities. Here, we highlight that the findings of systematic assessments are often lost in a cacophony of voices to which politicians are more carefully attuned, such as media responses and opinion polls, implying that scientific evidence is simply 'overwritten' with other kinds of evidence representing alternative rationalities and priorities. Despite numerous shortcomings, the true value of systematic assessment practices lies in their potential to furnish ammunition to state and non-state actors interested in securing change.

KEY WORDS: Reflexive governance, evidence-based policy-making, science– policy interface, sustainable development, policy cycle

This article is concerned with the way governments in industrialized countries (particularly in Western Europe) have drawn on systematic assessment practices to further governance for sustainable development. It considers how such practices relate to the cycles of policy-making and electoral politics familiar in contemporary democratic states. This discussion is placed in the context of recent interest in evidence-based policy-making and science–policy interactions, and it focuses particularly on the role these assessment practices play in developing more reflexive modes of governance that encourage societal movement towards sustainability.

There is now a rapidly growing scholarly literature dealing with sustainability-related assessment practices. The largest body of work relates to the design and use of particular inquiry practices, such as Sustainability Impact Assessments (see e.g. Hertin et al., 2008) and Strategic Environmental Assessments (Bina, 2008), indicator-based monitoring (Steurer et al., 2010; Steurer & Hametner, 2013), and peer reviews (Lehtonen, 2007; Spangenberg & Niestroy, 2010). Some contributions address the role of particular organizations that conduct these activities, such as environmental agencies (Huiteima & Turnhout, 2009) or sustainable development councils (Niestroy, 2005, 2011). Finally, some scholars have begun to adopt a broader focus, exploring the complex relationships among different types of assessment-based feedback on the one hand and policy-making on the other hand (see for example, Juntti et al., 2009; Mickwitz, 2006; Nilsson et al., 2008; Pregernig et al., 2012). Although research in each of these strands is increasing, linkages among them or holistic perspectives that address multiple practices and their relation to policy-making are rare. The present paper aims to fill this gap by offering a wider interpretation of the significance of assessment practices, their connection to policy and political processes, and their potential contribution to reflexive governance for sustainable development.

Over the past two decades, an array of systematic inquiry practices has been associated with policy-making for sustainable development. These vary in many respects, including: their explicit rationale or purpose; their temporal, geographic and thematic scope; the type of organization responsible for commissioning and carrying out the work; the applied methodologies and criteria invoked to judge

performance; and the target groups for the results. Despite this diversity, the assessment practices with which we are concerned here are united by four common features:

- first, they are official, in the sense that they are directly or indirectly commissioned and funded by governmental organizations;
- second, they are systematic, in so far as they attempt to present empirical facts or evidence-based judgements about some issue or area of government activity;
- third, they are relatively independent—in the specific sense that their activity and reporting is supposed to be protected from direct political or administrative interference; and
- fourth, they are public, involving a clear commitment to the open publication of their findings.

We focus our analysis on assessment practices that share these characteristics because they represent formal mechanisms by which the state acquires feedback on societal/environmental interactions and policy orientations. In this sense they occupy a potentially privileged position, which is close to government and policymaking, but not directly controlled by policy-makers and politicians, and which can in principle contribute to improved government performance as well as accountability and to broader democratic discussion.

The systematic assessment practices that concern us here can be divided into the following five general types:

- **Monitoring.** Emphasis may be placed on monitoring developments (environmental pressures and states and economic and social trends) or policy impacts and outcomes. Comprehensive sets of indicators, 'state of the environment' reporting, and identifying and tracking particular phenomena (for example, air quality) are all pertinent here.
- **Policy evaluations.** Evidence-based examination of all kinds of policies can be *ex ante* and *ex post* and employ a range of methods, including scenario techniques or models, cost–benefit analyses, impact assessments, and risk-based perspectives (Nilsson et al., 2008). Criteria for assessing policy include 'effectiveness', 'efficiency', 'equity'/fairness', 'flexibility', 'predictability', 'acceptability'/ 'legitimacy', and 'transparency' (see e.g. Huitema et al., 2011; Mickwitz, 2006).
- **Formal audits.** Auditing for the environment and sustainability has emerged as an outgrowth of long established practices for financial audits of government activities (Leeuwen, 2004). Here the focus is typically on the evaluation of conduct and the assessment of performance with respect to objectives and guidelines stated by governments. Audits typically focus on specific organizations and the programmes for which they are responsible.
- **Peer reviews.** Such assessments are conducted by a panel of 'peers', appointed from outside the target jurisdiction or organization, but with professional experience managing similar sorts of policy challenge. They may be focused around a specific policy area (such as water management), on particular policy instruments (environmental taxation or sustainable development strategies), or more broadly on the overall parameters of environmental policy (see e.g. Spangenberg & Niestroy, 2010).
- **Specialist reports.** Besides the practices described above, governments also commission specialist reports, or they finance organizations or commissions who produce, among other things, specialist reports. While prominent reports— such as the Assessment Reports of the Intergovernmental Panel on Climate Change, the Stern review on the Economics of Climate Change (Stern, 2008), or the Report of the commission on the measurement of economic performance and social progress (Stiglitz et al., 2009)—reflect scientific reasoning, they are neither *ex ante* assessments nor *ex post* evaluations of particular policies, but broader analyses of policy-relevant issues.

Although there is obvious overlap among these types (for example, monitoring with indicators may play a role in all five activities), each represents a distinct form of practice with a discernible logic. Organizations active in these five areas of systematic inquiry commissioned or financed by governments include (but are not restricted to) national/federal ministries (in particular environmental ministries), environmental agencies, parliamentary committees, expert advisory panels, audit bodies, statistical offices, multi-stakeholder advisory commissions, and research institutes and consultancies. While some of the actors are associated with one particular type of systematic inquiry (e.g. audit bodies are concerned with formal audits and statistical offices with monitoring), some others (such as environmental agencies, multi-stakeholder organizations, and research institutes) can be engaged in a variety of practices.

It is important to note that two major forms of policy-related assessments lie outside the direct focus of the current discussion. First, processes that are entirely 'internal' to executive government—where the monitoring is carried out by officials directly responsible for the design or delivery of the programme under review (self-assessment) or by external consultants, and where the results of the inquiry are kept from the public. Second, we are not concerned with analyses that are entirely separate from government, such as studies conducted or commissioned by 'think tanks', university researchers, industrial associations, business corporations, environmental groups, political parties, and so on. Of course, these may generate findings of critical importance to policy-makers. But here we are dealing with assessment practices that are in some sense 'official'—that are part of state efforts to monitor societal development and government performance and to present these openly to the political community.

The discussion is organized into four sections. We start by considering three conceptual frames that can be invoked to illuminate the policy and political relevance of such assessments: (i) evidence-based policy-making; (ii) science– policy interactions; and (iii) reflexive governance, which provides the main anchor for our argument. The article then reviews the five main assessment practices outlined above, before moving on to consider their links to policy and politics cycles. Finally, we return to the three conceptual frames, to explore particularly the significance of these assessment practices from the perspective of reflexive governance.

1. Three Ways of Conceptualizing Policy-Related Assessment Practices

At its core, reflexive governance refers to the idea of a more self-aware and selfcritical mode of governance which remains open to multiple societal perspectives and to a continuous interrogation of both means and ends (Hendriks & Grin, 2008; Voss & Kemp, 2005). Thus, the careful assessments of current circumstances and future possibilities, of alternative viewpoints and modes of intervention, and of the 'situatedness' of governance approaches are all critical (Smith & Stirling, 2008). But just how helpful are current assessment practices on sustainable development in this regard? Do they really contribute to more reflexive governance? To answer this question it is necessary to consider in more detail what reflexive governance entails. But it will prove helpful to start with two alternative and somewhat more conventional ways of framing policy-related assessment practices: first, the contemporary emphasis on evidence-based policy-making and second, the long-standing concern with science–policy interfaces.

1.1 Evidence-Based Policy-Making

Over the past two decades there have been repeated calls—from the policy research community, but above all from the political and administrative sphere—for greater emphasis on evidence-based policy-making. The core idea is that in an ever more complex world public authorities need a sound knowledge base to support their activities. Basing decisions on a systematic consideration of evidence is expected to ensure more effective policy design and implementation, secure value for money, minimize unintended consequences, and prevent policy disasters (European Commission, 2001, p. 22, 2002; European Council, 2006, p. 5; Sanderson, 2002, p. 1). ‘Supply-push’ factors related to the rise of knowledge societies certainly played a role in increasing the emphasis on the systematic consideration of evidence and the proliferation of assessment exercises (Radaelli & De Francesco, 2007). For example, growing scientific communities were able and willing to devote more resources to policy-relevant research, and new means of communication opened additional channels to disseminate scientific findings. On the other hand, governments were the decisive ‘demand-pull’ factor, inter alia because they faced an increasingly sceptical electorate familiar with the impacts of policy failure and because they repeatedly experienced the limits of budgetary expansion and had to do more with less resources (Pregernig et al., 2012). Moreover, public sector reform movements such as the ‘New Public Management’, ‘good governance’, or ‘smart regulation’ that aimed to make the public sector more efficient, accountable, and strategic added momentum to the demand-pull for evidence-based policy-making and performance measurement (European Commission, 2006; Radaelli, 2004, 2007; Sanderson, 2002, p. 4).

In the environmental sector, pressure for evidence-based policy-making grew from the mid-1980s as side effects from the environmental regulation of 1960/ 1970s became manifest, and with the realization that environmental problems were continuing to accumulate despite 20 years of active policy intervention. More than a decade ago, the European Commission insisted that ‘a stronger culture of evaluation and feedback’ is ‘needed in order to learn from the successes and mistakes of the past’ (European Commission, 2001, p. 22). And in the renewed EU sustainable development strategy from 2006, the European heads of state agreed that sustainable development policies should be ‘developed, assessed and implemented on the basis of the best available knowledge’ (European Council, 2006, p. 5; see also European Commission, 2002).

The call for evidence-based policy-making is an implicit repudiation of ‘nonevidence-based’ public decisions: that is to say, policy choices based on inadequate information and understanding, the pursuit of electoral or bureaucratic interest, corruption, and so on. Who could be opposed to that? And yet talk of evidence-based policy-making begs the question of exactly what is to count as evidence and how decision-makers are to gain access to it. The standard assumption is that evidence denotes knowledge generated by systematic assessments, whether they rely on understandings rooted in the natural or social sciences or particular groups of professionals (such as lawyers, auditors, administrators, and so on). However, since policy-making is not just a matter of determining scientific accuracy, but also of expressing normative commitments and accommodating legitimate interests, what constitutes credible evidence will always be contested in the public sphere. The sustainability assessments with which this article is concerned are systematic enquiries that draw on particular fields of specialist knowledge which evidence-based policy is typically taken to embrace. From this perspective, the issue becomes one of the extent to which these practices actually contribute to evidence-based policy-making for sustainable development and how they relate to other, non-scientific forms of evidence.

1.2 Science–Policy Interfaces

Another perspective that can be used to understand the political relevance of sustainability-related assessments is the long-standing concern with science–policy interfaces. While a ‘rational’ model of the policy process suggests that scientific findings should continuously feed into the policy sphere, providing a firm foundation for decision-making, the real world typically deviates from this ‘knowledge society ideal’. Evaluation scholars often speak of ‘pathologies’ (Stockmann, 2007) or the ‘misutilisation’ (Christie & Alkin, 1999) of knowledge (Pregernig et al., 2012, p. 29), whereas policy-makers usually refer to various constraints that shaped their decisions. Decades of political science research have discredited the rational model of policy-making as ‘blatantly inadequate, if not to say naïve’ (Pregernig et al., 2012, p. 30; see also Grundmann, 2009; Juntti et al., 2009; Pregernig, 2005). The tenor of the critique is that while scientists may be eager to speak ‘truth to power’ (Price, 1981), more often than not ‘scientific truth’ loses out to a political rationality that follows its own rules (Pregernig, 2005, 3f; Grundmann, 2009). Key elements which have emerged from this discussion include an appreciation of:

- The ‘separate worlds’ of science and policy: the two communities operate on distinct planes with different cultures, values, languages, reward systems, and rationalities (Pregernig, 2005, p. 4). This makes meaningful exchange between the two constituencies difficult. What can help is ‘knowledge brokerage’, conducted, e.g. by ‘boundary organizations’ working at the interface between science and policy-making (such as environment agencies) (Huitema & Turnhout, 2009).
- The strategic use of science: policy-makers often use science not to solve problems but to gain strategic advantage over political competitors (Pregernig, 2005, p. 11; Pregernig et al., 2012). Evidence can help them, e.g. to improve the authority and legitimacy of their own policies (Sharman & Holmes, 2010), to discredit political opponents, and to delay or avoid action (for the climate change discourse in the USA which illustrates both points, see Grundmann, 2007). Often ‘knowledge’ is ‘nothing but a “hook” on which interests hang their case’ (Radaelli, 1995, p. 173; taken from Pregernig, 2005, p. 10).
- The contested status of scientific knowledge: While policy-makers may have a ‘hidden agenda’ when invoking (or failing to invoke) science, the fact that science is often not as objective as the rational model assumes when it speaks of ‘truth’, gives policy-makers a good reason to use it strategically or cautiously. This applies in particular to the natural science of complex environmental problems such as climate change (Crabbe & Leroy, 2008) and to ‘knowledge of the social world’ which is often ‘socially constructed and culturally and historically contingent’ (Sanderson, 2002, p. 6; see also Pregernig, 2005, p. 7f). Since scientific assessments of sustainable development policies often rely on value-laden assumptions that combine the limitations of natural and social science research, this limiting factor is particularly pronounced here (Huitema & Turnhout, 2009, p. 578f).
- The ritual or symbolic use of science: As Pregernig et al. (2012, p. 35f) illustrate with Austrian case studies, systematic inquiries can be commissioned for ritual or symbolic purposes, i.e. they are neither expected to produce valuable evidence that helps to improve policies nor are they suitable for strategic or political use as defined above. They are performed ritually so that certain procedural standards have been observed, or they symbolically communicate that an issue is politically important. In short, systematic inquiries and respective evidencebased feedback loops are sometimes nothing but ‘decorative symbols’ of knowledge societies (Pregernig et al., 2012, p. 36 who quote Pollitt, 1998).

The systematic sustainability assessments with which this article is concerned often draw on findings, or invoke methodologies, drawn from the natural and social sciences. The concern here is the extent to which these systematic state-sponsored assessment practices contribute to a constructive science–policy interface that helps to improve decision-making for sustainable development.

1.3 Reflexive Governance

The notion of ‘reflexivity’ emerged from discussion of the character of contemporary societies which are beset by problems of their own making, and where naive modernist visions—that place unlimited faith in science, bureaucracy, and instrumental rationality—are under sustained pressure (Beck, 1994; Giddens, 1990). Reflexive governance constitutes a response to this condition, potentially replacing traditional, hierarchical, and deterministic governance approaches with a more reflective, adaptive, and interactive stance (Grin, 2006; Voss et al., 2006). Voss and Kemp, for example, describe reflexive governance as ‘the organization (modulation) of recursive feedback relations between distributed steering activities’ and emphasize the opening up of ‘problem-handling processes’ to ‘diverse knowledge, values and resources of influence’ (2005). Reflexive governance is typically associated with a more self-critical and self-conscious awareness, an emphasis on interactive processes that mobilize the knowledge and resources of diverse societal actors, and a modest appreciation of what top-down government can do to resolve social problems. Above all, there is an understanding that part of the function of governance is to reflect continuously on the character of governance itself. Stirling (2006), for example, has suggested a threefold distinction between unreflective, reflective, and reflexive governance arrangements. While the first denotes narrow instrumentally driven decision contexts, the second is associated with more critical attempts to manage side effects and canvass multiple perspectives so that the best policy option can be implemented, and the third embraces a critical epistemological and ontological vision that acknowledges the interpenetration of governance subjects and objects.

Sustainability is perhaps the quintessential terrain of reflexive governance. The idea of sustainable development: (1) is made necessary by recursive problems of modernity, (2) embodies a reflective critique of the environmentally destructive and socially inequitable character of the current development pathway, and (3) requires an institutionalization of reflexive practices in order to orient change towards a more sustainable social trajectory. Because policy-relevant knowledge is increasingly dispersed among different societal actors, be they other legislatures, scientists, businesses, NGOs, or citizens (Pierre, 2000), the quest for reflexivity is closely related to the principle of participation. As Agenda 21 emphasized: ‘In sustainable development, everyone is a user and provider of information considered in the broad sense. That includes data, information, appropriately packaged experience and knowledge’ (UNCED, 1992, paragraph 40.1). Indeed, from the vantage point of sustainable development, the encouragement of knowledge inputs and participation from across society is not just an instrumental imperative (that without such inputs it will not be possible to effect social change), but above all an ethical imperative: for only on the basis of interactive governance will it be possible to elaborate a development trajectory that reflects the authentic needs of society as a whole (Meadowcroft, 2007a).

When discussing reflexive governance for sustainability, some authors emphasize specific governance strategies such as ‘trans-disciplinary knowledge production’ or ‘iterative participatory goal formation’ (Voss & Kemp, 2006). There is also concern with the development of particular tools to promote reflexivity and with sites where it can gain purchase. For example, transition management scholars point to the importance of ‘visioning’ and ‘transition arenas’ (Loorbach, 2007; Rotmans et al., 2001). The large literature on the design of deliberative and participatory mechanisms to resolve environment and sustainability-related challenges can be interpreted in a similar way (Meadowcroft, 2004; Renn et al., 1995). Other theorists stress linkages to broader societal processes (Feindt, 2012). Hendriks and Grin (2008, 145), for example, worry that discussion of reflexive governance may neglect the dynamic

political context in which novel reflexive practices are embedded, pass over the crude reality of policy-making that includes the play of established interests, or miss the critical role of civil society when ‘the radical voices in civil society’ often ‘provide the necessary background conditions for wide-scale social transformation’.

The requirements for reflexive governance can be demanding. Smith and Stirling (2008), for example, list features of an ‘ideal type’ of ‘governance on the inside’ embodying reflexivity that include: accepting ‘system ambiguity’ and the ‘contested nature of sustainability’; acknowledging the incommensurability of perspectives and that all options are conditional and situated; recognizing that legitimacy is negotiated (rather than being conferred by ‘authority’ or ‘objectivity’), that interventions are always power laden, and that closure depends on wider political discourse; and so on. Yet their underlying message is rather straightforward: there is a clear contrast between approaching governance as a ‘largely instrumental managerial function’ (even if of a pluralist and reflective kind) and as a ‘fundamentally engaged political process’.

From this, two points should already be clear. The first is that the sorts of systematic state-sponsored assessment practices considered in this article will have a problematic relationship with reflexive governance. On the one hand, they do relate to the reflexive project of sustainable development, providing various forms of feedback to decision-makers. They may incorporate different forms of knowledge and interrogate existing practices. And yet, they are also tied to the formal governmental sphere (at the very least through funding mechanisms that influence their scope), rely above all on expert (elite) judgement, and can embody a narrow instrumentality that takes established governance relationships as given.

The second is that ‘reflexivity’ not only offers a different perspective on assessment practices that do the frames of ‘evidence-based policy-making’ and the ‘science–policy interface’, but also that it can be invoked to interrogate the adequacy of these frames. Thus, appeals to evidence-based decision-making can sometimes appear as a ploy to close down discussion (for example, opposing application of the precautionary principle on the grounds that regulatory decisions must be based on ‘sound science’ or promoting obligatory comprehensive cost/benefit analysis of proposed regulatory rules in order to slow down environmental policy initiatives). The science/policy discussion can be overly focused on instrumental uses of particular knowledge products while neglecting the wider social contexts in which scientific expertise is deployed.

2. The Menagerie of Systematic Assessments and Sustainable Development

We now provide a critical review of practices that fall into the five broad classes of sustainable development-related assessments described at the outset. Because of the enormous range of activities that can be included we have in each case chosen to highlight important practices that provide insight into the class as a whole. For monitoring, we focus on comprehensive indicator sets used to track progress towards sustainable development; for assessment, we target ex ante integrated assessments; for audits, we privilege those focused on the effectiveness of sustainable development policies; for peer review, we consider peer reviews of sustainable development strategies; and for specialist reports, we discuss analyses produced by sustainable development councils.

2.1 Monitoring

Some of the most widespread and regularly applied forms of monitoring for sustainable development are reporting schemes based on comprehensive indicator sets. Due to the strong emphasis on performance monitoring in public sector reform movements such as the new public management, and to guidance provided by the UN (e.g. in the Agenda 21, see UNCED, 1992) and the OECD (OECD, 2002; OECD-DAC, 2001) on how to pursue sustainable development, all EU member states, the European Commission, and many more countries around the world started to monitor selected key issues of sustainable development in the early 2000s as part of their national sustainable development strategies (Steurer, 2008; Steurer & Hametner, 2013).

Examples of widely used indicators covering the economic, social, and environmental dimensions of sustainable development are GDP per capita, the poverty rate, and CO₂ emissions. Although sustainable development scholars have developed several aggregated indices of human welfare (such as the Index of Sustainable Economic Welfare or the Genuine Progress Indicator, see e.g. Bossel, 1999; Daly & Cobb, 1990; Wackernagel et al., 2002), methodological concerns have prevented these and other aggregated measures from playing a significant role in sustainable development policy-making. So far, only four countries (Finland, Italy, Slovakia, and Slovenia) employ aggregated indices such as the Human Development Index or the Ecological Footprint (Steurer & Hametner, 2013).

As Steurer and Hametner (2013) show for 24 EU Member States, the size of indicator sets differ considerably: while some countries monitor their SD strategies with a small set of 12 (France) or 28 (Germany) indicators, others employ rather comprehensive sets with more than 100 indicators (see e.g. Denmark, Italy, Latvia, Switzerland, and the UK). As large sustainable development indicator sets may be too complex to provide meaningful feedback for policymakers, Denmark and the UK complement their comprehensive indicator sets with a smaller set of so-called headline indicators (OECD, 2003; Pinté'r et al., 2005). Usually, the indicator data is collected by the public sector itself (e.g. by statistical offices, environment ministries, and environment agencies), and it is published in indicator reports either by the ministry that leads the sustainable development strategy process (most often the environment ministry) or by the agencies who collect the data (sometimes with the support of researchers, see e.g. Eurostat, 2011).

A key strength of indicator monitoring is that it provides a comprehensive (although by no means unambiguous) picture of positive and negative trends in sustainable development. However, the language that indicators speak also has its limits. Although a key purpose of indicators is to make the pursuit of sustainable development more transparent, very few countries justify their choice of indicators. Furthermore, although a key purpose of most indicator sets is to monitor the implementation of sustainable development strategies, and although most countries developed their indicators based on policy objectives, the linkages among strategy objectives and indicators are rarely made explicit and are usually weak (Steurer & Hametner, 2013). Finally, compared to single economic indicators such as GDP growth or unemployment statistics, SD indicator reports rarely obtain significant public or political attention (for Switzerland, see Wachter, 2010).

2.2 Policy Evaluations

Policy evaluations are applied analyses of planned (ex ante) or existing (ex post) policies in which policy goals are either taken for granted (goal-based) or explicitly questioned (goal-free) (Davidson, 2005). As an empirical analysis for the climate policy field highlights, most evaluations focus their judgement on effectiveness (or goal-attainment), followed by efficiency and cost effectiveness (Huitema et al., 2011). Even when we narrow the wide field of policy evaluation to ex ante integrated assessments, it is difficult to provide a brief and general overview of recent developments because practices exploded in the 2000s. On the one hand, the proliferation of integrated assessments was due to the rise of the sustainable development concept itself and on the other hand, it was driven by the good governance and better regulation reform movements, in particular in Europe (Hertin et al., 2008; Kidd & Fischer, 2007). Even the terminology used to describe integrated assessments is difficult to follow. Whether referred to as 'integrated impact assessment', 'sustainability (impact) assessment', 'sustainability (impact) appraisal', or 'strategic impact assessment'—the practices have in common that they 'are the bringing together of environmental, social, and economic considerations and the balancing of these different substantive concerns in a single appraisal exercise' (Kidd & Fischer, 2007, p. 233). Appraisals can be concerned with cross-sectoral or sectoral policies, they can be conducted internally by public administrators and externally by researchers or consultants, and their conduct can be voluntary or mandatory (the latter is e.g. the case for all items included in the European Commission's Legislative and Work Programme). However, a closer look into the assessment practices of various countries and the European Commission shows that the frequency of integrated assessments is below what was expected based on the rhetoric and guidance from governments (for the EU, see e.g. European Commission, 2002), and that in practice these assessments are less 'integrated' than they claim to be. Whether countries add environmental aspects to their conventional regulatory impact assessments (such as Canada or the Netherlands) or apply newly developed impact assessment methods (such as the European Commission or the UK): empirical analyses show that assessment practices in both systems usually focus on short-term economic costs and benefits while often downplaying (or ignoring) environmental issues (Hertin et al., 2008; for the UK see Russel & Jordan, 2007). Thus, Hertin et al. (2008, p. 130) see the risk 'that integrated policy appraisal becomes a new label for what is in practice little more than conventional RIA'. So, the actual practice of integrated assessments dilute one of their key theoretical strengths, i.e. that they have the potential to provide balanced judgements and recommendations based on scientific methodologies.

2.3 Peer Reviews

Peer reviews not only have a long-standing tradition in science but also among public administrators, e.g. in the OECD and the EU contexts. As Lehtonen (2007, p. 15) emphasizes,

If there is a working method, a 'trademark', typical of the Organisation for Economic Co-operation and Development (OECD), it is peer review [...]. Carried out on member countries' economic policies since OECD's creation in 1961—and later in practically all areas of OECD activity, including environmental policy¹—peer reviews are a key instrument by which the OECD exercises power over member countries.

Lehtonen (2007, p. 19) defines peer reviews as processes 'whereby members of a profession evaluate activities of other members of the same profession, criteria of judgement being drawn directly from the profession's own quality criteria'. If peer reviews are carried out by colleagues working on similar issues in another country or organization, not only the reviewed but also the reviewers learn. More recently, peer reviewing became a popular type of ex post assessment in the context of sustainable development strategies. In 2005, France was the first country to conduct a thorough peer review of its sustainable development strategy. Since then, Norway, the Netherlands, the Slovak Republic, and Germany have also preferred peer reviews to researcher-led ex post evaluations (Berger & Steurer, 2006; Spangenberg & Niestroy, 2010). All peer reviews have been conducted by public administrators and other experts from two or more peer countries, supported by policy consultants. While consultants often prepare the background material and provide support in drafting the peer review report, the actual reviewing as well as the formulation of recommendations is a timeconsuming task left to the peers. A key strength of peer reviews is that peers know the inner workings of public administration well, so their recommendations take practical limitations into consideration. The downside of this strength, however, is that their recommendations are usually less critical or demanding than those generated by non-peer experts.

2.4 Formal Audits

Audits are the most formalized of assessment approaches, with clear professional guidelines and internationally recognized practices. National audit organizations are respected state institutions with a long history. Their original focus was financial probity—ensuring government funds were properly managed—but over time their work extended into 'performance audits'. The focus is on the evaluation of conduct and the assessment of performance with respect to stated objectives and guidelines. Important here are notions of a 'decision chain' (linking original decisions to actions and perhaps ultimate outcomes) and an 'audit chain' (linking external auditors with the internal quality management systems of the organizations they are evaluating). Organizations such as the Netherlands Court of Audit, the UK National Audit Office, or the Canadian Parliamentary Commissioner for the Environment and Sustainable Development (part of the Office of the Auditor General of Canada) now routinely audit government programmes related to the environment and sustainability. At the international level, the Working Group on Environmental Auditing of the International Organization of Supreme Audit Institutions has developed guidance for auditors (INTOSAI, 2004), assists with international audits of cross border environmental issues, and the audit of international environmental agreements.

Environmental performance audits are usually conducted in close contact with those under review. They typically work on the following premise: 'given what government has stated its policy to be, how have public institutions performed?' Auditors may point to inconsistencies among policies, to cases where policies are inadequate to achieve stated goals, and to implementation failures. They may note administrative weaknesses and poor procedures. But they do not try to tell policy-makers how to make policy or administrators how to run their departments. Internationally, this is among the most successful of assessment practices with thousands of audits on environmental policy issues around the world. Their strengths are close attention to detail, good grounding in policy reality, connection to decision-makers, and high legitimacy. The weakness is that they cannot challenge the fundamental orientation of policy or the assumptions on which it rests.

2.5 Specialist Reports

Since specialist reports are the least formalized type of feedback described here, they can take on many shapes and they can be produced by a variety of institutions. In the context of sustainable development, national councils for sustainable development are the designated 'watchdogs' that are supposed to produce policy-related specialist reports (Niestroy, 2005, 2011). Based on the assumption that stakeholder participation helps to reconcile conflicting interests and, in turn, facilitate policy integration across sectors, most national councils for sustainable development are multi-stakeholder organizations rather than scientific think tanks, bringing together business leaders, academics, and NGO representatives. Many European countries established them in the early 2000s in accordance with UN and OECD recommendations (European Environment Agency, 2005a, p. 29; Niestroy, 2005, 2011). A key purpose of sustainable development councils is to critically monitor and discuss national sustainable development policies and to advise governments on sustainable development-related issues, inter alia with specialist reports. The most prominent examples for 'watchdog-like' councils in Europe are the German Council for Sustainable Development (RNE) and—from 2000 until March 2011—the UK Sustainable Development Commission (SDC).² As the archive website of the SDC reads, it 'held Government to account to ensure the needs of society, the economy and the environment were properly balanced in the decisions it made and the way it ran itself'.³ The SDC reported to the Prime Minister and the First Ministers in the devolved regions. The still operational German RNE was established in 2001 as a 'by-product' of the German sustainable development strategy, and it advises the Federal Government on all items of the National Sustainability Strategy, pinpoints specific areas of action and projects, and tries to facilitate the public discussion on sustainability.⁴ In contrast to the watchdog model pursued in Germany (and, until 2011, in the UK), Finland and some Eastern European countries have established 'mixed councils' in which high-level politicians interact directly with stakeholders (Berger & Steurer, 2006; European Environment Agency, 2005a, pp. 29, 37). Irrespective of how a council for sustainable development is set up, the (theoretical) strengths of the specialist reports they produce are that they have the freedom to address politically hot topics and that they often represent a consensus of diverse groups of actors brought together by the councils. Overall, however, only a few councils have been able to establish themselves as critical interlocutors in the discussion of policy, and most achieved only modest levels of visibility (usually restricted to expert communities in science and public administration).

3. Systematic Assessments and the Policy and Politics Cycles

The preceding discussion should have already established two points: first, the diversity and multiplicity of systematic assessment practices that have been brought to bear on sustainable development and second, the limited impact which they appear actually to have had on ongoing rounds of policy-making. To gain insight into the second point, it is helpful to consider not only the characteristics of individual approaches but also the policy and political cycles which define the context within which they operate. Although both cycles are familiar to students of contemporary democracies, the politics cycle has received less attention from policy scholars.

The heuristic of the policy cycle has a long pedigree. It disaggregates the policy process into four basic stages: (i) 'agenda setting' (determining the issues to be addressed); (ii) 'policy formulation' (assessing options and designing the policy response); (iii) 'implementation' (carrying policy forward in practice); and (iv) 'evaluation' (assessing results and considering adjustments) (Howlett & Ramesh, 2003; Lasswell, 1956). The process is understood as iterative, with each round influenced by the experience of previous cycles. Of course, the real world is messier. The policy system will be handling many issues simultaneously and some of these will be enmeshed in multiple policy domains; phases of policymaking may overlap; the agenda may move on before policies are formulated; new policies may be abandoned before they are ever implemented; and established policies may keep being renewed without any

serious evaluation. Indeed, the timing of policy reassessment may be driven more by contingent circumstances (disasters, political scandal, external shocks, and so on) than by any institutionalized process of policy learning (Howlett & Ramesh, 2003; Jann & Wegrich, 2003). This said, the notion of the policy cycle provides a convenient point of reference for considering the operation of systematic feedback processes.

Several observations can be made about the ways monitoring and assessment practices for sustainability feed into the policy cycle (for an illustration see Figure 1). First, these exercises can be relevant at any point in the policy process. Although the model implies that the 'evaluation phase' is the privileged entry point for feedback, these practices can also influence the selection of topics that make it on to the policy agenda (for example, by highlighting problems in one area), inform decisions about policy design (by establishing what has and has not worked elsewhere), and encourage adjustment of implementation strategies (by identifying implementation 'gaps'). Second, specific monitoring and assessment practices have a stronger affinity for particular phases of the policy cycle. Obviously, ex ante assessments are most appropriate at the policy formulation phase, but they can also be applied at the implementation phase. Ex post evaluations and formal audits are typically focused on the efficacy of implementation and some aspects of policy formulation (assessing performance in relation to established goals and procedures). Peer reviews look backward on policy goals, policy design, and implementation. On the other hand, monitoring (for example, indicator reporting) and 'specialist reports' provide feedback at any point in the cycle. Third and most importantly, various monitoring and assessment practices both complement and compete with one another. They are complementary because they use different forms of expertise to provide different types of information to policy-makers at different points in the policy cycle. Competition, on the other hand, may take place through competing institutions (some

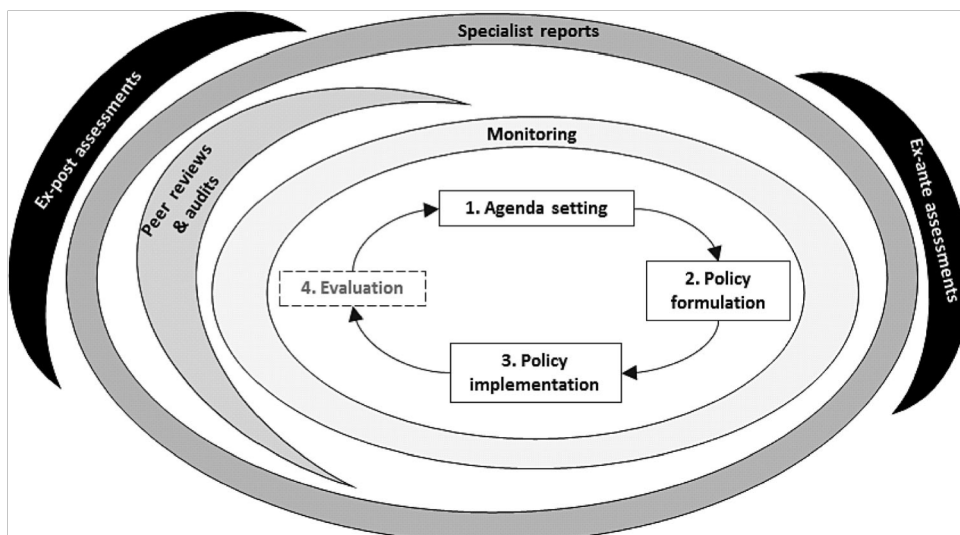


Figure 1. Systematic assessments around the policy cycle.

evaluators carry more authority than others),⁵ competing topics (some issues gain more interest than others, and there are more issues in play than publics and decision-makers can track), competing forms of expertise or methods, and finally through divergent findings and conclusions.

This leads us to another cycle central to policy-making where competition among different types of evidence is still more evident: the ‘politics cycle’. Through the politics cycle, those who exercise political power in democratic states are elected, discharge their responsibilities, and are eventually replaced. The basic rhythm of the politics cycle is determined by the periodic conduct of general elections which populate representative institutions, underpin the formation of governments, and help to define political priorities. Coalition negotiations, the distribution of cabinet posts, and the establishment of ministerial and departmental mandates follow elections but may occur more frequently. Within an election period, the politics cycle involves the regular sessions of the legislature and meetings of the executive, the continuous formation of government priorities, and the annual budget process through which state expenditure is authorized. This cycle, and the everyday business of politics that invigorate it, involves continuous competition among parties within coalition governments and between government and opposition parties. This competition, which peaks during election campaigns, conditions the management of policy issues (i.e. the contents of the many policy cycles) in general and the use of systematic assessments in policy-making in particular. More precisely, competition in the politics cycle draws attention to some kinds of evidence (including systematic assessments) and away from others, rules some potential solutions in and others out, enables progress on some questions and leads to stalemate on others.

Four main points about the ways existing sustainability assessment practices relate to this politics cycle stand out (for an illustration see [Figure 2](#)). First, systematic assessments are of potential significance throughout the politics cycle. Since systematic assessments provide feedback on social and environmental

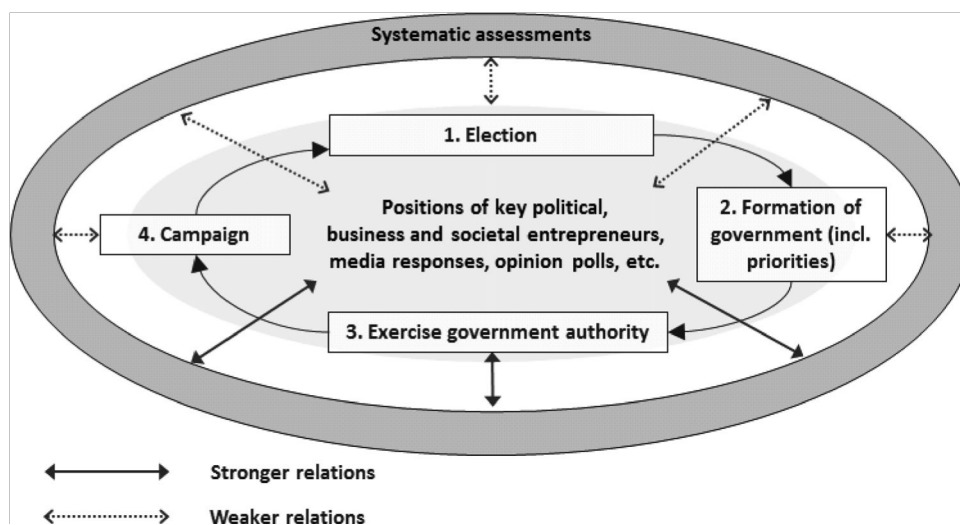


Figure 2. Systematic assessments and other kinds of evidence in and around the politics cycle.

trends and on the effectiveness (or otherwise) of policy designs and implementation, they can link into broader political debates about the appropriate orientation of public policy and the competence of the current political authorities. If the responsible government bodies fail to address the findings of systematic assessments adequately these can be taken up by other governmental actors (e.g. in other ministries or at other levels of government), by opposition parties, or by non-state actors (such as environmental NGOs). These may call publically (or lobby privately) for a policy response to the assessment. In addition, the media can transmit the findings to larger audiences, and weigh-in with editorials and commentary advocating policy action, and/or linking the issue to a broader evaluation of government performance. Since politicians must ultimately face the electorate, they can be sensitive to public controversies which could damage their reputation for effective, efficient, and fair administration.

So assessments may gain leverage in the political arena and influence how the issue is subsequently handled by officials in the policy cycle.

Second, systematic assessments of the type discussed here typically play only a marginal role in active electoral campaigning and elections. As the literature on electoral politics makes clear, elections are not just about policy performance (let alone about systematic assessment thereof), but also about ideological orientation, party mobilization, campaign funding, symbolism, personalities, and emotions (Glaser & Salovey, 1998; Glazer et al., 1998; Iversen, 1994). And to the extent that performance does matter, economic conditions usually eclipse other domains. Typically, there is simply too much going on in these sharp partisan contests for particular sustainability-related assessments to play much of a role at the decision point where voters actually determine the fate of governments (and indirectly the tenor of subsequent rounds of policymaking).⁶ In exceptional cases, a particularly damning assessment that undermines public confidence in a government can take centre stage in a campaign, but most examples relate to systematic corruption, large-scale waste of public funds, or foreign policy fiascos rather than the less well established field of sustainable development.

Third, systematic assessments are obviously only one among many different types of evidence or feedback to which politicians pay attention; and in political terms, they are usually far from the most potent ones that take centre stage in the politics cycle. Between elections (the ultimate feedback from voters to political leaders), the most important forms of political 'feedback' are provided by party organization, major interest groups, media, other elections (e.g. at the sub-national level), and opinion polls. Party links political leaders with their activist base and financial supporters. Major interest groups (especially business organization) have direct channels to government and provide constant commentary of government performance. The media reflect, but also help to shape, public discourse, and other (e.g. sub-national) election results and opinion polls provide the clearest indicator of future voting intentions. All these forms of 'societal feedback' are typically much more revealing of the currents effecting politicians' futures than the sort of systematic assessments discussed above.

Fourth, even if systematic assessments do manage to gain political salience they do not necessarily lead to substantial policy changes because they can be politically 'managed' in other ways. Politicians can dispute certain findings; they can 'accept' certain findings but make only symbolic adjustments; or they can defer judgement, pending the results of further inquiries. Since the media as well as voters have a short attention span, other issues will soon clamour for public attention.

What we witness in the politics cycle is not only competition among political actors but also among different kinds of evidence. It is obvious but often overlooked that actual events and societal as well as political feedback to decisionmakers (e.g. from citizens, media outlets, business, and political groups) also represent evidence, e.g. about the state of society or respective subsystems. Since 'Science never compels just one political outcome' (Pielke, 2004, p. 406), and since the plurality of science competes with many other kinds of evidence which can all be interpreted or managed in different ways, what counts as credible evidence will always be contested. Thus, the challenge for knowledge-based policy-making, organizing effective science-policy interfaces and/or enhancing reflexive governance in the political sphere is to make sense of a cacophony of voices in democratically transparent, accountable, and legitimate ways.

4. Competition, Cacophony, and Reflexivity

In the previous two sections, we have surveyed different systematic assessment practices and explored their relationship to the policy and politics cycles in the context of sustainable development. Our basic observation is that, in spite of the proliferation of countless systematic state-sponsored assessments, their influence remains limited. Indeed, it seems that the assessment practices most explicitly associated with sustainable development (such as national indicator sets, integrated sustainability assessments, and the specialist reports of multi-stakeholder councils or sustainability watchdogs), as opposed to the practices focused on the many specific issues of which sustainable development is composed (such as particular environmental problems), are among the least influential exercises. The difficulty of operationalizing and communicating sustainable development, the inherent complexity of the socio-economic and environmental issues involved, and the relative inexperience of the institutions engaged in such assessments may go some way to explaining this tendency. But it should also be recognized that the messages these assessments convey are often unwelcome to political and policy actors. Addressing sustainability requires a significant break with existing societal practice. Responding substantively to assessments that identify negative socio-environmental trends or point to the inadequacy of the existing government response usually implies politically challenging policy adjustments that require a reallocation of resources, disturb established entitlements, or threaten entrenched economic interests. So, it is not surprising that politicians and officials often choose to defer action, and let awkward assessments be buried by the constant flow of new evidence that lands in their in-trays.

What then can be said in relation to the three perspectives outlined at the outset? From the standpoint of evidence-based policy-making, the array of sustainability-related assessments appears to constitute exactly the type of systematic, expert-based evidence of which this perspective approves. What it often overlooks, however, is that evidence produced by systematic assessments represents only a small fraction of the evidence that actually must be weighed by government actors on a daily basis, and which includes also the impact of actual events as well as feedback coming from interest groups, political parties, the media, opinion polling, and so on. As the announcement of the German 'Energiewende' (which included the phase out of nuclear energy in Germany) in spring 2011 illustrates, real events (in this case, the nuclear accident in Fukushima) coupled with media coverage, public opinion, and political pressure by societal organizations usually represent more powerful pieces of evidence than do comparatively abstract assessments. Thus, we think that discussions of evidencebased policy-making often obscure the crucial political issue of what counts and/or is recognized as credible evidence. Poor decisions may have less to do with the presence or absence of systematically gathered evidence, but more with the power resources that influence what sources and types of evidence prevail over others.

Turning to science-policy interfaces, many of the assessment practices reviewed above aim to act as 'bridging mechanisms' linking science to the policy domain. While they cannot entirely reconcile the 'two worlds' of science and policy, they can at least help to bring some science-based methodologies, analyses, and metrics into the policy realm. But what of the three specific issues of the 'strategic use of science', the 'ritual or symbolic use of science', and the 'contested status of scientific knowledge'? Obviously, all three patterns are present in the context of sustainable development, in particular the rejection of disturbing scientific findings (e.g. on climate change) and the ritual use of systematic assessments. From what we have seen so far, it is clear that many of the reports on sustainability (and the processes through which they are generated) have a largely 'symbolic' function. In most countries, this seems to be the case with respect to monitoring, evaluating, and reviewing sustainable development strategies (Casado-Asensio & Steurer, in press; Meadowcroft, 2007b; Steurer, 2008). These and other assessments are carried out because it is politically expedient to do so, but for the most part decision-makers do not expect them to lead to significant change. However, since symbols and rituals can acquire political salience in certain circumstances, doing something for symbolic reasons

is not equivalent to not doing it at all (Steurer, 2008). What our approach adds to the literature on science–policy interfaces is the idea of competition among different systematic assessments around the policy cycle and among these and other types of evidence particularly in the politics cycle. The notion of competition suggests that it is not necessarily a lack of evidence, but rather its abundance and the resulting cacophony, that both forces policy-makers to struggle and allows them to cherry pick what suits them best. Thus, we hypothesize that in political life, scientific evidence is often neither used strategically nor contested systematically, but simply ‘overwritten’ with other kinds of evidence representing alternative rationalities and priorities.

This brings us to the issue of reflexive governance. Earlier we noted the potentially problematic relationship between the systematic sustainability assessments considered here and reflexive governance, because the former remain closely tied to the governmental sphere, rely primarily on expert (elite) judgement, and often embody a narrow instrumental approach that takes existing governance relationships as given. One possible conclusion is that these practices are of somewhat marginal importance for reflexive governance—that they are more an extension of the modernist impulse for ‘technocratic programming’ than a dynamic mechanism to advance genuine societal reflexivity. Let us return to the five forms of sustainability assessment in light of the preoccupation with reflexivity.

- . Monitoring with quantitative indicators can provide a useful (although by no means unambiguous) picture of trends in sustainable development. But as applied to date, these indicators typically have weak (or inexistent) links to strategic policy objectives and decision-making. These loose connections mean that the assessments scarcely provide a simple feedback loop for decision-makers, let alone facilitating a more transparent and reflective process intimated by reflexivity—that would encourage regular interrogation of the plural, dynamic, and conflicting social goals associated with sustainable development and the difficulties of measuring them.
- . Ex ante strategic impact assessments are potentially important tools to inform decision-making, but governments apply them less frequently than promised and often ignore the social and (especially) environmental dimensions. Thus, they are too often more formal than substantial. Even if governments implemented them more consistently and in good faith, fundamental flaws, in particular the absence of any systematic connections to ex post assessment years after major development consents, and/or lesson drawing about strategic assessments themselves, mean they fall well short of the ambitions of reflexive governance.
- . Peer reviews of government policy for the environment and sustainable development are sensitive to the constraints facing public administration, and may facilitate some reflective learning as those conducting the assessment draw lessons from the target jurisdictions and integrate civil society perspectives. But the socialization of most participants in established routines of government and the close collaboration with those reviewed means their conclusions are typically less critical than those of non-peer experts and more hesitant to adopt alternative frames and approaches.
- . Audits are also characterized by goal-based grounding in policy reality, and national audit bodies typically enjoy high legitimacy. Auditors may appeal to perspectives from outside the ambit of government, but their mandate typically precludes any challenge to the fundamental orientation of policy or the assumptions on which it rests.
- . Specialist reports have the freedom to address politically hot topics, and authors may deliberately bring multiple perspectives to bear on an issue. And yet assessments focused on sustainable development often garnish only low levels of public and political attention, in particular when published by comparatively weak institutions such as (most) sustainable development councils.

Thus, each type of state-sponsored assessment appears to fall short of the rather demanding requirements of reflexivity. Limits to the diversity of perspectives that are considered and to the depth of self-critical interrogation, as well as weak connections to subsequent rounds of decision-making, place them far from the reflexive governance ideal. Yet such a conclusion may be too hasty. In the first place, many of these efforts do try to capture multiple perspectives (or even different modes of social rationality). Indeed, individual assessments often display laudable 'integrative' ambitions (across environmental media, policy domains, organizational boundaries, economic/social/economic dimensions, knowledge types, and so on), while to some extent acknowledging value pluralism and the contestability of policy goals and processes. Of course, it could be argued that this represents at best a 'reflective' rather than a truly 'reflexive' stance, which would demand deeper appreciation of the mutual constitution of governance subjects and objects (Stirling, 2006). And yet such 'reflection'—developing an openness to a multiplicity of perspectives and the contingency of goals—is generally a precondition to a more elaborate reflexivity, and in this sense may represent a contribution towards it.

More importantly, in any given policy sphere we are confronted not with isolated assessment events, but with an ongoing stream of assessment activities of different kinds, covering overlapping topics, employing varied methodologies, integrating alternative perspectives, and so on. And here the multiplicity of assessments can begin to approximate the plural epistemological and ontological stances associated with reflexivity. We have already noted the competition among assessment exercises in the policy and (particularly) the political cycles, which can obscure potentially significant findings as players manoeuvre to favour one issue or perspective over another. Certainly, the jostling surrounding assessments in the policy and political cycles is far from the measured juxtaposition and provisional reconciliation of perspectives intimated by the ideal of reflexive governance. And yet it does provide an opening into which societal actors can take up struggles for the recognition of neglected viewpoints and to drive forward more consequent reform.

This potential, and an appreciation of the extent to which both the policy and the politics cycles are infused with politics, can lead one to turn back on the discussion of reflexive governance itself, and encourage one to reflect on 'reflexivity' in a more critical way. In the first place, we can note that reflexive governance is not an all or nothing proposition, but rather a matter of degree: something that is more or less manifest in specific social circumstances. Second, we should be careful not to get carried away with an ideal construct of reflexivity that makes almost impossible demands on societal institutions that are supposed continuously to interrogate their own foundations, relativize their judgements, remain ready to re-open any question at any time, appreciate the irony of governance and the limits of any intervention, view all commitments as conditional, while learning continuously from practice, and so on. Ultimately, such a 'reflexivity' is more an unreflective construct of social theorists (who after all are paid to reflect and not govern) than a reflexivity that could actually be embodied in governance practices. Most importantly, we should see reflexive governance not primarily as a property of novel institutional forums (specialized decision niches), but as emerging iteratively from broad patterns of interaction across a policy domain or indeed across the political system as a whole. This directs attention away from the design of specialized techniques or novel sites for enhancing reflexivity (although these can play important roles), and more towards the consideration of how reflexivity emerges as a product of interactions among actors in broader policy and political processes. In particular, it raises the issue of reflexivity at the interstices of the political and policy systems: how can political systems consciously be adapted to enhance reflexive governance potentials? After all, politics is the main avenue through which society can collectively 'bend back upon itself' to reform deliberately its socio-economic foundations. Thus, democratic politics, which in one sense is about closure and the exercise of power, must be recognized simultaneously as the subject, object, and medium of reflexivity.

In this context, the true value of state-sponsored assessment practices of the type considered here lies in their potential to open up the political sphere. Despite their many shortcomings, and the regularity with which pertinent findings are ignored in the policy and political systems, the analyses and insights they provide can furnish ammunition to state and non-state actors interested in securing change. In this way, such systematic assessments can potentially bridge not just science and policy, but also state and society.

Acknowledgements

We thank three anonymous reviewers from the Journal of Environmental Policy & Planning and Michael Pregernig for critical comments on a previous draft of the present paper. They propelled us not simply to revise but to rewrite large parts of it. James Meadowcroft acknowledges the support of the Canada Research Chair programme for supporting the research on which this article is based.

Notes

1. Environmental Performance Reviews are carried out by the OECD since 1992 (Lehtonen, 2007, p. 16).
2. For the countless specialist reports produced by the SDC, see the archive website at <http://www.sd-commission.org.uk/publications.php>; for the various publications produced by the German RNE, see <http://www.sd-commission.org.uk/publications.php>.
3. <http://www.sd-commission.org.uk/presslist.php/119/what-next-for-sustainable-development> (accessed 10 December 2011).
4. <http://www.nachhaltigkeitsrat.de/en/the-council/fact-sheet/?size¼%EF%BF%BDblstr%25D> (accessed 10 December 2011).
5. National audit institutions enjoy particularly high legitimacy. Financial audits have a tradition going back several hundred years and are important in ensuring the responsibility of the executive to parliament. Parliamentary committees have considerable legitimacy and their reports derive additional authority if they reflect cross-party support. Statistical offices and government agencies and departments are also generally well regarded. The institutions active in this area with perhaps the weakest authority are sustainable development watchdogs and stakeholder councils. They are relatively recent creations, far from the seat of power. Of course, a reputation for sound assessment can be built (or eroded), and over time institutional actors can increase (or lose) influence.
6. A rare exception to this rule can be found in the Netherlands where it is customary since 1986 for the Bureau for Economic Policy Analysis and the Environment Assessment Agency to assess party election manifestos with regard to their economic and environmental impacts during the election campaign phase since 1986 (Huitema & Turnhout, 2009, pp. 585–590).

References

- Beck, U. (1994) The reinvention of politics: Towards a theory of reflexive modernization, in: U. Beck, A. Giddens & S. Lash (Eds) *Reflexive Modernization: Politics, Tradition and Aesthetics in the Modern Social Order*, pp. 1–55 (Cambridge: Polity Press).
- Berger, G. & Steurer, R. (2006) The Finnish National Commission on Sustainable Development (SD) and the UK's SD Commission: Two Distinct Models of Involving Stakeholders in SD Policy Making, ESDN Quarterly Report December 2006. Available at http://www.sd-network.eu/?k¼quarterly%20reports&report_id¼43 (accessed 7 January 2013).
- Bina, O. (2008) Strategic environmental assessment, in: A. Jordan & A. Lenschow (Eds) *Innovation in Environmental Policy*, pp. 134–151 (Cheltenham: Edward Elgar).
- Bossel, H. (1999) *Indicators for Sustainable Development: Theory, Method, Applications* (Winnipeg, Canada: IISD).
- Casado-Asensio, J. & Steurer, R. (in press). Integrated strategies on sustainable development, climate change mitigation and adaptation in Western Europe: Communication rather than coordination, *Journal of Public Policy*.
- Christie, C. A. & Alkin, M. C. (1999) Further reflections on evaluation misutilization, *Studies in Educational Evaluation*, 25(1), pp. 1–10.
- Crabbe, A. & Leroy, P. (2008) *The Handbook of Environmental Policy Evaluation* (London: Earthscan).
- Daly, H. & Cobb, J. (1990) *For the Common Good. Redirecting the Economy Toward Community, the Environment, and a Sustainable Future* (London: Green Print).
- Davidson, E. J. (2005) *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation* (Thousand Oaks, CA: Sage).
- European Commission (2001) *European Governance: A White Paper*, COM(2001) 428 final.
- European Commission (2002) *Impact Assessment*, COM(2002) 276 final.
- European Commission (2006) *A Strategic Review of Better Regulation in the European Commission* (Brussels: European Commission).
- European Council (2006) *Renewed EU Sustainable Development Strategy*, 10917/06.
- European Environment Agency (2005) *Environmental policy integration in Europe: State of play and an evaluation Framework* (Copenhagen: European Environment Agency).
- Eurostat (2011) *Sustainable Development in the European Union: 2011 monitoring report of the EU sustainable development strategy* (Luxembourg: Eurostat).
- Feindt, P. (2012) Reflexive governance and multilevel decision making in agricultural policy: Conceptual reflections and empirical evidence, in: E. Brousseau, T. Dedeurwaerdere & B. Siebenhu"ner (Eds) *Reflexive Governance for Global Public Goods*, pp. 159–178 (Cambridge, MA and London: MIT Press).
- Giddens, A. (1990) *The Consequences of Modernity* (Oxford: Polity Press).
- Glaser, J. & Salovey, P. (1998) Affect in electoral politics, *Personality and Social Psychology Review*, 2(3), pp. 156–172.
- Glazer, A., Gradstein, M. & Konrad, K. (1998) The electoral politics of extreme policies, *The Economic Journal*, 108(451), pp. 1677–1685.
- Grin, J. (2006) Reflexive modernization as a governance issue, or designing and shaping Restructuration, in: J. Voss, D. Bauknecht & R. Kemp (Eds) *Reflexive Governance for Sustainable Development*, pp. 54–81 (Cheltenham: Edward Elgar).
- Grundmann, R. (2007) Climate change and knowledge politics, *Environmental Politics*, 16(3), pp. 414–432.
- Grundmann, R. (2009) The role of expertise in governance processes, *Forest Policy and Economics*, 11(5), pp. 398–403.
- Hendriks, C. & Grin, J. (2008) Contextualising reflexive governance: The politics of Dutch transitions to sustainability, in: J. Newig, J.-P. Voss & J. Monstadt (Eds) *Governance for Sustainable Development: Coping with Ambivalence, Uncertainty and Distributed Power*, pp. 141–159 (London and New York: Routledge).
- Hertin, J., Jacob, K. & Volkery, A. (2008) Policy appraisal, in: A. Jordan & A. Lenschow (Eds) *Innovation in Environmental Policy*, pp. 114–133 (Cheltenham: Edward Elgar).
- Howlett, M. & Ramesh, M. (2003) *Studying Public Policy. Policy Cycles and Policy Subsystems*, 2nd ed. (Oxford: Oxford University Press).
- Huitema, D. & Turnhout, E. (2009) Working at the science–policy interface: A discursive analysis of boundary work at the Netherlands Environmental Assessment Agency, *Environmental Politics*, 18(4), pp. 576–594.
- Huitema, D., Jordan, A., Massey, E., Rayner, T., van Asselt, H., Haug, C., . . . Strippel, J. (2011) The evaluation of climate policy: Theory and emerging practice in Europe, *Policy Sciences*, 44(2), pp. 179–198.
- INTOSAI (2004) *Sustainable Development: The Role of Supreme Audit Institutions*, INTOSAI Working Group on Environmental Auditing. Available at <http://www.environmental-auditing.org/LinkClick.aspx?fileticket¼cFPPm1WGwRY%3D&tabid¼473&mid¼4602> (accessed 7 January 2013).
- Iversen, T. (1994) The logics of electoral politics, *Comparative Political Studies*, 27(2), pp. 155–189.
- Jann, W. & Wegrich, K. (2003) Phasenmodelle und Politikprozesse: Der Policy Cycle, in: K. Schubert & N. C. Bandelow (Eds) *Lehrbuch der Politikfeldanalyse*, pp. 71–106 (Mun"chen/Wien: R. Oldenbourg Verlag).
- Juntti, M., Russel, D. & Turnpenny, J. (2009) Evidence, politics and power in public policy for the environment, *Environmental Science & Policy*, 12(3), pp. 207–215.
- Kidd, S. & Fischer, T. B. (2007) Towards sustainability: Is integrated appraisal a step in the right direction? *Environment and Planning C: Government and Policy*, 25(2), pp. 233–249.
- Lasswell, H. (1956) *The Decision Making Process* (College Park: University of Maryland Press).
- Leeuwen, S. Van. (2004) Developments in environmental auditing by supreme audit institutions, *Environmental Management*, 33(2), pp. 163–172.
- Lehtonen, M. (2007) Environmental policy integration through OECD peer reviews: Integrating the economy with the environment or the environment with the economy? *Environmental Politics*, 16(1), pp. 15–35.
- Loorbach, D. (2007) *Transition Management: New Mode of Governance for Sustainable Development* (Utrecht: International Books).
- Meadowcroft, J. (2004) Deliberative democracy, in: R. Durant, D. Fiorino & R. O'Leary (Eds) *Environmental Governance Reconsidered: Challenges, Choices and Opportunities*, pp. 183–218 (Cambridge, Mass.: MIT Press).
- Meadowcroft, J. (2007a) Who is in charge here? Governance for sustainable development in a complex world, *Journal of Environmental Policy and Planning*, 9(3–4), pp. 299–314.
- Meadowcroft, J. (2007b) National sustainable development strategies: Features, challenges and reflexivity, *European Environment*, 17(3), pp. 152–163.
- Mickwitz, P. (2006) *Environmental Policy Evaluation: Concepts and Practice* (Helsinki: Finnish Society of Sciences and Letters).

- Niestroy, I. (2005) *Sustaining Sustainability: A Benchmark Study on National Strategies Towards Sustainable Development and the Impact of Councils in Nine EU Member States* (Utrecht: Lemma).
- Niestroy, I. (2011) Sustainable development councils at national and sub-national levels stimulating informed debate: Stocktaking. *SDG 2012*. Available at <http://www.earthsummit2012.org/projects/sdg2012/think-pieces> (accessed 30 August 2013).
- Nilsson, M., Jordan, A., Turnpenny, J., Hertin, J., Nykvist, B. & Russel, D. (2008) The use and non-use of policy appraisal tools in public policy making: An analysis of three European countries and the European Union, *Policy Sciences*, 41(4), pp. 335–355.
- OECD (2002) *Improving Policy Coherence and Integration for Sustainable Development: A Checklist*. Available at <http://www.oecd.org/greengrowth/tools-evaluation/2763153.pdf> (accessed 30 August 2013).
- OECD (2003) *OECD Environment Indicators. Development, Measurement and Use* (Paris: OECD).
- OECD-DAC (2001) *Strategies for Sustainable Development: Practical Guidance for Development Cooperation* (Paris: OECD). Available at <http://www.nssd.net/pdf/guse.pdf> (20 October 2002).
- Pielke, R. A. (2004) When scientists politicize science: Making sense of controversy over the skeptical environmentalist, *Environmental Science & Policy*, 7(5), pp. 405–417.
- Pierre, J. (ed.) (2000) *Debating Governance: Authority, Steering, and Democracy* (Oxford: Oxford University Press).
- Pintor, L., Hardi, P. & Bartelmus, P. (2005) *Sustainable Development Indicators. Proposals for a Way Forward* (Winnipeg, Canada: IISD).
- Politt, C. (1998) Evaluation in Europe: Boom or Bubble? *Evaluation*, 4(2), pp. 214–224.
- Pregemig, M. (2005) *Between knowledge utilisation and accountable expertise: Studies on the roles of science and participation in environmental and natural resource policy*, Habilitation thesis.
- Pregemig, M., Hogl, K. & Nordbeck, R. (2012) The Politics of Sustainability Evaluation: Analysis of Three Austrian Strategies for Sustainable Development, in: M. Sedlacco & A. Martinuzzi (Eds) *Governance by Evaluation for Sustainable Development: Institutional Capacities and Learning*, pp. 21–44 (Cheltenham: Edward Elgar).
- Price, D. K. (1981) The spectrum from truth to power, in: T. J. Kuehn & A. L. Porter (Eds) *Science, Technology, and National Policy*, pp. 95–131 (Ithaca, NY; London: Cornell University Press).
- Radaelli, C. M. (1995) The role of knowledge in the policy process, *Journal of European Public Policy*, 2(2), pp. 160–183.
- Radaelli, C. (2004) The diffusion of regulatory impact analysis: Best practice or lesson-drawing? *European Journal of Political Research*, 43(5), pp. 723–747.
- Radaelli, C. (2007) Whither better regulation for the Lisbon agenda? *Journal of European Public Policy*, 14(2), pp. 190–207.
- Radaelli, C. M. & De Francesco, F. (2007) *Regulatory Quality in Europe: Concepts, Measures and Policy Processes* (Manchester: Manchester University Press).
- Renn, O., Webler, T. & Wiedemann, P. (eds) (1995) *Fairness and Competence in Citizen Participation: Evaluating Models for Environmental Discourse* (Dordrecht: Kluwer Academic).
- Rotmans, J., Kemp, R. & Van Asselt, M. (2001) More evolution than revolution: Transition management in public policy, *Foresight* 3(1), pp. 15–31.
- Russel, D. & Jordan, A. (2007) Gearing-up governance for sustainable development: Patterns of policy appraisal in UK central government, *Journal of Environmental Planning and Management*, 50(1), pp. 1–21.
- Sanderson, I. (2002) Evaluation, policy learning and evidence-based policy making, *Public Administration*, 80(1), pp. 1–22.
- Sharman, A. & Holmes, J. (2010) Evidence-based policy or policy-based evidence gathering? Biofuels, the EU and the 10% target, *Environmental Policy and Governance*, 20(5), pp. 309–321.
- Smith, A. & Stirling, A. (2008) Moving outside or inside? Objectification and reflexivity in the governance of socio-technical systems, in: J. Newig, J.-P. Voss & J. Monstadt (Eds) *Governance for Sustainable Development: Coping with Ambivalence, Uncertainty and Distributed Power*, pp. 159–182 (London and New York: Routledge).
- Spangenberg, J. H. & Niestroy, I. (2010) Politische Lernprozesse durch Peer Reviews, in: R. Steurer & R. Trutnigg (Eds) *Nachhaltigkeit regieren: Eine Bilanz zu Governance-Prinzipien und -Praktiken*, pp. 215–238 (München: Oekom).
- Stern, N. (2008) *Stern Review on the Economics of Climate Change* (London: HM Treasury).
- Steurer, R. (2008) Sustainable development strategies, in: A. Jordan & A. Lenschow (Eds) *Innovation in Environmental Policy? Integrating the Environment for Sustainability*, pp. 93–113 (London: Edward Elgar).
- Steurer, R. & Hametner, M. (2013) Objectives and indicators in sustainable development strategies: Similarities and variances across Europe, *Sustainable Development*, 21(4), pp. 224–241.
- Steurer, R., Berger, G. & Hametner, M. (2010) The vertical integration of Lisbon and sustainable development strategies across the EU: How different governance architectures shape the European coherence of policy documents, *Natural Resources Forum*, 34(1), pp. 71–84.
- Stiglitz, J. E., Sen, A. & Fitoussi, J.-P. (2009) *Report by the Commission on the Measurement of Economic Performance and Social Progress* (Paris). Available at http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf (accessed 30 August 2013).
- Stirling, A. (2006) Precaution, foresight, sustainability: Reflection and reflexivity in the governance of science and technology, in: J.-P. Voss, D. Bauknecht & R. Kemp (Eds) *Reflexive Governance for Sustainable Development*, pp. 225–272 (Cheltenham: Edward Elgar).
- Stockmann, R. (2007) Einführung in die evaluation, in: R. Stockmann (Ed) *Handbuch zur Evaluation: Eine praktische Handlungsanleitung*, pp. 24–70 (Münster: Waxmann).
- UNCED (1992) *Agenda 21*. New York. Available at <http://www.un.org/esa/sustdev/agenda21.htm> (8 October 2002).
- Voss, J.-P. & Kemp, R. (2005) Reflexive governance for sustainable development—incorporating feedback in social problem solving, *IHDP open meeting*, Bonn, October 9–13.
- Voss, J.-P. & Kemp, R. (2006) Sustainability and reflexive governance: Introduction, in: J.-P. Voss, D. Bauknecht & R. Kemp (Eds) *Reflexive Governance for Sustainable Development*, pp. 3–28 (Cheltenham: Edward Elgar).
- Voss, J.-P., Bauknecht, D. & Kemp, R. (Eds.) (2006) *Reflexive Governance for Sustainable Development* (Cheltenham: Edward Elgar).
- Wachter, D. (2010) Politischer Nutzen von Evaluationen und Monitoring am Beispiel der Schweizer Nachhaltigkeitsstrategie, in: R. Steurer & R. Trutnigg (Eds) *Nachhaltigkeit regieren: Eine Bilanz zu Governance-Prinzipien und -Praktiken*, pp. 215–238 (München: Oekom).
- Wackernagel, M., Schulz, N. B., Deumling, D., Callejas Linares, A., Jenkins, M., Kapos, V., Monfreda, C., Loh, J., Myers, N., Norgaard, R. & Randers, J. (2002) Tracking the ecological overshoot of the human economy, *Proceedings of the National Academy of Sciences*, 99(14), pp. 9266–9271.