



Environmental policy evaluation in the EU: between learning, accountability, and political opportunities?

Jonas J. Schoenefeld & Andrew J. Jordan

To cite this article: Jonas J. Schoenefeld & Andrew J. Jordan (2019) Environmental policy evaluation in the EU: between learning, accountability, and political opportunities?, *Environmental Politics*, 28:2, 365-384, DOI: [10.1080/09644016.2019.1549782](https://doi.org/10.1080/09644016.2019.1549782)

To link to this article: <https://doi.org/10.1080/09644016.2019.1549782>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 17 Jan 2019.



Submit your article to this journal [↗](#)



Article views: 433



View Crossmark data [↗](#)





Citing articles: 2 View citing articles [↗](#)



OPEN ACCESS



Environmental policy evaluation in the EU: between learning, accountability, and political opportunities?

Jonas J. Schoenefeld ^{a,b} and Andrew J. Jordan ^b

^aInstitute of Political Science, Technische Universität Darmstadt, Germany; ^bTyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, UK

ABSTRACT

Policy evaluation has grown significantly in the EU environmental sector since the 1990s. In identifying and exploring the putative drivers behind its rise – a desire to learn, a quest for greater accountability, and a wish to manipulate political opportunity structures – new ground is broken by examining how and why the existing literatures on these drivers have largely studied them in isolation. The complementarities and potential tensions between the three drivers are then addressed in order to advance existing research, drawing on emerging empirical examples in climate policy, a very dynamic area of evaluation activity in the EU. The conclusions suggest that future studies should explore the interactions between the three drivers to open up new and exciting research opportunities in order to comprehend contemporary environmental policy and politics in the EU.

KEYWORDS Evaluation; accountability; learning; environmental policy; evaluation governance; EU

Introduction

In the 25 years since *Environmental Politics* published its seminal special issue on European Union (EU) environmental policy (Judge 1992), policy evaluation (hereafter ‘evaluation’) has flourished. Our contribution seeks to identify the core drivers that lie behind the EU’s increasing proclivity to evaluate its environmental policies. Doing so matters because the resources committed to evaluation are substantial. In 2007, the European Commission employed 140 full-time staff in this area and spent 45 million Euros (Højlund 2015, p. 36). These investments are generating significant outputs in the form of new, policy-relevant knowledge. Mastebroek *et al.* (2016) found that the European Commission initiated 216 *ex-post* legislative evaluations between 2000 and 2012, with significant growth in more recent years: ‘nearly 200 evaluations were published [by the Commission] ... between January 2015 and mid-October 2016 [...]’ Schrefler (2016, p. 6). These numbers only

CONTACT Jonas J. Schoenefeld  schoenefeld@pg.tu-darmstadt.de

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

partially represent total evaluation output, given that other institutions, such as the European Court of Auditors (Stephenson 2015) and the European Parliament, as well as non-governmental organisations and industry associations, also evaluate (see Schoenefeld and Jordan 2017). The European Environment Agency (EEA) recently wrote that '[t]he evaluation of environment and climate policies is, today, a well-established discipline' (2016, p. 4). Textbooks on EU environmental policy now incorporate chapters on evaluation (e.g. Mickwitz 2013); a meta-analysis conducted at a time when climate policy outputs were growing rapidly, found over 250 evaluations in the sub-area of climate policy (Haug *et al.* 2010, Huitema *et al.* 2011).

While evaluation has become an established feature of EU environmental policymaking, it is much less clear *why* this has occurred. What are the key motivations of those engaging in evaluation? And to what extent can evaluation fulfil their aspirations? We follow Vedung (1997, p. 3) in defining evaluation as a 'careful retrospective assessment of the merit, worth and value of administration, output and outcome of government interventions, which is intended to play a role in future, practical action situations'. The key word is 'retrospective'; we focus here on *ex-post* evaluations (see also Mickwitz 2006, Crabbé and Leroy 2008), rather than on their *ex-nunc* (ongoing – see Crabbé and Leroy 2008) or *ex-ante* (prospective) elements (see Adelle *et al.* 2012, Turnpenny *et al.* 2016).

Until now, evaluation scholars have mainly concentrated on developing evaluation methods (e.g. Vedung 1997, Pawson and Tilley 2014), including on environmental policy (Mickwitz 2003, Crabbé and Leroy 2008). Work exploring the underlying drivers of evaluation has emerged only quite recently. Even though the 1992 special issue considered implementation (Collins and Earnshaw 1992), it did not address evaluation. Very few scholars have worked specifically on environmental evaluation in the EU (but see Mickwitz 2013).

The general neglect of evaluation matters because there are multiple reasons why actors may advocate, commission, fund, undertake, enact, and/or respond to evaluation. Scholars such as Radaelli (2010) and Adelle *et al.* (2012) have explored the politics of *ex-ante* impact assessment, but this has been much less the case for *ex-post* evaluation. Another key shortcoming in the *ex-post* evaluation literatures is that few scholars have considered multiple evaluation drivers together. Most existing accounts analyse the drivers in isolation (e.g. Bovens *et al.* 2006). For example, even though the prominent evaluation scholar Elinor Chelimsky asserts that '[m]y point is that claiming a unique purpose for evaluation flies in the face of past and current practice' (2006, p. 36), she neglects political aspects in her own review of the field. Our core aim is to incorporate all three drivers of evaluation, namely: a quest for learning; an enabler of accountability; and a way to manipulate political opportunity structures.

We proceed as follows. The next section reviews the emergence of evaluation (and especially environmental evaluation) in the EU. It focuses on evaluation's role in the EU Environmental Action Programmes (EAP), which the EU publishes regularly in order to guide and frame its environmental policy work. Their strategic nature makes them a suitable indicator of deeper shifts in EU environmental policy-making (see Mickwitz 2013). The third section returns to the three evaluation drivers outlined above and explores them theoretically, drawing on new empirical insights which are beginning to appear in the literature. The fourth section conceptualises the interaction between the drivers, drawing on emerging empirical evidence. The fifth draws together the main findings, concludes, and identifies new research needs.

Emergence of environmental policy evaluation in the EU

Most histories of evaluation identify its origins in the USA where actors assessed social policy in the 1960s (Toulemonde 2000, Stame 2003). About two decades later, the rise of New Public Management, which aims at more efficient and effective policy-making, proved influential in popularising evaluation in Europe (Pattyn 2014, Pattyn *et al.* 2018). Other factors include EU enlargement, and a perceived need to evaluate the effectiveness of structural and cohesion funding that was increasingly being dispersed east and southwards (e.g. Batterbury 2006) as well as encouragement from the OECD and the World Bank (Toulemonde 2000, Uitto 2016). More recently, scholars have also noted 'better regulation' initiatives and concerns over policy effectiveness in a world of dwindling public budgets as potential drivers (see EEA 2016).

Environment and climate change evaluation only emerged in the mid-1990s in the EU – largely following similar earlier trends in the USA (Knaap and Kim 1998, p. 23, see also Feldman and Wilt 1996). One reason for this lag may be that, as Toulemonde writes, 'professional [evaluation] networks have remained highly compartmentalised and hardly inclined to bridge the gap with other sectors' (2000, p. 351). Professional evaluators have typically focused on the EU fields where evaluation first developed, notably structural funds and research policy. However, policymakers were equally slow to demand evaluations of environmental policy. While evaluation first centred on spending policies, environmental policy was, and to a large extent remains, a regulatory affair in order to avoid distortions in the common market (Knill and Liefferink 2013) and has thus often been subjected to less evaluation. Regulation tends to be less political because the benefits it generates tend to be diffuse and slow to appear (see Majone 1994). It also took time for the environmental *acquis* to expand enough to generate effects that demanded evaluating (as happened in the USA, where environmental evaluation only really emerged a decade or so after significant legislation had been adopted – see Knaap and Kim 1998, p. 23).

However, EU environmental policy could not escape these broader trends forever (e.g. Toulemonde 2000, Mickwitz 2006, 2013, Stame 2008, EEA 2016). Evaluation did not suddenly appear in the environmental sector; rather, it gradually built over time. Some of its origins lie in earlier practices such as regulatory impact assessment. To trace this development, it is worth exploring evaluation's rising prominence in the EU's EAPs over time. These programmes identify strategic priorities for EU environmental policy, including in evaluation (see Mickwitz 2006). Table 1 summarises the appearance of 'assessment' and 'evaluation' in the seven EAPs to date.

Table 1 reveals that references to policy assessment date back to the first EAP, but have strengthened and become more common over time. The excerpts reveal how successive EAPs have defined assessments more concretely with more specific language on methodology. The focus has evolved from

Table 1. Evaluation in the EU's environmental action programmes.

EAP Number & Year	References to evaluation
1st (1973)	The protection of the environment [...] inevitably involves various kinds of expenditure [...] It is essential that the authorities make accurate assessments of the size of this expenditure in order to have a clear idea of what the economic, financial and social repercussions of proposed decisions are likely to be [...] (p. 31; emphasis added)
2nd (1977)	The Commission will try to find a method of costing anti-pollution measures... (p. 37; emphasis added)
3rd (1983)	Environmental impact assessment is the prime instrument for ensuring that environmental data is taken into account in the decision-making process. (p. 6; emphasis added)
4th (1987)	[...] Community environment actions shall take account of the potential benefits and costs of action or of lack of action. The Commission will endeavour to develop methods of assessment which will facilitate this task and which will, so far as possible, ensure the preparation of an adequate cost benefit analysis as a basis for environmental proposals. (p. 14; emphasis added)
5th (1993)	[...] the design and the choice of environmental priorities must be elaborated, based on the fullest possible assessment of all relevant costs and benefits . (p. 97; emphasis added)
6th (2002)	[...] improvement of the process of policy making through: <ul style="list-style-type: none"> – ex-ante evaluation of the possible impacts, in particular the environmental impacts, of new policies including the alternative of no action and of the proposals for legislation and publication of the results; – ex-post evaluation of the effectiveness of existing measures in meeting their environmental objectives [...] (p. 13–14; emphasis added)
7th (2013)	In order to improve environmental integration and policy coherence, the 7th EAP shall ensure that by 2020: [...] <p>This requires, in particular:</p> <ul style="list-style-type: none"> (i) integrating environmental and climate-related conditionalities [...] in policy initiatives, including reviews and reforms of existing policy, as well as new initiatives, at Union and Member State level; (ii) carrying out ex-ante assessments of the environmental, social and economic impacts of policy initiatives [...] to ensure their coherence and effectiveness; <p>[...]</p> <ul style="list-style-type: none"> (iv) using ex-post evaluation information relating to experience with implementation of the environment acquis in order to improve its consistency and coherence [...] (p. 195–196; emphasis added)

environment-related expenditure to incorporating economic effects, including costs and benefits, and, from 1987, the costs of inaction. However, explicit references to *ex-post* evaluation only emerged in the 6th and 7th EAPs (see also Mickwitz 2013), which was about ten years after evaluation became a standard part of the policy repertoire in the USA. A 1996 Communication by the Commission sought to systematise evaluation practices in the EU (European Commission 1996). Recent research shows that compared to other policy areas, DG Environment initiated a moderate level (around 40) of legislative *ex-post* evaluations between 2000 and 2014 (van Voorst and Mastenbroek 2017). Evaluation practices have thus grown rapidly in the EU since the late 1990s, often in the absence of a clear blueprint from the Commission or the Member States. But what are the deeper drivers of this trend? The following section identifies and explores three core drivers in detail.

Evaluation drivers: existing debates

Academic literatures on EU evaluation have, over time, consistently and repeatedly stressed three underlying drivers of evaluation. Two drivers that often feature in evaluation debates are accountability and learning – the latter often starting from the idea of evaluation as the last ‘stage’ in a stylised ‘policy cycle’ (Hanberger 2012, Vo and Christie 2015). However, actors may use evaluation in order to manipulate political opportunity structures, from using evaluation to delay processes through to legitimising pre-existing policy actions (Hanberger 2012). This section assesses these debates with a view to identifying what we know about the three drivers (see also Vedung 1997, p. 13).

Accountability

Many existing literatures focus on evaluation as an accountability mechanism. Bovens (2010) explains that meanings of accountability incorporate normative visions of transparency and virtue, and potentially organisational mechanisms through which agents answer to their principals – an important issue for climate change policy, which often involves numerous actors at various governance levels (Feldman and Wilt 1996, Jordan *et al.* 2015, 2018, Schoenefeld and Jordan 2017). In seeking to link evaluation and accountability, the relevant literatures mainly focus on the latter function, envisaging evaluation as an enabler of accountability (Stame 2003, Hanberger 2012) through processes of policy surveillance (see Aldy 2014). As Alkin and Christie assume: ‘[t]he need and desire for accountability presents a need for evaluation’ (2004, p. 12). To fulfil this role, many scholars emphasise the need for ‘independent’ evaluations that are removed from the turmoil of everyday politics (Weiss 1993, Feldman and Wilt 1996). For example, Chelimsky envisions evaluation as being largely external to government, emphasising that ‘[a]fter all, evaluation exists to report

on government, not to be a part of it' (2009, p. 65). Relatedly, Hildén (2011) stresses that powerful governmental actors may constrain government-sponsored or -produced evaluations. Taken together, evaluation may support key accountability mechanisms within states (Hanberger 2012).

However, there is a growing recognition that, particularly in an EU context, hierarchical, state-like structures have, in part, given way to more networked (e.g. Rhodes 1996) and, especially in the case of climate change, increasingly polycentric governance arrangements (see Dorsch and Flachsland 2017, Jordan *et al.* 2018). High levels of complexity and multiple actors in environmental governance make it especially difficult to ascertain who should be held accountable for which policy outcomes (van der Meer and Edelenbos 2006); a very politicised activity since holding organisations accountable for their actions is highly visible compared to, for example, the potentially subtler politics within *ex-ante* assessment of exploring *potential* impacts. New forms of accountability have thus emerged, such as horizontal accountability to a range of actors, including civil society (Bovens 2007, Hertting and Vedung 2012). This has profound implications for evaluation: if there is not one but many principals, evaluation may require broader approaches and multiple criteria (Hanberger 2012), as well as the involvement of numerous stakeholders (Hertting and Vedung 2012). Recent debates have thus focused on multiple criteria and triangulation in EU environmental policy evaluation (Mickwitz 2013). Scholars have often envisioned evaluation as an enabler of accountability in state-like and increasingly networked governance through knowledge provision. Some even argue that evaluation enables democratic processes by stimulating debate (Toulemonde 2000, Stame 2006).

Thus far we have discussed a normative case for evaluation as an accountability mechanism, but what do we know about the extent to which such accountability functions actually materialise in the EU? Recent evidence casts some doubt on these optimistic, normative visions for evaluation as an accountability mechanism. In a study of 220 legislative evaluations, Zwaan *et al.* (2016) found that only 16% were discussed in the European Parliament; even then, the main motivation appears to have been agenda-setting rather than holding the European Commission to account. While their study only considers evaluations carried out for the European Commission, it points to the need to further empirically investigate the assumed accountability functions of evaluation.

Learning

The second commonly discussed evaluation driver is policy improvement or policy learning. However, does a desire to learn actually stimulate evaluation? Much like accountability, policy learning is a contested concept,

involving many different forms (Zito and Schout 2009). In turn, these often arise from different perspectives on the nature of EU governance and its functions (see Radaelli and Dunlop 2013). Evaluation is often assumed to deliver critical inputs to stimulate learning (the so-called objectivist view) or to facilitate a process through which participants learn (the more argumentative view) (Borrás and Højlund 2015; see also Hildén 2011). Thus, as Haug argues, '[e]x-post evaluation of programmes or policies [...] is a widely applied group of approaches aimed at stimulating learning in environmental governance' (2015, p. 5). Hanberger (2012) develops the objectivist view to propose that a more hierarchical state-like organisation would benefit from information on policy effectiveness, whereas more network-like settings require evaluation that focuses on how collaboration works. States would thus learn from their elites using evaluation, while networks would learn from collective processes that evaluation enables (Hanberger 2012). This demonstrates that many evaluation and governance literatures still envision learning as a more or less direct 'feedback loop', meaning that actors learn through the knowledge they receive from evaluation. Relatedly, the WWF argued in its Climate Policy Tracker for the EU that '[t]he evaluation of past performance is important to verify the effectiveness and efficiency of measures, to learn about their driving forces and adjust policies accordingly' (2010, p. 16). This goes hand-in-glove with a rational 'evidence-based' policy-making view (for a fuller discussion, see Sanderson 2002).

Nevertheless, do such normative beliefs about evaluation's role in learning actually materialise in practice, or are references to learning largely rhetorical devices in order to justify evaluation done for other, more political, reasons? Long ago, evaluation scholars realised that the direct, linear use of evaluation is extremely rare (Weiss 1999). They stress how learning works through more nuanced and indirect processes (Zito and Schout 2009), such as Weiss's (1999) 'enlightenment', or situations where evaluation and monitoring exercises may perform more of a 'radar' function (Radaelli and Dunlop 2013). This is at least in part because evaluation is by no means the only source of knowledge and pressure on policy-makers (Weiss 1999). Challengingly, learning as improvement vis-à-vis evaluation ultimately requires consensus on policy values meaning that things are 'improving' on dimensions that particular actors deem relevant, important, and thus worthy of action.

This state of affairs generates at least two pertinent research questions: first, what do we know about the extent to which environmental evaluations facilitate learning? Focusing on 'government learning', Borrás and Højlund (2015) investigated the learning arising from three evaluations commissioned by the European Commission (two focusing on environmental policy). Based

on intensive interview research, they found that learning did take place, with programme or unit officers and external evaluators among the prime learners. It ranged from gaining a fuller overview of the policy area to learning about new evaluation methodologies, although interviewees stressed the incremental nature of both processes (Borrás and Højlund 2015). Focusing on climate policy in Finland, Hildén (2011) highlighted multiple forms of learning, but also detected political and rhetorical learning.

Together, these findings indicate that evaluation may indeed contribute to learning at EU level, but more research is required to assess what kind and under what conditions learning occurs as a function of evaluation. Second, it is pertinent to ask why the simplistic, linear view of learning appears to persist among evaluation scholars and especially practitioners, and what may be the alternative (e.g. more political) drivers of evaluation? The next section addresses these questions.

Political opportunity structures

Actors also use evaluation in order to manipulate political opportunity structures (McAdam 1996) as part of much broader political struggles (Bovens *et al.* 2006, Weiss 1993, Vedung 1997). Evaluation may expand or reduce the ‘scope of political conflict’ (Schattschneider 1975) by bringing certain actors into policy discussions, for example, through direct participation in an evaluation as a ‘stakeholder’ or by using evaluation results in public debates (see the introduction to this special issue, Zito *et al.* 2019). The more polycentric climate governance ‘opportunity structure’ emerging from the Paris Agreement (UNFCCC 2015) and its application in the EU (see Tosun and Schoenefeld 2017, Ringel and Knodt 2018) is likely to expand such access points. However, evaluations can also exclude actors or end discussion by delegating debates to evaluators or by effectively delaying political processes (Pollitt 1998). For some actors, engagement in evaluation has little to do with enabling accountability or fostering learning through evaluation; rather, it is a way to manipulate opportunity structures in order to advance their political goals. We thus should not understand evaluation as a ‘clinical, experimental science’, but something that is part and parcel of wider political processes (see Weiss 1993).

At a fairly basic level, evaluation may allow certain actors to participate in governance processes, and potentially shut out others (and thus affect the participatory structure). Manipulating political opportunity structures may furthermore involve shifting power relations among actors (see Schoenefeld and Jordan 2017) by legitimising certain actions, actors, or ideas, and delegitimising others (Hanberger 2012). Actors may commission evaluations simply to appear legitimate (i.e. by claiming that their decisions are evidence-based), with little interest in the results of the exercise. Alternatively,

they may use evaluations for political or symbolic, rather than more substantial, forms of learning. Actors may also utilise evaluation in order to avoid political conflict and/or escape blame (see Howlett 2014). The creation of evaluation units within the European Commission, and since 2014 a dedicated Commissioner for better regulation, have certainly sent a strong political signal by strengthening the institutional basis of evaluation. There are, however, other elements of evaluation that relate to the governance questions noted above: actors may decide to evaluate (or not) based on their pre-conceptions of policy success, or they may seek to influence the evaluation process so that the results suit pre-defined policy objectives (as a form of policy-based evidence). Policy-makers may even seek to legitimise certain policies by subjecting them to repeated evaluations.

A common response to these issues has often been to devise mechanisms that protect evaluators and their organisations from political pressure, for example, by creating independent evaluation units (Chelimsky 2009). There have thus been multiple attempts to organise the politics out of evaluation. These attempts to depoliticise completely evaluation have often been futile however, as evaluation always involves making value judgements (see Vedung 1997). Working towards a fuller understanding of manipulating political opportunity structures through evaluation requires an understanding of the various actors involved in pursuing, financing, commissioning, and/or conducting evaluations, and their core motivations. To date, our knowledge of actor motivation is at best patchy and at worst non-existent in the area of environment and climate policy in the EU (but see Schoenefeld and Jordan 2017). It is then also useful to understand the nature of the evaluation processes, the outputs and ideas they generate, and the usage of the outcomes – for example, in agenda-setting and policy formulation. In recent years, scholars have endeavoured to address these important empirical questions generally, and also with regard to environment and climate evaluation. This work details the growing evaluation activities in the European Commission as one key actor in pursuit of evaluation (e.g. Højlund 2015, Mastenbroek *et al.* 2016), but also increasingly the European Court of Auditors (e.g. Stephenson 2015), the EEA (EEA 2016, Schoenefeld *et al.* 2018), or across the EU as a whole (e.g. Stern 2009, Jacob *et al.* 2015).

Extant work has often focused on mapping the evaluation outputs of particular EU institutions. For example, Mastenbroek *et al.* (2016) found 216 studies focusing on *ex-post* legislative evaluation conducted by the European Commission and van Voorst and Mastenbroek (2017) have elaborated on it to test causal models. Warren (2014) conducted a meta-evaluation of experiences with demand-side energy policy. One of the key challenges of these literatures is that the sampling criteria for collecting evaluations vary widely, so that it is difficult if not impossible to compare their results, let alone explore the reasons for evaluation growth. For example, Huitema *et al.* (2011) included academic

articles as ‘evaluations’ in their study, while Mastebroek *et al.* (2016) only focused on evaluations by the European Commission. By contrast, Warren (2014) drew on academic databases in his analysis, but neglected evaluations published in other venues, such as those identified by Mastebroek *et al.* (2016). In sum, working towards clearer concepts that can be operationalised is a key first step in advancing this field towards more causal explanations of its politics (and the other drivers).

An emerging and important line of research considers the relationship between those who commission evaluations and those who conduct them. In a survey, Hayward *et al.* (2014) showed how members of the British government frequently aimed to influence the evaluators that they had commissioned to conduct evaluations, whether through influencing their methodologies or during the final write-up; Pleger and Sager (2016) have discovered similar dynamics in Germany and Switzerland. Even though earlier studies demonstrate that EU-level actors frequently commission environment and especially climate policy evaluations (Huitema *et al.* 2011), these dynamics have not yet been sufficiently explored.

Next steps

A key shortcoming of the existing literatures reviewed above is that they have hardly considered the drivers of evaluation side-by-side, either theoretically or empirically, especially in the case of EU environmental policy. As a first step, this section begins to work across them theoretically, a key endeavour in order to enable new theory-driven, empirical explorations. Second, it looks at recent empirical work that has begun to lay bare potential overlaps, as well as tensions, between the drivers.

Working across the drivers theoretically

As a first step towards building a more comprehensive understanding, Figure 1 maps the theoretical concepts and identifies their main overlaps and/or tensions.

Figure 1 depicts each driver in a circle in order to identify tensions and/or overlap between them, drawing on existing evaluation literatures. The figure affords significant space to each of the three drivers in order to propose that it appears conceptually helpful to understanding them individually (i.e. we found no indication of a perfect overlap between any two drivers in the evaluation literatures). Equally, we identified areas with significant potential conceptual overlap between the drivers. For example, instances where ‘learning’ and ‘manipulating political opportunity structures’ occur simultaneously may lead to ‘political learning’. Similarly, when accountability and manipulating political opportunity structures overlap, a theoretical result may become some form of

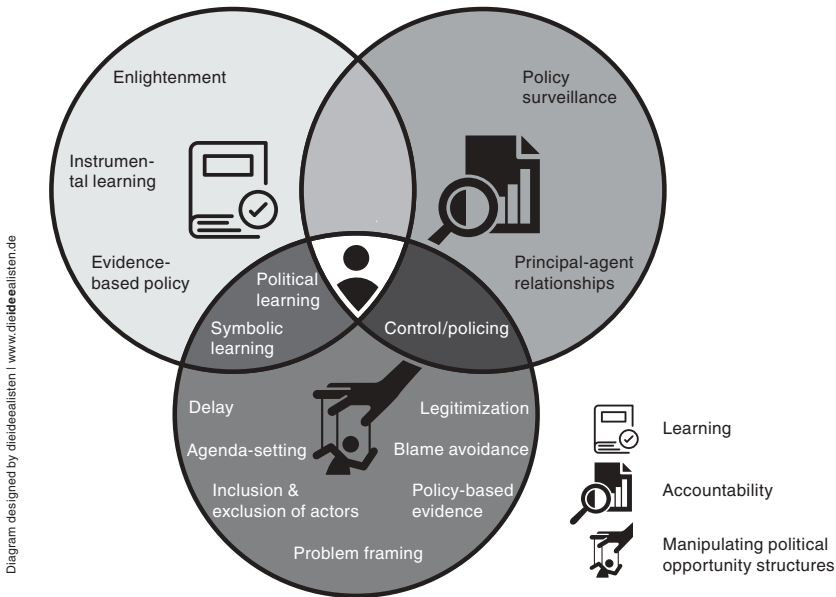


Figure 1. Learning, accountability, and political opportunities through evaluation.

policing or control. Last, a potential conceptual overlap between learning and accountability is less clear and points to tension between the two, but Regeer *et al.* (2016) argue for extending the concept of accountability in order to make learning a sub-set of it – but they acknowledge that evaluation focuses more on accountability than on learning. Stame (2003) writes that potential overlaps between accountability and learning depend on the relationship between principals and agents at the outset; if organisational goals are similar, the evaluation can serve a productive role, both in enabling accountability and potentially learning (complementarity – see also Sabel 1994, OECD 2001); however, if organisational goals differ, accountability functions may come at the cost of reduced or even no learning effects (antagonism). One concept that has been advanced to incorporate both (while arguable ignoring the potential tensions) is policy surveillance, which Aldy (2018, p. 211) argues includes

reporting and monitoring of relevant climate policy performance data, as well as the analysis and evaluation of those data. Doing so can facilitate learning about the efficacy of mitigation efforts...

However, particularly where few concepts point to an overlap between the drivers in Figure 1, tensions may emerge that may prevent the simultaneous manifestation of the normative ideals articulated through the three evaluation drivers as a consequence of evaluation. This is because theoretically one could expect significant antagonism or tension between the drivers. For example, if an actor conducts an evaluation in order to delay political processes, they

would want the evaluation to take long enough (from their perspective), and put lower emphasis on the usefulness of the evaluation in order to, for example, stimulate learning or enable accountability (both of which would benefit from evaluation insights becoming available at suitable times).

Tension is especially likely to emerge between concepts that only feature in a single circle. For example, a key theoretical tension may emerge between accountability and manipulating political opportunity structures. If actors use evaluation in order to manipulate political opportunity structures (e.g. to delay processes or bring in new actors) and thus deeply implicate evaluation with the related governance processes, it is all but impossible to conceptualise evaluation, *at the same time*, as an 'external' accountability mechanism. Ironically, evaluation may thus become subject to accountability pressures, such as when different organisations commission competing evaluations in order to support or delegitimise certain ideas. In sum, the extent to which these processes are antagonistic or complementary still requires more conceptual exploration.

Working across the drivers empirically

Exposing the theoretical and often deeply normative arguments outlined in the previous sections to empirical scrutiny is a key, but only partially realised, objective. Empirical analyses of interactions between the three drivers have often concentrated on just a few of the wide range of potential overlaps and/or tensions. The interstitials between learning and accountability attract most attention. Sabel (1994) has noted that monitoring could lead to learning if institutions forge mutual interest among various actors, who then come to understand an ongoing conversation about monitoring standards and outcomes as a learning opportunity. However, others argue that different evaluation approaches either suit learning or accountability: Højlund (2015), for example, explains that summative evaluations (at the end of a policy) are more accountability-oriented, whereas formative evaluations (occurring while a policy is being implemented) may be more geared towards learning. Tensions between accountability and learning may also emerge because evaluations can be used to name and shame governance actors, and they are effectively a control function, which can erode trust and a willingness to consider seriously potential improvements (Hermans 2009). Similarly, van der Meer and Edelenbos have highlighted that:

Evaluations that primarily have an accountability function tend to be public and are often performed by external evaluators. Units whose policies, management or implementation activities are being evaluated will often be inclined to defend their actions and achievements. (2006, p. 209)

Højlund (2015) has documented how evaluation in the European Commission has oscillated between accountability and learning. More streamlining since

the 1990s has usually meant a greater focus on accountability/control, and less on learning. While accountability and learning may not always be opposed to each other, more empirical evidence is needed to investigate their relationships. This is especially because high political hopes are currently investing in the accountability and learning functions of evaluation in the EU. For example, the Research Service of the European Parliament concludes that

Evaluation is an *important element for the proper functioning of the policy cycle*. It serves many purposes, for instance *assessing how a particular policy intervention has performed in comparison with expectations* [...]. Evaluation is also a means of *fostering transparency and accountability* towards citizens and stakeholders. Last but not least, evaluation provides *evidence for policy-makers in deciding whether to continue, modify or terminate a policy intervention*. (Schrefler 2016, 5 – emphasis added)

The EEA has recently expressed similar expectations about environment and climate evaluation in the EU (EEA 2016), which is again a call to researchers to assess whether these hopes materialise. We contend that it is very much an open question – both theoretically and empirically – whether evaluation can and does fulfil these high expectations in the areas of environment and climate policy. The implicit assumption in the quote above is that evaluation can fulfil all these roles *at the same time*. There is little recognition of potential or real tensions between the accountability and learning functions of evaluation, or that evaluation could also serve as ammunition in political battles. The hopes in the quote thus contrast sharply with emerging theoretical debates and empirical evidence on these dynamics.

Work is also emerging on the overlap between manipulating political opportunity structures and accountability. Schoenefeld *et al.* (2018) show how EU Member States were reluctant to strengthen climate policy monitoring in 2013 lest it increased the power of EU level actors. As Stame (2003) writes, monitoring is a less ‘intrusive’ activity than evaluation – in the case of the latter, concerns over control may be even stronger. As Schoenefeld *et al.* (2018) discuss, this may have severe ramifications for learning, because patchy and insufficient knowledge as well as limited indicators have disabled broader climate governance debates. Finally, the overlap between manipulating political opportunity structures and learning with a view to improving EU environmental policy evaluation is ripe for deeper empirical exploration.

While exploring the three drivers in pairs is certainly helpful, ultimately it would be useful to explore empirically all three in conjunction; that is, exploring the central area in Figure 1. A good place to begin addressing this research challenge is the EEA, which is central to many environmental policy knowledge development and dissemination activities (Martens 2010). Whereas the European Commission and Member States have generally

preferred the EEA to focus on data collection, the European Parliament prefers a stronger role in policy analysis in order to hold the Commission and the Council to account (Martens 2010). These tensions have certainly manifested in the EEA's approach to monitoring climate policies. A recent analysis of the outputs of the EU's Monitoring Mechanism on climate change – which the EEA manages – identified a range of political conflicts and tensions, ranging from Member State concern over reporting costs to fears of losing political control over knowledge generation and sharing and, potentially, even future target setting (Schoenefeld *et al.* 2018). By the same token, we know much less about the activities of non-governmental actors in evaluation (see Hildén *et al.* 2014).

There are other examples of new research that has worked across the three drivers. For instance, van Voorst and Mastebroek (2017) empirically tested motivations for evaluation – including aspects of accountability, learning, and politics – finding that the Commission is most likely to evaluate in order to enforce legislation (hence more accountability than learning), and when evaluation capacities are high. They did not find that politicisation in the Council (i.e. the apex of decision-making and hence the most openly political level) affected the initiation and framing of particular evaluations.

Conclusions

Since the original special issue on EU environmental policy (Judge 1992), evaluation has become an important element of environmental policymaking in the EU, and hence should be accounted for in any attempt to take stock of EU environmental policy and governance. There has undoubtedly been a clear change at EU level, expressed by the steep growth in political support and demand for evaluation, resource investments, and evaluation outputs. This change has been gradual over time, with an initial international impetus in the 1960s, followed by a strengthening of evaluation-related language in the EU Environmental Action Programs. In the last ten years, the institutionalisation of evaluation has materialised with many more evaluations published, together with leadership from the relevant European institutions and new guidance on best practice. Drawing on various existing literatures, we argue that evaluation has emerged from three core underlying drivers: a quest to foster policy learning; a perceived need for greater accountability; and a desire to manipulate political opportunity structures. We unpacked each driver theoretically and explored its empirical relevance before making a first attempt to work across the three drivers theoretically. We then drew on the emerging literatures on evaluation in order to explore the extent to which some of the overlaps and/or tensions between the drivers have been empirically explored.

High levels of complexity and uncertainty typically characterise environmental (and especially climate change) policy (Mickwitz 2013) and make

the conceptualisation of accountability, learning, and politics especially challenging and related effects empirically hard to detect. This is especially because many environmental issues including climate change do not neatly coincide with existing political jurisdictions (Bruyninckx 2009). The nature of the field thus invites multiple evaluation approaches and diversity in evaluation (Mickwitz 2006). Unlike more mature areas of EU evaluation such as structural funding and international development in which it is abundantly clear who has an active interest in evaluation (the Member States, as donors), in the fields of environment and climate policy the overall picture is murkier (see Schoenefeld and Jordan 2017). While we have some knowledge about which actors have in the past advocated evaluations and for what reasons, there has been a marked reluctance to open up the associated political and institutional aspects to further scrutiny.

Future research is necessary in order to further disentangle the different drivers and especially, to assess their empirical relevance (e.g. does a rhetorical emphasis on accountability and/or learning functions of evaluation really materialise empirically, or are political factors more prominent?), and thereby test the relationships we proposed in Figure 1. Where can we identify further evidence of tension and/or overlap between the drivers and the sub-concepts included in the Figure? In the area of environmental policy it matters who is undertaking evaluation, where, when, why, how, for what reasons, and with what consequences. This includes paying close attention to evaluation across different governance levels (i.e. is evaluation done at EU level, in the Member States, or elsewhere, including the relationships between different evaluation actors?). The world is clearly anxious to evaluate whether or not it is on a development path that avoids catastrophic climate change (see EEA 2016). It is telling that the performance of wholly new governance initiatives such as the post-Paris Review Process (Christoff 2016) and EU energy governance ride on monitoring and evaluation exercises that are relatively novel and still finding their feet (Schoenefeld and Jordan 2017, Ringel and Knodt 2018). The currently uneven patterns of empirical insight result in part from very limited data availability; even very simple evaluation databases remain rare.

The fact that evaluation is on the rise in EU environmental governance, both theoretically and empirically, does not necessarily imply that this is unequivocally a normatively desirable development. Even though it is possible to document the rise of evaluation, research on its use, effects, and governance is only just emerging (see Schoenefeld and Jordan 2017). Furthermore, researchers could also engage with factors beyond the three arguably functional drivers considered here, including issues of power (Partzsch 2017) and the role of state (Duit *et al.* 2016) and non-state actors (Bäckstrand *et al.* 2017) in evaluation (Schoenefeld and Jordan 2017), as well as at the level of individual organisations (see Pattyn 2014). It would also be helpful to know how far the three core drivers and their relationships in the

EU environmental sector are applicable to other sectors and to other levels of governance. Comparative theoretical and empirical explorations that work across a range of different policy sectors could thus be very illuminating. In all these different ways, researchers stand to learn about environmental policy and politics by reflecting on policy evaluation.

Acknowledgments

We thank the special issue editors and two anonymous reviewers for very helpful comments on previous drafts. We presented earlier versions at ECPR Joint Sessions in Pisa in 2016, at the Regulatory Governance Conference in Tilburg in 2016, and at the University of Gothenburg in 2017. The members of the PPE group, including notably John Turnpenny (UEA) and the EKV Group (TU Darmstadt), provided helpful feedback.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

JS and AJ benefitted financially from the COST Action INOGOV (IS1309). JS received financial support from a PhD studentship at UEA, and the German Federal Ministry of Education and Research (Reference: 03SFK4P0, Consortium ENavi, Kopernikus). AJ's contribution was supported by the 'UK in a Changing Europe Initiative' (ESRC PO 4030006272).

ORCID

Jonas J. Schoenefeld  <http://orcid.org/0000-0002-9451-9174>

Andrew J. Jordan  <http://orcid.org/0000-0001-7678-1024>

References

- Adelle, C., Jordan, A., and Turnpenny, J., 2012. Proceeding in parallel or drifting apart? A systematic review of policy appraisal research and practices. *Environment and Planning C: Government and Policy*, 30 (3), 401–415.
- Aldy, J.E., 2014. The crucial role of policy surveillance in international climate policy. *Climatic Change*, 126 (3–4), 279–292.
- Aldy, J.E., 2018. Policy surveillance: its role in monitoring, reporting, evaluating and learning. In: A. Jordan, D. Huitema, H. van Asselt, and J. Forster, eds. *Governing climate change: polycentricity in action?* Cambridge: Cambridge University Press, 210–227.
- Alkin, M.C. and Christie, C.A., 2004. An evaluation theory tree. In: M.C. Alkin, ed. *Evaluation roots: tracing theorists' views and influences*. Thousand Oaks: Sage, 12–65.

- Bäckstrand, K., *et al.*, 2017. Non-state actors in global climate governance: from Copenhagen to Paris and beyond. *Environmental Politics*, 26 (4), 561–579.
- Batterbury, S.C., 2006. Principles and purposes of European Union cohesion policy evaluation. *Regional Studies*, 40 (2), 179–188.
- Borrás, S. and Højlund, S., 2015. Evaluation and policy learning: the learners' perspective. *European Journal of Political Research*, 54 (1), 99–120.
- Bovens, M., Hart, P., and Kuipers, S., 2006. The politics of policy evaluation. In: M. Moran, M. Rein, and R.E. Goodin, eds. *The Oxford handbook of public policy*. Oxford: Oxford University Press, 319–335.
- Bovens, M., 2007. New forms of accountability and EU-governance. *Comparative European Politics*, 5 (1), 104–120.
- Bovens, M., 2010. Two concepts of accountability: accountability as a virtue and as a mechanism. *West European Politics*, 33 (5), 946–967.
- Bruyninckx, H., 2009. Environmental evaluation practices and the issue of scale. *New Directions for Evaluation*, 2009 (122), 31–39.
- Chelimsky, E., 2006. The purposes of evaluation in a democratic society. In: I. Shaw, J. Greene, and M. Mark, eds. *The SAGE handbook of evaluation*. London: Sage Publications, 33–55.
- Chelimsky, E., 2009. Integrating evaluation units into the political environment of government: the role of evaluation policy. *New Directions for Evaluation*, 123, 51–66.
- Christoff, P., 2016. The promissory note: COP 21 and the Paris Climate Agreement. *Environmental Politics*, 25 (5), 765–787.
- Collins, K. and Earnshaw, D., 1992. The implementation and enforcement of European community environment legislation. *Environmental Politics*, 1 (4), 213–249.
- Crabbé, A. and Leroy, P., 2008. *The handbook of environmental policy evaluation*. London: Earthscan.
- Dorsch, M.J. and Flachsland, C., 2017. A polycentric approach to global climate governance. *Global Environmental Politics*, 17 (2), 45–64.
- Duit, A., Feindt, P.H., and Meadowcroft, J., 2016. Greening Leviathan: the rise of the environmental state? *Environmental Politics*, 25 (1), 1–23.
- European Commission, 1996. *Evaluation: concrete steps towards best practice across the commission*. Brussels: European Commission.
- European Environment Agency, 2016. *Environment and climate policy evaluation*. Copenhagen: European Environment Agency.
- Feldman, D.L. and Wilt, C.A., 1996. Evaluating the implementation of state-level global climate change programs. *Journal of Environment and Development*, 5 (1), 46–72.
- Hanberger, A., 2012. Framework for exploring the interplay of governance and evaluation. *Scandinavian Journal of Public Administration*, 16 (3), 9–27.
- Haug, C., *et al.*, 2010. Navigating the dilemmas of climate policy in Europe: evidence from policy evaluation studies. *Climatic Change*, 101 (3–4), 427–445.
- Haug, C., 2015. *Unpacking learning: conceptualising and measuring the effects of two policy exercises on climate governance*. Thesis (PhD). Vrije Universiteit Amsterdam.
- Hayward, R.J., *et al.*, 2014. Evaluation under contract: government pressure and the production of policy research. *Public Administration*, 92 (1), 224–239.
- Hermans, L.M., 2009. *A paradox of policy learning: evaluation, learning and accountability*. Available from: <https://ecpr.eu/Filestore/PaperProposal/a3847a8d-4df2-4b74-a8a5-2fdb96d8fc48.pdf>
- Hertting, N. and Vedung, E., 2012. Purposes and criteria in network governance evaluation: how far does standard evaluation vocabulary takes us? *Evaluation*, 18 (1), 27–46.

- Hildén, M., 2011. The evolution of climate policies – the role of learning and evaluations. *Journal of Cleaner Production*, 19 (16), 1798–1811.
- Hildén, M., Jordan, A.J., and Rayner, T., 2014. Climate policy innovation: developing an evaluation perspective. *Environmental Politics*, 23 (5), 884–905.
- Højlund, S., 2015. Evaluation in the European Commission. *European Journal of Risk Regulation*, 1, 35–46.
- Howlett, M., 2014. Why are policy innovations rare and so often negative? Blame avoidance and problem denial in climate change policy-making. *Global Environmental Change*, 29, 395–403.
- Huitema, D., et al., 2011. The evaluation of climate policy: theory and emerging practice in Europe. *Policy Sciences*, 44 (2), 179–198.
- Jacob, S., Speer, S., and Furubo, J., 2015. The institutionalization of evaluation matters: updating the International Atlas of Evaluation 10 years later. *Evaluation*, 21 (1), 6–31.
- Jordan, A.J., et al., 2015. Emergence of polycentric climate governance and its future prospects. *Nature Climate Change*, 5, 977–982.
- Jordan, A.J., et al., 2018. Governing climate change polycentrically: setting the scene. In: A.J. Jordan, D. Huitema, H. van Asselt, and J. Forster, eds. *Governing climate change: polycentricity in action?* Cambridge: Cambridge University Press, 3–25.
- Judge, D., 1992. A green dimension for the European community? *Environmental Politics*, 1 (4), 1–9.
- Knaap, G.J. and Kim, T.J., 1998. *Environmental program evaluation: a primer*. Urbana: University of Illinois Press.
- Knill, C. and Liefferink, D., 2013. The establishment of EU environmental policy. In: A. Jordan and C. Adelle, eds. *Environmental policy in the EU*. 3rd ed. London: Routledge, 13–31.
- Majone, G., 1994. The rise of the regulatory state in Europe. *West European Politics*, 17 (3), 77–101.
- Martens, M., 2010. Voice or loyalty? The evolution of the European Environment Agency (EEA). *Journal of Common Market Studies*, 48 (4), 881–901.
- Mastenbroek, E., van Voorst, S., and Meuwese, A., 2016. Closing the regulatory cycle? A meta evaluation of ex-post legislative evaluations by the European Commission. *Journal of European Public Policy*, 23 (9), 1329–1348.
- McAdam, D., 1996. Conceptual origins, current problems, future directions. In: D. McAdam, J.D. McCarthy, and M.N. Zald, eds. *Comparative perspectives on social movements: political opportunities, mobilizing structures, and cultural framings*. Cambridge: Cambridge University Press, 23–40.
- Mickwitz, P., 2003. A framework for evaluating environmental policy instruments context and key concepts. *Evaluation*, 9 (4), 415–436.
- Mickwitz, P., 2006. *Environmental policy evaluation: concepts and practice*. Thesis (PhD). University of Tampere.
- Mickwitz, P., 2013. Policy evaluation. In: A. Jordan and C. Adelle, eds. *Environmental policy in the EU: actors, institutions and processes*. London: Routledge, 267–286.
- OECD, 2001. *Evaluation feedback for effective learning and accountability*. Available from: <https://www.oecd.org/dac/evaluation/2667326.pdf>
- Partzsch, L., 2017. ‘Power with’ and ‘power to’ in environmental politics and the transition to sustainability. *Environmental Politics*, 26 (2), 193–211.

- Pattyn, V., 2014. Why organizations (do not) evaluate? Explaining evaluation activity through the lens of configurational comparative methods. *Evaluation*, 20 (3), 348–367.
- Pattyn, V., et al., 2018. Policy evaluation in Europe. In: E. Ongaro and S. van Thiel, eds. *The Palgrave Handbook of Public Administration and Management in Europe*. London: Palgrave Macmillan, 577–593.
- Pawson, R. and Tilley, N., 2014. *Realistic evaluation*. London: Sage.
- Pleger, L. and Sager, F., 2016. Betterment, undermining, support and distortion: a heuristic model for the analysis of pressure on evaluators. *Evaluation and Program Planning*. doi:10.1016/j.evalprogplan.2016.09.002
- Pollitt, C., 1998. Evaluation in Europe: boom or bubble? *Evaluation*, 4 (2), 214–224.
- Radaelli, C.M., 2010. Rationality, power, management and symbols: four images of regulatory impact assessment. *Scandinavian Political Studies*, 33 (2), 164–188.
- Radaelli, C.M. and Dunlop, C.A., 2013. Learning in the European Union: theoretical lenses and meta-theory. *Journal of European Public Policy*, 20 (6), 923–940.
- Regeer, B.J., et al., 2016. Exploring ways to reconcile accountability and learning in the evaluation of niche experiments. *Evaluation*, 22 (1), 6–28.
- Rhodes, R.A.W., 1996. The new governance: governing without government. *Political Studies*, 44 (4), 652–667.
- Ringel, M. and Knodt, M., 2018. The governance of the European Energy Union: efficiency, effectiveness and acceptance of the Winter Package 2016. *Energy Policy*, 112, 209–220.
- Sabel, C.F., 1994. Learning by monitoring: the institutions of economic development. In: N.J. Smelser and R. Swedberg, eds. *The handbook of economic sociology*. Princeton: Princeton University Press, 137–165.
- Sanderson, I., 2002. Evaluation, policy learning and evidence-based policy making. *Public Administration*, 80 (1), 1–22.
- Schattschneider, E.E., 1975. *The semi-sovereign people: a realist's view of democracy in America*. New York: Rinehart & Winston.
- Schoenefeld, J.J., Hildén, M., and Jordan, A.J., 2018. The challenges of monitoring national climate policy: learning lessons from the EU. *Climate Policy*, 18 (1), 118–128.
- Schoenefeld, J.J. and Jordan, A.J., 2017. Governing policy evaluation? Towards a new typology. *Evaluation*, 23 (3), 274–293.
- Schrefler, L., 2016. *Evaluation in the European Commission: rolling check-list and state of play*. Brussels: European Parliamentary Research Service.
- Stame, N., 2003. Evaluation and the policy context: the European experience. *Evaluation Journal of Australasia*, 3 (2), 37–43.
- Stame, N., 2006. Governance, democracy and evaluation. *Evaluation*, 12 (1), 7–16.
- Stame, N., 2008. The European project, federalism and evaluation. *Evaluation*, 14 (2), 117–140.
- Stephenson, P., 2015. Reconciling audit and evaluation? The shift to performance and effectiveness at the European court of auditors. *European Journal of Risk Regulation*, 6 (1), 79–89.
- Stern, E., 2009. Evaluation policy in the European Union and its institutions. *New Directions for Evaluation*, 2009 (123), 67–85.
- Tosun, J. and Schoenefeld, J.J., 2017. Collective climate action and networked climate governance. *Wiley Interdisciplinary Reviews: Climate Change*, 8 (1). doi:10.1002/wcc.440

- Toulemonde, J., 2000. Evaluation culture (s) in Europe: differences and convergence between national practices. *Vierteljahrshefte zur Wirtschaftsforschung*, 69 (3), 350–357.
- Turnpenny, J., et al., 2016. Environment. In: C.A. Dunlop and C.M. Radaelli, eds. *Handbook of regulatory impact assessment*. Cheltenham: Edward Elgar, 193–208.
- Uitto, J.I., 2016. Evaluating the environment as a global public good. *Evaluation*, 22 (1), 108–115.
- UNFCCC, 2015. *Adoption of the Paris agreement*. Available from: <http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf>
- van der Meer, F. and Edelenbos, J., 2006. Evaluation in multi-actor policy processes: accountability, learning and co-operation. *Evaluation*, 12 (2), 201–218.
- van Voorst, S. and Mastenbroek, E., 2017. Enforcement tool or strategic instrument? The initiation of ex-post legislative evaluations by the European Commission. *European Union Politics*, 18 (4), 640–657.
- Vedung, E., 1997. *Public policy and program evaluation*. New Brunswick, NJ: Transaction Publishers.
- Vo, A.T. and Christie, C.A., 2015. Advancing research on evaluation through the study of context. *New Directions for Evaluation*, 2015 (148), 43–55.
- Warren, P., 2014. A review of demand-side management policy in the UK. *Renewable and Sustainable Energy Reviews*, 29, 941–951.
- Weiss, C.H., 1993. Where politics and evaluation research meet. *Evaluation Practice*, 14 (1), 93–106.
- Weiss, C.H., 1999. The interface between evaluation and public policy. *Evaluation*, 5 (4), 468–486.
- WWF, 2010. *Climate policy tracker for the European Union*. Brussels: World Wide Fund for Nature.
- Zito, A.R., Burns, C., and Lenschow, A., 2019. Is the trajectory of European Union environmental policy less certain? *Environmental Politics*, 28 (2). [this issue].
- Zito, A.R. and Schout, A., 2009. Learning theory reconsidered: EU integration theories and learning. *Journal of European Public Policy*, 16 (8), 1103–1123.
- Zwaan, P., van Voorst, S., and Mastenbroek, E., 2016. Ex post legislative evaluation in the European Union: questioning the usage of evaluations as instruments for accountability. *International Review of Administrative Sciences*, 82 (4), 674–693.