

Policy Learning and Policy Evaluation

Claire A. Dunlop and Claudio M. Radaelli

To appear in Handbook of Public Policy Evaluation Elgar, 2023 (Frédéric Varone, Steve Jacob and Pirmin Bundi Eds.)

Introduction

Policy learning is one of the most important promises of policy evaluation (Patton, 1998). Indeed, since its origin, policy evaluation has been geared towards the objective of learning through the development of theories of practice (Shadish, Cook and Leviton, 1991). Bundi and Trein (forthcoming, 2022) also note conceptual similarities and elective affinities between the literature on policy evaluation and the literature on policy learning. Already, two decades ago, Sanderson (2002) linked evaluation and policy learning through the systematic adoption of evidence-based reform agendas in government, observing however the difference between expectations and reality. Similarly, for Head (2016) the goal of learning from evidence (often contained in evaluations) has led to institutionalize policy evaluation. But the goal is reached only when both supply (of robust evidence, both from inside and outside government) and demand (among decision-makers) match. Among practitioners, the focus is on the design of evaluation at the expense of learning: get the evaluation design ‘right’ and learning will follow. Only recently we have seen empirical studies of how decision-makers learn from evaluations (or ignore them), mostly in the field of evaluations of legislation discussed in parliamentary assemblies (Bundi and Trein, 2022; Zwaan et al., 2016). The policy and politics that underpin the policy learning / evaluation intersection in different decision-making arenas are still opaque. This chapter aims to bring these dynamics into clear focus.

As concepts that span disciplinary boundaries across the social sciences, the definitional literatures for both policy evaluation and learning are huge with enough material to generate potential chapters of their own. Here, we simplify matters by following Pattyn (2014) (inspired by Scrivens’ [1991] widely adopted stance) who treats policy evaluation as ‘the scientific analysis of a certain policy or part of a policy, aimed at determining the merit or worth of the policy on the basis of certain criteria’ (Pattyn, 2014: 44). Thinking about when, in the policy process evaluation, learning happens, we take an inclusive approach – it covers the policy life-course from *ex ante* impact assessments in policy formulation, to *ex durante* monitoring throughout delivery¹, to *ex post* evaluation and audits following implementation.

¹ This is also known as ‘andante’ evaluation (see Hulsbergen, 2012 and Builelaar and Sorel, 2009).

We treat learning as the updating of beliefs about policies resulting from the interaction of a wide range of policy actors involved in evaluation processes; most obviously epistemic communities, citizens as policy targets, interest groups and institutional authorities as part of implementation or compliance (Dunlop and Radaelli, 2013; Dunlop, 2014; Preskill, 2008). Though policy evaluations are frequently formal, in that they are meticulously planned, the bounded rationality means the learning that flows from evaluations is often non-linear, unplanned and frequently delayed. In situations of high uncertainty and fast-burning crises, policy-makers learn from the association of episodes and behavioural responses – not from the association of an evaluation and a specific problem (Kamkhaji and Radaelli, 2017). Even in ‘normal times’, learning from evaluation is affected by heuristics: a decision-maker can draw on a certain evaluation because of proximity and analogy, not because it is the most suitable repository of evidence and lessons to learn (Cairney, 2016).

Further down the line, explicating the link between learning and policy change both conceptually and methodologically is underdeveloped in the literature (Moyson, Scholten and Weible, 2018). Addressing this weakness is beyond the scope of this chapter-

How do we explore the intersection of evaluation and learning? Specifically, when we look at the literature, how do we model causation? We take a process approach to evaluation and learning. Both policy evaluations and lessons can of course be products we detect empirically in documentation of analysis and reports, recommendations, and proposal for legislative and policy changes (see Forss et al., 2002 problematising the order of these). But, these products or objects emerge from processes. To get at these processes, we are interested in the idea of evaluation as an activator which triggers certain learning mechanisms in the policy process. In this light, evaluations themselves – their design, methods, format, participants, duration etc – are events that stimulate certain mechanisms of learning. These mechanisms capture the response of individuals, organizations and systems to the deployment of the evaluation – for example, teaching from authoritative expertise or dialogue via citizen participation. These learning mechanisms are activated by evaluation as part of a wider political context. This context is contingent. Following a realistic ontology of the social sciences: a mechanism generates an outcome in a given issue, time or space context, but not in others (Pawson, 2006; Pawson and Tilley, 1997). For example, a mechanism of teaching through evidence-based rationality activated by a systematic review may generate fundamental policy change in a healthcare-related policy issue but be less transformative in issues around climate change.

With our definitions and approach outlined, we proceed as follows. In the first section, we examine what is being learned in policy evaluation processes focussing specifically on depth of learning and policy change. Thinking about the depth of learning and how this is shaped by organisational contexts

helps us get under the skin of the policy and politics of policy evaluation and appraisal in general (Turnpenny et al., 2009: 640). Specifically, we can see that limits on learning can be ‘baked into’ evaluation design and timing. Section two then expands our contextual approach exploring how evaluation activates four distinct learning mechanisms – teaching from authoritative expertise, dialogue via participation, exchange through consultation, and, scrutiny via monitoring. Finally, we ask what this learning is good for – addressing what success and failure look like in evaluation driven learning, what it means for accountability and finally, for the policy cycle as a whole. Our argument is condensed in table 1.

Section 1 What learning is likely from policy evaluation?

If the aim of appraisal processes and tools is to ‘make institutions think differently’ (Radaelli, 2007: 13), we need to know what is being learned in policy evaluation. The policy appraisal literature has two main ways of addressing the ‘what’ of learning: i) the rational-analytical strand which focuses on the epistemic products resulting from specific analytical tools and methods (for examples of tools and methods, see Dunlop and Radaelli, 2016; Nilsson et al., 2008; Turnpenny et al., 2008, 2009), and ii) the organizational literature which takes a contextual approach, focussing on the different depths of learning made possible by prevailing institutional norms and politics (Argyris and Schön, 1974, 1978; Boswell, 2009; Rogers and Williams, 2006). Given the gap between the learning potential of evaluation tools and the reality of the ‘policy mess’ which mediates decision-makers’ learning (Hertin et al., 2009), here we drill down on the organizational approach to evaluation and learning. The seminal organizational learning model (Argyris and Schön, 1974, 1978) exposes three specific depths of learning that can be triggered by evaluation processes. Using the language of ‘loops’, learning can be: adaptive and incremental – single-loop; involve changes in goals – double; and, less frequently, systemic, where organizations learn about their own learning processes – triple loop.

For Argyris and Schön (1974, 1978), human action in organisations is guided by what they term ‘theories-in-use’. Contrasted with the idealised worldviews and self-conceptions organisations may espouse, theories-in-use underpin what actually takes place in reality. In relation to policy action, theories-in-use have three components:

- Governing variables – the objective or policy goal to be achieved,
- Action strategies – the policy instruments and tools deployed to deliver those objectives, and
- Consequences, both intended and unintended, that result from the goals set and action taken to reach them.

When the consequences match the policy goal, an organization's theory-in-use is confirmed and learning peripheral. Where there is a mismatch between intention and outcome, the learning window opens. Argyris and Schön's is a prescriptive account, where double-loop learning *should* be the goal for every learning organization. Yet, when we consider the messy reality of policy-making, more often, evaluations trigger the shallower single-loop learning variety where the focus is on 'doing things better' as opposed to 'doing things differently' (Hayes and Allinson, 1998). Bounded rationality and the associated difficulties in identifying the consequences of new policy means and ends (assuming a ready supply of these exists!) ensures decision-makers approach the evaluation process with the aim of 'satisficing' (Simon 1957). Considered simply, it is easier to design an evaluation with operational changes in mind. Fine tuning can be achieved through inquiring into what works in a policy, for example, the development of more detailed guidelines, clarification of time lines, deeper target group analysis etc. By contrast, designing evaluations towards deeper strategic and structural level reforms require political 'windows of opportunity' (Balthasar and Rieder, 2000). In politics what makes 'good' policy, 'what works' and constitutes 'policy success' (Lindblom, 1959; Marsh and McConnell, 2008; Parsons, 2004) are better served by small incremental changes in approach where competing interests are balanced as opposed to wholesale goal reformation (see Borrás and Højlund [2015] and Hilden [2011] for empirical examples of this incremental and political nature of learning).

For pluralists like Lindblom, adaptive learning is not merely reality but prescriptively the only way forward in democracies. In their firm rebuke to social scientists' 'mistaken pursuit of authoritativeness', Cohen and Lindblom (1979) emphasize scientific (in their case social scientific) evaluation is just one element in the informational mix to address complex social problems. On the flip side, there are those who argue single-loop learning is more an adjustment of tactics and policy instruments as opposed to learning involving disruption or reconsideration of beliefs, norms and assumptions (Haas, 1990: chapter 1).

Learning loops are contextually contingent – organisationally, politically and temporally. The level of learning possible in any given issue is mediated by the prevailing norms in the organisation – i.e. different departments' ways of working and structures; the boundaries of what it is actually possible to learn in political terms – and instrument adjustment the order of the day only. The synchronicity of the lessons being learned and the policy-making timelines – where legislation has raced ahead of epistemic knowledge – means evidence-based reversals can be slow. For example, these dynamics were all in play in UK policy-making around biofuels. Despite mounting evaluative evidence suggesting ambitious targets around first generation biofuels would result in deleterious consequences for the environment and actually increase CO₂ emissions, single loop learning was the pre-eminent response

where targets were adjusted but not abolished. It took time and external pressure for a rethink of the entire enterprise (Dunlop, 2010).

The organisational learning loops approach does of course have its limits. While there is a good deal of prescriptive discussion about how to design systems for triple loop or 'deutero' learning (notably in healthcare – for example, Nuño-Solinís [2017]), there are analytical biases that work against finding this higher form of learning. First, where systemic changes are made to an organisation's learning processes – say ones that require ongoing monitoring – research over the long-term is required to really pinpoint impact. Moreover, there is the challenge of making this visible. Would we truly be able to recognise meta-learning forms if we saw them? It is easy to imagine tracing a new early warning system back to a learning about a specific accident. But, beyond these 'big bang' examples of reform, where systems have learned how to learn over many years' worth of day-to-day interactions, issues are intertwined and actors multiple, the parts of the process that make the difference will be harder to discern. This is the classic attribution problem of evaluation. Beyond these methodological challenges, there are those who warn against treating triple loop learning as a holy grail to be found. Actually designing-in higher level learning systems may be impossible given the dynamic nature of the human mind and organisations (for an excellent discussion of this see Tosey et al., 2012).

Section 2 How does policy evaluation trigger policy learning?

We have looked at the 'what' of policy learning and how evaluations are associated with different depths of learning around different aspects of policies themselves and policy-making processes in which they are embedded. Now, we turn to how learning mechanisms are triggered by policy evaluation processes. To unearth the ways in which evaluation activates learning and the form that learning takes, we need to organise the evaluation contexts which recur in the literature. Drawing on Dunlop and Radaelli's (2013) modes of learning approach which systematizes a large body of the learning literature², we identify four evaluation contexts differentiated by the main actors involved and the types of questions the evaluation seeks to address:

1. Epistemic contexts are dominated by experts holding authoritative knowledge engaged in evaluations where knowledge is highly technical and uncertainty is high

² The learning literature is not short of analytical approaches. We have selected the modes approach but Heikkilä and Gerlak's (2013) collective learning framework offers an equally congenial lens through which to view the intersection with evaluation.

2. Reflexive contexts address issues that are ambiguous and contentious where there is no socially certified set of experts but rather evaluations are required to engage a range of citizen voices in dialogue and debate
3. Bargaining contexts most often address routine issues which where the evaluation challenge is to consult and accommodate a range of interest-based actors' preferences to ensure the continued smooth running of policy
4. Hierarchical contexts involve institutions with non-epistemic authority – for example, courts, inquiries, audits – who work to ensure evaluations enable policy compliance or sanction poor performance.

Brought together, these four contexts capture many of the central aspects of Campbell's (1971) ideal typical 'experimenting society'. Drawing on Haworth (1960), Campbell envisaged the experimenting society as at once: hard-headed on scientific objectivity (epistemic), popularly responsive (reflexive), accountable (bargaining) and challengeable (hierarchy) (Campbell, 1971: 224-226). We now explore each offering empirical examples that expose their learning mechanisms.

Epistemic

Where evaluations are underpinned by epistemic learning dynamics, scientist and/or experts are cast in the role of 'teacher' commissioned to provide authoritative policy knowledge. In these situations, the challenge for evaluators is to absorb as much knowledge as possible to get a basic grip on the main dimensions of how an intervention or policy technology is performing. In its ideal typical version, the use of this evaluative knowledge will be instrumental (Weiss, 1977).

Learning in highly professionalised and selective settings may not only be from expert to evaluator; experts also teach each other. Take for example Karlsson et al's (2008) study of Swedish evaluation workshops which delivered two purposes. First, they provided expert knowledge about the delivery of welfare policies in western Sweden. But, more than this, by bringing professionals together to share their different experiences a community of practice was created which worked to build capacity and learning within the experts themselves. Similar mutual learning and teaching has been found across diverse and even opposing groups of experts – see for example, Launsøn et al (2007) on medical doctors and alternative therapists treating respiratory conditions. As such epistemic contexts, while they may at first appear as reified zones of authority, in reality the house of science has many rooms (Yanow, 2005). Consequently, evaluations can help us turn hard-headed objectivity into valuable and transferable lessons where capacity and bridges are built.

Reflexive

In contrast to epistemic contexts, a good deal of evaluation practices involve a much wider range of social actors engaged in participatory approaches. Absence of consensus around the nature of the problem and lack of any singularly authoritative knowledge form create the conditions for reflexive and messy learning where fundamental questions about the desirability of a policy programme, its goals and instruments are debated. Guba and Lincoln (1989: 56) capture this exploratory quality of these evaluation contexts well, this is where a wide range of actors ‘come to understand their own construction better, and to revise it in ways that make it more informed and sophisticated than it was prior to the evaluation experience’. And indeed, the literature offers insights for how this positive learning, where the dialogue activated by participation generates lessons, can be achieved.

Much of the focus is on evaluation design to build social reality (and the difficulties of doing so in a focussed way) – the need for inclusivity in its design, construction of genuinely open questions; the use of language and lexicon which fits the local context; being adequately resourced; and, occurring far enough upstream in the policy process that the outcomes can genuinely make a difference to the future (Bezzi, 2006; van der Knaap, 2004). Ian Sanderson (2002) – an evaluation scholar turned government evaluator³ – goes further to pinpoint particular evaluation methods – pilots and policy experiments as particularly fruitful ways to create reflexive dialogues with a wide range of publics about ‘what works’. Sanderson goes further still. Inspired by John Dewey’s focus on the life experiences of citizens where the central challenge of his age was to connect often distracted publics not interested in public policy problems with the essence of democracy (Dewey, 1927), Sanderson (1998) explores how local government can use participatory evaluation tools to transcend narrow audit regimes and renew engagement with their publics.

An example of re-tooling the evaluation tools to promote reflexivity is Lorenzo Allio’s analysis of the instrumentation of policy appraisal in light of design thinking. By promoting empathy, iterative reflection on prototypes, experimentation and multiple framings of policy problems, design thinking can become resource in evaluation processes, including those that are less stable, need innovation or are geared towards building resiliency after a catastrophe (Allio, 2014: chapter 3). To achieve that – Allio notes – the traditional instrumentation and the attitude of evaluators and their clients have to change dramatically; this is the essence of re-tooling evaluation.

³ In the mid-2000s, Professor Sanderson moved from the Policy Research Institute at Leeds Metropolitan University to become director of the Scottish Government’s Corporate Analytical Services Division.

Bargaining

Evaluations are not only sites of learning, they are also the ground where political conflicts are played out (Weiss, 1993; Vedung, 1997). As such, evaluator processes can delineate which policy actors have a voice and provide platforms for the advancement of particular interests, ideas and values. In short, they are political opportunity structures (Hanberger, 2012; Schoenefeld and Jordan, 2019).

Attempts to open-up evaluations beyond the usual suspects can be fraught with difficulty. Forss et al (2006) memorably capture this in the idea of 'chasing a ghost'. In their account of UNESCO's response to HIV/AIDS and attempts to generate learning beyond policy-makers and commissioned evaluators in stakeholder settings, they found that resourcing conflicts often dominated. By opening things up, exchange-based interactions around budgets, incentive systems, skills and training for NGO teams came into sharp focus. Learning did occur in these consultations but as a by-product of capacity and preference-based encounters. Similarly, Baldwin's (2020: 27) evaluation of implementors in inter-agency partnerships casts learning as 'negotiated work with conflicts' rooted in policy practitioners' differing professional approaches and perspectives. Sometimes the actual widening of who counts as a stakeholder in evaluation can be emancipatory in itself (Jacobson et al., 2011). Documenting their experience as an evaluator, Fjellström (2008) demonstrates redefining the interests and stakeholders around the design of undergraduate medical education in Sweden's universities can be the starting point of new cooperation and enable unexpected re-composition of policy preferences. What recurs, in bargaining contexts, is the presence of an evaluator working to open-up the sphere of decision-making and re-purpose different resources and exchanges between implementors to create opportunities for cooperative exchange which is a precondition of expansive learning (see Suárez-Herrera et al., 2009 for a detailed discussion of the issues evaluators can take into account when convening such spaces to produce intentional change).

Hierarchy

Discussions of evaluations in hierarchical settings is central to considerations of learning and accountability (Picciotto, 2002, 2018). Rightly, learning for accountability is treated with scepticism – is real learning ever possible in the shadow of hierarchy where tick in the box exercises of powering matter more than force-free puzzling (Kogen, 2018; Regeer et al., 2016)? Added to this, evaluations in the form of inquiries and audits are often focussed around understanding failures. As such, learning be mandated and forced.

Weiss et al's (2005) case the Drug Abuse Resistance Education (DARE) program for United States (US) schools stands as an exemplar of how policy evaluations can be set in contexts where they trigger monitoring and compliance processes resulting in lessons being driven downward hierarchically. Faced with federally determined 'principles' of effectiveness', school districts had to adopt policy programs from an approved list held by the US Department of Education. Though hitherto widely adopted, the DARE program did not make the list. Following the evaluation, DARE was dropped in many school areas as a result. Where evaluations become accountability tools in this way, we have a fourth type of knowledge utilisation of 'imposed use' (to add to the well-known instrumental, conceptual and political/symbolic) (Weiss et al., 2005).

In the DARE case, the quality of learning was narrow. – Some district policy-makers complied because they had no choice and others appeared to update their beliefs on the basis of the evaluation. Certainly, a superordinate body, working in a results and accountability oriented political climate, created rules that pushed out DARE. Yet, more voluntaristic learning experiences are possible. For example, in their analysis of policy practitioner experiences receiving reports from Norway's Supreme Audit Institution, Reichborn-Kjennerud and Vabo (2017) report the majority of civil servants surveyed found audits to be positive and useful to drive policy reform and improvement. Huber (2006) reports similar findings when comparing EU cohesion policy mandated reporting with the networked evaluations built-in to Austrian regional policy programmes. Thus, while the depth of learning may be more adaptational and single loop, learning does happen in these settings and reforms that stimulate deeper learning experiences are evident (see van der Knaap, 2011).

Section 3 Learning for what?

Learning linked to evaluation processes is often assumed to be a 'good thing' (Dahler-Larsen, 2012). We do not make any normative assumptions about policy learning; policy learning might be seen as sub-optimal in terms of its depth and improvements generated; under certain conditions constellations of policy actors learn the wrong lesson, or lessons that are not acceptable from the point of view of standards of good governance (Dunlop and Radaelli, 2018; Preskill, 2008). Assessing the impact of evaluation on learning processes (Henry and Marks, 2003; Lumino and Gambardella, 2020) allows us to pull back from the optimistic view and produces a more nuanced picture of what success looks like on the one hand and dysfunction on the other.

Epistemic

Evaluative learning via epistemic enlightenment is useful to constellations of actors in environments that are so uncertain that policy preferences can be tough to identify. By reducing uncertainty, epistemic evaluative learning allows deliberations and choices to be made. The emphasis here is not on controlling and planning, but rather in the utilization of empirical knowledge, including shedding light on corners that are often not seen by the client (either because they are politically problematic or because the client approaches the problem with a narrow vision). There is therefore a strong relationship between the production of usable knowledge and evaluation as a tool of evidence-informed policy, and, in long term, policy enlightenment.

There are several policy instruments available. They all somewhat intersect or interact with the domain of policy evaluation: quantitative risk assessment, the requirement to submit policy proposals to an explicit benefit-cost test, health technology assessments, and the obligation to ask for opinions delivered by scientific bodies before the bureaucracy completes an appraisal. An example of the last type of instrument is the obligation for the European Commission to act only after having gathered the opinion of scientific bodies such as the European Food Safety Authority (EFSA) and the European Medicine Agency (EMA).

Science has driven the emergence and development of whole policies via *ex ante* appraisals, trials and *ex post* evaluation. We find examples in the historical evolution of policy sectors such as tobacco control, the regulation of emissions, foodstuff regulation and public health. In 2020, the United Nations Committee on Economic, Social and Cultural Rights published General Comment no.25 pointing to the operationalization and fruition of the right to science, mentioning the fundamental role of freedom of academic research, international cooperation, and (getting closer to our subject matter) evidence and science-based policy instruments of policy and technology appraisal (UN Committee on Economic Social and Cultural Rights, 2020). However, this historical process of science-in-policy has not been a steady march of enlightened progress. Although the Comment clarifies the core obligations of states in relation to science and science policy, as of today there is no UN Rapporteur on the right to science and therefore no mechanism through which states are accountable for how they have implemented this right.

Epistemic evaluation has its own politics that cannot be ignored (Cairney, 2015). Science is a process of conjectures and confutations, hence epistemic learning should also be open to the possibility of being falsified by the next evaluation. We should not be naïve about the interests of those who contribute epistemically to the process of policy evaluation – the same way we are not naïve about the attention of politicians for evidence and science. Social and natural scientists engage with public policy processes for different purposes: to produce usable knowledge is one of these purposes, but

also in the mix are budget incentives for research institutes, matters of prestige in the profession, and access to the media.

Epistemic accountability is questionable when the preferences of experts displace democratic standards of public deliberation. Another dysfunctionality is the choice of the wrong experts – for example, by empowering by institutional choice a certain category of experts as evaluators, or by generating a governance architecture for evaluation that gives priority to a ‘school of evaluative thought’ attached to a method that is contested within the profession. In France, in the early 2010s the controversy over randomized control trials as the primary standard for evaluation created conflict within the community of evaluators and between the government and the professional national association of evaluators (Devaux-Spatarakis, 2014).

Reflexive

Evaluative learning for reflexivity is the opposite of planning policies in stable environments with fixed constellations of actors. Evaluations triggers different forms interactions. They can also change the composition of the constellations of actors. For example, carrying out an evaluation with questions about sustainability, gender and biodiversity can re-compose the network of the stakeholders of the evaluation, allowing new ideas and new actors to have a say. Learning draws strength from the inclusion of a plurality of actors and arguments. Arkesteijn et al (2015) add a complexity perspective, arguing that reflexive evaluation must challenge systemic stability to support processes of learning and, in time, institutional change. In their words ‘[R]eflexive evaluation approaches build on the assumption that dominant problematic social practices are historically grown and institutionally embedded and need to be challenged as such’ (Arkesteijn et al., 2015: 110).

Thus, what is this learning good for? Reflexivity is not limited to the time-constraints of a single evaluation leading to a decision on a policy program. It is particularly useful in post-deliberation mode, transforming a single game into a sequential game (Majone, 1993: 18). This type of learning is suitable for turbulent environments, but at the same time turbulence may disrupt the institutionalization of what has been learned (Dente, 1999). Results will not be generated in the short time period of one or few evaluations and this may frustrate the initial expectations of policymakers (Dente, 1999).

Here we find the wisdom of procedures created exactly to cope with the problem of generating as much public deliberation as possible without compromising the need to reach decisions and stabilize the results of learning. Majone (1993: 19) reminds us of the role of administrative, judicial,

parliamentary and electoral procedures that emerged as ‘fruit of centuries of experience in coping with practical problems of public deliberation’.

The challenge is to develop similar governance architectures in the field of policy evaluation. Historically, an attempt along these lines is the shift from policy evaluation to the whole-of-cycle approach called better regulation (OECD, 2021; on ‘closing the policy cycle’ in the EU see van Golen and van Voorst, 2016). Essentially, better regulation is a meta-policy (Radaelli, 2007; Scott, 2010) that coordinates a range of appraisal and deliberation tools from the early stages of policy appraisals of proposals to *ex post* evaluation, with the aim of closing the policy cycle (Listorti et al., 2020 for a review of the EU experience). It also injects *ex post* evaluation into the policy appraisal of new proposals with the principle of ‘evaluate first’ – that is, one should start appraisal work on new regulations and laws in a given domain only after having completed and learned from the findings of evaluations of existing legislation (van Golen and van Voorst, 2016). In relation to what Majone (1993) wrote nearly thirty years ago about procedures and institutionalization, we find something similar to institutionalization in today’s better regulation architecture. In the OECD (2021), we find examples of countries that have their own actors, processes, champions and rules of procedure (including cabinet-level and parliamentary procedure) to embed impact assessment, consultation and legislative evaluation in public decision-making processes. Further, in some countries better regulation contemplates regulatory oversight institutions that provide stability to the learning and deliberation processes. Whether these regulatory institutions genuinely promote reflexive learning or political control is an empirical question that has been debated for a long time with reference to regulatory oversight in the USA and more recently in the case of the European Union (Radaelli and De Francesco 2010; Senninger and Blom-Hansen, 2021).

Of course, reflexive evaluation settings are not always positive places. The design of evaluations may be responsive to one populous but ignore another – such is the tyranny of the majority. Reflexive evaluation settings are far from power free. Issue salience and complexity affect the quality of learning processes (see Bundi and Trein, 2022, forthcoming on the use of evaluations in parliamentary assemblies in Switzerland). Though there may be no socially certified authoritative expert who can dominate, highly politicized and contested issues often result in evaluations being little more than ‘dialogues of the deaf’ where the lack of genuine findings and change only serve to increase scepticism in the power of evaluation (see for example Stewart and Jarvie [2015] on repeated evaluation failure to generate productive dialogue regarding indigenous community development in Australia).

Bargaining

What does evaluative learning for bargaining look like, and what is it good for? Dente (1999) noted that, in many ways, this type of learning mimics the scientific method. Evaluators are supposed to choose the program with the best *ex post* results (as identified via evaluation) in order to assign resources. He observed that negotiation is the prevailing mode. Bargaining is functional and delivers results when the environment is stable and public expenditure is not constrained (Dente, 1999). One important condition for this mode to work is the presence of a relatively stable constellation of actors in a given policy domain, with shared policy beliefs. It follows that the model becomes dysfunctional if we cast evaluation in a context where diffuse interests, radical innovation and changing paradigms matter or ought to matter, such as the ecological or digital transition and the search for resilient economies after the pandemic. When there are significant policy shifts, like in present times, the utility of bargaining is limited, and the whole mode of learning becomes inherently fragile.

Where bargaining settings for evaluation fail to deliver learning dividends, the conditions for functional exchange have broken down. This may be due to a single actor directing the evaluation – often at the department or funder. Take, for example, Biott and Cook's (2000) account of the UK National Early Years Excellence Centres. Here an evaluator framework of pre-populated fields constrained proceedings in such a way that performance management was the order of the day rather than exchange driven learning. This is a classic case of learning in the wrong mode. Rather than take an audit inspired evaluator approach where one actor dominates, the pretence of exchange and bargaining generated resentment from local providers for whom the exercise was one of bureaucratic burden.

In their account of dealing with distrust among stakeholders in evaluations, Baur et al (2010) use two case studies of client councils representing vulnerable populations to remind us where power relations are asymmetric, finding learning within exchanges may require skilled evaluators. Working as mediators challenges evaluators to find ways of using tension and conflict as learning activators to ensure that the same groups do not always perceive themselves to have 'lost'. Where this does come to pass, we have degenerative policy systems (Schneider and Ingram, 1993) where ultimately policy participants – implementors, target group advocates or target groups themselves – may simply exit the process altogether. The stakes are especially high in developing country contexts where the possibility for power asymmetries between local and donor organisations is especially pronounced (Lennie et al., 2015)

Hierarchy

Finally, in hierarchical mode evaluative learning is particularly suited to control and planning, as well as to deliver on predictability. The pre-conditions for this type of learning to work are a stable environment, strong legal base (for evaluation), independent evaluators and shared norms of evaluation as sanctioning mechanism (Dente, 1999). Under these conditions, evaluation assists a constellation of actors in learning about how to plan by finding dysfunctions and irregularities. It also allocates pay-offs with predictability, when the actions envisaged *ex ante* are carried out correctly. In some instances, there are good arguments for adopting explicitly this mode of learning for control. In the US, federal executive agencies respond to the Presidential administration. It follows that the requirement to carry out cost-benefit analysis of new rules proposed by these agencies is consistent with the Presidential obligation to control the regulatory activities of the whole executive. The argument is much less solid when the same requirement to evaluate benefits and costs is imposed on legislative initiatives that are processed by parliaments (Radaelli, 2010; Radaelli and De Francesco, 2010).

But, this type of learning too comes with possible dysfunctionalities. In the end, its logic echoes what Majone and Wildavsky wrote about ‘implementation as control’ in 1979. Control – they wrote – ‘has the intuitive appeal of all teleological or means-end theories, which seem to embody the very essence of rational action. As description, it leaves out the detours, the blind alleys, the discarded hypotheses, the constraints tightened and loosened, the lumpy stuff of life in favor of a pre-digested formula consisting of a ranking of objectives, a considering of alternatives, and a criterion that chooses among them’ (Majone and Wildavsky, 1979: 165).

INSERT TABLE 1 HERE

Conclusions

The theory and practice of evaluation have often been contingent on the assumption that evaluation is a learning activity. Learning however has not found a systematic place in evaluation. Is learning an objective or an outcome of evaluation? We have examined the policy and politics that define how evaluation and learning come together. To explain the relationship between the two, we have taken four modes of learning, discussed how they can appear in evaluation, their mechanisms and dysfunctionalities. Each mode of learning can be ‘good for’ some purposes, given the right contextual and institutional conditions (see table 1). Each comes with its conditions for success in terms of constellations of actors, characteristics of the environment, and governance architectures. Indeed,

one important lesson is to accompany learning in evaluation with the design of appropriate governance architectures. Without careful institutional design, even the most desirable mechanisms of learning may expose fragility. In turn, governance architectures can be designed to control, to facilitate negotiation, to maximize epistemic input or to encourage reflection. The other side of the coin, which complements institutional design, is the evaluative culture (Quadrant Conseil, 2018) – that is, the attitude of public managers and evaluators, their curiosity and predisposition for learning (or lack thereof). This cannot be imposed from the top or by institutional design. The challenge of culture is therefore addressed to schools of government, communities of practice, and professional associations that orient the practice of evaluation. We insist: all too often we tend to think that the solution is to produce new governmental guidance, impose standards, certify content and so on – without paying attention to attitude and practice. Finally, by looking at the question learning for what we have found that evaluative learning for planning, control, and negotiation (which assumes clearly identifiable pay-offs from alternative actions) may be less attractive when societies are poised for major paradigmatic changes, such as the ecological-digital transition and the building of resiliency in the post-pandemic climate.

Whatever the challenges of a particular age, the role of evaluators remains the same thirty-three years after Majone identified it as follows:

‘[...] analysts can contribute to societal learning by refining the standards of appraisal of public policies and by encouraging a more sophisticated understanding of public policies than is possible from a single perspective. The need today is less to develop “objective” measures of outcomes – the traditional aim of evaluation research – than to facilitate a wide-ranging dialogue among advocates of different criteria’ (Majone, 1989, 183).

Acknowledgements: we gratefully acknowledge the support of the European Research Council, Advanced Grant ‘Procedural Tools for Effective Governance’ (PROTEGO), grant number 694632. We wish to say thank you to Elliot Stern for his suggestions on the literature and to the editors for their precious comments.

References

- Allio, L. (2014). *Design Thinking for Public Sector Excellence*. Singapore: UNDP Global Centre for Public Service Excellence.
- Argyris, C., & Schön, D.A. (1974). *Theory in Practice*. San Francisco, CA: Jossey Bass.
- Argyris, C., & Schön, D.A. (1978). *Organizational Learning: A Theory of Action Perspective*. Reading, MA: Addison-Wesley.
- Arkesteijn, M., Van Mierlo, B., & Leeuwis, C. (2015). The need for reflexive evaluation approaches in development cooperation. *Evaluation*, 21(1), 99-115.
- Baldwin, C. K. (2020). Implementer learning in developmental evaluation: A cultural historical activity theory and expansive learning analysis. *Evaluation*, 26(1), 27-48.
- Balthasar, A., & Rieder, S. (2000). Learning from evaluations: Effects of the evaluation of the Swiss Energy 2000 Programme. *Evaluation*, 6(3), 245-260.
- Bauer, V.E., van Elteren, A.H.G., Nierse, C.J., & Abma, T.A. (2010). Dealing with distrust and power dynamics: Asymmetric relations among stakeholders in responsive evaluation. *Evaluation*, 16(3), 233-248.
- Bezzi, C. (2006). Evaluation pragmatics. *Evaluation*, 12(1): 56-76.
- Biott, C., & Cook, T. (2000). Local evaluation in a National Early Years Excellence Centres pilot programme: Integrating performance management and participatory evaluation. *Evaluation*, 6(4), 399-413.
- Borrás, S. (2011). Policy learning and organizational capacities in innovation policies. *Science and Public Policy*, 38(9), 725-734.
- Borrás, S., & Højlund, S. (2015). Evaluation and policy learning: The learners' perspective. *European Journal of Political Research*, 54(1), 99-120.
- Boswell, C. (2009). *The Political Uses of Expert Knowledge: Immigration Policy and Social Research*. Cambridge: Cambridge University Press.
- Buitelaar, E., & Sorel N. (2009). *Ex-durante evaluatie Wro. Evaluatie ontwerp en nulmeting*. Den Haag: Planbureau voor de Leefomgeving <https://www.pbl.nl/publicaties/ex-durante-evaluatie-wro> (date accessed 2 November 2022)
- Bundi, P., & Trein, P. (2022). Evaluation use and learning in public policy. *Policy Sciences*, 55, 283-309.
- Cairney, P. (2015). *The Politics of Evidence-Based Policymaking*. Basingstoke: Palgrave.
- Campbell, D.T. (1991). Methods for the experimenting society. *Evaluation Practice*, 12(3), 223-260.
- Dente, B. (1999). La politique d'évaluation des politiques publiques. In Kessler, M.-C., Lascombes, P., Setbon, M., & Thoenig, J.-C. (Eds.). *Évaluation Des Politiques Publiques* (pp. 147-155) Paris: L'Harmattan.

- Devaux-Spatarakis, A. (2014). *La méthode expérimentale par assignation aléatoire: un instrument de recomposition de l'interaction entre sciences sociales et action publique en France?* PhD Dissertation. Université de Bordeaux, Bordeaux: École doctorale Sociétés, politique, santé publique <https://tel.archives-ouvertes.fr/tel-01234599> (date accessed 15 November 2022)
- Dewey, J. (1927). *The Public and its Problems*. New York, NY: Holt.
- Dunlop, C.A. (2010). The temporal dimension of knowledge and the limits of policy appraisal. *Policy Sciences*, 43(4), 343-363.
- Dunlop, C.A. (2014). The possible experts: How epistemic communities negotiate barriers to knowledge use in ecosystems services policy. *Environment and Planning C – Government and Policy*, 32(2), 208-228.
- Dunlop, C.A., & Radaelli, C.M. (2013). Systematizing policy learning: From monolith to dimensions. *Political Studies*, 61(3), 599-619
- Fjellström, M. (2008). A learner-focused evaluation strategy: Developing medical education through a deliberative dialogue with stakeholders. *Evaluation*, 14(1), 91-106.
- Forss, K., Kruse, S.E., Taut, S., & Tende, E. (2006). Chasing a ghost? An essay on participatory evaluation and capacity development. *Evaluation*, 12(1), 128-144.
- Forss, K., Rebien, C.C., & Carlsson, J. (2002). Process use of evaluations: Types of use that precede lessons learned and feedback. *Evaluation*, 8(1), 29-45.
- Haas, E.B. (1990). *When Knowledge is Power*. Berkeley, CA: University of California Press.
- Hanberger, A. (2012). Framework for exploring the interplay of governance and evaluation. *Scandinavian Journal of Public Administration*, 16(3), 9-27.
- Haworth, L. (1960). The experimenting society: Dewey and Jordan. *Ethics*, 71(1), 27-40.
- Hayes, J., & Allinson, C.W. (1998). Cognitive style and the theory and practice of individual and collective learning in organisations. *Human Relations*, 51(7), 847-871.
- Head, B.W. (2016). Toward more 'evidence-informed' policy making? *Public Administration Review*, 76(3), 472-484.
- Heikkilä, T., & Gerlak, A.K. (2013). Building a conceptual approach to collective learning: Lessons for public policy scholars. *Policy Studies Journal*, 41(3), 484-511.
- Hertin, J., Turnpenny J., Jordan, A., Nilsson, M., Russel, D., & Nykvist, B. (2009). Rationalising the policy mess? *Ex ante* policy assessment and the utilisation of knowledge in the policy process. *Environment and Planning A-Economy and Space*, 41(5), 1185-1200
- Hildén, M. (2011). The evolution of climate policies – the role of learning and evaluations. *Journal of Cleaner Production* 19(16), 1798-1811.
- Huber, W. (2006). Evaluation of European Union cohesion policy: Window-dressing, formal exercise or coordinated learning process? *Regional Studies* 40(2), 277-280.

- Hulsbergen, E. (2012). More emphasis on crucial evaluation indicators needed: A critical look at the new charter of Athens. *Built Environment*, 38(4), 540-550.
- Jacobson, C., Carter, R.W., Hockings, M., & Kelman, J. (2011). Maximizing conservation evaluation utilization. *Evaluation*, 17(1), 53-71.
- Kamkhaji, J.C., & Radaelli, C.M. (2017). Crisis, learning and policy change in the European Union. *Journal of European Public Policy*, 24(5), 714-734.
- Karlsson, P.-Å., Beijer, E., Eriksson, B., & Leissner, T. (2008). Evaluation workshops for capacity building in welfare work: Some Swedish examples. *Evaluation*, 14(4), 483-498.
- Kogen, L. (2018). What have we learned here? Questioning accountability in aid policy and practice. *Evaluation*, 24(1), 98-112.
- Launsø, L., Rieper, J., & Rieper, O. (2007). Evaluative feedback as a contribution to learning between groups of professionals. *Evaluation*, 13(3), 306-322.
- Lennie, J., Tacchi, J., Wilmore, M., & Koirala, B. (2015). A holistic, learning-centred approach to building evaluation capacity in development organizations. *Evaluation*, 21(3), 325-343.
- Lindblom, C.E. (1959). The science of muddling through. *Public Administration Review*, 19(2), 79-88.
- Lindblom, C.E. (1965) *The Intelligence of Democracy*. New York, NY: The Free Press
- Lindblom, C.E., & Cohen, D.K. (1979). *Usable knowledge: Social science and social problem solving*. New Haven, CT: Yale University Press.
- Listorti, G., Basyte-Ferrari, E., Acs, S., & Smits, P. (2020). Towards an evidence-based and integrated policy cycle in the EU: A review of the debate on the Better Regulation agenda. *JCMS: Journal of Common Market Studies*, 58(6), 1558-1577.
- Lumino, R., & Gambardella, D. (2020). Re-framing accountability and learning through evaluation: Insights from the Italian higher education evaluation system. *Evaluation* 26(2), 147-165.
- Majone, G.D. (1989). *Evidence, Argument and Persuasion in the Policy Process*. New Haven, CT: Yale University Press.
- Majone, G.D. (1993). *When does policy deliberation matter?* EUI Working Paper SPS 93-12. Florence: European University Institute. <https://cadmus.eui.eu/handle/1814/242> (date accessed 12 November 2021)
- Majone, G.D., & Wildavsky, A. (1979). Implementation as evolution. In Pressman, J.L., & Wildavsky, A. (Eds.). *Implementation*. 3rd edn Berkeley CA: University of California Press.
- Nuño-Solinís, R. (2017). Revisiting organisational learning in integrated care. *International Journal of Integrated Care*, 17(4), 1-6.
- OECD (2021) *Regulatory Policy Outlook*. Paris: OECD. <https://www.oecd.org/gov/oecd-regulatory-policy-outlook-2021-38b0fdb1-en.htm> (date accessed 12 November 2021)
- Patton, M.Q. (2008). *Utilization-focused evaluation* 4th edn. Thousand Oaks, CA: SAGE

- Picciotto, R. (2002). The logic of mainstreaming: A development evaluation perspective. *Evaluation*, 8(3), 322-339.
- Picciotto, R. (2018). Accountability and learning in development evaluation: A commentary on Lauren Kogen's thesis. *Evaluation*, 24(3), 363-371.
- Quadrant Conseil (2018). *Consultation sur le bilan de Mieux Légiférer*. Paris. https://www.quadrant-conseil.fr/ressources/documents/Quadrant_Mieux_Legiferer.pdf (date accessed 17 November 2021).
- Radaelli, C.M. (2007). Whither better regulation for the Lisbon agenda?. *Journal of European Public Policy*, 14(2), 190-207.
- Radaelli, C.M. (2010). Regulating rule-making via impact assessment. *Governance*, 23(1), 89-108.
- Radaelli, C.M., & De Francesco, F. (2010). Regulatory impact assessment. In Baldwin, R., Cave, M., & Lodge, M. (Eds.). *The Oxford Handbook of Regulation* (pp. 279-301). Oxford, Oxford University Press.
- Regeer, B.J., De Wildt-Liesveld, R., Van Mierlo, B., & Bunders, J.F.G. (2016). Exploring ways to reconcile accountability and learning in the evaluation of niche experiments. *Evaluation* 22(1), 6-28
- Reichborn-Kjennerud, K., & Vabo, S.I. (2017). Performance audit as a contributor to change and improvement in public administration. *Evaluation*, 23(1), 6-23.
- Sanderson, I. (1998). Beyond performance measurement? Assessing 'value' in local government. *Local Government Studies* 24(4), 1-25.
- Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public Administration* 80(1), 1-22.
- Schoenefeld, J.J., & Jordan, A.J. (2019). Environmental policy evaluation in the EU: Between learning, accountability, and political opportunities? *Environmental Politics* 28(2), 365-384.
- Scott, C. (2010). Reflexive governance, regulation, and meta-regulation: Control or learning? In Lenoble, J., & De Schutter, O. (Eds.). *Reflexive Governance: Redefining the Public Interest in a Pluralistic World* (pp. 43-66). Oxford: Hart.
- Scriven, M. (1991). *Evaluation Thesaurus*. Thousand Oaks, CA: SAGE.
- Senninger, R., & Blom-Hansen, J. (2021). Meet the critics: Analyzing the EU 'Commission's Regulatory Scrutiny Board through quantitative text analysis. *Regulation & Governance* 15(4) 😊, 1436-1453.
- Shulock, N. (1999). The paradox of policy analysis: If it is not used, why do we produce so much of it? *Journal of Policy Analysis and Management*, 18(2), 226-244.
- Stewart, J., & Jarvie, W. (2015). 'Haven't we been this way before? Evaluation and the impediments to "policy learning. *Australian Journal of Public Administration*, 74(2), 114-127.
- Suárez-Herrera, J.C., Springett, J., & Kagan, C. (2009). Critical connections between participatory evaluation, organizational learning and intentional change in "pluralistic organizations. *Evaluation* 15(3), 321-342.

- Tosey, P., Visser, M., & Saunders, M.N. (2012). The origins and conceptualizations of 'triple-loop' learning: A critical review. *Management Learning*, 43(3), 291-307.
- United Nations, Committee on Economic Social and Cultural Rights. (2020) *General comment No. 25 (2020) on science and economic, social and cultural rights (article 15 (1) (b), (2), (3) and (4) of the International Covenant on Economic, Social and Cultural Rights)*. https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=E%2fC.12%2fGC%2f25&Lang=en (date accessed 21 November 2022)
- van der Knaap, P. (2004). Theory-based evaluation and learning: Possibilities and "challenges". *Evaluation* 10(1), 16-34.
- van der Knaap, P. (2011). Sense and complexity: Initiatives in responsive performance audits". *Evaluation* 17(4), 351-363.
- van Golen, T., & van Voorst, S. (2016). Towards a regulatory cycle? The use of evaluative information in impact assessments and ex-post evaluations in the European Union. *European Journal of Risk Regulation*, 7(2), 388-403.
- Vedung, E. (1997). *Public Policy and Program Evaluation*. New Brunswick, NJ: Transaction Publishers.
- Weiss, C.H., Murphy-Graham, E., & Birkeland, S. (2005). An Alternate Route to Policy Influence: How Evaluations Affect D.A.R.E. *American Journal of Evaluation* 26(1), 12-30.
- Yanow, D. (2005). In the house of 'science', there are many rooms: Perestroika and the 'science studies' turn. In Renwick Monroe, K. (Ed.), *Perestroika! The raucous rebellion in political science* (pp. 200-217). New Haven, CT: Yale University Press.
- Zwaan, P., van Voorst, S., & 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.

Table 1 Policy Learning and Policy Evaluation

	Epistemic	Reflexive	Bargaining	Hierarchical
Guiding question of the evaluation process	How does evaluation provide solutions to social problems?	How are citizens' voices heard in evaluation learning?	In what ways does evaluation create value at the negotiation table?	How do evaluations support monitoring and oversight?
Key actors driving the process of learning	Expert communities	Citizens, issue networks	Interest actors with resources central to policy delivery	Ministers Programme designers in departments Independent regulators
Mechanism of action	Authoritative expertise	Dialogue	Exchange	Authority
Who is it good for?	Reduction of uncertainty and identification of causal chains	Re-composition of issue networks Post-deliberation	Identification of the correct choice / efficient selection process	Detection of disfunctions and irregularities Predictability
Dysfunctions	Violation of democratic standards (technocracy) Conflicting recommendations arising from evaluations with different epistemic perspectives Choice of the wrong experts	Power disparities skew learning processes Institutionalization of learning is disrupted by turbulence	Reflexive Learning does not take place in situations of i. Sharply different preferences ii. Radical innovation iii. Changing paradigms iv. Public expenditure is constrained	Its requirements are often absent: i. Stable environment ii. Strong legal base (for evaluation), iii. Independent evaluators iv. Shared norms