Participatory vs expert evaluation styles

Fritz Sager and Céline Mavrot

Abstract

This chapter focuses on policy evaluation, defined as the assessment of a public policy to determine whether it has achieved its objectives. We discuss two types of policy evaluation: expert evaluation and participatory evaluation. While expert styles insist on the technicality of policy evaluations and on scientific sophistication, participatory styles focus on the integration of policy stakeholders and laypersons during the evaluative process. Because expert evaluation styles are marked by their scientific character and independence, they enjoy a high internal reliability and a result-based legitimacy. In contrast, core features of participatory evaluations styles are their representativeness and transparency, which is associated with external reliability and a process-based legitimacy. The chapter outlines the historical development of these two evaluation styles, and discusses their respective strengths and weaknesses in the light of contemporary decision-making processes. Finally, we discuss the risks associated with technocratic policy evaluation styles on the one hand, and politicized participatory evaluations on the other hand, in the context of post-truth democracy.

Introduction

In the field of policy research, evaluation most generally denominates the second-last stage of the policy cycle, that is, the assessment of whether a policy has achieved its goals and is thus to be continued or whether it has failed its objectives and consequently is to be reformed or terminated (Lasswell 1956). This broad delineation of policy evaluation remains murky with regard to two questions: first, who assesses policy success, and second, what actually constitutes policy success? Today's understanding of policy evaluation provides presumably clear answers to the two questions: ascertaining policy success is the task of a particular disciplinary caste, the policy evaluators, and policy success is narrowly defined as what Marsh and McConnell (2010) term program success, that is, the achievement of the goals predefined in the policy concept. The two notions of policy evaluation largely differ in both respects: while the first, broad notion leaves policy evaluation open to everybody and consequently does need transparent criteria and replicable methods and data, the second, narrow definition restricts policy evaluation to a well-trained group of experts with a common take on policy success. They employ a positivist approach and empirically assess success as performance, compliance and goal attainment measurable by adequate indicators and sound methodology. While the first view of policy evaluation is participatory, the second is technical or expert-based. Between the two ideal-types of expert and participatory policy evaluation styles is a wide range of approaches situated on different points of the continuum.
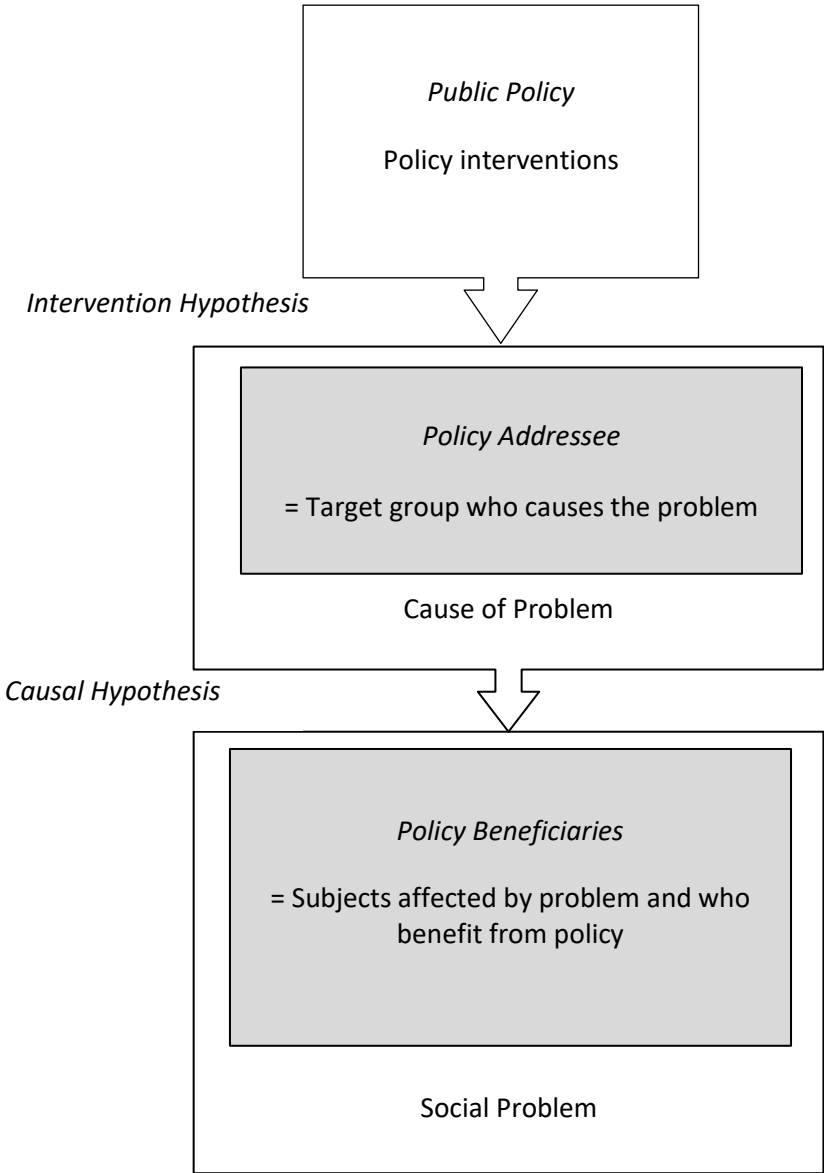
In the following, we will not oppose the two notions as two equally employed styles of evaluation praxis. Much to the contrary, today's world of policy evaluation is a world of social scientific, empirical and methodologically sophisticated policy evaluation delivering robust information for policy makers and implementing agencies on how to achieve their goals. Dahler-Larsen (2012) speaks of the "evaluation society". However, scientific empirical policy evaluation embraces different styles that oscillate between more expert and more participatory approaches. Both types of evaluation can achieve a high degree of technicality. What mostly distinguishes them is the degree of policy stakeholders' involvement during the evaluation process. These different schools not so much relate to countries, traditions or policy sectors than they link to different preferences emerging in the development of policy evaluation as a "trans-discipline" (Scriven 2008). Consequently, the chapter proceeds as follows: in the next section, we present the emergence of policy evaluation as a scientific endeavor. We show how through authors like Rossi et al. (1999) it at the same time adopts its scientific claim from basic research and emancipates from the latter in its self-awareness as applied research with authors such as MacRae (1975). The result of this development is policy evaluation's coming of age as a technical and expert style of assessing policy. The section that follows looks at the responses to the increasing sophistication of policy evaluation in the form of more participatory approaches. Authors like Weiss, Alkin and Patton shifted policy evaluation's focus on the usefulness, the usability and ultimately the use of evaluations next to their scientific quality. Authors like Stufflebeam (2000, 2003) called for a stronger participation of evaluators in the policy they evaluated, while others promote a practice where policy actors participate in the evaluation of their own policy. In today's literature, numerous

approaches tend to build on the strengths of both evaluation styles, drawing on mixed methods that allow for scientific rigor and democratic participation (Greene 2005; Park et al. 2014; Torrance 2012). The critical gap between the two styles lies in the question of whether a policy evaluation needs to be independent in order to be of value for decision makers. We address this question in the consecutive section where we discuss the role of evaluation in times of evidence-based policy-making. We show both the peril of pressure on evaluators and the increasing hybridity between science and advocacy in times of contested scientific evidence, as well as the benefit of participatory approaches in terms of information quality, acceptance and utilization of the evaluation results. We conclude the chapter with a discussion of the role of evaluation in post-truth democracy and of what both the expert and the participatory styles can do to reestablish empirical evidence as a ground for sound policy-making.

The expert style of policy evaluation

Public policy is political problem solution. As such, we can understand policy as the technical causal chain depicted in Figure 30.1. We lay out this causal chain and its elements in the following before we show how this technical understanding not only leads to but also necessitates an expert style of policy evaluation.

Figure 30.1: The causal model of public policy

```
                    ┌─────────────────────────────────┐
                    │                                 │
                    │          Public Policy          │
                    │                                 │
                    │        Policy interventions      │
                    │                                 │
                    └─────────────────┐    ┌──────────┘
                                      │    │
Intervention Hypothesis               ▼    ▼
                    ┌─────────────────────────────────┐
                    │  ┌───────────────────────────┐  │
                    │  │                           │  │
                    │  │      Policy Addressee      │  │
                    │  │                           │  │
                    │  │ = Target group who causes  │  │
                    │  │        the problem         │  │
                    │  └───────────────────────────┘  │
                    │          Cause of Problem       │
                    └─────────────────┐    ┌──────────┘
                                      │    │
Causal Hypothesis                     ▼    ▼
                    ┌─────────────────────────────────┐
                    │  ┌───────────────────────────┐  │
                    │  │                           │  │
                    │  │     Policy Beneficiaries    │  │
                    │  │                           │  │
                    │  │ = Subjects affected by      │  │
                    │  │  problem and who benefit    │  │
                    │  │        from policy         │  │
                    │  └───────────────────────────┘  │
                    │          Social Problem         │
                    └─────────────────────────────────┘
```

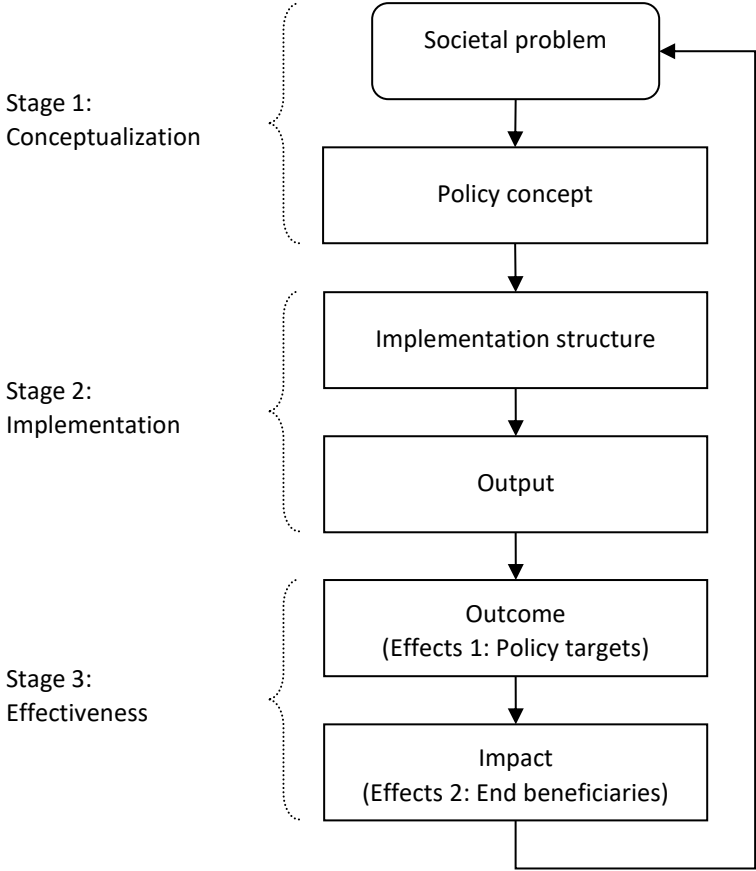Source : Pleger, Lutz and Sager, 2018, based on Rossi, Lipsey and Henry, 2019, p. 88

A policy intervention aims at eliminating the causes of a situation or a phenomenon that a democratic decision has deemed societally undesirable and hence a problem. In order to eliminate the problem, policy makers rely on two hypotheses. First, the causal hypothesis informs about the sources of the societal problem. Second, the intervention hypothesis informs about how to suspend the identified sources of the societal problem. Public policy thus intends to change a situation. Change by itself is an action, and as public policy intends to stimulate change with an intervention, the response must be deliberate and hence human in nature. Public policy thus always is a social intervention that seeks a response in a target group. The causal hypothesis therefore identifies not only causes but also causers of the problem the policy can address. These causers are the policy targets, and the policy intervention aims at changing their behavior so they no longer cause the problem. Rossi et al.

(2019, p. 59) speak of a "program's theory" that consists of a "theory of change" (i.e., the causal hypothesis) and a "theory of action" (i.e., the intervention hypothesis) (2019, p. 77). While public policy is a common if not the core subject of everyday political debates, the language of this short account of what public policy ontologically entails shows that it takes sophisticated expertise to conceptually capture and empirically test the social mechanisms of a public policy in order to evaluate it.

As the matrix of evaluation has mainly been the academic community, evaluation evolved into a scientific subdiscipline, and expert evaluation has historically become the gold standard of policy evaluation in modern society. Within the academic world, however, policy evaluation soon distanced itself from purely fundamental research by underlining its applied purpose: "evaluation research . . . derives its problems from the goals of organizations or of research sponsors who wish to evaluate them" (MacRae 1975, p. 363). The most common model of policy evaluation employs the causal model of public policy and turns it into a testable linear model of policy effects. Chen and Rossi (1987) introduced the label of "theory-driven evaluation" to assemble these forms of evaluation. Theory-driven evaluation models usually build upon a reconstruction of the implicit or explicit causal model of the public policy or intervention they evaluate.

The classic linear model of policy program evaluation, as for instance in Weiss (1997), distinguishes three basic stages of the policy process considered relevant for problem resolution: the concept, the implementation and the effects.

Figure 30.2: The linear model of policy evaluation

```
Stage 1:              ┌─────────────────────────┐
Conceptualization     │     Societal problem    │◄──────┐
                ⎰     └────────────┬────────────┘       │
                ⎱                  ▼                     │
                      ┌─────────────────────────┐       │
                      │      Policy concept      │      │
                      └────────────┬────────────┘       │
                                   ▼                     │
                      ┌─────────────────────────┐       │
                      │  Implementation structure │      │
Stage 2:        ⎰     └────────────┬────────────┘       │
Implementation  ⎱                  ▼                     │
                      ┌─────────────────────────┐       │
                      │          Output          │      │
                      └────────────┬────────────┘       │
                                   ▼                     │
                      ┌─────────────────────────┐       │
                      │         Outcome          │      │
                      │ (Effects 1: Policy targets) │   │
Stage 3:        ⎰     └────────────┬────────────┘       │
Effectiveness   ⎱                  ▼                     │
                      ┌─────────────────────────┐       │
                      │          Impact          │      │
                      │ (Effects 2: End beneficiaries) │ │
                      └────────────┬────────────┘       │
                                   └──────────────────────┘
```

Source: based on Knoepfel and Bussmann (1997) and Ledermann *et al.* (2006, p. 3)

The first stage of the linear model is the concept, which entails the definition of the societal problem the policy addresses and the delineation of the share of this problem the policy sets out to resolve. The problem itself is not part of the expert evaluation, as it is a social construct of the political decision process. As such, it is the result of value- and interest-based deliberation that the evaluators cannot empirically assess. What evaluation can do, however, is to assess how well the policy design responds to the problem as it was politically defined, that is, whether the causal model of the policy is fit to resolve the problem and whether the planned implementation structure is apt to perform the policy. The study of the policy concept allows identifying potential policy failure due to bad design even before implementation starts.

Implementation is the second stage of the linear model. It consists of two elements, the implementation structure and the output. The implementation structure is the organizational basis of service delivery once the policy has taken off, that is, it is the organizational part of the policy design put into practice. The output denominates all products and services this organization delivers.

The third stage of the linear model is effectiveness. This last stage hence measures what the policy is all about, namely the problem resolution. In order to evaluate effectiveness, the model distinguishes the two elements of the causal hypothesis in the causal model of a public policy: the outcome and the impact. The outcome means the target population's response to the policy output delivered in implementation. The impact means the problem resolution due to the target population's response to the policy. Both variables are assessed with the help of effectiveness indicators measuring the degree to which change in target group behavior and change of the problem situation correspond to the expectations of the policy design.

The question of effectiveness implies causal analysis. In fact, the whole model does so and, correspondingly, causal analysis is the gold standard of expert policy evaluation. The establishment of causality distinguishes policy evaluation from mere controlling accounts that also measure performance. Controlling attributes outcome performance to the policy intervention it controls but fails to empirically substantiate the asserted causal link between policy intervention and impact.

Causal analysis allows telling concept failure, or bad design, from implementation failure, or bad execution (Linder and Peters 1987). This is crucial when it comes to attribute accountability: implementing agencies are not responsible for bad design; policy makers are not responsible for poor service provision. Policy evaluation therefore depends on causal analysis and must not content itself with mere controlling. Establishing causality, however, is a sophisticated task and requires solid training in social scientific methodology and research experience in data collection and analysis.

Both the conceptualization of public policy as a causal chain and its empirical test therefore demand a considerable degree of expertise. Evaluation rooted in basic research has grown closer to practical questions of policy practice, but during this process it did not strip off its original scientific epistemology nor its standards of sound social scientific research. The dominance of experts in evaluation practice has led to the formation of a discipline in its own right with its own outlets, community and standards (Pawson and Tilley 1994). The dominant expert style of evaluation and its increasing sophistication, however, also bear the risk of missing the chief objective of evaluation research, that is, informing the practice how to improve policy performance. The better the research, the more difficult it is for practitioners alien to academic logic to understand it. Rigor comes at the expense of relevance. While exclusiveness is not necessarily a deficiency of basic research, it is detrimental for evaluation. Applied research like policy evaluation only makes sense if the formulated recommendations are implemented. If its theoretical or methodological sophistication prevents its usage, evaluation fails to fulfill its purpose. This insight drove evaluation scholars of the 1980s to shift their focus from modelling and methods to the use of evaluation findings. We show in the next section how this shift led to new, more participatory models of evaluation like Stufflebeam's (2003) famous CIPP model (context, input, process, product) and the new paradigm of utilization-focused evaluation (Patton 1997, 2011, 2012), up to recent discussions

on mixed evaluation models aimed at integrating scientific legitimacy and democratic responsiveness (Chouinard 2013).

The participatory style of policy evaluation

In the 1980s, evaluation scholars increasingly started to focus their attention on the use of evaluation results in addition to the evaluation's scientific and methodological quality. The symbol of this usage turn of evaluation research was Patton's (1997) *Utilization-Focused Evaluation*, written in 1978 and now in its fourth edition. Utilization scholars shifted the dominating paradigm of evaluation research from evaluation as a product to evaluation as a consumer good. The quality of an evaluation no longer was an inherent characteristic, such as scientific quality, but an attribution from the outside, such as its usefulness. The context of an evaluation, in particular its potential audience, gained in importance in the judgement of whether an evaluation is good or not. The research started to focus on the different kinds of evaluation use and to elaborate analytical framework to capture the factors affecting utilization (Alkin and Daillak 1979). This historical turn in evaluation research can be pictured by the fact that authors like Patton and Alkin are categorized within the "social accountability" branch of the "evaluation theory tree" (Alkin and Christie 2013, p. 12). Similarly, evaluation research started to distinguish between the different "classes of users", such as program sponsors, program directors, program staff, client groups and the broader civil society (Weiss 1998). The interest for evaluation use can be traced back to the 1960s (Alkin and Taut 2003). Evaluation thus moved back to its original meaning in the policy cycle (May and Wildavsky 1978) to what Hendriks (2012, p. 434) describes

> as an act of judgment about the performance of a particular policy process, department or program based on its desirability, worth, or value. Evaluation is the phase of the policy process where there is ideally an explicit opportunity for learning, reflection and improvement. It represents the moment where the policy cycle ends but also restarts. (Althaus et al . 2007).

This broad understanding of policy evaluation brought affected target groups as well as ordinary citizens back in.

Cousins depicts the principle at the heart of participatory evaluation as follows: "the locus of judgment about program merit, worth, and significance should lie with *nonevaluator* stakeholders . . . located within the program community. A pivotal role for evaluators, then, is to foster informed judgment making on the part of stakeholder" (2013, p. 346). Models of participatory evaluation evolved, and the most prominent of them is Stufflebeam's (2000, 2003; Stufflebeam and Shinkfield 2007) CIPP model that we present in the following before turning to additional models of participatory evaluation and to the question of how participatory evaluation fosters use and usability of evaluation results.

Stufflebeam's (2000) core idea is that evaluators should participate in the programs they evaluate to fully understand what is going on in them. In this vein, context – the "C" in

the CIPP model – means that evaluators help the affected groups, the commissioners or even the evaluees understand the problem situation and its causes. Input, the "I" in CIPP, then is the solution the evaluator develops in cooperation with the commissioners. "P" for process regards the implementation, where the evaluator accompanies the implementation of the solution in a formative evaluation.[i] "P" for product then is the effect of the implemented solution, which again the evaluator assesses together with the involved actors. The process thus is highly participative. The evaluator basically is always there. This omnipresence makes it obvious that the model is hard to apply in public policy where strong political interest and values are at stake and evaluation by no means is the dominant actor. The CIPP model was not designed for policy evaluation, though, but for organizational contexts such as schools or hospitals addressing concrete issues with specific programs. However, the strong presence of the evaluator also in these cases makes it difficult to distinguish evaluation from intervention and the two may blur. In the end, the evaluators evaluate their own solution.

The main argument for the CIPP model, however, is its strong integration of all involved actors when it comes to determining whether a program is a success. The purest manifestation of participative evaluation is self-evaluation, in which evaluees evaluate their own activities and ultimately themselves (Kirkhart et al. 1991). Balthasar (2011) takes the poor quality of self-evaluations as a point of departure for what he labels the "critical friend approach". Balthasar's approach is a milder form of the CIPP model. He argues that the constant availability of an evaluator in an advisory capacity to program managers is an adequate way for evaluators to respond to three partly contradictory criteria evaluations: methodological soundness (Rossi et al. 1999), practical relevance (Patton 1997) and a transparent evaluation process (Scriven 1991). According to Balthasar, one of the crucial benefits of formative evaluation is its ability to "quickly inform practice" (2011, p. 201).

Similarly, important debates have taken place on the question of evaluation criteria. The choice of evaluation criteria is a crucial step of the evaluative process and determines the understanding of a policy success, which can be considered under various perspectives. In this regard, criticism can be raised concerning the narrowness of evaluation criteria used in classical top-down approaches where political authorities mainly aim at testing the effectiveness and the efficiency of public interventions. Some authors question the fact that political actors and policy experts are the only ones to define evaluation criteria. According to Enzensberger, Wietschel and Rentz, "While these criteria are helpful and commonly used to assess and to compare policy instruments from the point of view of a policy maker, they neglect in many cases the interests of other important stakeholder groups" (2002, p. 796). Indeed, a whole range of evaluation criteria that better reflect the diversity of the concerned policy stakeholder has emerged, especially in policy fields confronted with acute and transversal society challenges such as environmental and health policies. These evaluation criteria can cover gender responsiveness (Peersman 2014), equal opportunities (Chianca 2008), equity (Cabugueira 2001), capacity building (Grant and Curtis 2014), geographic coverage (Vidal Legaz et al. 2017), additionality (Lucas et al. 2008) and the economic, environmental and social dimensions of sustainability (Kowalski et al. 2009). The importance

of triangulating different kinds of evaluation criteria has been underlined (Crabbé and Leroy 2008). In a participatory approach, stakeholders from civil society can thus be included in the reflection regarding the choice of evaluation criteria, in order to achieve a representative understanding of policy success.

In recent discussions, a way of combining technical approaches and stakeholder participation was identified in mixed-method evaluation designs. It was argued that method and data triangulation, as core components of mixed approaches, allows for an inclusive evaluation process, a more comprehensive analysis of the evaluation object as well as sound respondent validation mechanisms. In this view, mixed models enhance not only the democratic legitimacy of evaluations but also their scientific robustness (Torrance 2012). Moreover, mixed-method approaches are seen as a way of striking a balance between outcome and process analysis (Park et al. 2014) as well as between efficacy and quality evaluation (Greene 2005). In this sense, by engaging with methodological diversity, mixed approaches constitute "an important counterpoint to the contemporary debate about what constitutes valid, rigorous, and 'scientific' research'" (Greene 2005, p. 207). Much more generally, Hendriks takes sides for a stronger role of citizen deliberation in evaluation. She states,

> *ex post* policy evaluation can be greatly enriched by the application of more innovative forms of public participation. Processes that emphasize inclusion and deliberation can help to elicit important perspectives on policy programs and agencies, and expose relevant arguments, perspectives, and values to public scrutiny. Participation can also help public managers to negotiate the increasing complexity of policy issues by bringing those affected by policies into the evaluation process. Most importantly the inclusion of affected publics helps to secure the democratic legitimacy of an evaluative process and its outcomes. (Hendriks 2012, p. 443)

Hendriks's approach does not go as far as the CIPP model or the critical friend approach. Participation is an important part in assessing policy success but it does not substitute for full evaluation. Her grip on participative evaluation styles therefore definitely is more realistic in a policy landscape that largely depends on expertise. Still, this form of participative evaluation must also strike the difficult balance between evaluee involvement and independent assessment of a policy.

Shortcomings and strengths of the two evaluation styles

The two evaluation styles evolved for good reasons. Neither one is perfect; each needs a complement. The core reproach to expert evaluations is technocracy referring to an overly strong political bureaucracy (Rosser and Mavrot 2016), or even worse, expertocracy (e.g., Grek 2013) referring to an unduly political role of experts out of governmental control and public accountability. Both reproaches are relatively easy to counter in that policy evaluations can play many roles in the policy process but they hardly ever are decisive in democratic decision-making. Rather, political stakeholders may use evaluations politically to support their

arguments. Symbolic or political use or evaluations, however, are not expertocracy, rather the opposite: it is the politically motivated and deliberate ignorance of the actual findings of an evaluation for the strategic selection of convenient arguments (Ledermann 2011). Schlaufer et al. (2018) and Sager et al. (2017) show that even the political use of evaluations leads to an increase in the discourse quality of political debates. The danger of expertocracy due to evidence gained from evaluations, therefore, is more of a strawman than a real threat to democratic policy-making.

However, Hernando et al. (2018) make an important point as to the relationship of experts and politicians in today's politics: the more contested policy findings become, the more policy experts become political actors and the less the two spheres of expertise and politics are distinct. The authors refer to the role of think tanks in the global economic crisis, but the same argument applies to highly salient issues like climate change, economic inequality and vaccination in public health. If the public can no longer tell experts from politicians, evidence-based policy-making is a difficult ideal to achieve in reality. The argument, however, does not apply to the same degree to expert evaluations that focus much more on specific questions of implementation and effects than basic research that addresses the large societal questions. We posit that evaluators can contribute to policy debates but based on their research cannot take the role of political stakeholders. Expert evaluations, however, may run the risk of being too technical to be of use for their audience in administration and politics. Ledermann (2011) shows that a high scientific quality of evaluations only under very specific conditions leads to use. More important seems to be that they pass the truth test by their audience readers. Ledermann's (2011) findings corroborate Patton's (1988, p. 5) claim for the "evaluator's responsibility for utilization".

However, the contribution of expert evaluations in the deliberative public sphere of contemporary democracies is highly valuable. Their distance from the studied policy subfield guarantees that the focus lies primarily on scientific considerations and accounts for a high process robustness. In spite of the fact that they do not include stakeholders to the same degree that participatory approaches do, the core principles of expert evaluations are the explicit disclosure of evaluation criteria and the replicability of the enquiry, which ensures a fully transparent process. Finally, as much as the inclusive promise of a participatory stance to policy evaluation is tempting, the required equilibrium between involvement and independence is highly difficult to achieve in practice. In contrast, the posture of expert evaluation is unambiguous. An expert distance to all evaluees ultimately also works as a guarantee of equitable treatment. The inevitable proximity of evaluators to their direct commissioners, however, remains a challenge.

As seen, a close relationship between evaluators and evaluees may foster perceived usefulness and hence utilization of evaluation results. A high proximity between the two roles, however, may impair the independence of the evaluation and thus its credibility. This in turn may be detrimental to the usability of evaluation results. Pleger et al. (2017, p. 315) show for four advanced Western democracies with high national evaluation activity how consistently strong the attempts of policy stakeholders are to put pressure on evaluators. The particular

role of evaluation commissioners must be underlined: "Moreover, the person who commissioned the evaluator for evaluation is identified by all studies as the primary influencing stakeholder in the evaluation process." Independence of evaluations thus clearly is an issue in current evaluation praxis. This finding is an impediment for participatory evaluation styles. Pleger et al. (2017) propose evaluations systems to more explicitly refer to evaluation standards to prevent pressure on evaluators. Such standards tend to draw a clear line between evaluators and evaluees and hence run counter to ideals of participative evaluation. Participatory evaluations, however, strive to avoid excessive asymmetry between the role of commissioners and the role of evaluees or other important stakeholders by actively including the latter in the process.

In fact, not all influence is bad. In line with the claim of participative evaluation, stakeholder attempts to influence evaluation findings can also be informative, constructive and consequently highly valuable for an evaluation to come up with robust and usable findings and recommendations. Pleger and Sager (2018, p. 166) therefore propose a

> heuristic model of influence on evaluations that does justice to this ambivalence of influence: the BUSD-model (betterment, undermining, support, distortion). The model is based on the distinction of two dimensions, namely "explicitness of pressure" and "direction of influence". [. . . The model offers a] tool to distinguish positive from negative influence in the form of three so-called differentiators (awareness, accordance, intention).

Awareness asks whether the stakeholder knows that they put pressure on the evaluator. Accordance asks whether the influence impairs scientific standards. Intention asks whether the influencer means to decrease the overall quality of the evaluation. Pleger et al. (2018) claims that the three differentiators support evaluators to tell negative from positive influence and help them to take advantage of the benefits of participative evaluation styles while avoiding the pitfalls of undue distortion of results. Finally, a crucial factor in sorting between positive and negative influence on evaluations is the timing. When commissioners and stakeholders are included in the reflection regarding research questions, the research design and evaluation criteria during the early stages of the process, their opinions can be highly valuable to optimize the conception of the evaluation. In contrast, late interventions aimed at modifying evaluation criteria in retrospect or at influencing the presentation of results are detrimental.

Apart from the risk deriving from the proximity of evaluators to stakeholders (Villeval et al. 2016, p. 1032), further challenges of participatory evaluation approaches have been identified, such as their high costs and their time-consuming nature (Plottu and Plottu 2009). However, participative evaluations have many further advantages. Their long-term approach centered on stakeholders allows for individual and organizational capacity-building (King 2013), empowerment processes and a better appropriation and use of the evaluation results (Papineau and Kiely 1996). Moreover, the integration of stakeholders can lead to a better quality of the information gathered, and the field knowledge of stakeholders can be used to

secure appropriate interpretation of the data (Viswanathan et al. 2004). Finally, the transparency of the process enhances the acceptance of evaluations (Taut 2008). Table 30.1 summarizes the respective advantages of both evaluation styles. The dimensions highlighted in the table are to be understood as patterns and not as fixed characteristics of each evaluation style. In fact, a wide range of mixed models exists in evaluation practice.

Table 30.1 – Respective strengths of expert and participative evaluation styles

Expert-based Evaluation – Top-down		Participatory Evaluation – Bottom-up	
Strengths	*Weaknesses*	*Strengths*	*Weaknesses*
Objectivity	Non-inclusiveness	Stakeholder ownership	Subjectivity
Scientificity	Risk of illegitimacy	Representativeness	Risk of partiality
Independence	Black box	Transparency	Influence
Knowledge transfer	Top-down process	Feedback loops	Lack of control
Cost-effectiveness	Single-shot	Durability	Resource-consuming
Result-based legitimacy	Internal validity	Process-based legitimacy	External validity

Conclusion: the role of evaluation in post-truth democracy

> *The man who wears the shoe knows best that it pinches and where it pinches, even if the expert shoemaker is the best judge of how the trouble is to be remedied*. (Dewey 1927, p. 224)

It is not arbitrary to begin the conclusion of a chapter on expert versus participatory evaluation styles with a quote from the great American pragmatist John Dewey. Pragmatism, after all, is the philosophy of problem solving as overarching ethical principle of social action. Policy evaluation is nothing else but the empirical quest for the best political solution to a societal problem. Rogers (2010, p. 81) sums up Dewey's stance on experts versus involved actors:

> For Dewey, the hypotheses we form for responding to political problems are only as good as the methods we employ – that is, the extent to which the methods make us receptive to data from various parts of the environment. But problems themselves, as he argues, frame and guide our inquiry; they imply the existence of a complex horizon of value and meaning that is now fractured and in need of creative valuation to restore continuity. So Dewey's point is not simply that without the input of the wearer of shoes the shoemaker will respond in a way that would not address the existing pinch.

Rather, without input from the individual experiencing the pinch, the expert shoemaker will not have the subject matter to initiate or guide his inquiry.

Applied to the case of policy evaluation, Dewey thus takes us back to the challenge of striking the right balance between stakeholder participation on one side and evaluation independence and scientific integrity on the other. In this chapter, we discuss various ways to strike this balance. The current political context, however, may be an additional hurdle for these attempts. Oxford dictionaries voted "post-truth" as their word of the year 2016.[ii] A post-truth democracy is a problem for the role of evaluation as an evidence-based fundament for sound policy-making. As we have seen earlier, the peril of stakeholder pressure lurks even in countries with a solid evaluation culture. At the same time, truth becomes a negotiable entity when there no longer exists a shared understanding of what counts as robust evidence. Political pressure on the one side and a lack of acceptance of scientific evidence on the other may therefore result in heavily distorted evaluations that not only will not be accepted as relevant information by their users but also ultimately will be worthless from a scientific point of view. A way out of such a dead end is in order.

We make the case for a pragmatist blend of evaluation styles to overcome post-truth deadlock. We see great merit in the classic expert evaluation tradition. Evaluation knowledge needs replicability to be worthwhile for practitioners to consider it in decision-making. We also must stress the value of expert evaluations for accountability purposes. Expert evaluations distinguish design failure from implementation failure and hence allow telling whether mistakes took place at the political or at the administrative arena. However, such findings need acceptance and terrain credibility to find usage. Participative evaluation styles achieve this goal. We follow Hendrik's (2012) positive account of deliberative citizen participation in policy evaluation. These approaches go well together with systematic social scientific research if we clearly delineate scope conditions. We therefore conclude with suggesting a division of labor that merges the two styles in the following oversimplified manner: expert evaluation takes care of the analytical model and the causal analysis, while data collection and, most importantly, data assessment follow a participative evaluation style. This proposition ideally allows keeping the advantage of causal analysis while benefitting from a realistic and terrain prone value judgement of policy success. We cannot always have the cake and eat it, too, but our proposal may help us get a bit closer to doing both.

References

Alkin, M.C. and Christie, C.A., 2013. An Evaluation Theory Tree. *In*: M.C. Alkin, ed. *Evaluation Roots: A Wider Perspective of Theorists' Views and Influences*. 2nd ed. Los Angeles: Sage, 12–58.

Alkin, M.C. and Daillak, R.H., 1979. A Study of Evaluation Utilization. *Educational Evaluation and Policy Analysis*, 1 (4), 41–49.

Alkin, M.C. and Taut, S.M., 2003. Unbundling Evaluation Use. *Studies in Educational Evaluation*, 29 (1), 1–12.

Althaus, C., Bridgman, P. and Davis, G., 2007. *The Australian Policy Handbook*. 4th ed. Crows Nest: Allen & Unwin.

Balthasar, A., 2011. Critical Friend Approach: Policy Evaluation Between Methodological Soundness, Practical Relevance, and Transparency of the Evaluation Process. *German Policy Studies*, 7 (3), 187–231.

Cabugueira, M.F.M., 2001. Voluntary Agreements as an Environmental Policy Instrument: Evaluation Criteria. *Journal of Cleaner Production*, 9 (2), 121–133.

Chen, H.-T. and Rossi, P.H., 1987. The Theory-driven Approach to Validity. *Evaluation and Program Planning*, 10 (1), 95–103.

Chianca, T., 2008. The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement. *Journal of MultiDisciplinary Evaluation*, 5 (9), 41–51.

Chouinard, J.A., 2013. The Case for Participatory Evaluation in an Era of Accountability. *American Journal of Evaluation*, 34 (2), 237–253.

Cousins, J.B., 2013. Privileging Empiricism in Our Profession: Understanding Use Through Systematic Inquiry. *In*: M.C. Alkin, ed. *Evaluation Roots: A Wider Perspective of Theorists' Views and Influences*. 2nd ed. Los Angeles: Sage, 344–352.

Crabbé, A. and Leroy, P., 2008. *The Handbook of Environmental Policy Evaluation*. London: Routledge.

Dahler-Larsen, P., 2012. *The Evaluation Society*. Stanford: Stanford University Press.

Dehar, M.-A., Casswell, S. and Duignan, P., 1993. Formative and Process Evaluation of Health Promotion and Disease Prevention Programs. *Evaluation Review*, 17 (2), 204–220.

Dewey, J., 1927. *The Public and Its Problems*. Chicago: Swallow Press.

Enzensberger, N., Wietschel, M. and Rentz, O., 2002. Policy Instruments Fostering Wind Energy Projects: A Multi-Perspective Evaluation Approach. *Energy Policy*, 30 (9), 793–801.

Grant, A. and Curtis, A., 2014. Refining Evaluation Criteria for Public Participation Using Stakeholder Perspectives of Process and Outcomes. *Rural Society*, 14 (2), 142–162.

Greene, J.C., 2005. The Generative Potential of Mixed Methods Inquiry. *International Journal of Research & Method in Education*, 28 (2), 207–211.

Grek, S., 2013. Expert Moves: International Comparative Testing and the Rise of Expertocracy. *Journal of Education Policy*, 28 (5), 695–709.

Hendriks, C.M., 2012. Policy Evaluation and Public Participation. *In*: E. Araral *et al.*, eds. *Routledge Handbook of Public Policy.* Abingdon: Routledge, 434–448.

Hernando, M.G., Pautz, H. and Stone, D., 2018. Think Tanks in 'Hard Times': The Global Financial Crisis and Economic Advice. *Policy and Society*, 37 (2), 125–139.

King, J.A., 2013. Getting People Involved: The Origin of Interactive Evaluation Practice. *In*: M.C. Alkin, ed. *Evaluation Roots: A Wider Perspective of Theorists' Views and Influences*. 2nd ed. Los Angeles: Sage, 334–343.

Kirkhart, K.E., Morgan, R.O. and Sincavage, J., 1991. Assessing Evaluation Performance and Use. *Evaluation Review*, 15 (4), 482–502.

Knoepfel, P. and Bussmann, W., 1997. Die öffentliche Politik als Evaluationsobjekt. *In*: W. Bussmann, U. Klöti and P. Knoepfel, eds. *Einführung in die Politikevaluation.* Basel: Helbing und Lichtenhahn, 58–77.

Kowalski, K. *et al.*, 2009. Sustainable Energy Futures: Methodological Challenges in Combining Scenarios and Participatory Multi-Criteria Analysis. *European Journal of Operational Research*, 197 (3), 1063–1074.

Lasswell, H.D., 1956. *The Decision Process: Seven Categories of Functional Analysis*. College Park, MD: University of Maryland Press.

Ledermann, S., 2011. Exploring the Necessary Conditions for Evaluation Use in Program Change. *American Journal of Evaluation*, 33 (2), 159–178.

Ledermann, S. *et al.*, 2006. *Evaluation der Strategie Migration und Gesundheit 2002–2006*. Bern: Büro Vatter.

Linder, S.H. and Peters, B.G., 1987. A Design Perspective on Policy Implementation: The Fallacies of Misplaced Prescription. *Review of Policy Research*, 6 (3), 459–475.

Lucas, K. *et al.*, 2008. Promoting Pro-Environmental Behaviour: Existing Evidence and Policy Implications. *Environmental Science & Policy*, 11 (5), 456–466.

MacRae, D., 1975. Policy Analysis as an Applied Social Science Discipline. *Administration & Society*, 6 (4), 363–388.

Marsh, D. and McConnell, A., 2010. Towards a Framework for Establishing Policy Success. *Public Administration*, 88 (2), 564–583.

May, J.V. and Wildavsky, A.B., eds., 1978. *The Policy Cycle*. Beverly Hills: Sage.

Papineau, D. and Kiely, M.C., 1996. Participatory Evaluation in a Community Organization: Fostering Stakeholder Empowerment and Utilization. *Evaluation and Program Planning*, 19 (1), 79–93.

Park, M.M. *et al.*, 2014. Transforming Mental Health Services: A Participatory Mixed Methods Study to Promote and Evaluate the Implementation of Recovery-Oriented Services. *Implementation Science*, 9 (1), 119.

Patton, M.Q., 1988. The Evaluator's Responsibility for Utilization. *Evaluation Practice*, 9 (2), 5–24.

Patton, M.Q., 1997. *Utilization-Focused Evaluation: The New Century Text*. 3rd ed. Thousand Oaks: Sage.

Patton, M.Q., 2011. *Essentials of Utilization-Focused Evaluation*. Los Angeles: Sage.

Patton, M.Q., 2012. A Utilization-Focused Approach to Contribution Analysis. *Evaluation*, 18 (3), 364–377.

Pawson, R. and Tilley, N., 1994. What Works in Evaluation Research? *The British Journal of Criminology*, 34 (3), 291–306.

Peersman, G., 2014. *Evaluative Criteria: Methodological Briefs, Impact Evaluation No. 3*. Florence: UNICEF Office of Research.

Pleger, L. and Sager, F., 2018. Betterment, Undermining, Support and Distortion: A Heuristic Model for the Analysis of Pressure on Evaluators. *Evaluation and Program Planning*, 69, 166–172.

Pleger, L. *et al.*, 2017. Are Some Countries More Prone to Pressure Evaluators Than Others? Comparing Findings From the United States, United Kingdom, Germany, and Switzerland. *American Journal of Evaluation*, 38 (3), 315–328.

Pleger, L., Lutz, P. and Sager, F., 2018. Public Acceptance of Incentive-Based Spatial Planning Policies: A Framing Experiment. *Land Use Policy*, 73, 225–238.

Plottu, B. and Plottu, E., 2009. Approaches to Participation in Evaluation: Some Conditions for Implementation. *Evaluation*, 15 (3), 343–359.

Rogers, M.L., 2010. Dewey and His Vision of Democracy. *Contemporary Pragmatism*, 7 (1), 69–91.

Rosser, C. and Mavrot, C., 2016. Questioning the Constitutional Order: A Comparison of the French and the U.S. Politics – Administration Dichotomy Controversies After World War II. *The American Review of Public Administration*, 47 (7), 737–751.

Rossi, P.H., Freeman, H.E. and Lipsey, M.W., 1999. *Evaluation: A Systematic Approach*. 6th ed. Thousand Oaks: Sage.

Rossi, P.H., Lipsey, M.W. and Henry, G.T., 2019. *Evaluation: A Systematic Approach*. Los Angeles: Sage.

Sager, F., Widmer, T. and Balthasar, A., eds., 2017. *Evaluation im politischen System der Schweiz: Entwicklung, Bedeutung und Wechselwirkungen*. Zürich: NZZ Libro.

Schlaufer, C., Stucki, I. and Sager, F., 2018. The Political Use of Evidence and Its Contribution to Democratic Discourse. *Public Administration Review*, 78 (4), 645–649.

Scriven, M., 1991. *Evaluation Thesaurus*. 4th ed. Thousand Oaks: Sage.

Scriven, M., 2008. The Concept of a Transdiscipline: And of Evaluation as a Transdiscipline. *Journal of MultiDisciplinary Evaluation*, 5 (10), 65–66.

Stufflebeam, D.L., 2000. The CIPP Model for Evaluation. *In*: D.L. Stufflebeam, G.F. Madaus, and T. Kellaghan, eds. *Evaluation Models: Viewpoints on Educational and Human Services Evaluation.* Dordrecht: Kluwer Academic Publishers, 279–317.

Stufflebeam, D.L., 2003. The CIPP Model for Evaluation. *In*: T. Kellaghan and D.L. Stufflebeam, eds. *International Handbook of Educational Evaluation*. Dordrecht: Springer, 31–62.

Stufflebeam, D.L. and Shinkfield, A.J., 2007. *Evaluation Theory, Models, and Applications*. San Francisco, CA: Jossey-Bass.

Taut, S., 2008. What Have we Learned about Stakeholder Involvement in Program Evaluation? *Studies in Educational Evaluation*, 34 (4), 224–230.

Torrance, H., 2012. Triangulation, Respondent Validation, and Democratic Participation in Mixed Methods Research. *Journal of Mixed Methods Research*, 6 (2), 111–123.

Vidal Legaz, B. *et al.*, 2017. Soil Quality, Properties, and Functions in Life Cycle Assessment: An Evaluation of Models. *Journal of Cleaner Production*, 140, 502–515.

Villeval, M. *et al.*, 2016. Enabling the Transferability of Complex Interventions: Exploring the Combination of an Intervention's Key Functions and Implementation. *International Journal of Public Health*, 61 (9), 1031–1038.

Viswanathan, M. *et al.*, 2004. Community-Based Participatory Research: Assessing the Evidence. *Evidence Report/technology Assessment (Summary)*, 99, 1–8.

Weiss, C.H., 1997. *Evaluation: Methods for Studying Programs and Policies*. 2nd ed. Upper Saddle River: Prentice Hall.

Weiss, C.H., 1998. Have we Learned Anything New About the Use of Evaluation? *The American Journal of Evaluation*, 19 (1), 21–33.

Notes

[i] As opposed to "summative evaluation" mainly focused on outcome measurement,

"formative evaluation" is defined as "an ongoing process", which "provides

assessment information within a feedback loop. This assessment identifies the

strengths and weaknesses of the project as it progresses" (Evans *et al.* 1989, p. 230,

cited in Dehar et al. 1993, p. 213).

[ii] https://languages.oup.com/word-of-the-year/word-of-the-year-2016