



Theory-based Evaluation and Learning: Possibilities and Challenges

PETER VAN DER KNAAP

Netherlands Court of Audits, The Hague, The Netherlands

Originated in the 1970s, decried in the 1990s, theory-based evaluation recaptured debate at the European Evaluation Society conference in 2002. The promise is that the set-up, implementation, delivery and utilization of evaluation research may be facilitated by taking the assumptions and objectives of public policy as a starting point. In addition, a theory-based approach is considered to be valuable for a government that wants to learn, whilst potentially reconciling positivist and constructivist approaches. This article explores the rationale for a theory-based approach in policy development, debate, learning and evaluation utilization. It does so by confronting the – often paradoxical – dilemmas that haunt every evaluator (and policy maker). How can we do justice to societal complexity yet still maintain focus? How can we enter apparently closed policy systems? And how do we find a balance between mere improvement and innovative learning?

KEYWORDS: constructivism; dialogue; learning; positivism; theory-based evaluation; utilization

Theory-based Evaluation: A Rediscovered Promise

Both amongst scholars and practitioners, theory-based evaluation appears to be promising again. At the 2002 European Evaluation Society (EES) conference in Seville, where the concept was featured in the trinity of ideas ‘Movements in contemporary evaluation: learning, theory and evidence’, Lam aptly summarized the reasons for this:

Theory-based evaluation and objective-based evaluation have gained tremendous popularity in the last decade because of its power to augment an evaluation’s capacity to assess program impact without strong experimental designs and the contemporary reliance on setting standards. (Lam, 2002: 1)

Deliberately taking the assumptions and objectives on which policy programmes are based as a starting point, may enhance the set-up, implementation, delivery and utilization of evaluation research.

Theory-based evaluation can be described as the analysis and valuation of the contribution of intervention strategies to resolving or controlling social problems. The traditional starting point of theory-based evaluation is provided by the objectives and assumptions on which a certain policy programme is based (see Leeuw, 1983; Hoogerwerf, 1984). The pursuit of policy objectives may thereby be regarded as deliberately initiating or supervising processes of social change. These processes are based on rational assumptions: a link is assumed between objectives and the manner in which these objectives might be attained by the deployment of resources. The collection of assumptions on which policy measures are based constitutes the policy theory: a system of values, norms, assumptions regarding causal links between actions and the results of actions and preferences.

Theory-based evaluation often in the first instance revolves around the analysis and assessment of the extent to which policy programmes have resulted in their original objectives: are the initial expectations met? A second characteristic of theory-based evaluation is a better understanding of the underlying causal mechanisms. In such a case, the central question is: were the assumptions on which the policy programme was based 'right' or not? The objectives, the causal assumptions as well as the implementation of the policy may then be judged and discussed on the basis of the answers to these questions (Leeuw, 1983).

The Focus of this Article

This article addresses the possibilities and challenges of a theory-based approach for a government that wants to learn from evaluation and evaluators that want to make a difference. To do this, I will focus on the impact of 'theory' in policy development, debate, implementation and learning, and evaluation utilization.

It must be stressed that theory-based evaluation in the year 2004 is on no account, as in the positivist ideal, expected to rely on policy theories with universal validity. On the contrary: the local valuation and implementation of policy initiatives are most relevant: 'The same programme will have different outcomes in different situations' (Pawson, 2002). The review of policy theories therefore always revolves around the multiple question: 'What works for whom in what circumstances?' (Pawson and Tilley, 1997: 85). In other words, despite the fact that rationality is central to the notion of theory in public policy, contextuality and subjective judgements are never dismissed as mere methodological flaws: rather they form the framework in which policy theories can be useful, discussed and valued in the first place.

Accepting the latter premise, an intriguing question presents itself: what role can theory-based evaluation play in reconciling positivist and constructivist approaches to evaluation? To this end, an assessment is made of the benefits and limitations a theory-based perspective can offer the evaluator who tries to be goal-oriented and at the same time open to innovation and, hence, the perspectives of stakeholders. Such an evaluator, also, strives to meet the demand for feedback information whilst seeking to make a contribution to policy-oriented learning by way of policy-based debate and argumentation. Having assessed the

possibilities and limitations of theory-based evaluation an attempt is made to bridge paradigms.

Benefits and Challenges of a Theory-based Approach: A Closer Look at Policy Development, Debate, Learning and Evaluation Utilization

Phase 1: Policy Development – Complexity and Focus

Policy theories guide perception, thought and action. They give focus but may also distort our worldview. Obviously, the term ‘theory’ is not exclusively reserved for public policy or evaluation. The notion touches one of the most important insights of modern cognitive psychology: the ways in which people organize their knowledge of the world – and thereby the world itself (Taylor and Crocker, 1981). People develop generalizing notions and theories of regularities or links between phenomena they have observed or experienced. Schematic knowledge and assumptions embodied in theories serve as the basis of our understanding of the world, including public policy programmes. Weick probably best describes the twofold function of theoretic schemes: ‘A schema is an abridged, generalized, corrigible organization of experience that serves as an initial frame of reference for action and perception’ (Weick, 1979: 50).

Schematic knowledge and theories are by nature logical abstractions of reality. The theory of logic (*logikè*) deliberately *excludes* a great wealth of ideas. In this way, people get focus. Government policies often set out to achieve something specific: to solve – or control – a problem.¹ A policy theory expresses the ‘promise’ of the – logically assumed – causality between means, instruments and those specific objectives: ‘If the actions we recommend are undertaken, good (intended) consequences rather than bad (unintended) ones actually will come about’ (Wildavsky, 1987: 35). By definition, any policy theory gives a simplified version of reality. Moreover, every such policy theory, in particular one that is well articulated and accepted, directs perception, interpretation and, hence, assessment.

The last 50 years or so, many have pointed out the risks of this simplifying effect, especially where they are applied in a so-called result-based management model. Schwandt formulates it as follows:

One normative ideal [of theory-based performativity] is that monitoring systems and accountability ought to replace the complex social–political processes entailed in the design and delivery of social and educational services. (Schwandt, 2002: 9)

And indeed it is difficult to underestimate the simplifying effect of a policy theory and the effects of measurability and monitoring. Where too much emphasis is placed on theoretical causality and measurability, this may lead to tunnel vision, rigidity and fear of innovation. A real risk is that everything that cannot be expressed in theories, performance data and objectives may escape the attention of both decision makers and evaluators. Especially where objectives and the performances to be conducted are described in detail, there is a risk that the implementation of a certain task turns into a rigid, inflexible implementation of

provisions. Both tunnel vision and rigidity may pose an obstacle to understanding, and also to innovation and improvement of policy.

Notions such as simplification, tunnel vision and rigidity imply that a theory-based approach to policy making only concerns negative aspects. This is not always the case: experience with the introduction of goals-oriented budgeting demonstrates that many government organizations can profit from looking at social problems, policy options and their own performance with a little more of the above-mentioned 'focus'. In a world that seems to become more complex by the minute, policy makers need to manage their sense making and understanding by distinguishing the significant from the trivial. During the policy-development stage, an explicit outline of the assumptions and expectations on which policy measures will be based is functional in that regard. The systematic look it provides at the classifications and linkages may even help the critical individual to resist the automatic classifying pressures of our public policy institutions. If a policy maker is aware of the negative effects mentioned earlier, the combination of focus and the opportunity to re-examine proposals may enable better-informed decisions on policy proposals. This prospect is realized most visibly when ex ante evaluations are fed into programmes.

Phase 2: Policy-oriented Debates – Openness and Closure

An explicit policy theory is clear to others: it may be referred to within a policy debate. By emphatically 'articulating' a policy theory, a policy maker adopts a position. The policy theory not only provides guidelines to thought and action (and the evaluation of both – see below), but also sets a framework of reference for argumentation. The focusing effect of such an articulated policy theory is functional where it offers a clearly recognizable starting point for a policy-oriented debate.

In our multi-form societies, actors often pursue divergent interests. They therefore have specific opinions about and images of what is real and what is to be desired. Areas for special attention, perspectives and rationalities often diverge: 'Realities exist in the form of multiple mental constructions, socially and experientially based, local and specific, dependent for their form and content on the persons who hold them' (Guba, 1990: 27). Herein, it is now widely accepted that policy makers do not have a monopoly to 'right' or 'true' realities – no matter how perfectly they are polished or presented. Also, in our multi-form societies, the same actors that entertain different definitions of reality are often stuck with one another for the solution of 'problems' (which may include the policy proposals themselves and on which definition, under the rule of democracy, there *can* exist monopolies). In most cases, this 'being stuck with one another' means: to seek to reach a workable agreement through dialogue.

In a constructive dialogue the policy actors involved should be open to the possibility that their own definitions of reality and policy theories will be challenged (Rein and Schön, 1993; see paragraph 3.3 for defensive mechanisms). It is conceivable that, consciously or unconsciously, they revise certain normative starting points and objectives as a result of the dialogue. The added value of policy development by means of open-argumentation dialogue is the possibility

Evaluation 10(1)

of reconciling or even superseding individual, branch or ideological contradictions. In sum, by means of a communicative dialogue the parties may:

- have their own vision and insights confronted with those of other parties;
- attain a more or less shared perception of a phenomenon or problem;
- take well-substantiated decisions on the policy measures to be taken; and
- look forward to a broader level of support for the measures to be implemented.

In short: by entering into a constructive dialogue a process of argumentation may commence through which actors may develop a more or less shared notion of what can and should be done (Majone, 1989).

Three comments must be added to all this. First, if theories guide perception, thought and action it is inevitable that this 'social construction of reality' will be subject to prejudice, bias and closure (see also the defensive mechanisms described in Phase 4).

Secondly, public policy must be legitimized in a democratic manner. This means that a government may at no time participate in any serious dialogue without any obligations attached: it must account for its intentions and/or the manner in which these have been turned into actions. This may mean that policy-oriented debates are subject to certain preconditions. For example, it must be indicated that the consultations may or may not concern objectives and have to be confined to implementation modalities. In this respect, it is often useful to temporarily 'freeze' a policy theory by 'closing the discussion' (e.g. in a policy memorandum, discussion document, white paper etc.). In my view, such a frozen articulation is not only important where it concerns the proper approach to the political primate, but also in order to preclude wrong expectations on the part of the discussion partners. In order to maintain trust (a prerequisite for true dialogue), one should be open about this relative reticence.

The third comment partly follows from the first: power and knowledge are hardly ever equally divided (Wildavsky, 1987). It would be naïve to assume that this unequal division does not influence the outcome of a debate. While knowledge may be power, power also influences – the distribution of – information and knowledge. A possible role for evaluators is to 'correct' this unequal division by bringing existing knowledge up for discussion and by making new knowledge available to the parties involved. Leeuw illustrates the dilemmas of this opening-up possibility very well by his reflections on the advantages and risks of publishing school results of educational institutions by way of ranks in order to stimulate schools to achieve a higher quality (Leeuw, 2002).

Phase 3: Policy Implementation and Learning – Seeking Improvement or Innovation, Judging Success or Contents

Undeniably, the concept of learning has a strong positive connotation, i.e. the inevitability of a certain belief in progress. It refers to those processes of knowledge production that result in a 'better' understanding or 'improved' intelligence. A 'learning' government can be described as a government that aims to improve its policies – and does so with a certain degree of success.

Policy-oriented learning is a process in which policy actors try to improve public policy measures, policy objectives and underlying normative assumptions. They can do so by detecting and correcting perceived imperfections. Herein, evaluation may contribute using theories as a starting point:

Evaluation research tries to discover whether programmes work. Programmes are theories. Therefore it follows that evaluation is theory-testing. A theory tells us where to look for evidence, theories drive learning. (Pawson, 2002)²

A policy theory has normative layers; for example, the Dutch integration policy is based on the substantive value that all legal newcomers must be provided with the opportunity to assume a valuable position within Dutch society. This value is subsequently translated into, inter alia, the operational objective or norm that newcomers must be able to find a 'suitable working environment' within three years of arrival. The integration programmes are subsequently introduced as an instrument. Based on the causal assumption that mastering a country's language will enhance one's chances to find a good job, these programmes contain an important language component.

The layers of a policy theory correspond with the different levels of policy-oriented learning. In general, learning concerns the development of wisdom, knowledge and skills. In policy-oriented learning two levels can be distinguished: success and contents. First of all, there is learning in support of the efficiency, effectiveness and successful implementation of policy. This 'improvement learning' often takes place during policy implementation. It revolves around the improvement of existing policy programmes and/or measures by detecting and correcting errors or shortcomings.³ Where the contents of policy, its underlying policy theory and the policy objectives embodied in this theory are themselves points of discussion and appraisal, however, more elementary questions emerge. Do we still think all legal newcomers must be provided with integration programmes? Are the norms used in the integration policy and its causal assumptions themselves valid? Given that this pursuit of improvements affects the basic principles of the policy and often leads to new points of departure and objectives, this can be referred to as 'innovative learning' (Van der Knaap, 1995; see Argyris and Schön, 1978).

The day-to-day practice is often confused by the fact that values, norms, assumptions and preferences are not easy to separate in practice (see also below). In many cases, they are a complex mixture of rational and emotional convictions. Even the parties involved may find it at times difficult to indicate the lines between normative convictions and instrumental preferences (Sabatier and Jenkins-Smith, 1993). In the example of the disproportionate labour participation of ethnic minorities, the large-scale refusal of employees to have 'race' or 'country of origin' registered, for example, has greatly reduced the employers' willingness to co-operate with the policy (Verweij, 1995). This touches on the tension described by March and Olsen (1989) between the logic of consequence (relating to the causal link between policy and objectives) and the logic more determined by the situation: the logic of appropriateness (relating to the acceptability and feasibility of action alternatives in specific situation).

Departing from a formal policy theory might provide a solution in this respect. As the values, norms, assumptions and preferences of much government policy are explicitly indicated in objectives and instruments, both the policy maker and the evaluator may aim at the specific levels of learning (e.g. the final objective or the interim objectives, the instrumentation method or the implementation manner). In this context Teisman et al. (2002) referred to ‘combination rules’. By means of these rules the policy makers and evaluators might review which use of evaluation most closely corresponds to the specific phase of the policy and regarding which points a connection must be made between assessing and reflecting evaluation. In this model, ‘evaluation for improvement’ focuses on existing frameworks of policy, while the ‘evaluation for innovation’ naturally provides more room for dynamics. Table 1 provides an overview of the characteristics of both evaluation types, which may be helpful to gear evaluations to specific needs.

Phase 4: Evaluation and its Utilization – Information Identification and Defensive Routines

Ever since the birth of evaluation, the utilization of evaluation findings has been an issue (Weiss, 1990). Taking the policy theory – or its absence! – as a starting point for an evaluation provides the evaluator with the opportunity to:

1. participate in the debate conducted in relation to a certain policy; and/or
2. make the evaluation findings part of the policy debate and learning.

This works on two dimensions (see Patton, 2002). The first dimension is of a *cognitive-constructivist* nature. As argued above, the policy theory supported by a policy maker determines what she expects and – subsequently – observes. Objective, non-prejudiced perception and interpretation are not possible (see Popper, 1972). Where an evaluation produces information with respect to the extent in which expectations are met, the policy maker – either consciously or unconsciously – is more susceptible to this than where she is unable to ‘classify’ the information easily (Weiss, 2001). The information is in line with an existing ‘general understanding of how the world works’ (Weiss, 2001). As a consequence, the recipient identifies with it. This even applies in cases where findings as such are at odds with the policy theory. After all, the conclusion that A has *not* led to B is also relevant.⁴

By opting for the ‘official’ policy theory as the starting point for research and assessment, an evaluator may connect more easily with the policy maker’s

Table 1. Tailoring Evaluation for Learning

<i>Evaluation for improvement</i>	<i>Evaluation for innovation</i>
Primarily judging success	Judging success and content
Focus is on existing policy theory, often quantitative	Focus is on complexity, often qualitative
Fits relatively closed policy systems	Requires openness, which is often painful after resistance
Often leads to identification	Often leads to resistance

framework of reference. In this context, Mabry convincingly illustrated that a non-conformist approach of an evaluation often leads to a mutual lack of understanding between the client and evaluator and – subsequently – to mutual disappointment: ‘Modernist expectations and yearnings are not easily allayed, even when evaluators and clients try to step beyond them’ (Mabry, 2002: 149). Not only the policy theory itself, but also the criticism of it thereby provides a good carrier for exchanging convictions, experiences and knowledge, including the findings and conclusions of an evaluation report (Pawson, 2002). This is especially true if an evaluation also plays a role as an instrument of accountability.

As in policy development, the risks of tunnel vision and rigidity present themselves at this stage too. In order to influence debate and evaluation, it may be a good strategy for a policy maker to strongly articulate and express one’s own policy theory. In their structural policy the European Commission succeeded in ‘framing’ to a large extent the discussion on the merits of the structural funds.⁵ The assessment of the European regional policy by the European Court of Auditors and other evaluative parties is determined to such a large extent by the general principles of the policy, that the lessons have for the most part been confined to the bandwidths of the existing policy theory (Van der Knaap, 1997). By the selection and convincing presentation of issues, the formulation of questions and by determining the sequence of meetings, it has succeeded in directing – the debate on – the outcome of the evaluation in a certain, subservient direction. Hence, any evaluator must be aware of the imperative effect of an explicitly formulated and convincingly presented policy theory.⁶

This bias effect brings me to the second dimension, which is of a *social-constructivist* nature. As a policy theory is increasingly ‘expressed publicly’ and becomes part of the dominant policy debate, it will be more difficult for a policy maker to ignore relevant criticism.⁷ This applies in particular to evaluation reports that have been made public, e.g. by the media. The principle is simple: a policy maker who takes the fundamentals of her policy seriously *and* has publicly announced it, knows that she is committed to this. If evaluation conclusions relate to such a policy theory, she will be more inclined to respond to the findings and advice, which may lead to utilization.

This does not mean that all criticism will be publicly accepted. In part, this is due to the fact that our knowledge – once it has solidified – will be defended against new insights by ‘dynamic conservative’ mechanisms (e.g. Schön, 1982). Criticism may also lead to open defensive conduct as a result of which a dialogue will, of course, lose its constructive nature. An evaluation will at all times pose at least some threat to a policy maker. This is in particular true where the evaluation results (will) form part of a public debate. Many policy makers display the understandable reaction of defending both the basic principles and the implementation of ‘their policy’. From the alleged interest of one’s own organization – or, further to this: the policy – a protective wall is constructed. Many reactions – such as ‘the evaluator has misjudged; used the wrong methods; drawn her conclusions too early’ – are therefore easy to understand (e.g. Van der Knaap, 1997). In this context, Argyris (1990) refers to ‘defensive routines’: mechanisms that protect

one against threatening confrontations and the ancillary feelings of insecurity, embarrassing shyness and threatened positions and reputations.

The paradox here is that a defensive dialogue may still lead to the learning processes an evaluator seeks to bring about. The defensive attitude of a policy maker during an institutionalized dialogue (in reply to evaluation and policy advice) might seem pathological, but it does not prevent elements of the criticism and advice from being used when the policy is being revised. In the case of the structural policies, the policy maker in question seems to adopt a considerably more defensive position in her replies to critical evaluation reports than is really necessary – given the later change to the policy. The European Commission did accept many of the improvements suggested by the European Court of Auditors and other parties after they have at first been reviled. The interweaving of the cognitive and social dimension in this process is demonstrated by the fact that key respondents within the Commission experience the improvements above all as ‘following from their own advancing insights’ (Van der Knaap, 1997: 262).

Utilization of evaluation findings often goes unnoticed. True learning sometimes requires a place away from the spotlight. The opportunity to learn from evaluation depends inter alia on the extent to which a policy maker might learn from a policy-based dialogue without feeling threatened. The paradox is that both constructive dialogues and defensive routines may provide just that.

Summary

To summarize: the function of a policy theory in a policy-based dialogue is fourfold. In policy development, a theory gives focus. It reduces complexity, which enables the policy maker to give her attention to the most important issues and to, indeed, develop a clear vision on how to achieve what. The drawback here is that the guidelines to thought and action a theory provides may come to distort attention and interpretation. Simplification, tunnel vision and rigidity are never far away from theories and beliefs. The second and third functions of a policy theory lie in the argumentation, implementation and learning processes around policy programmes. Expressing a policy theory gives the policy maker a clear position in the debate and does provide a framework of reference to all stakeholders and other parties involved in the policy-oriented debate. This use of theories allows for a well-considered architecture of policy argumentation. After debate, as the agreed directive for future policy, a policy theory constitutes the starting point of collective action on the basis of which:

1. interim adjustments by way of monitoring might be made; and
2. ‘learning for improvement’ or ‘learning for innovation’ may take place.

All the three functions summarized above are relevant to the evaluator. To her, opting for the ‘official’ policy theory as the starting point for research and assessment, may not only offer a logical starting-point for assessing success and contents, but also a good opportunity to connect with the policy maker’s thoughts and ambitions (whether publicly expressed or not). Table 2 summarizes the benefits and challenges of using policy theories as a starting point in the different phases of the policy cycle.

Policies on Quicksand: The Challenge of Vague Objectives and Hidden Assumptions to the Theory-based Approach

A phenomenon every evaluator will be familiar with is the apparent lack of theoretic background some policy programmes display. It is not exceptional for objectives to be stated in such vague terms as to be virtually meaningless. In addition, the mechanisms through which policy measures are expected to achieve stated outcomes are frequently left unspoken. This obviously poses a problem if the evaluation is to be based on theory. But does this mean that, in these cases, a theory-based approach is useless? I think not.

Firstly, the question to pose in such instances is: can we still call it policy, good policy? The question has semantics on its side: in our languages definition of the word 'policy' is tied up with ratio. Regarding differing definitions of 'policy':

- the standard Dutch dictionary (Van Dale) refers to 'well-considered' and 'deliberate' and links this to 'principles for action' (and also to 'consultations');
- the Oxford English Dictionary mentions 'prudent conduct'; whereas
- the Webster's Dictionary defines policy as 'a selected, planned line of conduct in the light of which decisions are made and coordination achieved'.

The whole concept of policy presumes a rational, almost positivist connection between the deployment of resources and the objectives to be attained. To put it succinctly: where no connection between the deployment of resources, instruments and intended results is presumed, managers and the persons managed will regard the attainment of the policy objectives merely as a stroke of luck.

I would hesitate to use the label 'policy' – or 'programme' for that matter – for every ill-considered wild guess initiative or blind experiment. True, there can be very good reasons for 'just doing something'. Except for symbolic or desperate (political) reasons, the main point – as I see it – then should be to learn more about the mechanisms at work in the social domain we seek to change. Depending on things like urgency and personal taste, this urge for better understanding can be of an inductive or deductive nature. If this experimental approach is lacking, (proposed) policy measures may well deserve criticism for just that. On

Table 2. Functions of Policy Theories in Development, Debate, Implementation and Learning, and Evaluation Utilization

<i>Phase in the policy cycle</i>	<i>Benefits of a theory-based approach</i>	<i>Challenges</i>
Development	Focus	Simplification, tunnel vision and rigidity
Debate	Framework of reference	Closure, reticence and power distortions
Implementation and learning	Deliberate choice between improvement and innovation	Wrong level learning, foggy theories
Evaluation (utilization)	Relevance, identification	Framing, defensive routines

the other hand, according to my view, some policy processes deserve approval: those in which real-time feedback on 'how things went' has led to lessons about stakeholders' desires, possible goals, and relations between action and results.

Secondly, there is always the possibility that there is a hidden theory behind a set of measures. Sometimes, at the start of a policy programme, it is considered not wise to be specific about intentions (e.g. the policy maker wants to avoid obstruction or opposition). It may also be that people just did not bother to write down the assumptions and expectations behind a programme. As an example, the idea that giving a subsidy may trigger desired behaviour (by industries, schools or other governments) may be considered too obvious to amplify. The obvious thing for the evaluator then is to try to reconstruct the 'innate' theories that guided decision making and action.

Theory-based Evaluation: Prospects to Reconcile Positivist and Constructivist Approaches

While assessing the possibilities and limitations of theory-based evaluation, we touched many of the grounds covered by the positivist–constructivist debate. Now let's return to the intriguing question presented at the beginning of this article: what role can theory-based evaluation play in reconciling positivist and constructivist approaches to evaluation?

The interpretation of trends in evaluation research is always precarious. On the one hand, the rather positivist call for 'insight into specific objectives and results attained' continues unabated. Formal frameworks continue to focus primarily on the quantification of efficiency and effectiveness (e.g. Ministry of Finance, 2001; UK Treasury, 2002). On the other hand, both the academic world and practitioners are distancing themselves from mere quantitative approaches. After all, formal facts and figures mean little without an underlying insight into the complex 'world views', interests and preferences of the parties involved (compare Weiss, 2001). Moreover, not everything can be measured. At the same time, however, highly quantitative 'goal-free' reconstructions of how a policy programme has been implemented and how this may be assessed from the perspectives of a wide range of interested parties, are increasingly subject to criticism as well (e.g. Shaw, 1999). The attention increasingly shifts towards the seemingly pragmatic question of whether policy, given the policy objectives, is 'working' or not and how this may be explained using 'evidence' (Pawson and Tilley, 1997; Weiss, 2001).⁸

Although there are more ways to classify thoughts on policy and research, many authors characterize thoughts on evaluation as a debate between the two 'antagonistic schools' of positivism and constructivism (Swanborn, 1999; also see Pawson and Tilley, 1997; Patton, 2002). Some even claim jokily that 'the "grand debate" over inquiry paradigms is a metaphysical exercise which merely serves to distract evaluators from practical work' (Miles and Huberman, 1988: 63). The extensive comparison of both approaches has been done relatively often. Still, elements from both approaches are essential for the pragmatic or 'realist' approach to theory and evaluation I am trying to establish. Before drawing

conclusions, I will therefore briefly revisit the most important insights from the trenches.

The Positivist, Rational-analytical School

Traditional ‘positivists’ may be found at one end of the evaluation scale. On the basis of optimal methods and techniques, they believe that the goal of evaluation – or every input of knowledge in the policy domain, for that matter – consists of seeking ‘the truth’. By way of the production of ‘true knowledge’ regarding the efficiency and effectiveness of the policy pursued, evaluation research may contribute to the quality of the decision making with respect to policy. According to positivists, it must be possible to deal with social problems by means of the ‘right’ policy theory. After all, the ‘right theory’ ensures that the ‘right instruments’ are deployed in the ‘right manner’ thereby efficiently attaining the ‘desired objectives’. Proceeding on the assumption that efficiency and effectiveness of policy will increase if the policy theory is better, the policy theory must be harmoniously brought in line with the situation to be influenced. In this rational vision of policy, the evaluation has an instrumental function: evaluation research must lead to an improvement of the existing policy theory or the replacement by a better one. Often, this is based on a model whereby the policy theory of a central actor is dominant.

The Constructivist, Responsive School

At the other end of the scale, we find ‘constructivists’. According to constructivists there can be no universally applicable claims to ‘truth’ or ‘reality’. Evaluation forms part of the continuous processes driven by political and other interests that may at their best lead to some agreement on ‘images of realities’ (see Weick, 1979). Here, the function of evaluation is to contribute to the quality of such processes. It is thereby often not possible to reduce uncertainties: the ambiguity of meanings may even be valued positively (e.g. Abma, 1996).

The objections of more responsive approaches to the positivist vision may be divided into two types:

1. the normative approach is too simplistic; and
2. the normative approach suffers from a defect as it wrongly holds on to a central decision maker and ignores the importance of argumentation.

Simon with his notion of ‘bounded rationality’ has perhaps most convincingly formulated the first objection. As early as 1957 Simon stated that there is a gap between the cognitive capacities of policy makers on the one hand, and the complexity of policy problems on the other hand: ‘The capacity of the human mind for formulating and solving complex problems is very small compared with the size of problems whose solution is required for objectively rational behaviour in the real world’ (Simon, 1957).⁹

The second objection of constructivists concerns the fact that the monocentric approach of policy and direction does not do justice to the complex, multi-form nature of society (In ‘t Veld et al., 1991). Policy development occurs – or should occur – in social networks within which stakeholders must reach acceptable

solutions by means of negotiation and argumentation (see Sabatier and Jenkins-Smith, 1993; Majone, 1989). Given that opinions and solutions differ, there can never be a single ‘best solution’ for a certain policy problem. According to constructivists evaluation may lead to learning processes, but they explicitly distance themselves from the ideal that evaluation may contribute to monocentrically, government defined, universally applicable policy theories: the use of causal assumptions and the valuation of objectives and effects of policy cannot be monopolized (Schwandt, 2001). In my opinion, this is not only a realistic but also a normatively superior vision.¹⁰

Towards Reconciliation: A Theory-based Approach That Is Social, Interactive and of a Dialogical Nature

As indicated, the dispute between positivists and constructionists has antagonistic traits. In my view, the latter occasionally seem to close themselves off from argumentation – under the guise of alleged superiority, but difficult to understand from their own point of view. This is illustrated well by the title of the influential and rich ‘Fourth Generation Evaluation’ by Guba and Lincoln (1989). Closer inspection shows that the first three generations of evaluation described in the book are fairly simplistic caricatures of older rationalist models. Opposite counting, describing and assessing, they state the responsive approach. They do not shy away from radical positions, which lead to equally strong responses. For example, Pawson and Tilley provocatively write: ‘When push comes to shove we have to admit an inclination to dismiss constructivist/naturalistic/fourth-generation evaluation as mere casuistry’ (1997: 20).

Of course: it may also be done differently. For example, Hoppe once stated that rationality must also include being open to learn, dialogue and doubts and that the – in his definition – post-positivist approach must continue to provide room for the technical verification of assumptions (Fischer, 1980; Hoppe, 1998: 38). Schwandt rightly states:

articulating schemes offers a kind of pragmatic purchase on how we construe the world, discuss our differences, act on some sense of what it is best to do, and evaluate the consequences of our action. (Schwandt, personal communication)¹¹

Patton has argued that both approaches must above all be dealt with pragmatically and that, where possible, one must use a context-specific combination of – methods of – both schools (1996, 2000). As said, Teisman et al. have recently developed a balanced evaluation arrangement to learn in the field of spatial planning that does just that (Teisman et al., 2002). Also in practice, increasingly more hybrid forms of evaluation research can be recognized (Ministry of Finance, 2002). Quantitative information on the attainment of goals is linked to qualitative research into the motives and opinions of the policy target groups and stakeholders. The general picture that the approaches are mutually exclusive must therefore be corrected.

There is nothing new under the sun. These wise albeit disputable words of Ecclesiastes apply particularly well to the view that in most cases most people are more or less *rational* beings. We have a certain intellectual capacity and are

fixated on the development of all kinds of ideas and theories on how to control, change or – increasingly more, fortunately – to preserve the world. From this perspective, the maxim that nothing is as practical as a good theory has traditionally applied forcefully to both policy making and evaluation. Especially in an age where result-based management and accounting for one's actions are at the top of the political list of morals and customs, the policy theory provides a useful and serviceable medium for both the development and the systematic assessment of policy. The fact that in the Netherlands the state budget and policy evaluation programming on a national level are now centred around 150 policy objectives, may illustrate this (see Van der Knaap, 2000).

The above, of course, can only be entirely true in a world of stainless steel, paper reality and honest politicians. Terms such as 'order', 'truth' and 'knowledge' have a relative meaning and can not be defined unilaterally. For evaluation, a mere departure from unilaterally established, instrumentalist policy theories does not suffice. In looking for the answers to the question of: 'Why does a policy programme work, where and for whom?', policy makers as well as evaluators should be aware of:

1. the relativity of knowledge and truth;
2. the restricting effect of policy theories; and
3. the value of thorough qualitative evaluation research.¹²

Any true constructivist approach must *incorporate* attempts to interpret or evaluate phenomena, including both 'naïve' and 'enlightened' positivist approaches to evaluation (Van der Knaap, 1997).¹³ Perhaps Patton is right when he says the evaluation community should not make principal choices: 'The point is to be pro-meaningfulness . . . the issue is how to combine the strength of each approach' (Patton, 2002: 573–4). In my opinion, the benefits and limitations of the theory-based approach demonstrate foremost that it is time for a broader concept of rationality. Rationality no longer merely depends on the cognitive capacity to develop deductive-theoretical systems of laws – whether or not supported by a strong empirical policy analysis *ex ante* – the capacity to substantiate these presumed laws and to thereby learn from others is decisive (see Hoppe, 1998). A person who develops, debates or evaluates public policy in a 'rational manner' in 2003 must be open to doubts, dialogue and learning. Or, in the words of Habermas: 'cognitive-analytical rationality can only exist on the basis of rationality that is primarily focused on understanding the other; it must therefore be social, interactive and of a dialogical nature' (Hoppe, 1998: 13).

It may often be rewarding from a constructivist point of view to depart from a – not necessarily 'the central' – rational argumentation of policy interventions. By acknowledging the simplifying and focused effect of a democratically legitimized policy theory within processes of policy argumentation beforehand, one may subsequently examine its specific merits and limitations. The manner in which a (policy) theory usefully orders complexity is thereby the most important criterion. This has two dimensions. Firstly, it is impossible to incorporate all exceptional situations, all contradictory expectations, each different interpretation in a policy theory. Secondly, even the best evaluator will fail to do justice

to this diversity.¹⁴ Both sides of the policy cycle have insufficient cognitive capacities to fully comprehend the discrepancy between ‘the real thing and its representation’ (Schwandt, 2002: 19). Just as policy theories will never be perfect, the best evaluation research will at no time be able to lead to final, true assessments: ‘We are self-interpreting, meaning-making beings, and the task of interpreting the value of our activities and actions is always contingent, complex, contested and never finished’ (Schwandt, 2002: 14).

Closing Remarks

Next to ‘theory’, ‘learning’ is probably the most constructivist notion there is. Still, if someone refers to ‘learning’ on the basis of evaluation and links this to ‘looking for better policies’, she partly draws on the remnants of the positive–positivist worldview of modern political theory (see Hoppe, 1998: 39). The role of the government may not be without obligations in this respect. It concerns the management of the many, considerable problems which a society faces: crime, violence, poverty, pollution of the environment etc. In my view, policy makers and evaluators alike have the moral duty to pursue intelligent policy: policy that focuses on influencing social development in a well-considered and well-balanced manner in the direction ‘desired’ by the majority of voters (Van der Knaap, 1997). Evaluation must thereby seek to let ‘government’ be the master of its policy measures: not from a perspective of political power, but rather from the realization that political responsibility cannot exist without acknowledgment of the diverging views and mutual dependencies.

A well-balanced theory-based approach to policy and evaluation, i.e. one that is true to focal objectives and programmes, provides but also draws attention to the context-specific diversity of stakeholders’ interests and preferences. Wildavsky has formulated the rationale for such a constructivist approach to policy-oriented learning most clearly:

In an aspiring democracy, the truth we speak is partial. There is always more than one version of the truth. . . . This is not only democracy’s truth, it is also democracy’s dogma. (Wildavsky, 1987: 404)

In a world where policy decisions have enormous impact on the planet and those who live on it, there is a constant need for doubt, dialogue and learning. Herein, policy makers and evaluators must always be modest but never afraid to speak out. Paradoxically, the ‘formal’ starting points and objectives of public policy may provide a fruitful basis to define the limits of doubtful modesty and strong recommendation. It – literally – provides a framework of reference to zoom into or to derogate from; to hold on to or to let go.

Notes

The author would like to thank Tineke Abma, Rob Hoppe, Arno Korsten, Frans Leeuw, Frans-Bauke van der Meer, Thomas Schwandt and Elliot Stern for their kind comments on an earlier version of this article.

1. Einstein once expressed it as follows: 'Logic takes one from A to B, imagination takes one everywhere'.
2. This statement derives from Kurt Lewin.
3. In literature, the term single-loop learning usually refers to this type of learning: 'When the error detected and corrected permits the organization to carry on its present objectives, then that error-detection-and-correction process is single-loop learning . . . It is primarily concerned with effectiveness' (Argyris and Schön, 1978: 3, 21).
4. This also applies in a management model whereby a policy theory is not merely of importance at 'start and finish' but whereby adjustments are systematically made on the basis of interim monitoring information concerning the deployment of instruments, performances obtained or objectives attained. After all, where an evaluation is used as a supplement it will be based on interim information, but it will in particular focus on the actual contribution of those instruments or performances to attaining the objectives.
5. To frame means both to structure and to trick someone into something.
6. Wildavsky describes this aptly: 'Public policy remains a world we never made, consciously or entirely. Policies, acting as their own causes, drive as well as being driven. Like ideas or theories, policies, once promulgated, exist independently of their origins' (1987: 404).
7. Naturally, the distinction between the 'espoused theory' and the 'theory-in-use' made by Festinger (1957), later applied by Argyris and Schön, is relevant in this context.
8. The description 'theory-based' already indicates that a certain actor perspective is not dominant. Naturally, reasoning from actors, what works as a policy programme for policy maker *A* or interested party *B* does not have to work from the point of view of implementer *C*. Theory-based evaluation revolves around the objectives of the policy itself, whereby – both from the point of view of the level of support and democracy – it is desired that the policy is formulated, wherever possible, in consultation with the interested parties.
9. This opinion may be illustrated by means of a practical policy issue: must a municipality introduce one-way traffic for one particular street? It seems impossible for the policy makers to know all effects in advance. For example, how exactly will it affect the traffic circulation in the adjoining streets? What will be the effect on the turnover of shops? Lindblom and Woodhouse (1993: 5) conclude that the contribution of the rational analysis is always bounded, 'Even the best analytic techniques . . . could answer these questions only with inconclusive estimates that might turn out to be misleading'.
10. However, it must thereby be noted that taking account of the complexity, diversity and reflective capacity of society and *therefore* – from an instrumentalist point of view – the importance of dialogue must be considered as purely 'rational'.
11. Schwandt continued: 'The irony here is that this way of thinking is of course quite constructivist in orientation – a sin, I fear that many of those who endorse a theory-based view of evaluation would be loath to admit!'.
12. Swanborn (1999) once expressed the difference between quantitative research and qualitative evaluation research as 'counting coffee beans' versus 'tasting the coffee'. Every restaurateur knows that 'how coffee tastes' apart from the number and quality of the beans is also determined by several – inherently subjective – factors.
13. The view that 'radical positivism' (the only true reality exists; evaluation pursues the objective of finding this by means of scientific methods) also fits in a constructivist view as a *strategy* may be defended. My perception is that most supporters of a rational-analytical approach are essentially in favour of some form of 'enlightened positivism': the acknowledgment that 'the only true reality' does not exist is

widespread. At the same time, they hold on to the ideal of a methodological optimally sound (re) construction of the ‘true picture’ regarding a certain policy programme. 14. Pawson (2002) refers to this as ‘the greatest bugbear of evaluation’.

References

- Abma, T. A. (1996) *Responsief evalueren – Discoursen, controversen en allianties in het postmoderne* [Responsive Evaluation – Discussions, Controversies and Alliances in the Postmodern]. Delft: Eburon.
- Argyris, C. (1990) *Overcoming Organizational Defenses, Facilitating Organizational Learning*. Needham Heights, MA: Allyn and Bacon.
- Argyris, C. and D. A. Schön (1978) *Organizational Learning: A Theory for Action Perspective*. Reading, MA: Addison-Wesley.
- Festinger, L. (1957) *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
- Fischer, F. (1980) *Politics, Values and Public Policy: The Problem of Methodology*. Boulder, CO: Westview Press.
- Guba, E. G. (1990) *The Paradigm Dialogue*. London: Sage.
- Guba, E. G. and Y. S. Lincoln (1989) *Fourth Generation Evaluation*. Newbury Park, CA: Sage.
- Hoogerwerf, A. (1984) ‘Beleid berust op veronderstellingen: De beleidstheorie’ [‘Policy is Based on Assumptions: The Policy Theory’], *Acta Politica* 14(4): 493–531.
- Hoppe, R. (1998) ‘De broosheid van debat en argumentatieve beleidsanalyse’ [‘The Fragility of Debate and Argumentative Policy Analysis’], inaugural address delivered upon his appointment as professor of policy science. Twente University, Enschede.
- In ‘t, Veld, R. J., L. Schaap, C. J. A. M. Termeer and M. J. W. van Twist, eds (1991) *Autopoiesis and Configuration Theory: New Approaches to Societal Steering*. Dordrecht: Kluwer Academic.
- Lam, T. C. M. (2002) ‘Theory-Based Evaluation and Objective-Based Evaluation: An Integration of the Two Approaches’, paper presented at the European Evaluation Society conference, Seville, Spain, 12 October.
- Leeuw, F. L. (1983) *Overheid en bevolkinggroei: een evaluatie van beleidstheorieën* [Government and Population Growth: An Evaluation of Policy Theories]. Leyden: University of Leyden.
- Leeuw, F. L. (2002) ‘Evaluating and the Ranking of Higher Education Studies’, in J. A. A. M. Bevers and M. J. F. Hulshof (eds) *Willem en wetens*. Nijmegen: IOWO.
- Lindblom, C. E. and E. J. Woodhouse (1993) *The Policy-Making Process*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.
- Mabry, L. (2002) ‘Postmodern Evaluation – or Not?’, *American Journal of Evaluation* 23(2): 141–58.
- Majone, G. (1989) *Evidence, Argument and Persuasion in the Policy Process*. New Haven, CT: Yale University Press.
- March, J. G. and J. P. Olsen (1989) *Rediscovering Institutions: The Organizational Basis of Politics*. New York: Free Press.
- Miles, M. B. and A. M. Huberman (1988) ‘Drawing Valid Meaning from Qualitative Data: Toward a Shared Craft’, in D. M. Fetterman (ed.) *Qualitative Approaches to Evaluation in Education: The Silent Scientific Revolution*. New York: Praeger.
- Ministry of Finance (2001) *Regeling prestatiegegevens en evaluatieonderzoek Rijksoverheid* [Central Government Performance Data and Evaluation Research Arrangements]. The Hague: Ministry of Finance.

- Ministry of Finance (2002) *Evaluatieoverzicht Rijksoverheid 2003* [Central Government Evaluation Overview 2003]. The Hague: Ministry of Finance.
- Patton, M. Q. (1996) 'A World Larger than Formative and Summative', *Evaluation Practice* 17(2): 131–44.
- Patton, M. Q. (2000) *Utilization-focused Evaluation*. Thousand Oaks, CA: Sage.
- Patton, M. Q. (2002) *Qualitative Research and Evaluation Methods*. Thousand Oaks, CA: Sage.
- Pawson, R. (2002) 'Nothing as Practical as a Good Theory', plenary address presented at the European Evaluation Society conference, Seville, Spain: 12 October.
- Pawson, R. and N. Tilley (1997) *Realist Evaluation*. London: Sage.
- Popper, K. R. (1972) *Objective Knowledge*. Oxford: Oxford University Press.
- Rein, M. and D. A. Schön (1993) 'Reframing Policy Discourse', in F. Fisher and J. Forester (eds) *The Argumentative Turn in Policy Analysis and Planning*. London: UCL Press.
- Sabatier, P. A. and H. C. Jenkins-Smith (1993) *Policy Change and Learning – An Advocacy Coalition Approach*. Boulder, CO: Westview Press.
- Schön, D. A. (1982) *The Reflective Practitioner*. Cambridge, MA: Basic Books.
- Schwandt, T. A. (2001) *Evaluation Practice Reconsidered*. New York: Peter Lang.
- Schwandt, T. A. (2002) 'Democratization, Globalization and Performativity: Implications for Learning in and from Evaluation', keynote address from the European Evaluation Society conference, Seville, Spain, 12 October.
- Shaw, I. F. (1999) *Qualitative Evaluation*. London: Sage.
- Simon, H. A. (1957) *Models of Man*. New York: Wiley.
- Swanborn, P. G. (1999) *Evalueren* [Evaluating]. Amsterdam: Boom.
- Taylor, S. E. and J. Crocker (1981) 'Schematic Bases of Social Information Processing', in E. T. Higgins, C. P. Herman and M. P. Zanna (eds) *Social Cognition*. Hillsdale, NJ: Lawrence Erlbaum.
- Teisman, G. R., F. B. van der Meer, E. H. Klijn, J. Edelenbos, H. L. Klaassen and M. A. Reudink (2002) *Evalueren om te leren – naar een evaluatiearrangement voor de Vijfde Nota RO* [Evaluation to Learn – Towards an Evaluation Arrangement for the Fifth Policy Document on Spatial Planning]. Rotterdam: Erasmus University.
- UK Treasury (2002) *2002 Spending Review – Public Service Agreements 2003–2006*. London: HM Treasury.
- Van der Knaap, P. (1995) 'Policy Evaluation and Learning – Feedback, Enlightenment or Argumentation', *Evaluation* 1(2): 193–220.
- Van der Knaap, P. (1997) *Lerende overheid, intelligent beleid – De lessen van beleidsevaluatie en beleidsadvisering voor de structuurfondsen van de Europese Unie* [Learning Government, Intelligent Policy – The Lessons of Evaluation and Policy Advice for the Structural Funds]. The Hague: Phaedrus.
- Van der Knaap, P. (2000) 'Performance Management and Policy Evaluation in the Netherlands: Towards an Integrated Approach', *Evaluation* 6(3): 335–50.
- Verweij, A. O. (1995) *Een kwestie van organiseren, De invoering van een etnische registratie in arbeidsorganisaties* [A Question of Organizing, The Introduction of an Ethnic Registration in Labour Organizations]. Rotterdam: Institute for Socio-Economic Research.
- Weick, K. E. (1979) 'Cognitive Processes in Organizations', *Research in Organizational Behavior* 1: 41–74.
- Weiss, C. H. (1990) 'Evaluation Research for Decisions: Is Anybody There? Does Anybody Care?', in M. C. Alkin (ed.) *Debates on Evaluation*, pp. 171–84. Newbury Park, CA: Sage. Reprinted from *Evaluation Practice* 9(1).
- Weiss, C. H. (2001) 'What Kind of Evidence in Evidence-based Policy?', paper presented

Evaluation 10(1)

at the Third International Evidence-Based Policies and Indicator Systems conference, Durham, UK, 4–7 July. Available at <http://www.cemcentre.org/eb2003/> (site visited 05/12/03).

Wildavsky, A. (1987) *Speaking Truth to Power – The Art and Craft of Policy Analysis*. New Brunswick, NJ: Transaction Books.

PETER VAN DER KNAAP is director of research directorate II of the Netherlands Court of Audit. This article was written when he was head of the Policy Evaluation Department of the Netherlands Ministry of Finance's Budget Affairs Directorate. Prior to this he worked at the Erasmus University of Rotterdam, where he investigated the relation between policy evaluation and policy-oriented learning in the domain of EU structural policy. Please address correspondence to: Netherlands Court of Audit, PO Box 20015, NL-2500 EA The Hague, The Netherlands. [email: peter.vanderknaap@rekenkamer.nl and peter.vanderknaap@wanadoo.nl]