Citizens' Control of Evaluations

Formulating and Assessing Alternatives

RICHARD MURRAY¹

Swedish Agency for Public Management, Sweden

Democratic and participatory evaluation raises questions of power. Power lies not only in agenda setting and problem definition but also in formulating alternatives. The latter seems to have been forgotten in the literature on democratic evaluation. This may be partly due to the general neglect of assessing alternatives in evaluations, whether *ex ante* or *ex post*. The article distinguishes different forms of democracy and specifies a model of evaluation that is intended to fit within representative democracy. The model is exemplified by an evaluation of a large infrastructure programme in Stockholm, in which environmental organizations took part.

KEYWORDS: alternatives; democratic evaluation; power; representative democracy

Introduction

If government is to include citizens, so also must evaluations. But democracy and evaluation may take very different forms; and the particular form will determine the balance of power between citizens, bureaucrats and politicians. That the issue of 'bringing evaluation to the people' is a question of power is highlighted by the time-honoured political science topic of agenda setting. That agenda setting and generating alternative courses of action invoke power is not surprising but is nevertheless almost completely neglected in evaluation literature.²

This article describes an attempt to include citizens in the process of traffic planning in greater Stockholm. For some time, the politicians had been talking in general terms of including citizens in the planning process; and for some time, citizens had wanted to be involved in the process. Eventually, politicians admitted one group of citizens – environmental organizations – into the process. The story is of interest because, unusually, citizens exercised considerable control over evaluations that were at the heart of the planning process.

Beginning with the theory of evaluation as a democratic process (House and Howe, 1999), the article focuses on the relationship between the concept of democracy and the way citizens participate in evaluation. By way of introduction

there is a brief discussion of the literature on democratic or participatory evaluation.

The article also considers various ways to integrate an evaluation process, which is controlled to some extent by citizens, with a political decision process. Different conceptions of democracy differ as to how difficult this integration is to achieve. The aim of this article is not to dismiss representative democracy but rather to make it work from a citizens' viewpoint. If representative democracy aims at citizens' participation it will have much to gain from participatory evaluations, properly designed.

Two main points will be made. A main line of argument concerns the importance and implications of considering alternatives in evaluation. First I would argue that any evaluation has to be based on alternatives, i.e. the action or programme evaluated has to be compared with some other course of action. An evaluation aims to find out whether a specific action was (ex post) or will be (ex ante) a good or a bad thing. This requires finding out what the consequences are of the action undertaken or proposed, i.e. causal relationships have to be verified. If effects are searched for, more often than not the effect caused by the action is assessed by comparing it with what would have happened without any action at all, i.e. the *counter-factual*. This is seldom reasonable. If roads are overcrowded, the impact of building a new road should be compared with what would have happened if the road was not built – which is not the same thing as doing nothing. Without a new road most people would have started travelling on buses instead and, probably, more buses would have been supplied which constitutes an alternative course of action. So, the reduction in congestion resulting from a new road would be overstated if compared with no action at all. And even more so, if compared with a concerted effort to improve public transport.

In order to assess whether an action has positive or negative consequences, it needs to be compared with the consequences of some other action. If compared with no action, for example, the building of the road might look like the best solution, but if compared with providing more buses, it might prove to be an inferior solution. Too often, evaluations seek to make absolute judgements: that the action was good or bad, successful or unsuccessful.³ Yet often evaluations can only lead to relative judgements: better or worse, rather than good or bad.

Second, in order for citizens to have a say in evaluations, they must be given the chance to formulate alternatives. Formulating alternative courses of action is very important in the quest for power. By only formulating very similar alternatives for citizens to choose from, politicians and bureaucrats (henceforth: decision makers) are able to exercise a large measure of control over decisions. This may also be achieved by formulating two very different alternatives, if one of the alternatives is obviously implausible. Surprisingly little attention is paid to alternatives in ex post evaluations. In ex ante evaluations it is, for obvious reasons, natural and common to include alternatives (as in Floc'hlay and Plottu, 1998). However, it ought to be just as natural to include alternatives in ex post evaluations. Any assessment of past performance must be measured against an alternative – could anything better have been done?

These are strong statements, and the article does not claim to comprehensively

review existing theoretical and practitioner literatures. The underpinning of these statements comes from many years of reading evaluation reports; from a review of some of the most frequently referenced standard books on evaluation; and from some recent articles in prominent evaluation journals.⁴

Participation of Citizens in Evaluation: Why and How?

Why should stakeholders be included in evaluations? Greene (2000) gives the following three answers (similar to Gregory, 2000).

1. Pragmatic aspect – including stakeholders will increase utilization of evaluation findings and enhance organizational learning and development.
2. Emancipatory reasons – it aims to give a voice to and empower disenfranchised groups and to promote social change based on equity and justice.
3. Deliberative aims – in a democratic society evaluations should be based on democratic values, include all stakeholders in a fair and equitable way and promote dialogue.

Including stakeholders in evaluations can be done in many ways and be called many things. Democratic or participatory evaluation are common catch-all names. These kinds of evaluations differ from conventional evaluations in several respects. Rebien (1996) draws the line in this way:

The purpose of participatory evaluation is to empower people, who should therefore help design the evaluation and the indicators for measuring project success and should be engaged in self-evaluation activities. In contrast, the purpose of conventional evaluation is to provide accountability, in order to determine whether funding should continue. In conventional evaluation, the project is evaluated according to predetermined success indicators from the project document and the evaluation is conducted *impartially* by *independent* evaluators. (1996: 152)

Further, Rebien states that 'there is no single accepted version of participatory evaluation, which is not surprising considering that there is quite some confusion surrounding the participation concept itself' (1996: 152). Participation may vary from total control over the evaluation by participants to nearly total manipulation by evaluators or programme administrators.⁵

The notion of participatory or democratic evaluation, judging from definitions often given in the literature, also seems to be rather blurred. Carol Weiss (1998), among others, uses the term generically to include: advocacy, empowerment, collaborative, stakeholder and client-orientated models of evaluation (See also Floc'hlay and Plottu, 1998; Rossi et al., 1999).

These models differ according to the number of groups who participate in the evaluation and the amount of control over the process that they have. At one extreme of the control continuum is an advocacy model in which the ultimate beneficiaries design and conduct the evaluation themselves, sometimes of their own projects. This may be done with little or no assistance from professional evaluators (Fetterman et al., 1996; Weiss, 1998). At the opposite end of the continuum, the stakeholder model reduces participation to keeping the evaluator informed

about stakeholders' opinions and preferences. In between we have the collaborative evaluation model, in which the stakeholders and the evaluator share equal control.⁶

Rebien (1996) posits the following criteria for evaluation to be truly participatory.

1. Important stakeholders must be represented.
2. Stakeholders must have an active role as subjects rather than being mere sources of data.
3. Stakeholders should participate in at least designing the terms of reference, interpreting the data and using evaluation information.

In this article, the third criterion will be shown to require further elaboration in order for citizens to have real control.

Democracy and Democratic Evaluation

The approach to public participation varies and has over time shifted its focus, from a policy maker's and programme manager's point of view to a citizens' perspective. At a more general level, this change can be traced, at least partly, to an increasing stress on direct democratic participation as a complement to 'traditional' participation through parliamentary channels (e.g. Floc'hlay and Plottu, 1998; Dahlberg and Vedung, 2001). This shift is consistent with the emerging view of society and knowledge as pluralistic, multi-valued and subjective. Weiss (1998: 100–1) testifies to this change: '[p]eople in different locations in the social system construe knowledge, truth, and relevance in markedly different ways, each of them legitimate and worthy'. From this perspective the evaluator places different beliefs and values on a par with each other when planning and conducting an evaluation.

Deliberative democratic evaluation may be identified with genuine democracy, that is, with what democracy requires when properly analysed and understood in its full context. (House and Howe, 1999: 97)

Democratic evaluation is coupled here with a specific understanding of democracy. In order to state a preference for a specific model of evaluation, one needs a set of values. In the context of democratic evaluation these values must be related to a preferred form of democracy. As Hanberger (2000: 14) points out: 'Evaluators need to make their own position on democracy explicit, because the assumptions will permeate the evaluation, particularly any recommendations'.

Representative democracy is the general framework for the kind of political decision making that is the focus of this article, i.e. the model to be developed for evaluations controlled by the citizens is not intended to replace representative democracy or change it into something different. But representative democracy can take many forms. The definition of representative democracy used here is admittedly a narrow one, but it captures the essential elements: *politicians elected in general democratic elections take or delegate decisions within a sphere by consensus considered to be their domain – the political domain.*

Only if politicians delegate to others – administrators or citizens – may these others take decisions. The decision-making prerogative remains squarely with elected representatives.

What then is the role of citizens besides electing representatives? This is where representative democracy may vary, from no other role than voting to broad participation in decision making. Without challenging the decision-making prerogative of the politicians, citizens may engage in public debate and criticize their elected representatives and their decisions. The politicians may – without giving up their prerogative – engage citizens in consultations, debate specific issues and even delegate decisions to them. They may, for example, delegate decisions on local issues to a neighbourhood council. Referenda are another way of delegating decisions, sometimes even on issues of national importance.

In this section the aim will be to define and distinguish different forms of democracy and different models of evaluation, whilst also relating them to each other. This poses major problems of definition for both evaluation and democracy. Thus, for evaluation, words such as 'participatory' and 'democratic' are used interchangeably and with varying connotations by different authors, alongside a host of other labels, such as 'empowerment', 'advocacy', 'stakeholder' and many more. Terms such as 'participatory' and 'deliberative' are both used to characterize various forms of both democracy and evaluation models, which adds to the confusion. In the face of this confusion, a new classification of democratic evaluation models was needed.

The earlier quotation, from House and Howe (1999) identifies only one (ideal) type of democracy that they define as 'genuine' democracy. In order to be able to discuss various forms of evaluation and how they are related to democracy we need to distinguish between various forms of representative democracy.

House and Howe offer three criteria to define 'genuine' democracy: inclusion, dialogue and deliberation. These criteria help mainly to disqualify and distinguish elitist versions of representative democracy from 'genuine' democracy. According to Dryzek (1996) democracy has three dimensions: franchise (which corresponds to inclusion but refers to voting), scope (i.e. the extent of the political domain) and authenticity (i.e. the degree of involvement by citizens, which corresponds to dialogue and deliberation).

Following Hanberger (2000) one may distinguish three main types of democracy: elitist, participatory and deliberative. *Elitist* democracy is synonymous with representative democracy, carried to its extreme, i.e. the role of citizens is limited to voting for representatives once every third or fourth year. Franchise may be wide, authenticity is low and scope may vary. The characteristic of *participatory* democracy is participation of citizens in policy formation. It aims to be inclusive and, if successful, increases the authenticity of democracy. The way in which citizens are included doesn't matter: success is measured only through members participating.

Deliberative democracy (or discourse theory of democracy) also aims to increase the participation of citizens in policy making but in a more active way than in participatory democracy. Its salient characteristic is 'a common mode of reasoning' – rational and impartial – to which all affected and their representatives will

Evaluation 8(1)

be committed (Hanberger, 2000: 12). Through such reasoning 'democracy revolves around the transformation rather than simply the aggregation of preferences' (Elster, 1998: 1). This type of democracy stresses the importance of authentic participation rather than simply counting the number of participants.

As pointed out above (and by Hanberger), there are many more varieties of democracy. One variety that deserves mentioning is *direct* democracy or self-government. It is characterized by the people involved making decisions. The difference between direct and deliberative democracy, as defined here, is that in direct democracy the citizens make the decision, not their representatives. The role of elected officials is only to mediate and facilitate this collective decision making. This mode is totally inclusive and authenticity depends on the amount of deliberation. Scope is delineated by that which elected representatives delegate and more often than not, is very limited compared with other forms of democracy.

Still another form may be referred to as *transformative* democracy. This form aims to achieve social change with means other than rational and impartial reasoning and voting. Elected representatives thereby mobilize a specific group of citizens and give them the instruments to implement societal change.

Since there are many varieties of democracy, there are at least as many varieties of democratic evaluation (Table 1).

The elitist notion of democracy may be supported by purely *technocratic* evaluation, conducted by experts (what Rebien [1996] calls 'conventional evaluations'). The evaluation problem and success criteria are defined by decision makers. Citizens are only sources of information.

As participatory forms of democracy strive to engage the citizens between elections, attitude surveys are not sufficient. Participation in the evaluation includes stakeholders setting criteria for assessments (*stakeholder* evaluation). There could be an effort to bring out the opinions of underprivileged and politically passive groups, by various techniques to stimulate participation and articulate needs and hopes. This is often referred to as *empowerment* evaluation.

Just as the deliberative form of democracy may take many different forms, so may *deliberative* evaluation. House and Howe (1999: 113) listed 10 criteria that, if adhered to, constitute deliberative democracy. Among these are:

1. the inclusion of all relevant or at least major (affected) parties;
2. a fair balance of power and power-balancing procedures;
3. sufficient time and involvement for participation; and
4. sufficient time and reflection in deliberation.

Table 1. Types of Democracy and of Evaluation

<i>Types of democracy</i>	<i>Types of evaluation</i>
Elitist	Technocratic
Participatory	Stakeholder/Empowerment
Deliberative	Deliberative
Direct	Deliberative
Transformative	Emancipatory/Advocating

Deliberation should include deliberation on conclusions, but may also include deliberation on other stages in the evaluation process, such as problem definition, specifying evaluation methods and writing reports. Deliberation on conclusions may identify new evaluation problems, possibly pointed out by an evaluator acting as a 'critical interrogator who tries to help the participants to see possibilities for development' (Karlsson, 1996: 414). To be truly deliberative discussions should include decision makers.

If an evaluation is to serve direct democracy it ought, ideally, to be deliberative. It should involve a deliberation process between different interest groups that is facilitated by the evaluator and officials. Citizens involved in the evaluation will have full control of the evaluation, draw conclusions themselves and take consequential decisions.

Finally, what is termed *emancipatory* evaluation aims, not only to give due voice to disenfranchized groups, but to promote social change based on equity and justice. Therefore, it might be adequate to include what is termed *advocacy* evaluation. In emancipatory evaluations, representation may be unbalanced. I take it to correspond to a form of democracy that permits groups of citizens to transcend the rules of the game and therefore I call it 'transformative'. The role of the citizens is not clear, since this form of evaluation permits the evaluator to take on a very active, interpretative role. It may also be one of complete involvement of (some) citizens from beginning to end of the evaluation.

Agenda Setting and Power

If citizen participation is to be more than cosmetic, participatory evaluation must be formulated as a question of power. Questions of power are embedded in procedural issues: Who is to participate? With what resources? These and other similar questions are dealt with in the literature on democratic evaluation.

Nevertheless, it seems that the power perspective is not the focus of most of the literature on democratic evaluation. For example, the following three issues in the theory of power and power relationships are almost completely absent from this literature.

- Agenda setting.
- Problem definition.
- Formulating alternatives.

That power lies in *agenda setting* has long been understood in political science since the publishing of Bachrach and Baratz's *Power and Poverty* in 1970. These authors criticized prevailing theory on power for focusing solely on decisions taken and neglecting non-decisions. Anyone who has been politically active has experienced the difficulties in raising an issue. Powerful forces, foremost the main actors of the big political parties, have the prerogative to decide what issues should be prioritized. This may be achieved just by raising a dispute – other issues will vanish when the media become involved.

Socio-economic forces that shape the political agenda have been studied at

length (e.g. Rochefort and Cobb, 1993) but in relation to evaluations little seems to be known.

What programmes are chosen for evaluation? And who decides? Usually decision makers themselves call for evaluations, though sometimes the political opposition may do so. In some countries groups of citizens, independent audit agencies and specialized evaluation agencies may undertake evaluations of their own choice. With general rules requiring that no government decisions be taken without prior evaluations – such a rule has been in force for some time in Australia – or that all reforms shall be evaluated after a specified period of time – which is now becoming common – the established centres of power lose control over the agenda in favour of the general public. In some areas, such as infrastructure and town planning, it is becoming more frequent to require *ex ante* evaluations, often referred to as environmental impact assessments. These are important steps in bringing evaluations to the people, by shifting the power over the agenda away from the establishment.

Once the problem, programme or action to be evaluated is chosen it becomes necessary to *define the evaluation* problem. How to describe the programme? Which is (or was) its intended purpose? How is (was) it supposed to work? The way in which questions like these are answered has a strong impact on the outcome of the evaluation.

In the field of policy analysis it has been observed that problem definition 'serves as the overture to policy making' (Weiss, 1989: 97). Problem definition in a sense 'determines how people think about the problems that are (and are not) on the public agenda', and can be understood roughly as 'a package of ideas that includes, at least implicitly, an account of the causes and consequences of undesirable circumstances and a theory about how to improve them' (Weiss, 1989: 97). Krogstrup (1997) advocates a model of evaluation that aims to clarify and challenge the theory behind the programme. This means that 'users define the evaluation questions' (p. 209).

As already noted, during the last two decades the pluralistic view of society has been increasingly stressed. According to this view there are many ways in which to look at social issues, and to define social problems. The definition of a problem is of crucial importance in shaping the evaluation, involving considerations such as what criteria should be included and what causal relationships should be tested.

Earlier arguments and quotations direct our attention to *formulating alternative courses of action*. This is the third neglected area in democratic evaluation. A long time ago one author coined the phrase: 'The definition of the alternatives is the supreme instrument of power' (Schattschneider, 1960: 66). This issue will be discussed further below.

A Model for Citizens' Control of Evaluations

Participation of Citizens and Representative Democracy

Evaluation is just a basis for decisions. Decisions crucially depend on information; therefore control of the supply of information upon which decisions are

based is strategic for control of subsequent decisions. Evaluation highlights this issue. Furthermore, decision makers and citizens face a similar problem of controlling evaluations. Unless decision makers specify what should be evaluated and what alternatives it should be compared with, somebody else – the administration, the evaluator or the citizens – will do so, and hence control the outcome.

We can posit five levels of citizen participation, in ascending order of control:

1. debating the results of an evaluation;
2. giving their opinions (through answering questionnaires etc.) as part of the evaluation;
3. influencing the way the evaluation is designed and implemented, i.e.:
 - i. dimensions to be observed and criteria of assessment;
 - ii. methods of data retrieval and treatment;
 - iii. the specification of alternatives;
 - iv. problem definition;
 - v. conclusions;
4. taking the ensuing decisions;
5. deciding what programme to evaluate.

All five levels of citizen participation are compatible with representative democracy. The fourth level requires a deliberate decision by the politicians to delegate decisions while the fifth level is the role of citizens to scrutinize their elected representatives.

Usually citizens will not take the decisions. Bridging the gap from conclusions to decisions is always a problem when input into decision making comes from others than the decision makers themselves.

The Model

The model described below is consistent with what has been described as a deliberative evaluation model. It is intended to give citizens control over the evaluation process, while still retaining the rational and impartial characteristics of deliberation, and staying within the limits of representative democracy.

The problems of how the citizens are represented and how they are selected to become part of the process is not dealt with here. Those issues are dealt with at length in the literature on democratic evaluation. Instead the focus here is on what has previously been neglected.

Setting aside the issue of deciding which problems to evaluate, the following outlines the prerequisites of a deliberative evaluation model.

Problem Definition Citizens and decision makers should take part in discussions to define the problem. There will probably be different and competing problem definitions of the problem and the likelihood of agreement increases if the aims are looked for at a higher level. For example, if instead of alleviating traffic jams the aim is to speed up transportation, this opens up a much wider common ground for finding solutions, and hence the likelihood of a common definition of the problem increases. Different definitions of the problem will be

Evaluation 8(1)

associated with different evaluation criteria, data requirements and methods of analysis. This, of course, will complicate the evaluation but is necessary if citizens are to exercise control over an evaluation.

Organizing the Evaluation Process A suitable organization of evaluation work includes:

- breaking down the overall problem into separate sub-problems, to be handled by sub-groups;
- setting a time limit for the exercise, long enough for serious evaluations to take place, yet tight enough to meet the needs of decision making;
- allocating sufficient resources for the evaluation work, e.g. professional evaluation skills, data acquisition, compensation for citizens' loss of income;
- agreeing rules for control over the resources for evaluation – what influence will the different parties have if controversy arises over the allocation of resources, choice of evaluation methods etc.?
- agreeing access to information to ensure complete openness;
- setting aims for the handling of evaluation results, e.g. whether reports should be written together or separately;
- specifying protocols for open communication with the general public.

The last point is vitally important to ensure citizens' control.

Specifying Evaluation Criteria

Criteria such as time saving, costs, environmental impact and social distribution must be specified. If agreement cannot be reached the principle should be to include all criteria put forward by all participants.

Defining and Elaborating Alternatives Citizens, as well as political parties and special interests groups, may have formulated alternatives to the implemented or proposed programme. These alternatives should be put on a par with the proposed programme even if they appear amateurish, hazy and illogical. They should be worked on until it is possible to judge the merits of such alternatives in comparison with the proposed or implemented programme.

Prioritizing Studies Participants – including citizens – will have to decide which among a multitude of potential studies to undertake. Only a few of these will fit within the time and resource constraints.

The Conduct of Studies Monitoring by all the participants must be assured. Research methods, data collection etc. need to be transparent and open to influence by citizens. Where there is conflict over research methods and data collection, evaluators are only able to call on their professional skills and knowledge.

Writing Reports Reports should be as professional as possible, consistent with the possibility of citizens' influence. If necessary, in this model of evaluations, different parties may produce their own (different) reports.

Conclusions If possible, joint conclusions should be formulated. Otherwise separate conclusions should be presented.

Deliberation During the process of writing reports and formulating conclusions deliberation should take place, influencing reports and conclusions as well as final recommendations.

Dissemination of Evaluation Results Evaluation reports should be publicly disseminated; the results and conclusions should be marketed in an appropriate way to all concerned. The citizens' part of the report should be treated on a par with that of decision makers and experts.

Decisions Decisions should be taken by the decision makers (politicians) unless delegated in some manner.

The Prehistory of the Stockholm Traffic Scheme

For many years traffic and traffic planning have been a source of political conflict in Stockholm – as in other cities, large and small. There is no clear-cut division of opinions along ideological and party lines. This has left an unusually large area for citizens to act independently: it has become an 'establishment vs the people' issue. In turn, this has prompted the parties to open up a dialogue, though very cautiously, with the citizens. The topic is very technical and therefore also suitable for technocrats.

During the 1950s and 1960s car traffic increased rapidly. A centrally located thoroughway was completed in the early 1960s destroying important parts of the medieval part of the town. With the aim of creating a downtown central business area with modern offices and parking houses, large parts of the city centre were demolished. In 1973–4, after citizen protests against the heavy increase of car traffic in residential areas, the Social Democrats, in power at the time, arranged a consultation forum on traffic regulation. Interest groups from all walks of life participated but its conclusions were never implemented because of changes in government in Stockholm in 1976.

In 1989 a local environmental party formed a coalition with the Social Democrats with a road-pricing scheme to be implemented in Stockholm. In 1990 these efforts were defeated by the national Social Democratic government and the coalition broke up. Instead local Social Democrats formed an agreement with Conservatives and Liberals for a massive investment programme in motorways and public transportation. Road pricing was to be used for financing the motorway investments instead of for controlling traffic.

Several small parties, the green movement and large sections of the citizens of Stockholm were opposed to the programme from the outset. Demonstrations, petitions and lobbying followed. In the election of 1994 the Social Democrats formed a coalition with the Greens and the 'Left' party. One of the conditions of the coalition was to arrange for a thorough evaluation of the programme to which

Evaluation 8(1)

the Social Democrats were still committed. The green movement was to take part in this evaluation.

An Experiment in Democratic Evaluation

The following account risks bias since the author took part in the evaluation as one of the representatives of the green movement. Nevertheless, given that the experience is of such general interest it is important to 'tell the story'.

The account that follows will test the model set out above. The evaluations are documented in a series of reports (Evaluation Report, 1995a, 1995b, 1995c, 1996).

Empowerment In the literature on democratic evaluation, training and organization of citizens are a major theme, usually under the heading 'empowerment evaluation' (e.g. Fetterman et al., 1996).

At the first meeting of the in-depth evaluation the green movement acted in a concerted manner. A special association had already been formed by 22 different organizations for the sole purpose of taking part in the evaluation. All green organizations who wanted to take part in the work were welcomed, which included all important green organizations. The association had had a couple of meetings where a common platform and rules of conduct had been agreed. The rules of conduct specified that the participating organizations should as far as possible seek unanimity but that no organization could be bound by majority decisions. This very loose organization was kept together by holding several separate meetings during the evaluation process. It helped that several of the people involved knew each other from earlier campaigns. During the years, these organizations had formed somewhat similar ideas of how traffic problems in the Stockholm area should be solved.

There was no problem of empowerment as long as we accept that participation was limited to environmental organizations. An effort was made during the process to bring together a much larger representation of interest groups, including business and motorists' organizations in Stockholm. However, the purpose of this wider consultation was not made clear and after only one meeting the idea was abandoned. During the whole process, local politicians showed only limited interest.

Problem Definition The major parties already had a definition of the problem: increase accessibility, improve the environment and promote economic development. Stated this generally the green organizations had no objections. Their objections concerned the means to reach those ends, specifically the massive and very expensive motorway investments.

Without much discussion the environmental issues were identified as the most important. This was because of an implicit hierarchy of values. Reduced congestion, rightly achieved, promotes both environmental and economic development. This entails a special definition of accessibility. Had accessibility been defined only in terms of the number of trips by car and their average speed it would not have been possible to agree on aims. But since accessibility was defined in terms of the average speed of travel and the number of trips, regardless of the means

of transportation, less environmental damage could be won by increasing accessibility through reducing the number of trips by car. By giving priority to business travel and transportation, economic development was promoted. Congestion became the focal problem and the classical issue of whether you can eliminate congestion with more roads or not became the core issue of the evaluation.

Not all green organizations were prepared to accept that congestion was a problem, because that might lead to the conclusion that some roads ought to be built.

Organizing the Evaluation Process At the first meeting a tough timetable was presented by the politicians. The green organizations did not object to this even though subsequently it turned out to be grossly optimistic. The green organizations instead proposed a way of organizing the work. Since neither the politicians nor the administration had thought about this the green organizations' proposal was accepted.

They proposed a breakdown of the problem. First, there was the decomposition of the programme into specific projects, the most urgent being the North Link, a 2.5-km, 4–6-lanes motorway tunnel. Hence, the evaluation should start with that project. At the same time the proposed scheme for road pricing was to be evaluated. It constituted an overall approach to solving traffic problems and ought therefore to be considered at an early stage. Evaluating the South Link was to be undertaken later, since it was further down the timetable of planned implementation. These projects were chosen because they were seen as detrimental to the environment: the two motorways because increased capacity for car traffic would cause increased pollution, noise and accidents in the whole region of Stockholm plus intrusion into natural preserves; the road-pricing scheme because its construction was aimed at generating funds, not at controlling traffic, and would direct even more traffic through the centre of the city. In addition a public transportation project – a light-rail scheme – was chosen for evaluation because the green organizations thought it would be cost-ineffective.

The North Link was further decomposed into one study of the motorway as a whole and another study of its design.

In this way work process was agreed, which economized on the number of participants and the capacity of the administration, and still met with the timetable set by the politicians. In the event it also proved to be creative since different conclusions derived from the various evaluations.

Specifying Evaluation Criteria

Out of the problem definition, a set of criteria emerged. Noise, pollution and accidents are major problems in the inner parts of Stockholm. Reduction of these problems coincided with reduction of congestion in the inner city. Reduction of car traffic volumes in the inner parts of Stockholm therefore became the overriding criterion. However, the impact area included greater Stockholm. Consequently, pollution – mainly carbon dioxide and nitrogen oxides – as well as traffic accidents in the whole Stockholm area also qualified as criteria. These variables

Evaluation 8(1)

are proportional to the volume of traffic. Average speeds of travel by car and public transportation were another self-evident criterion.

Originally cost-effectiveness was not a criterion, but the green organizations wanted to include it. This would make comparisons possible between projects of very different sizes and additionally was relevant since several projects competed within a total investment budget of \$4 billion. Eventually cost-benefit calculations were performed for some of the projects.

Defining Alternatives

There had been ex ante evaluations earlier, commissioned by the three parties behind the programme, as a response to mounting criticism of what was proposed. These evaluations had calculated the consequences, specifically the amount of traffic in the centre of Stockholm and in the Stockholm region as a whole. They showed that the situation would be better with the programme than without the programme.

Due to the vociferous critiques of the programme an ambitious cost-benefit analysis of the whole programme had been undertaken. It also showed a surplus of social and economic benefits over costs when compared with doing nothing and having congestion piling up.

So what use was there of yet another evaluation? This is the crucial point. As long as it was accepted that the comparison should be between a situation with the programme and one without the programme, the previous evaluations were more or less flawless, in terms of e.g. methodology, data and criteria.⁷ Most evaluations are based upon this type of comparison, even though there is a difference between ex post and ex ante evaluations. In ex ante evaluation it is more obvious that there can be other possible programmes than the one proposed. But in ex post evaluation this is just as true: there have always been other possible programmes that could have been implemented other than the one undertaken.

The green organizations were not satisfied with the previous evaluations, because they had a vision of a completely different programme than the one proposed. Therefore, they asked for evaluations of their suggested actions to be compared with the proposed programme. 'Doing nothing' would clearly be unreasonable. The city would be gridlocked in 20 years, with low accessibility, hampered economic growth and a damaged environment.

The green organizations and some political parties had, for several years, advocated solving the traffic problems by the massive development of public transportation. In addition to this rather conventional, 'green' way of tackling traffic problems in cities, some parties and organizations had for some time advocated road pricing.

In order for evaluations to be based on specific, technically and economically feasible comparisons other than 'doing nothing', alternatives had to be defined. This was a crucial moment in the evaluation process. Could the green organizations define specific and viable alternatives? One of the green organizations had been investigating road-pricing schemes with technical and economic advice and was able to come up with an alternative scheme to the one proposed.

But for the alternative – a massive development of public transportation – the green organizations had to rely on technicians of the city of Stockholm.

The most immediate concern was the evaluation of the North Link. It was proposed as part of a ring road around inner Stockholm but it could also be seen in isolation as a solution to reducing traffic on a road through a residential area. What alternatives could there be?

The alternatives suggested and then worked out in collaboration between the green organizations and the administration's technical experts were as follows:

1. *Various scenarios for a reduction of car-traffic volume in the Stockholm region as a whole.* These scenarios were based on pessimistic predictions of economic development (the studies took place following the economic depression of 1992–5) and rising fuel prices. One scenario was the reduction of carbon dioxide by 80 percent, mentioned as a long-range political goal. The scenario was worked out as a combination of doubling fuel prices, doubling proposed road prices and a very rapid technological development, reducing emissions from cars. The reduction of traffic in the region was calculated to be only 15 percent by the year 2020 compared with 1993, despite these draconian measures.
2. *Alternative designs of the proposed motorway avoiding damage to parks and townscape.* This included doing away with some entrances and exits, limiting the capacity, trying different routes.
3. *Various alternatives to solve the ventilation problem of the 2.5-km tunnel.* The proposed scheme was to build smoke stacks along the road. The emissions from these chimneys and their visual impact were considered a problem.

Prioritizing Studies A timetable was constructed, specialists in various fields were called upon, and resources were allocated for modelling the impact on traffic volumes of various actions. The green organizations, very much in collaboration with the technicians, set up a series of studies that were to be undertaken during the time allotted. The consent of the politicians was needed with regard to time and money, and during discussions to gain consent the two coalition parties – the Green party and the Left party – played a very important role. The green organizations continued to be frustrated over what they saw as inadequacies of the studies.

The Conduct of Studies Various specialists performed the studies: evaluation of the health hazards of pollution levels in tunnels by medical experts; the spreading of pollution from the smoke stacks by meteorologists; and so on. The studies were monitored by working groups of representatives of the green organizations and the political parties. A precondition for an independent evaluation under these circumstances was that technicians wanted to do a fair job. Whilst the green organizations were at the complete mercy of these professionals, in the event very little criticism was directed at their contribution. On the contrary, many of these professionals showed great interest in the issues investigated and contributed creatively.

Evaluation 8(1)

The first parts of the evaluations took nine months to complete and included the North Link and the road-pricing scheme. For the participants it meant continuous work during the summer of 1995, with no compensation for the representatives of the Greens.

The most substantial and critical study concerned the effects of various actions on traffic volume. Two competing computerized models were available. Both have been developed over many years. They are models of trip generation, based on the places of residence and of work of the population, their incomes, ownership of cars, prices of gasoline, public transportation and of road capacity. This included all personal trips during the rush hour using cars and public transportation.

The models have been estimated and calibrated in order to explain past patterns of trips and traffic volume based on previous ex post evaluations. These ex post evaluations have assessed the impact of lowering the price of public transportation, raising the price of parking, increasing road capacity etc.

Using these models was an expensive part of the evaluation. Each alternative simulation cost \$500; a great many alternatives were assessed.

Writing Reports A report was to be written, relating the outcome of the various evaluations as objectively as possible. Various conclusions could be drawn from the results, but only those conclusions on which both the green organizations and the administrators could agree were included in the report.

In the first round of evaluations – the North Link – a high-ranking administrator was responsible for the final report and drafts were circulated to the green representatives for comments. As a result the final report had all the authority, objectivity and legitimacy of a regular report from the administration.

The politicians did not take part in writing the report. Therefore, they had no stake in it, and could not be said to share any of its conclusions. This was unfortunate and violates House and Howe's requirements for deliberation.

In the next round of evaluations – the road-pricing scheme, the South Link and the light-rail project – the politicians did not permit administrators to make the same contribution to writing reports. Instead they had to be written by the green organizations themselves. As a result the reports were not so well produced and had less credibility. Arguably politicians had learnt that having the administrators take responsibility for the report gave it too much credibility (author's interpretation).

Conclusions Several non-controversial conclusions emerged from the first two evaluation reports – but listed below are the more controversial conclusions.

A first set of conclusions was difficult for all the three political parties who had agreed to the infrastructure programme that was being evaluated. For example:

- In one scenario, which aimed for an 80 percent reduction of carbon dioxide, it turned out that the North Link was not needed to achieve the necessary reduction of traffic through the residential area in order to meet environmental standards.

- By a combination of road pricing and the intensive development of public transportation the same result could be achieved – and at a much lower cost (socially and financially).
- Both conclusions required what were deemed to be draconian road and fuel prices. If such prices were ruled out the North Link was needed to achieve environmental standards in the residential area as well as in the inner city of Stockholm, but not necessarily the North Link as proposed. If various entries and exits were omitted, there would be a saving of \$50 million and a substantial reduction in damage to parks and nature reserves.
- The programme as a whole would not meet the targets set for the reduction of traffic and increased accessibility in inner Stockholm, but the road-pricing scheme proposed by the green organizations would.
- Adding more motorways in the Stockholm area had the effect of increasing traffic and worsening the environmental impacts.

A second set of conclusions were especially awkward for the green organizations to accept:

- An overall conclusion undermined the belief of many of the green organizations that the development of public transportation alone would solve the traffic and environmental problems.
- If the draconian alternatives would not be accepted by public opinion, the North Link would be needed to achieve acceptable environmental standards for the residential area.
- It was also found that a four-lane rather than a two-lane link would be needed.
- Some green organizations had suggested a different solution, that of eliminating heavy vehicles on the road through the residential area. This, however, proved to be insufficient.
- Other routes for the tunnel than through the park were not technically feasible.
- The smoke stacks could not be replaced by other forms of ventilation.

Deliberation

From these and several other conclusions a fruitful discussion – deliberation – could have started. However, at this stage the process was aborted. The requirements – mentioned above – set up by House and Howe for a genuine democratic evaluation were not met. The low level of involvement of politicians and policy makers in the entire process now had its predictable effect. Had politicians and policy makers been more involved in the process it is more likely that they would have developed an interest in continuing the process, through a dialogue on policy.

Dissemination of Evaluation Results

Reports were distributed within the City Hall and among green organizations but not given any wide publicity; media coverage was scant. Most citizens remained uninformed about the evaluations.

Decisions

The reports were fed into the regular political decision-making process, and decisions were then taken at a rapid speed, with the reports having very little impact other than a couple of entrances to the South Link being altered slightly in line with report recommendations.

Postscript

Democratic evaluation cannot take place in a meaningful way without the possibility of citizens forming and evaluating alternatives along with the proposed or implemented programme. This remains true whether we talk of *ex ante* or *ex post* evaluations. If citizens are to take part in a critical *ex post* evaluation of an implemented programme, it must be compared with other programmes that could have been undertaken at the time.

Finally I would argue that if the process of involvement had been more extensive, i.e. not confined to green organizations, the evaluation would have been more likely to have had greater impact. Thus involving tenants, parents, motorists and businesses may have been cumbersome; however, the likelihood would then have been much greater that policy makers would have found a course of action that had gathered a broad consensus. Instead, today conflicts are as great as ever.

Notes

1. Ola Lejdhamre has assisted with researching evaluation and political science literature and with valuable comments.
2. An exception is Floc'hlay and Plottu (1998). They point out that in democratic *ex ante* evaluation there should be time and process for politics, population and experts to structure the problem of choice. 'Thus new projects . . . can emerge while decisions are being made' (p. 265). The point of their article is to construct a process assisted by a methodology to compare alternatives. Gregory (2000) speaks in general terms about power, stating that 'only through an appreciation of power can the problem of participation be addressed' (p. 194).
3. This is often due to a methodology that compares outcomes with stated goals. Actions that don't lead to the goal are deemed 'no good' and those that do 'good'. However, even if the achievement in terms of goal fulfilment is less than 100 percent, but that is the best that can be done, it is of course a better action than any other.
4. Weiss (1998) hardly deals with the issue. Only in connection with a short part on evaluation for decision making (*ex ante*) does she talk of assessing several alternatives, but never as part of the method of evaluation. Comparisons are between before and after and between those who were beneficiaries of a programme and those who were not. Mohr (1992) reminds the evaluator to think about the counterfactual. By that is meant what would have happened if no action had been taken. However, Mohr pays no attention to what would be the reasonable alternative to measure impacts against. Vedung (1997) doesn't relate evaluations to the decision problem at hand. If that were done it would become apparent that alternative actions would be the yardstick against which to measure outcomes of an action.
5. Arnstein (1971) has created an eight-step 'ladder of citizen participation', ranging from 'citizen control' to 'manipulation'.

6. These questions are implicitly being raised by Hofstee (1992: 277–85).
7. However, it had been shown by Jansson (1994) that the cost–benefit analysis was grossly overstating benefits. Instead of a surplus of 20–50 percent, benefits would be only 40 percent of costs.

References

- Arnstein, S. R. (1971) 'Eight Rungs on the Ladder of Participation', in E. S. Cahn and B. A. Passett (eds) *Citizen Participation: Effecting Community Change*. New York: Praeger.
- Bachrach, P. and M. S. Baratz (1970) *Power and Poverty. Theory and Practice*. New York: Oxford University Press.
- Dahlberg, M. and E. Vedung (2001) *Demokrati och Brukarutvärdering*. Lund: Studentlitteratur.
- Dryzek, J. S. (1996) *Democracy in Capitalist Times – Ideals, Limits, and Struggles*. New York: Oxford University Press.
- Elster, J. (1998) *Deliberative Democracy*. Cambridge: Cambridge University Press.
- Evaluation Report (1995a) *Miljömässig utformning av Norra Länken*. Rapport 1995-09-25. Stockholms stad: Stadsbyggnadskontoret.
- Evaluation Report (1995b) *Alternativ till Norra Länken*. Rapport 1995-09-28. Stockholms stad: Stadsbyggnadskontoret.
- Evaluation Report (1995c) *Betalstationer och betalsystem*. Rapport Oktober 1995. Stockholms stad: Stadsbyggnadskontoret.
- Evaluation Report (1996) *Alternativ till Södra länken*. Rapport 1996-02-29. Stockholms stad: Stadsbyggnadskontoret.
- Fetterman, D. M., S. J. Kaftarian and A. Wandersman (eds) (1996) *Empowerment Evaluation. Knowledge and Tools for Self-Assessment & Accountability*. Thousand Oaks, CA: Sage.
- Floc'hlay, B. and E. Plottu (1998) 'Democratic Evaluation. From Empowerment Evaluation to Public Decision-Making', *Evaluation* 4(3): 261–78.
- Greene, J. C. (2000) 'Challenges in Practising Deliberative Democratic Evaluation', in K. E. Ryan and L. DeStefano (eds) *Evaluation as a Democratic Process: Promoting Inclusion, Dialogue, and Deliberation. New Directions for Evaluation* 85: 13–26. San Francisco: Jossey-Bass.
- Gregory, A. (2000) Problematising Participation, *Evaluation* 6(2): 179–99.
- Hanberger, A. (2001) 'Policy and Program Evaluation, Civil Society and Democracy', *American Journal of Evaluation* 22(2): 211–28.
- Hofstee, W. K. B (1992) 'From the Citizen's Point of View: The Positioning of Program Evaluation', in J. Mayne, M. L. Bemelmans-Videc, J. Hudson and R. Connor (eds) *Advancing Public Policy Evaluation. Learning from International Studies*. Amsterdam: North-Holland.
- House, E. R. and K. R. Howe (1999) *Values in Evaluation and Social Research*. Thousand Oaks, CA: Sage.
- Jansson, K. (1994) *Dennispaketet – är det lönsamt?* Stockholm: KFB.
- Karlsson, O. (1996) 'The Political Dimension of the Stakeholder Model', *Evaluation* 2(4): 405–15.
- Krogstrup, H. K. (1997) 'User Participation in Quality Assessment', *Evaluation* 3(2): 205–24.
- Mohr, L. B. (1992) *Impact Analysis for Program Evaluation*. Thousand Oaks, CA: Sage.
- Rebien, C. C. (1996) 'Participatory Evaluation of Development Assistance: Dealing with Power and Facilitative Learning', *Evaluation* 2(2): 151–71.

Evaluation 8(1)

- Rocheort, D. A. and R. W. Cobb (1993) 'Problem Definition, Agenda Access, and Policy Choice', *Policy Studies Journal* 21(1): 56–71.
- Rossi, P. H., H. E. Freeman and M.W. Lipsey (1999) *Evaluation: A Systematic Approach*, 6th edn. Thousand Oaks, CA: Sage.
- Schattschneider, E. E. (1960) *The Semisovereign People: a Realist's View of Democracy in America*. New York: Holt, Rinehart and Winston.
- Vedung, E. (1997) *Public Policy and Program Evaluation*. New Brunswick, NJ: Transaction Publishers.
- Weiss, C. H. (1998) *Evaluation*, 2nd edn. Upper Saddle River, NJ: Prentice-Hall.
- Weiss, J. A. (1989) 'The Powers of Problem Definition: The Case of Government Paperwork', *Policy Sciences* 22(2): 97–121.

RICHARD MURRAY has a PhD in Economics from the Stockholm University. His present position is Chief Economist of the Swedish Agency for Public Management (Statskontoret). He has been working on quantitative analyses of public sector programmes – productivity measurements, cost–benefit analyses and evaluations. He is one of the founders of an environmental party in Stockholm – the Stockholm Party – and served as a council member for 10 years. Please address correspondence to: The Swedish Agency for Administrative Development, Box 2280, S-103 17 Stockholm, Sweden.
[email: richard.murray@statskontoret.se]