University of Groningen

Reconstructing Program Theories

Leeuw, Frans de

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Link to publication in University of Groningen/UMCG research database

American Journal of

Evaluation

Reconstructing Program Theories: Methods Available and Problems to be Solved

The online version of this article can be found at:

Published by:

$SAGE Publications

On behalf of:

American Evaluation Association

Additional services and information for *American Journal of Evaluation* can be found at:

Email Alerts: http://aje.sagepub.com/cgi/alerts

Subscriptions: http://aje.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations (this article cites 2 articles hosted on the
SAGE Journals Online and HighWire Press platforms):
http://aje.sagepub.com/cgi/content/abstract/24/1/5#BIBL

Articles should deal with topics applicable to the broad field of program evaluation. Articles may focus on evaluation methods, theory, practice, or findings. In all cases, implications for practicing evaluators should be clearly identified. Examples of contributions include, but are not limited to, reviews of new developments in evaluation, descriptions of a current evaluation study, critical reviews of some area of evaluation practice, and presentations of important new techniques. Manuscripts should follow APA format for references and style. Most submissions are 20–30 double-spaced typewritten pages in length; longer articles will also be published if their importance to AJE readers is judged to be high.

Reconstructing Program Theories: Methods Available and Problems to be Solved

FRANS L. LEEUW

ABSTRACT

This paper discusses methods for reconstructing theories underlying programs and policies. It describes three approaches. One is empirical–analytical in nature and focuses on interviews, documents and argumentational analysis. The second has strategic assessment, group dynamics, and dialogue as its core. The third has cognitive and organizational psychology as its foundation. For each of the three approaches, case-illustrations are given. These approaches can help to make the process of reconstructing underlying program theories more open for scrutiny. This is important because mis-reconstruction of policy and program theories is dangerous. All three approaches have a number of weaknesses to be remedied. The paper discusses these shortcomings and presents some suggestions for overcoming the limitations are presented.

WHY PAY ATTENTION TO METHODS FOR RECONSTRUCTING AND EVALUATING PROGRAM THEORIES?

With a history of over 50 years, policy researchers and evaluators have paid attention to the social and behavioral assumptions that underlie policy instruments, programs and strategies.

Frans L. Leeuw • Netherlands Inspectorate for Education & Professor, Evaluation Studies, University of Utrecht, P.O. Box 2730, 3500 GS Utrecht, The Netherlands; E-mails: f.leeuw@owinsp.nl, flleeuw@cuci.nl.

5

In the 1930s the sociologist Karl Mannheim (Mannheim, 1935/1967) published *Man and Society in an Age of Reconstruction*, in which he appealed for the articulation of assumptions underlying what was then called "social planning." In the 1950s a Dutch sociologist, Jacques van Doorn, coined the word "policy theory." Rogers, Hacsi, Petrosino, and Huebner (2000a, 2000b, p. 5) refer to a history of program theory of "over 30 years now" in the field of evaluation.

Sometimes underlying assumptions are referred to as policy theories, but more often the terms program theory or program logic are used. An important difference exists between program theory and program logic. Program logic often specifies the inputs and components of a program, as well as short-term and long-term outcomes, along with the assumed linkages among these. However, a program logic rarely outlines the underlying mechanisms that are presumed to be responsible for those linkages. In contrast, Rogers et al. (2000a, 2000b, p. 5) see a program theory as an explicit theory or model of *how* a program causes the intended or observed outcomes.

This paper focuses on *methods* for reconstructing or articulating the underlying and often implicit theory of programs and policies. Therefore, I do not look into the question how a program *should* or *could* work, according to the most relevant social science theories.

The first reason to pay attention to methods is to make "the inductive hunt for the causes of those changes" [produced by the program] more transparent (Davidson, 2000, p. 22). Although Davidson presents a five-stage process for tracking causes (i.e., underlying social mechanisms), this methodology is imprecise. For example, it is unclear how this "inductive hunt" takes place. Bickman (2000, p. 107) refers to the "lack of precision in most [program] theories"; he is of the opinion that the "program theory approach has exposed the impoverished nature of the theories that underlie many of the interventions we study." However, the question is how do we know what the "nature" of the underlying theory is if our reconstruction methodology is vague? Cook (2000, p. 29) refers to the problem of how to "find" the program theory, when "program theories not always are very explicit" and there is "a multiplicity of possible program theories which entails a large . . . problem for a theory of theory-based evaluation."

A second reason for focusing on methodological aspects is that, although in most of the textbooks on evaluation nowadays attention is paid to program theory (Rossi, Freeman, & Lipsey, 1999; Shaw, 2000; Swanborn, 1999; Vedung, 1998), rather limited attention is spent on methods for articulating underlying theories when they are not already made explicit by stakeholders themselves. Comparing different methods, for example, is not part and parcel of textbooks.

Finally, Tilley (1999, p. 53) describes a situation in which evaluators mis-reconstruct policy theories and hence provide poor advice to policy makers and society at large. One of the remedies might be to have better knowledge available on different methods how to reconstruct program and policy theories. This in particular is relevant now that the number of studies focusing on program theories is increasing.[1]

This paper describes three methods for reconstructing and assessing the theories behind strategies, policies and programs, when these theories are not well-specified by stakeholders or producers of the programs: (1) a policy–scientific approach, (2) a strategic assessment approach, and (3) an elicitation approach. These methods come not only from the evaluation literature as such. They are also adapted from studies in the field of organizational psychology and management that were originally designed for other purposes than articulating program theory. The first case-study presented below concerns an evaluation of a *new program of the World Bank*. The second case focuses on an *evaluation* of "*procedural arrangements*" of a *census bureau*. The third case looks into assumptions underlying a *strategy of competition* in industry.

RECONSTRUCTING PROGRAM THEORY:
A POLICY–SCIENTIFIC APPROACH

Background

Hoogerwerf (1987), Vedung (1998), Rossi et al. (1999), and this author (Leeuw, 1991; Leeuw & Van Gils, 1999) have outlined elements of a methodology that—when integrated—has the steps described below as its core. Because these authors mainly work in the field of public policy analysis and evaluation, I have labeled this approach as "policy–scientific."

Methods

There are six steps involved in articulating a program theory using the policy–scientific approach.

- Identify behavioral mechanisms expected to solve the problem. Search in formal and informal documents and in interview transcripts for statements that indicate: why it is believed necessary to solve the social, organizational, or policy problem; and what the goals are of the policy or program under review. These statements point to mechanisms; these can be considered the "engines" that drive the policies or programs and are believed to make them effective. Examples are manifold. They include determinants of innovation diffusion processes, mechanisms that underlie prisoner games, processes that produce social capital, cognitive dissonance, different types of learning behavior (like vicarious learning), and many more.
- Statements that have the following form are especially relevant for detecting these mechanisms:
  - "It is evident that x . . . will work,"
  - "In our opinion, the best way to go about this problem is to . . . ,"
  - "The only way to solve this problem is to . . . ,"
  - "Our institution's x years of experience tells us that . . . ."
- Compile a survey of these statements and link the mechanisms with the goals of the program or policy under review.
- Reformulate these statements in conditional "if–then" propositions or propositions of a similar structure ("the more x, the less y").
- Search for warrants, to identify missing links in or between different propositions through argumentational analysis. Argumentational analysis is a standard tool in logic and philosophy and has as one of its foundations Toulmin's (1958) *The Uses of Argument*. It describes a model for analyzing chains of arguments and it helps to reconstruct and "fill in" argumentations. A central concept is the warrant. According to Toulmin (1958) and Mason and Mitroff (1981), a warrant is the "because" part of an argument: it says that B follows from A because of a (generally) accepted principle. For example, "the organization's performance will not improve next year" follows from "the performance of this organization has not improved over the last 5 years," because of the principle, "past performance is the best predictor of future performance." The "because" part of such an argument is often not made explicit. Consequently, these warrants must be inferred by the person performing the analysis.

- Reformulate these warrants in terms of conditional "if–then" (or similar) propositions and draw a chart of the (mostly causal) links.
- Evaluate the validity of the propositions by looking into:
  - the logical consistency of the set of propositions;
  - their empirical content, that is, the extent to which the theory and in particular the assumed impact of the behavioral mechanisms correspond with the state of the art within the social/behavioral/economic sciences on these mechanisms; and
  - the extent to which the theory focuses on variables that can be "manipulated" or "steered" through policy programs.

Evaluation usually takes place on the basis of research reviews, interviews, and other data.

A Case-Illustration

A mid-term evaluation of anticorruption initiatives of the World Bank Institute (WBI; Leeuw & Van Gils, 1999) will illustrate the policy–scientific approach. It focuses on two countries, Tanzania and Uganda, and on one strand of the program (i.e., workshops). Later, several other countries were added. WBI developed the concept of national integrity systems as a means to identify and strengthen those institutions with a mandate to fight corruption. These institutions, known as the "pillars of integrity," include watchdog agencies, Parliament, civil society, media, and the judiciary. Central in this anticorruption initiative are two types of actions, integrity workshops and media workshops. The main purpose of integrity workshops is to formulate and agree upon an anticorruption program and in the process raise awareness of the costs of corruption and discuss the roles that the various pillars of integrity play in the fight against corruption. Workshops are also intended to create a partnership between members of the integrity pillars. As a result, participants' awareness is believed to be enhanced, which facilitates the further development and implementation of the program, which basically consists of public awareness campaigns, following up cases where corruption is reported, and implementing legal and institutional change. With regard to the media workshops, the media are seen as key players in informing the public about corruption and exposing corrupt practices. The workshops focus on awareness raising and on discussions of the media's role in the fight against corruption, but also on improving professional techniques. In the evaluator's language, the focus of this case-study was to assess, shortly after the implementation of the program in the first two countries, its strengths and weaknesses, with a particular focus on the quality of the underlying program theory.

In this case-study, the following steps were taken. In order to detect the underlying logic, documentary information on over 30 workshops was brought together and analyzed to answer the question: Why is it believed that workshops can and will help reduce the corruption problem in Tanzania and Uganda? The analysis also described the timing and the management of the workshops, their objectives, funding, number of participants, outcomes, and the institutional affiliation of participants. Articles appearing in seven newspapers and magazines in both countries over a 2-week period were also examined. Back to office (BTO) documents produced by the WBI were analyzed in terms of why it was believed that the program can be successful; BTO reports are produced after WB officials return from country missions. Next, "country

reports," which are produced by the regional offices of the Bank, were content analyzed to find underlying assumptions.

In addition, six interviews with WBI officials were held in Washington and some 30 interviews with stakeholders in both countries. In addition to a large number of questions focused on the implementation of the WBI project, topics included:

- the interviewee's assumptions about the importance of the integrity pillars;
- the WBI assumptions that fighting corruption and vesting pillars is important for curbing corruption; and
- the assumptions underlying the different workshops.

Argumentational analysis was the next step, which finally led to charting the program theory. The result is the program logic outlined in Figure 1.

This *program theory* consists of 26 propositions (Leeuw & Van Gils, 1999). Once reconstructed, the propositions were evaluated on the basis of a review of existing social, economic, and behavioral research findings (in some nine fields, ranging from educational sciences, criminology, and psychology, to economics and anthropology), field interviews, social mapping, and document analysis (Leeuw & Van Gils, 1999).[2] The evaluators concluded that:

> workshops . . . can be effective. However, we also encountered certain weaknesses in the approach. A first weakness is that there is no automatic progression from awareness of an unjust situation to intervening to bring it to an end. A second weakness in the approach [of the World Bank Institute] deals with the assumed trickle-down effect of policy dialogues to society at large. This is crucial because the general public is not an entity that can be empowered through workshops. In order for knowledge and awareness or concern with corruption to trickle down to regional or district segments of society, an infrastructure in terms of literacy levels and radio, TV and telephone facilities is essential. Data indicate that in both Tanzania and Uganda the illiteracy is very high, while the mass media facilities are extremely limited. Because of these circumstances, it is likely that the results of the workshops do not trickle down to the local populace (Leeuw & Van Gils, 1999, pp. 208–209)

Strength and Weaknesses of the Policy–Scientific Approach

One of the strengths of this approach is its focus on documents as well as interview data, which stimulates the use of multiple methods. The use of argumentational analysis also helps to validate the reconstruction process, while the diagrammatic presentation of the underlying theory helps dialogues with stakeholders and others. The use of research reviews as a source to assess the underlying theory reduces the possibility that the informative and empirical content of the program theory is being "negotiated." This means that, without research evidence, ideological or even "politically correct" statements about the content of the program theory might play a role when discussing the question of the validity of the program theory.

A weakness is that often no attention is paid to the social and behavioral dynamics involved in processes of articulation and evaluation of theories. People have political, economic, and social stakes in their personal theories, which makes relevant the question of how to get those people involved in an "assessment process." Another weakness is that field experiences show that this methodology can sometimes be rather cumbersome and time consuming. Lack of attention paid to differences in power positions of the stakeholders is also a potential weakness.
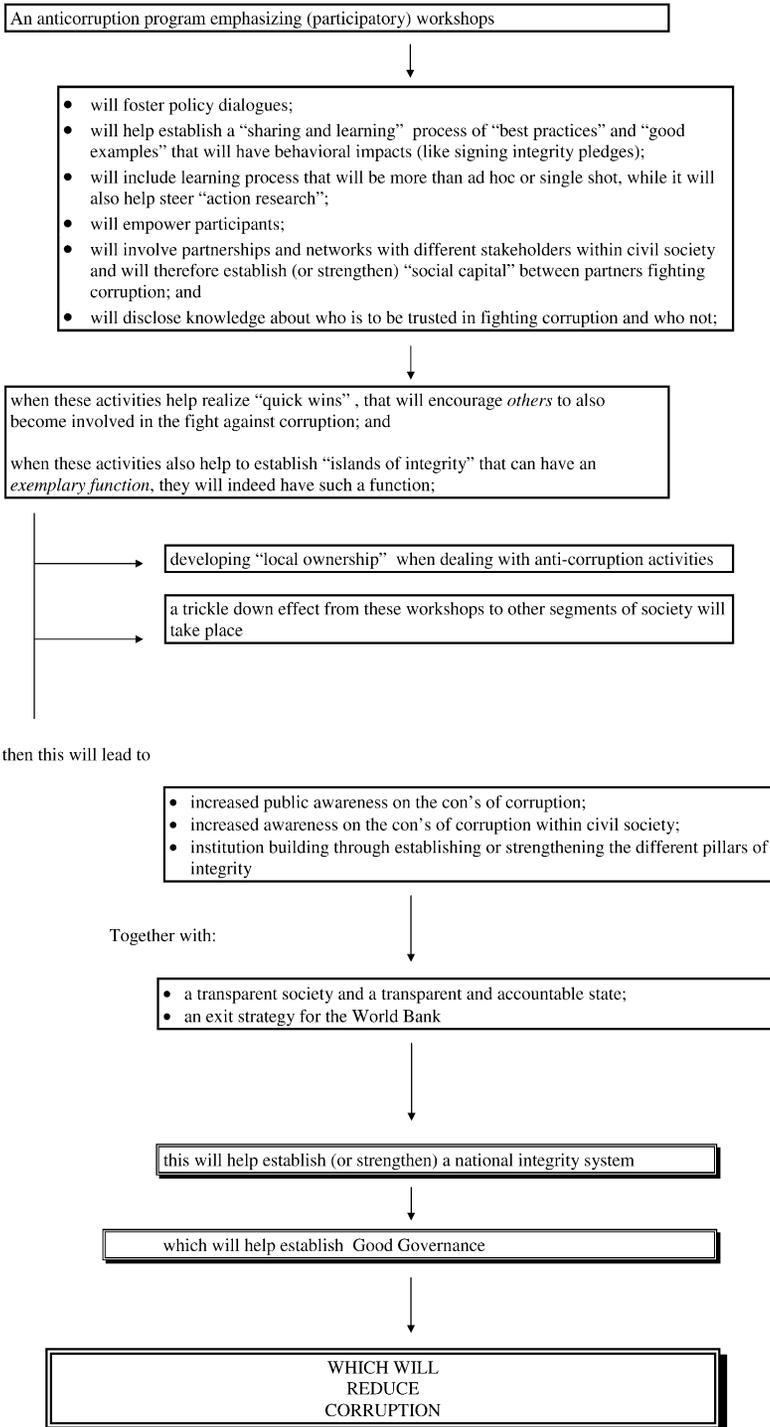
An anticorruption program emphasizing (participatory) workshops

- will foster policy dialogues;
- will help establish a "sharing and learning" process of "best practices" and "good examples" that will have behavioral impacts (like signing integrity pledges);
- will include learning process that will be more than ad hoc or single shot, while it will also help steer "action research";
- will empower participants;
- will involve partnerships and networks with different stakeholders within civil society and will therefore establish (or strengthen) "social capital" between partners fighting corruption; and
- will disclose knowledge about who is to be trusted in fighting corruption and who not;

when these activities help realize "quick wins" , that will encourage *others* to also become involved in the fight against corruption; and

when these activities also help to establish "islands of integrity" that can have an *exemplary function*, they will indeed have such a function;

developing "local ownership" when dealing with anti-corruption activities

a trickle down effect from these workshops to other segments of society will take place

then this will lead to

- increased public awareness on the con's of corruption;
- increased awareness on the con's of corruption within civil society;
- institution building through establishing or strengthening the different pillars of integrity

Together with:

- a transparent society and a transparent and accountable state;
- an exit strategy for the World Bank

this will help establish (or strengthen) a national integrity system

which will help establish Good Governance

WHICH WILL
REDUCE
CORRUPTION

Figure 1.   Schematic Representation of Core Elements of WBI's Underlying Program Logic.

A STRATEGIC ASSESSMENT APPROACH

Background

This approach, also called "assumptional analysis" can be found in a series of studies (Jackson, 1989), but has as its core knowledge basis Mason and Mitroff's (1981) book on *Challenging Strategic Planning Assumptions*. Dialogues form a crucial element in this approach, though their role is different from the roles they play in "dialogue-driven evaluations," as recently described in a special issue of *Evaluation* devoted to this topic (see Vol. 7, No. 2, April 2001). The role attached to dialogues in dialogue-driven evaluation is more seen from an empowerment perspective than from a perspective of "criticizing" (and even debunking) policy assumptions with the help of "argumentational analysis" (as in the strategic assessment approach). Schwandt (2001, p. 268) suggests that adherents would do well to use the "extensive literature on the logic of argument as reasoned dialogue," and not only focus on empowerment, participation, and democratization.

Methods

Central in the approach are four major stages: (1) group formation; (2) assumption surfacing; (3) dialectical debate; and (4) synthesis.

Stage 1: Group formation.   The aim of this stage is to structure groups so that the productive operation of the later stages of the methodology is facilitated. A wide cross-section of individuals with an interest in the relevant policy question should be involved. They are divided into groups, care being taken to maximize convergence of viewpoints within groups and to maximize divergence of perspectives between groups.

Stage 2: Assumption surfacing.   During this stage the different groups separately unearth the most significant assumptions that underpin their preferred policies and programs. Two techniques assume particular importance in assisting this process. The first, stakeholder analysis, asks each group to identify the key individuals or groups upon whom the success or failure of their preferred strategy would depend. This involves asking questions such as: Who is affected by the strategy? Who has an interest in it? Who can affect its adoption, execution, or implementation? And who cares about it? For the stakeholders identified, each group then lists what assumptions it is making about each of them in believing that its preferred strategy will succeed.

The second technique is "assumption rating." Initially one should find and list the assumptions. Mason and Mitroff (1981, p. 10) indicate that this involves searching for statements about symptoms of the problem (that have to be solved through a policy or program distinguishing them from statements about causes of the problem): This search is carried out by reviewing interoffice memos, corporate directives, magazine articles, and public announcements. If one applies this to a human and social services setting or to the deliverance of education programs, the focus would be on program proposals, procedure manuals, swot-analysis ("strengths, weaknesses, opportunities and threats") and mission statements.

For each of the listed assumptions, each group asks itself two questions: (1) How important is this assumption in terms of its influence on the success or failure of the strategy? And,
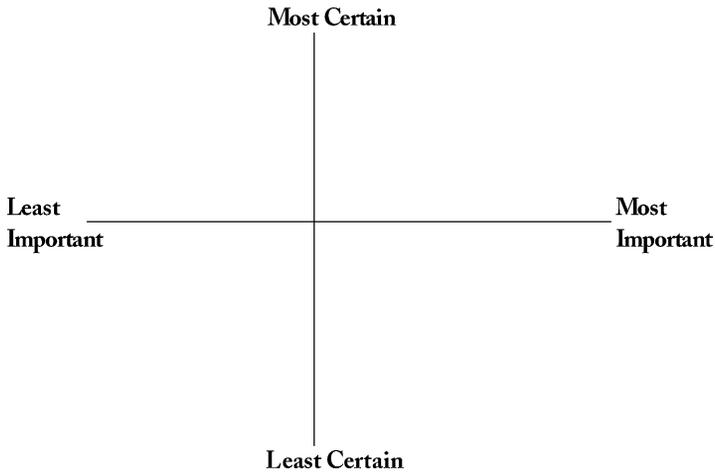
Most Certain

Least Important   **Most Important**

Least Certain

Figure 2.   Assumption Rating Chart.

(2) how certain are we that the assumption is justified? Here, in fact, the evaluation of the listed assumptions takes place, usually by using research reviews and similar documents. The results are recorded on a chart such as that shown in Figure 2. Each group then is able to identify a number of key assumptions—usually in the most important/least certain quadrant of the chart—upon which the success of its strategy rests.

Stage 3: Dialectical debate.   The groups are brought back together and each group makes the best possible case to the others for its preferred strategy, while identifying its key assumptions. Only points of information are allowed from other groups at this time. There is then an open, dialectical debate focusing on which assumptions are different between groups, which are rated differently, and which of the other groups' assumptions each group finds most troubling. Each group should develop a full understanding of the preferred strategies of the others and their key assumptions. This, to some extent, is linked to the idea of "educative authenticity" as described by Guba and Lincoln (1989) in their book on *Fourth Generation Evaluation*.

Stage 4: Synthesis.   An attempt at synthesis is then undertaken. Assumptions are nego-tiated and modifications to key assumptions made. Agreed-upon assumptions are noted; they can form the basis for consensus around a new strategy that bridges the gap between the old strategies and goes beyond them as well. If no synthesis can be achieved, points of disagree-ment are noted and the question of what research might be done to resolve these differences is discussed.

A Case-Illustration

The case-study can be found in Chapter 4 of Mason and Mitroff's (1981) *Challeng-ing Strategic Planning Assumptions*. It concerns the way in which the U.S. Bureau of the Census dealt with the problem of accuracy of its census and in particular with the problem of undercounting population numbers. "The undercounting varies significantly according to

ethnicity of the area. Whites tend to be undercounted by only 1.9%, whereas Blacks are undercounted by 7.7%" (Mason & Mitroff, 1981, p. 59). The problem was "to adjust or not to adjust [in the census] and, if so, how?" Also, the possible impacts of the adjustment(s) had to be assessed. In an evaluator's language, the focus was on an ex ante evaluation of adjustment procedures.

The first step set by the Census Bureau was to organize an "Undercount Adjustment Workshop" in which assumptions underlying different adjustment procedures were addressed. Four working groups were formed at the start of the conference, looking into different adjustment plans, which they were to support. Stakeholders were identified and assumptions were ranked for their importance and the certainty with which an assumption should be viewed. Next, four self-selected groups started to discuss these assumptions that underlined four "adjustment plans" (in practice, they were called "conventions").

The second day of the conference was devoted to discussion within each of the four groups, working alone. The main goal was to identify the assumptions that were necessary for each group's particular plan of action to be made feasible. Next, each group's assumptions and their relative importance and certainty were presented at a plenary session. During the session, clarifying questions were allowed, but no discussion of the merits of the assumptions was permitted. After this plenary presentation, the groups met individually to determine which of the assumptions that had been presented, if true, would be most damaging to their group's "convention." In another plenary session, the damaging assumptions were identified and a debate took place.

After identifying these crucial issues on the third morning, the plenary group determined what types of information might be necessary to remove uncertainties associated with each issue or assumption. The result was a list of about 20 measures. Mason and Mitroff (1981, p. 76) make clear that "the Undercount Adjustment Workshop was successful in surfacing assumptions, many of which had been implicit prior to the conference. As a result of this process, Bureau management personnel have a better understanding of the issues and assumptions involved in adjustment procedures."

Strengths and Weaknesses of the Strategic Assessment Approach

One of the strengths of the strategic assessment approach is the focus on group formation and group dynamics. Sharing knowledge and sharing perspectives when doing a reconstruction is central. The strategic assessment approach was compared to the Devil's Advocate (Schwenk, 1984) and "the results do not show that either the use of the Devil's Advocate approach and the [Mason & Mitroff] approach is a superior method for introducing conflict." Conflict is believed to be important because it helps stimulate critical ("dialectical") attention for assumptions. However, Midgley (1996, p. 77) is of the opinion that the Mason and Mitroff-approach "has enormous potential for helping to promote democratic decision-making and addressing openly declared conflicts within [service] systems."

A weakness is that the criteria used for assessing the validity of the assumptions remain largely unspecified, while the same goes for the role of existing research in the dialogues. Group think, yeah-saying, and similar social–psychological artifacts of group dynamics are not adequately considered. In addition, too little attention is paid to differences in the power positions of the participants and the consequences for their group behavior.

ELICITATION METHODOLOGY

Background

The literature shows that there are multiple systematic ways of eliciting mental models. These elicitation methods largely draw on the same theoretical literature, specifically, Eden and Spender (1998), Weick and Bougon (1986) and Argyris and Schön (1978). As policies and programs are developed and implemented by organizations, the "mental models" or "cognitive maps" of members of these organizations are important for understanding the anticipated impact of these policies or programs. Therefore, attention is paid to articulating assumptions not only within the policy sciences, but also within organization studies, including the psychology of organizations. In these studies the focus is on managerial and organizational cognitions (Eden & Spender, 1998). One of the central questions is what the relationships are between these cognitions and the outcomes of organizations. Managers, shareholders, stakeholders, and workers have "cognitions" (or "mental maps") about the organization and its environment. These maps of what is going on in their organization partly determine their behavior. Their content concerns the organizational strategies, their chances of success, the role power plays, their own roles and the relationships with the outside world. Following Weick and Bougon (1986, p. 131), these schemes "assimilate uncertain aspects to existing structures." Empirical studies reported by Eden and Spender (1998) show how crucial these mental maps are when trying to explain organizational outcomes.

It has also been shown how important the articulation of these mental maps is for organizations to become "learners." This partly is reflected in the idea that double loop learning always involves a critical assessment of the assumptions underlying organizational activities. Following Argyris and Schön (1978), it is wise to distinguish between theories-in-use versus espoused theories. Eden and Spender (1998, p. 15) argue that parts of these "maps" or "theories" are implicit and are almost "tacit knowledge," both on an individual level and on a collective level. By articulating these mental models, it is possible to compare them with evidence from scientific organization studies.

Methods

Van der Heijden and Eden (1998) have described some methods for reconstructing (or "eliciting") and assessing these mental or cognitive maps:

- Look at the concrete record of strategic intentions, through, for example, a study of the documentation which is designed to direct behavior;
- Look at decision-making in action, get involved in the organization (an anthropological observer approach). Watch decision-makers, listen to stories;
- Work with managers on strategic breakdown situations. Become immersed in the thinking and the social process of "strategic fire fighting";
- Use well-designed trigger questions in interview situations so that "theories in use" can be detected. Follow interviews with feedback to individuals and to the team. The "elicitation cycle" is built on responses to designed trigger questions. The process uses dialectic and non-conforming statements and is iterative. Van der Heyden and Eden specificy six "techniques";

- ○ Create an open-ended atmosphere in the interview;
- ○ Do away with formal language and create a "playful" atmosphere in which it is easier to deviate from the formal phraseology and the "official script";
- ○ Do "set the interviewees up against themselves";
- ○ Create dialectical tension by asking the interviewees to adopt unusual roles;
- ○ Listen very carefully for internal inconsistencies in what is being said;
- Apply data/content-analysis programs like NUD.Ist (Scolari) or other text analysis programs to the interview reports and documents; and
- Confront the results of these content-analysis activities with relevant (social) scientific research.

A Case-Illustration

The elicitation of mental models of managers about competition in the automobile industry will illustrate the elicitation process. Although this case-illustration relates to the corporate setting, the way in which Johnson and her colleagues elicited the "mental models" in evaluation settings of managers is also applicable to a more typical evaluation setting in the public sector. For example, attention could be focused on mental models underlying changes in the mission of an organization or its structure and operations.

Some 10 years ago the U.S. General Accounting Office introduced the concept of a program's underlying operational and conceptual model. A conceptual model concerns the social and behavioral *logic* behind a program, an operational model focuses on which actors are assumed to do what, and why, in order to make the program a success. The methods to be used to reconstruct both sets of logics, however, were rather unclear. Evaluators using an approach similar to the one Johnson and her colleagues applied to business might be better able to reconstruct the conceptual and operational models underlying a program.

In the study, the authors used the Virtual Card Sort technique (Daniels, De Chernatony, & Johnson, 1995). "This technique was selected as it makes minimal a priori assumptions concerning the characteristics of managers' knowledge . . . . [It] is administered by asking each participant to think about the international automotive industry as a whole and name all those organizations that they perceive as competitors to their organization" (Johnson et al., 1998, p. 136). "The names of the organizations including those of the participants' organizations, are written on small cards. These are then handed back to the managers who are asked to arrange them on the table in front of them to reflect the manner in which those organizations compete with each other. . . . The managers are allowed to remap [each] as many times as required. They are asked to explain the knowledge structures/maps they have just created. The justifications they verbalize are again written on small cards. This final set of cards, once given back to the managers and arranged, is considered to make up the construct map. The first map that the managers create of their competitive environment is considered the most salient and is taken forward for data-analysis" (Johnson et al., 1998, p. 36). In total, the study involved 22 managers from three international automotive industry organizations. Fifty-nine constructs were developed by them.

Strengths and Weaknesses of Elicitation Methodology

A strength is that this approach highlights the importance of observing managerial behavior in breakdown-situations, or critical circumstances, and not only in times of prosperity.

The more a decision-making process resembles a breakdown-situation, the larger the likelihood that major assumptions are articulated given the need to overcome that situation. A second strength is the almost anthropological orientation to unraveling the ways in which managers "work with" mental models.

With regard to weaknesses, a first one to mention is that the elicitation approach is unclear about the knowledge bases used and about the criteria. When asking trigger questions and when creating a "dialectic atmosphere," it is important to know what the knowledge is from which one starts. To establish a dialectic atmosphere on the basis of empirically *incorrect* assumptions is inefficient and ineffective. To be engaged in discussions about assumptions that clearly are untrue (i.e., are falsified by substantive empirical evidence) is not only a waste of time; it can also increase the likelihood that these untrue assumptions nevertheless start playing a role in decision-making processes. Also, the level of specificity how to create the "open atmosphere" is rather low. To *suggest* that a non-phraseaology atmosphere is created is less difficult than to *create* such an atmosphere in practice.

ORIGINS OF THE METHODS

The three methods have been labeled here largely because of their origins. One stems from applied social research, another from business and administrative sciences, while the third goes back to the cognitive tradition of "mapping" in psychology. Although they differ in data and precise interview techniques, all methods use documents (sometimes primarily official ones, sometimes informal interoffice memos), all work with interviews, and all make use of existing (social and behavioral) research to check for consistency and the empirical content of the reconstructed theory. In particular, the second approach has some variants, like the devil's advocate approach outlined by Schwenk (1984). Trochim's approach to concept mapping (Trochim, 1989) can be seen as a variant of the elicitation method, including the Q-sort-methodology. An evaluator following the policy–scientific approach can make use of the "trigger" questions outlined by Van der Heijden and Eden (1998).

In principle, the three methods can be applied to the same evaluation task, that is, to the reconstruction of program theories in general. However, it could be argued that the policy–scientific approach is best suited for ex post evaluations of programs and policies backed by documentary evidence (i.e., often public policies), while the strategic assessment and elicitation approach appear to be more relevant for ex ante evaluations under other conditions.

FUTURE PERSPECTIVES

To help reduce some of the weaknesses of the approaches and in particular the ones that deal with the social context of reconstructing program logic, a challenging new development comes from a relatively new field of interdisciplinary studies. This is Computer-Supported Collaborative Argumentation (CSCA). It brings together research from the areas of Computer-Supported Collaborative Learning (CSCL) and Design Argumentation, and in particular Design Rationale (DR). For "argumentation," one can read "assumptions underlying programs or policies." The new field is an information and communication technology (ICT)-driven approach to reconstruct argumentations underlying decisions of people and organizations. Virtual-graphic

representations are crucial in this approval; they function as (external) representations of the structure of the underlying assumptions. By using ICT, the transparency of the reconstruction process is greatly enhanced, which makes it possible to reduce some of the dangers involved in the social and group processes when reconstructing assumptions.

ICT-driven group decision rooms, for example, show (on the screen and in hard copy) all participants exactly which assumptions are believed by the group members. Moreover, this makes it possible to comment on the assumptions. Another advantage is that no comments or assumptions can get lost during the debates.

Examples can be found in systems such as SenseMaker (Bell, 1997) and Belvédère (Suthers, Toth, & Weiner, 1997; Suthers & Weiner, 1995). In Sensemaker, hypotheses and supporting evidence are brought together in so-called "claim frames." Sensemaker only uses "theory"; it does not use "objects" like "hypothesis" or "data." Moreover, Sensemaker does not make contradictions between evidence visible (http://www.kie.berkeley.edu/KIE/software/sensemaker_large.gif). Belvédère (http://lilt.ics.hawaii.edu/belvedere/index.html, http://lilt.ics.hawaii.edu/belvedere/index.html) is a synchronous system that supports collaborative inquiries. Learners are confronted with challenge problems (e.g., What caused the extinction of the Dinosaurs? Why does an anticorruption policy focused on the judiciary in Bolivia not work?) that need a (scientific) explanation. The Belvédère environment offers access to web-based material (that will also guide learners through the steps of the inquiry), a chat window, and a shared visual workspace where learners construct scientific explanations in so-called "evidence maps." The environment has a coach that comments on the structure of the evidence maps and that makes suggestions for improvements. The ontology of the evidence maps in Belvédère is defined in terms of the objects and relations that students may use when they create evidence maps. In the current version of Belvédère, the objects are "principle," "hypothesis," "data" and "unspecified." The relations are reduced to a basic set of "for," "against" and "and" relationships. Participants can express the strength of their beliefs in the objects and relations. This set of objects and relations obviously limits the scope of what Belvédère can express. Nevertheless, these methods and related ones to come may be worth exploring as tools for reconstructing program theory.

WHEN TO DO PROGRAM THEORY RECONSTRUCTION AND EVALUATION?

With regard to the question under which circumstances it is worth to take the resources and effort to construct a program theory, three answers can be given. The larger the assumed impact of a program, the wiser it is to reconstruct a program theory with rigorous methods. The larger the risks involved in the program, the more it is worth to carry out such an activity. Risks imply uncertainties with regard to the impact and (unintended) side-effects of programs. And, third, the more (public or private) money is involved in a program, the wiser it is to get involved in a reconstruction and evaluation of the underlying theory.

CONCLUSIONS

In order to make the reconstruction of underlying program theories transparent, attention has to be paid to methods. This paper describes three approaches. One is strongly linked

to mainstream evaluation methodology (i.e., the policy–scientific) and can be considered a further development of Rossi et al.'s (1999) discussion of "expressing and assessing program theory." It focuses on interviews, documents, and argumentational analysis. The second has strategic assessment, group dynamics, and dialogue as its core. The third has cognitive and organizational psychology as its foundation, together with group dynamics and an almost anthropological approach to managerial "mind sets."

These approaches can help to make the process of reconstructing underlying theories more open for scrutiny. This is important because, as Tilley (1999) showed, mis-reconstruction of policy and program theories is dangerous.

Nevertheless, all three approaches have a number of weaknesses that have to be remedied. Finally, some suggestions for overcoming the limitations referring to group dynamics issues are presented. As "dialogues in evaluation" have become "an increasingly popular yet problematic theme in the evaluation community" (Stern, 2001), we hope to add something to this debate by linking dialogue to argumentational analysis.

ACKNOWLEDGMENTS

NOTES

1. During the recent congress of the European Evaluation Society (October 2002) in 19 out of the approximately 150 sessions, papers were presented on program theories and theory-driven evaluations. Most of them did report about ongoing research. Compared to earlier EES conferences, this is a substantial number. Recent fascinating studies include Pawson's (2002) theory-reconstruction of Megan's Law (the U.S. sex offender notification and registration program) and Andrews (2002) analysis of the implicit theory underlying budget reforms.

2. In a World Bank paper and a journal article (Leeuw & Van Gils, 1999), further information on the methodology can be found in six appendices.

REFERENCES

Andrews, M. (2002). A theory-based approach to evaluating budget reforms. *International Public Management Journal*, *5*, 135–154.

Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addison-Wesley.

Bell, P. (1997). Using argument representations to make thinking visible for individuals and groups. In R. Hall, N. Miyake, & N. Enyedy (Eds.), *Proceedings of CSCL '97: The Second International Conference on Computer Support for Collaborative Learning* (pp. 10–19). Toronto: University of Toronto Press.

Bickman, L. (2000). Summing up program theory. In P. J. Rogers, T. A. Hacsi, A. Petrosino, & T. A. Huebner (Eds.), *Program theory in evaluation: Challenges and opportunities. New directions in evaluation* (Vol. 87, pp. 103–113). San Francisco, CA: Jossey-Bass.

Cook, T. D. (2000). The false choice between theory-based evaluation and experimentation. In P. J. Rogers, T. A. Hacsi, A. Petrosino, & T. A. Huebner (Eds.), *Program theory in evaluation: Challenges and opportunities. New directions in evaluation* (Vol. 87, pp. 27–35). San Francisco: Jossey-Bass.

Daniels, K., De Chernatony, L., & Johnson, G. (1995). Validating a method for mapping managers' mental models of competition. *Human Relations*, *48*, 971–975.

Davidson, E. J. (2000). Assserting causality in theory-based evaluation. In P. J. Rogers, T. A. Hacsi, A. Petrosino, & T. A. Huebner (Eds.), *Program theory in evaluation: Challenges and opportunities. New directions in evaluation* (Vol. 87, pp. 17–27). San Francisco, CA: Jossey-Bass.

Eden, C., & Spender, J.-C. (Eds.). (1998). *Managerial and organizational cognition. Theory, methods, and research*. London: Sage.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Hoogerwerf, A. (1987). Beleid berust op veronderstellingen: De beleidstheorie. *Acta Politica*, *19*, 493–531.

Jackson, M. C. (1989). Assumptional analysis. *Systems Practice*, *14*, 11–28.

Leeuw, F. L. (1991). Policy theories, knowledge utilization, and evaluation. *Knowledge and Policy*, *4*, 73–92.

Leeuw, F. L., & van Gils, G. H. C. (1999). Evaluating anti-corruption initiatives. Underlying logic and mid-term impact of a World Bank Program. *Evaluation*, *5*, 194–219.

Manheim, K. (1935/1967). *Man and society in an age of reconstruction: Studies in modern social structure*. New York: Hardcourt, Brace & World.

Mason, I., & Mitroff, I. (1981). *Challenging strategic planning assumptions*. New York: Wiley.

Midgley, G. (1996). Evaluating services for people with disabilities: A critical systems perspective. *Evaluation*, *2*, 67–85.

Pawson, R. (2002). *Does Megan's Law work? A theory-driven systematic review*. ESCR UK Centre for Evidence Based Policy and Practice Working Paper 8, Queen Mary University of London.

Rogers, P. J., Hacsi, T. A., Petrosino, A., & Huebner, T. A. (Eds.). (2000a). *Program theory evaluation: Practice, promise and problems. New directions in evaluation* (Vol. 87, pp. 5–17). San Francisco, CA: Jossey-Bass.

Rogers, P. J., Hacsi, T. A., Petrosino, A., & Huebner, T. A. (Eds.). (2000b). *Program theory in evaluation: Challenges and opportunities. New directions in evaluation* (Vol. 87). San Francisco, CA: Jossey-Bass.

Rossi, P., Freeman, H., & Lipsey, M. (1999). *Evaluation: A systematic approach*. Thousand Oaks, CA: Sage.

Schwandt, T. A. (2001). A postscript on thinking about dialogue. *Evaluation*, *7*, 264–277.

Schwenk, C. (1984). Devil's advocate in managerial decision-making. *Journal of Management Studies*, *21*, 153–168.

Shaw, I. (2000). *Evaluating public progammes: Contexts and issues*. Aldershot: Ashgate.

Stern, E. (2001). Editorial foreword to the special issue of evaluation: Dialogue in evaluation. *Evaluation*, *7*, 1–3.

Suthers, D., & Weiner, A. (1995). *Groupware for developing critical discussion skills*. http://www-cscl95.indiana.edu/cscl95/suthers.html.

Suthers, D. D., Toth, E. E., & Weiner, A. (1997). An integrated approach to implementing collaborative inquiry in the classroom. In R. Hall, N. Miyake, & N. Enyedy (Eds.), *Proceedings of CSCL '97: The Second International Conference on Computer Support for Collaborative Learning* (pp. 272–279). Toronto: University of Toronto Press.

Swanborn, P. G. (1999). *Evalueren*. Amsterdam: Boom Press.

Tilley, N. (1999). Evaluation and evidence-(mis)led policy. *Evaluation Journal of Australasia*, *11*, 48–64.

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

Trochim, W. (1989). Concept mapping. Soft science or hard art? *Evaluation and Program Planning*, *12*, 87–110.

Van der Heijden, K., & Eden, C. (1998). The theory and praxis of reflective learning in strategy making. In C. Eden & J.-C. Spender (Eds.), *Managerial and organizational cognition: Theory, methods and research* (pp. 58–76). London: Sage.

Vedung, E. (1998). *Public policy and program evaluation*. New Brunswick & London: Transaction.

Weick, K. E., & Bougon, M. G. (1986). Organizations as cognitive maps. Charting ways to success and failure. In H. P. Sims, et al. (Eds.), *The thinking organization: Dynamics of organizational and social cognition*. San Francisco, CA: Jossey-Bass.