Evaluation and the social construction of impacts

Frans-Bauke van der Meer
Erasmus University Rotterdam

Biographical note

Dr. F.B. van der Meer, associate professor of Public Administration, department of Social Sciences, Erasmus University, P.O.Box 1738, 3000DR Rotterdam, The Netherlands. Phone: +31 10 408 2137; Fax: +31 10 408 9099; E-mail: f.b.vandermeer@BSK.FSW.EUR.NL. His research interests include public and public-private management and organizational learning. He has also been working in the field of social simulation and on organizational aspects of informatization.

Abstract

The central question of this article is when, how and what do governmental agencies learn from evaluations. A structural constructivist theoretical framework is developed and applied to two case studies, in both of which a report of the Dutch Court of Audit is taken as a starting point. A reconstruction is made of the intra- and interorganizational processes in which the impacts of these evaluations were socially constructed. It appears that an evaluation hardly has any direct effect that can be unequivocally ascribed to it. Rather, evaluations seem to support or counteract debates, tendencies and options already present (or 'under construction') in the interaction among actors involved. Using a structural constructivist theoretical framework we identify mechanisms and conditions that enhance forms of learning processes.

The paper concludes with some hypotheses about the genesis of evaluation impact.

Keywords

organizational learning, evaluation, social construction, Court of Audit, privatization, defense procurement

Introduction

Evaluation studies in the public domain may have different functions. One of them certainly is improvement of the object evaluated, which may be a policy or a (semi-)governmental organizational unit. That is, evaluation may trigger changes in policy making, management, procedures, implementation strategies, and so on. But when and how do they do so? Even if goals are fixed and results can be assessed unequivocally, it is not at all self-evident how evaluation results will be utilized. When are they recognized as relevant? When do they trigger defensive or proactive responses? How are the translated into behaviors directed at 'improvement'? Moreover, very often, certainly in the sphere of public administration, both conditions (fixed goals and unequivocal results) are not met.

Research on utilization of evaluation has identified a number of conditions that appear to stimulate utilization in one way or another. Examples are the timing of reports in relation to the actual stage of a policy process (Mulder, Walraven et al., 1991) (Rist, 1994: 194–199) the source of evaluation data (internal or external), the credibility of the source, the way of communicating results (Rist, 1994: 200–203), the quality of the presentation (Pröpper and De Vries, 1995), and whether or not there is an institutionalized evaluation practice (Leeuw and Rozendal, 1994). Oh (1996) develops a model in which search and utilization of knowledge in governmental bureaucracies is explained from characteristics of political issues, the organization, the individual and the information concerned. Patton (1997) points at the importance of user involvement in formulation central evaluation questions, choosing methods, etc.

However, such an inventory of conditions in itself does not explain why evaluation results are used in specific instances and not in other ones. On the one hand, not all conditions mentioned need to be met for utilization to occur. On the other hand even when a number of favorable conditions are in force, utilization is not guaranteed. Moreover: what is 'utilization'? *How* are evaluation results made use of? Policy makers whose policy is being evaluated may react in a defensive way and try to avoid scrutiny of the evaluator. But they may also take evaluation studies serious and act accordingly, e.g. in conformance with suggestions the evaluator made. Maybe they take the evaluation serious, but draw quite different conclusions and consequences from it than the evaluator did. Evaluations may be used to legitimate and 'sell' policies or they may be used to change policies in a superficial or fundamental way. What does determine the actual reaction to and application of evaluation results?[1] I try to contribute to an answer by conceiving of the influence of evaluation in terms of learning processes. More specifically I will focus on the analysis and explanation of learning processes in governmental units induced or enhanced by Court of Audit evaluations.

In the next section, I will elaborate the concept of organizational learning, and argue why I think it fruitful in connection with the present research questions. Next, I introduce a 'structural-constructivist' theoretical framework, which will be used to describe and analyze learning processes in two case studies, which are summarized in the sections to follow. Both cases take an evaluation study of the Dutch Court of Audit as their starting point and try to reconstruct related learning processes in the governmental units most involved. Finally I draw some conclusions on the relation between evaluation and change on the organizational level.

Organizational learning

Evaluations may trigger learning on the individual level. However, in discussing the impact of evaluation in and on public administration, we should focus on transformation in the organization or behaviors of governmental units and hence on organizational learning.

Organizational learning (Cf. Argyris, 1992) may be defined as any process of change in organizational structures, codes, or practices that is triggered or reinforced by new experiences, new interactions or new information[2]. Note that this definition does not say anything on the process itself, but only, in part, about its inputs and its results. In latter sections we turn to the question how these types of inputs are used to produce these results.

Thinking about the impact of evaluation on governmental behavior in terms of learning processes has a number of advantages:

- evaluation generates information and perhaps new interactions or experiences for the unit evaluated. These 'inputs' are implied in the concept of organizational learning;
- the notion of learning processes draws our attention to the *processing* of inputs. Learning processes imply sensemaking and the development of new practices, often piecemeal and by trial and error;
- it directs our attention to changes at the *organizational* level. Individual learning may reinforce organizational learning, but will not necessarily do so. It is therefore relevant to gain insight in when and how.

There also is a disadvantage. It concerns the danger of too easily thinking of organizational change in anthropomorphic terms. Some (Senge, 1992) hold that organizational learning requires learning by individuals within the organization. But, individual learning is certainly not a sufficient condition for organizational learning and perhaps not even a necessary one. Hence, I do not

consider individual learning to be a valid metaphor for organizational transformation, but try to understand how collective processes of sensemaking en constructing of new behaviors take place[3].

A structural-constructivist approach

When and how do governmental units learn from evaluations related to their organization or performance? It seems useful to decompose this question in a number of smaller ones:

- when and how are evaluations identified?
- when and how are they read or listened to?
- how are they interpreted?
- how do they become related to organizational characteristics or organizational behavior?
- how do they eventually find their way to changes in organizational structure or behavior?

Implied in all these 'steps' are two important ideas. The first is the notion of *interpretation*, or perhaps better *the construction of meaning and behavior*. Giving meaning, if only implicitly, is essential, because it determines whether an evaluation is recognized as relevant, how ideas of how it may be used are generated and how these are translated into actual practices.

In research on the impact of evaluation, the role of giving meaning (or *sensemaking* (Weick, 1995)) and *social construction* of behavior, is often neglected. This is the case in research in which one tries to explain the impact of evaluation from characteristics of the evaluation itself (its 'quality', its presentation) or from structural conditions (timing, embedding in institutionalized procedures) (Cf. Leeuw and Rozendal, 1994; Rist, 1994). In such research giving meaning is (implicitly) considered irrelevant, or, more probable, unproblematic. In the latter case it is presupposed that evaluation results are unequivocal and have the same meaning for all actors involved. This not only is at variance with much recent social scientific insight, but in the case of the impact of Court of Audit reports, also demonstrable incorrect (see cases below).

The second basic idea underneath the steps above is that there is a connection between sensemaking and organizational patterns. These patterns condition and direct sensemaking processes (but do not determine them fully), while at the same time sensemaking generally will be reflected in some measure of change in organizational culture, procedures, technology or structure.

Now, my thesis is that, in order to explain the impact of evaluation, we need a better and more detailed understanding of processes of the construction of meaning and behavior within

(and between) organizations. We also need to understand the way in which these processes are influenced by and influence organizational patterns. Structural constructivism[4] offers a fruitful framework to pursue this mission.

Structural constructivism starts from the notion that actors (individuals, groups, organizations or even networks) can, at any moment, be characterized by a *repertoire*. Repertoires are defined as stabilized ways of thinking and acting (on the individual level) or stabilized codes, operations and technology[5] (on the other levels). The concept of repertoire is akin to 'frame' (Giddens, 1984: 87) 'scheme' and 'definition of the situation' (e.g. Hewitt, 1984: 75-85, 139-151). Nevertheless I prefer 'repertoire' since it has a stronger connotation with behavioral patterns next to cognitive ones. When dealing with patterns on (inter-)organizational, these behavioral components, as they are encapsulated in procedures and technology, seem to be the even more important than their cognitive reflections (Cf. March and Simon, 1958: 177, who also use the term repertoire more or less in this sense; Allison, 1971: 72).

The theoretical ideas about the genesis, interrelations, functioning and change of repertoires can be summarized as follows:

- *repertoires are the residuals of preceding interaction processes* in which meanings and behaviors are constructed and have acquired a measure of self-evidence. The Court of Audit is characterized by certain procedures, normative frameworks etc. that reflect a developmental process over time. Also governmental agencies have their acquired ideas on aims, policy instruments, limits of steering, role of other actors etc. , which will color their perception of and reaction to evaluations
- *different actors have different repertoires*, because they have different histories, different experiences, different positions, and different relations. Thus, the Court of Audit and a governmental agency may apply different criteria to judge the effectiveness or legitimacy of a policy.
- *repertoires are used in the process of sensemaking and construction of behavior*. Individuals and groups apply their existing ways of thinking and acting in order to give meaning to new inputs and to react upon them. This implies that impacts of evaluations are not determined by the logic of the evaluator, but by the repertoire based interpretations and (re)actions of the agent involved.
- Since different actors have different repertoires, *they continuously produce behaviors that are not self-evident for (some) other actors*. To a certain extent these other actors are confronted with equivocality. A governmental unit receiving an unexpected evaluation is

puzzled: what does it mean? The unit will be inclined to reduce this equivocality in order to give meaning to the evaluation and to react to it. By definition the existing repertoire is insufficient in this connection. The repertoire will change somewhat by consequence[6]. This process of change can be called 'learning'.

- *Repertoires are connected*. Individuals generally are involved in multiple, partly related social contexts. A minister functions in political circles in which party ideologies, parliamentary majorities and accountability probably are central in the dominant repertoire. On the other hand the minister also is an administrator and perhaps to a certain extent even a manager focused on economy and effectiveness. And he also is a partner in consultations with societal groups and organizations, which may involve repertoires in which cooperation, and harmony and support may be central themes. In reacting on evaluations, all these repertoires will play a role. On the one hand this implies that different repertoires become linked to each other. On the other hand it follows that, at the group or organizational or network level the dominant repertoire is never fully shared by the constituting individuals, subgroups, departments or representatives of organizations. We say that different individuals or subgroups are *included* in different repertoires to a larger or smaller extent. This implies that in processes of sensemaking at the group and organizational level – and perhaps even at the individual level – equivocality (or multiplicity) is not only coming from other actors, but also from within: the process of sensemaking (reducing equivocality) also produces equivocality.
- Repertoire elements are *anchored* in different ways. Repertoire-elements often will be not only anchored in the isolated repertoire concerned, but also in other repertoires and hence in relations with other actors. On the one hand this is a consequence of the interconnection between repertoires discussed above. The results of an evaluation study are not only made sense of from the dominant repertoire of the governmental unit, but also from individual repertoires (perhaps referring to the personal career) or from the perspective of consulting bodies in which the unit is represented. On the other hand the anchoring of repertoire elements in other repertoires is a consequence of the fact that third actors always play a role. Sensemaking and construction of reactions within the unit never takes place in isolation. At the very least it will anticipate upon reactions of other actors in the policy area. And when the evaluation study is public (as is generally the case at some point of time with Court of Audit investigations), actual reactions of other actors are available during the sensemaking and construction process, which will be influenced consequently.
- Learning requires loosening of anchors and/or the development of *new links*. New links in the sense of new connections between repertoire elements, or in the sense of including new

elements in the repertoire, seem to be more probable if there is a certain overlap between repertoires. In our case, this concerns the repertoire of the Court of Audit and the one of the governmental unit, e.g. concerning the framework for judgement of performance. But again, constructing the impact of evaluation is not a bilateral process. It involves other actors as well. One can imagine that a governmental unit considers the results of an evaluation rather irrelevant (because it comes from 'an other world'), but that the evaluation at the same time gets much positive attention in press or in Parliament. Changes in the repertoire of the focal unit may then not be a consequence of the content of the evaluation as such, but of meaningful reactions of third parties.

The implication of this argument is that we need to have an image of the major repertoires (by definition plural) in the policy area, if we are to understand when and how evaluation has which effects. But, although this may help to interpret results in retrospect, our ambition should be a bit higher. How is it determined which elements of evaluation get attention, what are the meanings they get, and when and how does which translation to organizational behavior take place? It is to these mechanisms and conditions that our[7] research project is devoted. Two of the cases we studied are described below. They will be explored in terms of the theoretical framework outlined above, in order to generate a number of more specific hypotheses on the impact of evaluation on governmental learning.

The Court of Audit

The Court of Audit (CoA)[8] is an independent institution, with constitutional auditing tasks and authority. Originally its main task was to ascertain whether government expenditures were in accordance with the law. Gradually new elements were introduced, such as assessments of efficiency, effectiveness, and effects of specific policies and audits on organization and management. These new types of activities have been anchored in successive adaptations of the Government Accounts Act, in part as a canonization of existing practices (Dolmans, 1989; Leeuw, 1992). The Court of Audit proper consists of three members. The supporting organization (almost 350 people) is divided in a number of sections, each heading a number of departmental bureaus. These latter units, which perform most of the basic research and maintain most of the contacts with the departments of government, are in most cases physically located inside those departments.

The usual procedure in efficiency/effectiveness/effect studies starts with a proposal for (further) study by the departmental bureau. Such a proposal is based on the one hand on their

contacts with the department, observations of what is going on, or monitoring what happens with previous CoA studies. On the other hand the strategic plan for a five-year period and the year plan of the CoA are the framework within which the proposals are made. If the Court endorses the proposal, the research is done, using written materials from the department, other internal or external evaluations and audits, interviews with actors within and outside the department, etc. The resulting report of findings is checked with collaborators of the department. This often leads to adaptations in the report. After the Court has given its consent, the Minister involved is asked to react. The results are summarized in periodical reports (such as the year report) or sometimes in a separate, more extensive publication. In both cases the CoA includes a short account of the reaction of the Minister and its own comments on this reaction in the final report, which is sent to the Parliament and sometimes gets substantial press coverage.

The Lynx[9]

In 1974 the Dutch Minister of Defense planned the procurement of 36 'standard' helicopters for the Royal Dutch Navy. In the same year the Westland WG-13 (Lynx) was chosen. Between 1975 and 1983 24 Lynx helicopters were actually bought, in three versions (A, B and C). The last 12 were postponed and later canceled because a new project to replace the Lynx by a new helicopter (NH-90) was already underway.

On March 10, 1987, the Court of Audit (CoA) published its report on procurement and availability of the Lynx. The findings of the CoA can be summarized as follows:

- There was no clear set of criteria and no clear plan in the procurement process.
- There was no real (price) competition between producers, because of too quick a choice for the Westland Lynx.
- Standardization was a failure, because three different versions were procured and no international coordination of procurement took place.
- Information to Parliament was fragmentary and incomplete. Initially NGL 192 million was budgeted for 24 helicopters, while eventually NGL 361 million was paid. These facts were only reported to Parliament in bits and pieces. It was unclear whether these amounts included sensors, weapons, spare parts, documentation, training etc. There never was a total overview.
- Bad arrangements concerning price adaptations.
- An alarming lack of readiness of the helicopters (less than 40%) and low availability of spare

parts.

Did the investigations of the CoA and their results evoke learning processes within the Ministry of Defense, and if so, which and with what observable results? At first sight this looks like an almost unanswerable question. If a change takes place after the publication of a report, it is not at all sure that it is a *consequence* of the report. One day after the report appeared the Ministry issued a situation report on the Lynx. But why did it do so? There may be other reasons then the CoA report. Even if there is a change that seems congruent with the findings of the report, a causal connection is not self-evident. On the other hand if a change precedes the report, that change might be the consequence of anticipation of it. Even if changes prior to or after the report are not congruent with its contents, they still may be triggered by the report. Moreover, policy changes, or changes in organizational practices hardly ever have only a single cause, while on the other hand a single input (the CoA report) may have differential – and even contradictory - effects on different actors or subgroups.

But the theoretical framework presented above does not suggest to search primarily for direct correspondence between the contents of an evaluation report and changes in policy or management, but draws our attention towards the social construction processes of these changes. Below I present a partial report[10] on the reconstruction that we made of the Lynx case. We proceeded as follows:

- From the CoA report, Defense Budgets, Parliamentary and other public documents, we built a picture of the major developments, debates and measures with respect to procurement and readiness of the Lynx, just to have a starting point and a 'context' for our further research.
- After that we studied the CoA-archive with respect to the report. From this we gained an image of the steps and interactions preceding the report, including the exchange of information and views with the Ministry and with the Navy.
- On the bases of the preceding two points we made tentative list of relevant actors and a list of issues we would like to have their view on, because these might mirror the relevant parts of their repertoires.
- Next we conducted interviews with different actors[11] in order to complement our images of the relevant developments, the relevant repertoires of the actors involved, their perceptions of and reactions to the CoA report and to the reactions of others. In doing so we also complemented our image of changes in repertoires and how they came about.
- Finally, based on the preceding steps and utilizing the structural constructivist theoretical

framework, we made a tentative explanation of the governmental learning processes related to the inputs generated by the CoA.

Below I summarize this reconstruction in two steps. First, I give a short description of the repertoires of a number of key actors, viz. the CoA (and more specifically Bureau Defense of the CoA), the Navy, and the political leadership of the Ministry of Defense. Second, I describe and explain the reactions of the latter two actors and their consequences, referring to the three repertoires. Of course there are more actors and more repertoires. Reference to these will be made when necessary. The same holds for sub-units within the actors just mentioned. But a reconstruction on the level of interaction between (and within) these three actors will suffice to demonstrate the approach and to generate a number of fruitful conclusions and hypotheses to be formulated by the end of this paper.

The tables below summarize some characteristics of the dominant repertoires that appeared to be relevant in our case study. We derived them from utterances of members of the units involved, and – as far as the CoA is concerned – also from written reflections (Bemelmans-Videc, 1998)

Court of Audit (Bureau Defense)
General: • The CoA is independent and bases its assessment on factual data. • The CoA chooses subjects, timing and forms of presentation that will enhance actual use of the reports in policy processes. • The main aim of 'value for money' investigations[12] of the CoA is to assess the extent to which initial policy goals are realized, the economy of spending, and the efficiency of implementation. • Clear information and control are essential to realize value for money. • CoA investigations have a preventive and sometimes a correcting effect. *Bureau Defense* • Good relations with the units evaluated (in this case the Ministry of Defense and the Navy) are necessary as a precondition both for obtaining relevant information and for a constructive reception of findings. And in fact the relations are considered good. • The Navy is inclined to give top priority to technological sophistication. This is understandable, but the CoA should remain critical at this point.

Navy
• Navy operations are core business. Of course, in a democracy, giving information to parliament and financial accountability etc. are necessary. But these should not preempt effective and efficient operations management. • Quality and efficiency means buying the best technology that can be afforded with available budgets. • The tactical and strategic concepts of the Navy are not mere ideas, but have been anchored in the past in the military hardware procured. • Procuring advanced military systems is not buying from the shelf. You hook on to a development process that is underway. It is a sellers market. Deviation from contractual specifications is often necessary, but unavoidably leads to much higher costs. • CoA evaluations generally are useful, though they often lack expertise.

Political leadership
• It is always necessary to weigh different interest (which may be military-strategic in nature, or related to foreign or industrial policy). • Parliamentary support is paramount. • (From the beginning of the eighties) reduction of budgets, control of spending. • Gradual introduction of new internal control mechanisms for defense procurement. • Fear for negative publicity. • Navy is very capable and has high expertise. But they are rather 'closed' and hard to control. • CoA investigations are primarily useful for civil service and perhaps for the Navy, but not so much for the political leadership.

General reactions to the CoA Lynx report

According to our Navy informants, the CoA report was a traumatic experience to the Navy[13]. There was a kind of double reaction. On the one hand terror: 'Is it so bad?' And on the other hand resentment 'We are judged far too negatively'. Also fear arose with respect to the public image of the Navy. These general feelings contributed, according these respondents, to an increased attention for presentation and accountability of procurement projects. There were also pressures from the political level in more or less the same direction: budget control and businesslike procedures. But still, these activities kept being seen as an extra burden that tends to interfere with the core operations.

These reactions from the Navy are quite understandable from the professional self-image anchored in its dominant repertoire. Although the Navy does not share the normative framework applied by the CoA (see also below), a number of CoA comments clearly hook on to key notions of good and reliable management and of the professional image the Navy is proud of. The consecutive behavioral change may, therefore, well be interpreted in terms of organizational learning in which part of the repertoire (concerning accountability, transparency, contacts with politicians and public relations) changes to maintain and strengthen other parts of the repertoire (professional control, advanced technology, quality and efficiency).

With the political leadership the report also induced a double reaction. On the one hand the Minister explicitly recognizes imperfections in the procurement process and level of readiness. On the other hand, with a view on the international and technological circumstances, the political leadership had the impression that an alternative course of events would not have been really possible. So the CoA report seems to have confirmed the idea that the Navy was hard to control and that improvements in the procurement process (introduced mainly because of the need for tighter financial control) justified. At the same time the repertoire of the political leadership implied much confidence in the Navy and a high sensitivity for difficulties in decision making and planning in a dynamic environment. From this it is quite understandable that much of the CoA critique was considered too exaggerated or even beside the point.

These general attitudes with respect to the CoA report seem to play an important role in the verbal and substantial (re)actions on specific points, which we found in documents and heard in interviews.

Planning

The CoA critique that there were no clear goals and by consequence no clear planning at the start of the Lynx procurement project is quite understandable from the CoA repertoire. Since there is no clear goal, its attainment can hardly be assessed. Besides, from the point of view of the CoA the procurement process is very whimsical, which means that it is hard to assess and evaluate the actual Lynx procurement process. This is even more problematic since there is no comprehensive financial overview.

To the Navy, on the contrary, a relatively open goal formulation at the outset is an asset, because it allows for flexible adaptation to new developments and circumstances. This is in conformity with the drive to procure the most advanced systems. Hence the Navy tries to keep this openness.

The Minister and his Deputy essentially agreed with the substance of the CoA critique at this point. They claimed however that improvements were underway. New procurement procedures for large projects were implemented (Dutch acronym: DMP). These improvements, according to the Minister, were initiated independent of the CoA investigations[14]. But the feeling that the Navy is hard to control still existed in the political leadership. Here we see that the CoA reinforces a repertoire change that is already underway, be it partly for other reasons.

Competition

The CoA report argued that economy in spending of public means implies organizing maximal competition between suppliers, and that, by consequence, an early choice for one vendor is ill advised. This position is consistent with the focus on economy, which is central in the CoA repertoire

The Navy considered this point nonsense: 'You don't buy a (military) helicopter in the supermarket, you invest in a development trajectory'. At the time there was, according to the Navy (and the then Minister), simply no alternative to the Lynx development trajectory. We could not trace any substantial action or change in or by the Navy related to this point. There seems to be no new connection between repertoire elements in this respect, which can be understood from the very large differences between the repertoires concerned

Although the political leadership was subtler in its reaction (they claimed that studies performed before the Lynx was chosen can be considered as applying competition to a certain extent), they too kept holding that there was no alternative.

Standardization

With respect to standardization there was a more or less clear goal at the start: both internal and international standardization should be striven for. From the CoA repertoire it is natural that the CoA made a point of the fact that neither form of standardization was realized.

The Navy consideres standardization as a value in its own right, but it is not given highest priority. Adaptation to military-strategic and technological developments is considered at least as important. The Navy feels that sensible spending of public money is buying the best you can get. This way of thinking is so strongly anchored in the repertoire that it often appears in mere *technical* terms: 'The Lynx-A *appeared* too light to carry the sensors and weapons *required*' (my emphasis – FBvdM). Therefore, even in retrospect, within the Navy one does not

see how the procurement process (three versions) could have been otherwise.

The political leadership had two somewhat contradictory reactions at this point. On the one hand it felt that the CoA was unfair in giving no due attention to technological and political developments during the procurement process. To their mind, the Navy should not be forced to work with systems that are technologically and conceptually outdated. On the other hand standardization is considered desirable from the point of view of efficiency, reliability, readiness and cost control. This combination of reactions, which *both* are clearly related to their repertoire, lead the policy makers to an effort to standardize ex post. In 1986 the Project Unit for the Maintenance of the Lynx[15] was charged with this task, which was also intended to extend the operational period of the Lynx. The CoA did not suggest this solution. Nevertheless the CoA investigations may well have played a role[16]. It is also probable that Parliament played a role in this connection. Parliament made many remarks on financial aspects of the Lynx project, but also on the standardization problem. It asked why the political leadership did not interfere in an earlier stage. From the repertoire of the political leadership, it seems natural to take measures by consequence, if only to restore the image and trustworthiness of the political leadership.

Information to Parliament

The CoA stated that Parliament never has had complete and clear information on the financial aspects of the Lynx procurement project. From the financial control role of the CoA (and Parliament!) it is clear that the CoA considers this a major point.

From the perspective of the Navy, on the contrary, this initially is a less relevant problem. Nevertheless the Navy recognized that Parliament has the right to get more information. Besides, responsible functionaries within the Navy perceived that Parliament and political leadership were quite sensitive to the CoA critique at this point. The Navy therefore felt forced to provide more information. Although they kept considering it as an extra burden, they gradually discovered an unexpected advantage of increased interaction with members of Parliament: it increased the lobby opportunities for the Navy and enhanced early commitment of key members of Parliament to Navy projects. Eventually a clear change in external information and relations management by the Navy can be observed. Again we see a change that is not explicitly advocated by the CoA, but seems to be enhanced, if not triggered by it.

The Minister claimed (at least officially) that information to Parliament had been sufficient in all stages. Nevertheless gradual changes in the amount and nature of information to Parliament can be observed, probably to maintain support of Parliament, which is a key issue in

the repertoire of the political leadership.

Price control

The control of spending and prices, was also a major point of attention in the CoA repertoire.

The Navy, however, was not impressed by CoA critique in this respect. It was used to substantial increases in prices, high cost of adaptations in design and high cost for building in equipment. This has to do with the nature of technology and technological development and with the suppliers controlled market. Moreover inflation and exchange rates play a substantial role. Finally the attitude of the Navy, in contrast with the standards of the CoA, is not to buy things as cheap as you can, but buying the best with the money that appears to be available.

The responsible politicians tried, both on their own account and under the perceived pressure of the CoA and Parliament, to gain more control of the financial aspects of procurement processes. The procedures of DMP are an example in case. Moreover, already in 1984 a Contract Commission was erected within the central core of the Ministry. This Commission played, according to the Deputy Minister, an increasing role in the negotiations of new contracts. This amounted to frictions between the central department and the Navy with respect to the question who actually is (responsible for) negotiating procurement contracts. Here we see that CoA critique (reinforced by Parliament) links easily with notions in the political leadership (budget control, control of the procurement process) and influences the relation between Ministry and Navy in the direction of more openness and control of the latter.

Readiness

The very low readiness of the Lynx reported by the CoA hitted the Navy in its core. The first reaction was one of unbelief and irritation. According to our Navy informants, the CoA appeared to lack expertise. For example, many disturbances and failures were caused by the relatively high trembling of the Lynx as a consequence of its rotor suspension, and so unavoidable. The CoA reacted by pointing out that this probably demonstrates that the Lynx was a bad choice. But the Navy was supported in its image of the lack of expertise of the CoA: the specific rotor suspension made the helicopter very maneuverable, a splendid machine! The high failure rate should be taken as necessary costs. The CoA critique on the bad stock keeping of spare parts was also rejected initially. Spare parts are very expensive, in part they were not even developed, and you cannot know in advance which parts will have to be replaced. The Navy saw the lack

of a sufficient number of crews as a consequence of training problems, certification and competition with civil aviation.

But still, the CoA critique hurted because the Navy cannot escape the conclusion that essentially the CoA was right: a readiness of about 40% is far too low. A number of measures were taken. The first one was primarily presentation: the definition of readiness was changed, which lead to somewhat higher percentages. But more fundamental was a large-scale reorganization of the maintenance activities in order to perform faster, well structured and mission oriented maintenance. Finally cooperation was sought with a number of other European countries to create a joint pool of spare parts. Here the CoA critique obviously links with the professional self-image of the Navy, which explains the dual reaction of discrediting the critique and reacting to it by changing organization and procedures.

Intermediate conclusion

The Lynx case shows how the theoretical framework of structural constructivism can be used to account for (lack of) organizational learning related to CoA evaluations. Based on the analysis a number of conditions and mechanisms producing certain forms of organizational learning can be identified. However, before doing so, it seems useful to enlarge the empirical bases for some tentative hypotheses by giving a condensed description of a second case study in which some other mechanisms appear to play a role.

The state museums[17]

The report

In October and November 1987 the CoA investigated the performance of the state museums with respect to conservation of cultural objects and with respect to their tasks directed at the general public. These were the main officially stated task of the museums. Next to that, the CoA explicitly devoted attention to the quality of the management of the museums.

Within a period of three weeks all state museums were investigated, on the one hand to prevent anticipation by the museums and on the other hand in order to be able to report quickly. This approach required working with parallel teams, in part consisting of CoA officials usually dealing with other policy fields.

The report, eventually published in September 1988, was quite critical. It mentions seri-

ous problems with respect to registration, conservation, management, security, personnel and financial relations with supportive foundations. Also the public service functions in a number of museums are judged to be below standard. The CoA asks for more attention for conservation, better management and uniform and transparent financial regulations. The CoA thinks a '(direction of) solutions for the problems identified' a necessary condition *before* museums can be granted more autonomy.

The CoA repertoire

The table below summarizes core CoA repertoire elements.

Court of Audit
• assessment of realization of initially formulated goals • good managerial and financial control mechanisms are an important guarantee for good policy implementation • uniform and transparent regulations are important • timing of reports to induce actual use and influence in the political process

It is interesting to note that during the eighties a clear development took place in the CoA repertoire[18]. Both 'political' timing and attention paid to management and control are new elements, which are reflected in the approach in the museums case. The timing of the report is directly and explicitly related to the actual discussion within the ministry on privatization of the state museums. The focus on the organization and management of the museums reflects both this current political issue and the growing CoA interest in institutionalized control.

The impact

Within the framework of this contribution I limit myself to an account of the reaction of the Minister to the report and the use he made of it. In doing so, I try to explain his policy from his repertoire and the administrative interactions in which he was involved.

In his formal reaction, the Minister agreed to a very large extent with the CoA report. He considered the arrears in conservation 'very alarming' and he did not deny the other problems identified by the CoA. However, he felt that the CoA to a large extent did not consider the causes of these problems, nor the impediments for their solution. He thought that the fact that

different Ministries are involved in the financial, control and policy aspects of the museums creates much inefficiency and lack of efficaciousness, because of bureaucratic procedures and equivocal steering. In general, the fact that the state museums are part of public administration hinders improvement, amongst others because of the lack of financial incentives for new creative initiatives and public oriented activities. From this interpretation of the problems, more autonomy appeared as a sensible solution. The Minister was determined to move in that direction.

Note that the Minister buys the problem assessment of the CoA, but is heading for a solution quite contrary to the CoA argument. Both elements of this reaction can be understood from the repertoire of the Minister and his staff and the meaning the CoA report acquired in their interpretation of it.

The Minister's reaction reflects changes in the repertoire of the Ministry taking place in the years preceding the CoA report. I summarize the main developments[19].

- *debate on core business*. From the beginning of the eighties, a debate developed, especially from within the Ministry, with respect to the question what the main tasks of the state are in relation to museums. Initially this debate was linked to efforts to separate the responsibilities of different levels of government in this connection, and to cuts in the state budget. But later on a new element came on the agenda. It concerned the growing uneasiness associated with the combination of a general responsibility for culture policy and a specific management responsibility for a selected number of museums. From these elements in the debate the notion was constructed that the Minister only has a responsibility at the *macro* level (including conservation of the total national cultural inheritance) and not on the *micro* level of museum management. Some groups within the Ministry however sticked to a more comprehensive role of the Ministry.
- *autonomy and privatization*. The idea of more autonomy for the state museums, increasingly discussed within the Ministry and more and more in public, was connected with budget cuts as well. A decision to privatize the Dutch Open-air Museum was primarily anchored in this consideration. But soon the idea that more autonomy may enhance the performance of museums and solve many problems accordingly, acquired a central place in the debate. The factions within the Department promoting privatization were gaining ground, although other factions resisted the idea. Within the circle of the state museums there was dissension: some museums were quite enthusiastic about the idea, others were more reluctant. The former expected reduction of bureaucracy and room for own policy development, the latter feared that government would drastically reduce its (financial) responsibility.

- *conservation*. Already in 1984 the Ministry recognized arrears in conservation of national heritage. In the following years, the awareness of the problems kept growing and the assessment of the situation became increasingly negative. The Ministry related the problem to the low management quality in the museums on the one hand and to the lack of financial means on the other.

In sum, in the course of the eighties, the repertoire of some groups within the Ministry developed in a direction that facilitated links with the definitions of the problems of the museums as the CoA identified them in 1988. However, at the same time a repertoire with respect to solutions developed, which was in some respects at variance with the argument the CoA developed. It seems, then, that the CoA assessment of the shortcomings of the museums strengthened the repertoire and the position of groups with similar problem definitions. Consequently the *solutions* generated by these groups gained strength in the internal debate, which eventually may have contributed to the decision in 1991 to privatize the museums. For, although the CoA had rejected that solution for the near future, the CoA report was frequently cited to support the privatization proposals.

Conclusions

In this paper I have examined when and in what way (outcomes of) evaluation studies are used in governmental learning processes. I described two cases, in which the impact of CoA reports on policy making, control and management was (partly) reconstructed, using a structural constructivist theoretical framework.

Both cases support the central ideas of the theory:

- The impacts of evaluations are constructed by actors, using their existing repertoires.
- These construction processes do not take place in isolation but in a context of other actors and other construction processes. For example the impact of the Lynx report is clearly related to the interaction between the Ministry and the Navy. The impact of the museum report must be understood in the context of ongoing interactions within the Ministry and between Ministry and museums, resulting in new repertoires.
- A decisive point in determining the substantial impact of evaluation lies in the linking of (new) elements in the evaluations to (elements in) existing repertoires and/or in their contribution to loosening existing anchors of repertoire elements. The CoA critique on readiness of the Lynx links to the professional self-image of the Navy, resulting in adaptations not suggested by the CoA. The museum report strengthens repertoires “under construction” within

the Ministry.

- Often elements of evaluation link to the repertoires of certain actors or factions. Consequently, changes in relations between these actors or factions and other groups show up. The consecutive interaction processes determine the eventual links and the further developments. Here the museum case provides a clear example.

In the case studies actual developments were interpreted ex post in terms of social construction processes. However, inspired by the insight the case analysis gave, it is possible to derive more general hypotheses that are consistent with the theory developed in this contribution. When operationalized adequately, these hypotheses can be tested. Here, I confine myself to a number of examples.

Evaluations will have *direct* effects if they link to the existing repertoire of the unit involved. As is clear from the case studies, such links often are partial. There may be some overlap with respect to performance criteria, or with respect to problem definitions, or with respect to policy options, or with respect to assessment of relevant contexts, etc., but not on all those dimensions. The direction and intensity of effects will probably depend on the nature and the measure of overlap. This gives rise to a first hypothesis:

> hypothesis 1: *direct effects of evaluation on units evaluated reflect the nature and measure of links between the repertoires involved: (a) arguments and recommendations based on overlapping repertoire elements will be taken up by the unit evaluated, at least insofar they are not perceived as contradictory with other repertoire elements; (b) assessments made in an evaluation that fit in the repertoire of the unit evaluated, will be taken as support and reinforce the repertoire, irrespective of how such assessments are used in arguments of the evaluator; (c) redefinition of repertoires of units evaluated is connected to overlapping parts of repertoires.*

From the cases it is also clear that units evaluated are no monoliths. This leads to a second hypothesis:

> hypothesis 2: *to the extent that assessments and recommendations of evaluations do link more closely to the repertoires of certain individuals or factions within the*

unit evaluated, or to developments in those repertoires, their effects will be stronger and in the direction advocated by these individuals or factions.

Insofar as my analysis is convincing, it may have a number of consequences for evaluation practices. As I said in the introduction, evaluation may have different purposes or functions. My suggestions (which in a sense are hypotheses as well) only pertain to the function of enhancing learning processes.

- If evaluation is to enhance organizational learning processes, the mode and content of the evaluation should link to the repertoire of one or more actors involved. Linkage can occur via the content of the evaluation, the standards used by the evaluator, anticipation by actors on reactions of other actors, or in other direct or indirect ways. Therefore, it may help evaluators if they gain knowledge of the repertoires of key actors.
- Knowledge of repertoires can also help to initiate explicit discussion on central presuppositions, value orientations etc. that are embedded in these repertoires. This may give rise to another type of learning process than mere discussion of specific behaviors, because it may enhance other interaction processes and other linkages.
- Moreover, knowledge of different repertoires may help to tailor evaluation to trigger unconventional interactions in the policy area, which may give rise to new types of learning processes.
- Evaluators, like other actors, perceive and analyze, using their own repertoire. They take substantive positions accordingly. In general there is nothing wrong with that. Such a position, however controversial, may trigger interaction and learning processes that give rise to changes that many may consider improvements. Looking for new links is something quite different from taking other actors positions without critique. But it is useful to be able to evaluate one's own positions from the point of view of other actors.
- An essential element of evaluation aiming at learning should be to take account of the relevant context and developments in it as perceived by the actors in the policy area. A number of comments in the CoA report on the Lynx (on lack of competition and standardization) were not taken very seriously within the Navy because they confirmed the idea that the CoA lived in another world. Again this point should not imply that the evaluator should take for granted the context as actors see it, but only that they acknowledge that that context is there in the present repertoires of the actors and hence will reinforce their ways of

(re)acting.

Literature

Allison, Graham T. (1971). Essence of decision: explaining the Cuban missile crisis, Harper Collins Publishers.

Argyris, Chris (1992). On organizational learning. Cambridge (Mass):.

Bemelmans-Videc, M.L. (1998). *De Algemene Rekenkamer: controlenormen en -stijlen in een veranderende bestuurlijke context.* In: M.L.M. Hertogh**,** N.J.H. Huls A.C.J.M. Wilthagen (eds). Omgaan met de onderhandelende overheid: rechtsstaat, onderhandelend bestuur en controle . Amsterdam, Amsterdam University Press: 89-117.

Dolmans, L. (1989). Naar supervisie en meer aandacht voor doelmatigheid: de Algemene Rekenkamer tussen 1945 en 1988. In: P.J. Magry, E.C. van Heukelom and A.J.R.M Linders (eds.). Van Camere van der Rekeninghen tot Algemene Rekenkamer. Den Haag, Sdu: 377-431.

Giddens, Anthony (1984). The constitution of society. Berkely, Los Angeles, University of California Press.

Hewitt, John P. (1984). Self and society: a symbolic interactionist social psychology. Boston, Ally and Bacon.

Leeuw, Frans L. (1992). Doelmatigheidsonderzoek bij de Algemene Rekenkamer. Bestuurswetenschappen 46(1)): . 8-24.

Leeuw, Frans L. en Piet J. Rozendal (1994). Policy evaluation and the Netherland's government: scope, utilization and organizational learning. In: F.L. Leeuw, R.C. Rist and R.C. Sonnichsen (eds.).Can governments learn? Comparative perspectives on evaluation and organizational learning. New Brunswick, London, Transaction Publishers: 67-86.

March, James G. en Herbert A. Simon (1958). Organizations. New York.: John Wiley & Sons.

Mastop, J.M. en A. Faludi (1993). Doorwerking van strategisch beleid in dagelijkse beleidsvoering. Beleidswetenschap 7(1): 71-90.

Mulder, H–P., G. Walraven, A. de Groot, F.S. Terpstra, P.Rozendal, O. Delsman, A. van Loon, R. Venderbosch, F.L. Leeuw: Gebruik van beleidsevaluatieonderzoek bij de rjksoverheid, in: *Beleidswetenschap*, 45(3), p., 1991.

Oh, Cheol H. (1996). Information searching in governmental bureaucracies: an integrated model. *American Review of Public Administration* 26(1): 41-70.

Patton, Michael Quinn (1997). Utilization-focused evaluation. Thousand Oaks, Sage.

Pröpper, Igno M.A.M. en Michiel S. De Vries (1995). Presentatie en gebruik van evaluatieon-

derzoek.In: J.Th.A. Bressersen A. Hoogerwerf (eds.) Beleidsevaluatie. Alphen aan den Rijn, Samson H.D. Tjeenk Willink: 158-177.

Rist, Ray C. (1994). The preconditions for learning: lessons from the public sector. In: Frans L. Leeuw e.a. (eds.). Can governments learn? Comparative perspectives on evaluation and organizational learning. New Brunswick, Transaction Publishers: 189-205.

Senge, Peter M. (1992). The fifth discipline: the art and practice of organizational learning. London, Century Press.

Stevers, Th. (1979). De Rekenkamer. Leiden, Stenfert Kroeze.

Van der Meer, F.B. en A.J. Roodink (1991). The dynamics of automation: a structural constructionist approach. Informatization and the Public Sector 1(2): 121-141.

Vissers, Geert A.N. (1994). The production of strategy. Delft, Eburon.

Weick, K.E. (1979). Social psychology of organizaing. Reading (Mass.), Addison-Wesley.

Weick, Karl E. (1995). Sensemaking in organizations. London, Sage.

Notes

[1] Mastop and Faludi (1993) suggest in an article on the impact of strategic policy, that because of the types of complications mentioned in the text, one should concentrate on 'use', irrespective of its direction. I do not agree, for I believe that both theoretical analysis and empirical research has more to offer.

[2] I refrain here from the question whether there exist still other types of organizational change. It may be suggested that some actors have the 'power' to force units evaluated to change, even if those units have not learned anything from the evaluation. But even then, organizational behavior

changes as a consequence of new experiences, information and/or interactions, which I defined as learning. Moreover, from the perspective adopted here, power is only 'operational' to the extent and in the way it is reproduced in interaction.

[3] Perhaps organizational learning is after all a good metaphor for individual learning processes, because the latter unavoidably take place in a social context. This question, however, is outside de scope of this article as well.

[4] For a more comprehensive discussion of the theory see (Van der Meer and Roodink, 1991; Vissers, 1994).

[5] In a sense technology in organizations can be considered as perhaps the most reified and anchored way of thinking and acting: it structures organizational behavior in a way that is often both subconscious and experienced as hard to change.

[6] Weick thinks this process the core of 'organizing' which he defines as 'a consensually validated grammar for reducing equivocality by sensible interlocked behaviors' (Weick, 1979: 3)

[7] In the research I have collaborated with my colleagues Geert Vissers and Gert Jan de Vries. We have investigated learning processes triggered by five Court of Audit evaluations.

[8] For factual information about the CoA and its development over time see Stevers (1979) Dolmans (1989) and Leeuw, (1992).

[9] The following account is based on Parliamentary Documents and interviews. The documents are categorized under the following numbers: 12994, nr. 2 (Defense White book 1974), 13100, nr. 7 (Defense Budget 1975), 14600 nr. 12 (Report of the Minister of Defense, 1978), 18169 nr. 2 and nr. 78 (Defense White Book 1984), 19282 nr. 2 to nr. 5 (Report of the Deputy Minister of Defense, 1985-1987) and 19897 nr. 1 to nr. 12 (CoA report on Lynx, 1987).

[10] A more comprehensive report is in preparation (in Dutch).

[11] Extensive interviews were conducted with a former Deputy Minister of Defense (1981-1989), the head air operations and a PR-officer of the Navy, and three members of the Defense

Bureau of the CoA that were involved in the Lynx report.

[12] In Dutch: 'doelmatigheidsonderzoek', i.e. assessment of efficiency, effectiveness and goal attainment.

[13] The same holds, to an even larger extent, for the CoA report on the procurement of the Walrus submarines, which appeared only one year and a half earlier.

[14] A Navy respondent, however, indicated that the Walrus report was the immediate cause of DMP.

[15] In Dutch: Projectbureau Instandhouding Lynx.

[16] The then Deputy Minister of Defense thinks that probably the CoA investigations were a very important direct reason for the erection of the Project Bureau.

[17] We used a.o. the Parliamentary Documents 19066 nr. 1 to nr. 35 (Memorandum on Museum Policy), 20697 nr. 1 to nr. 10 (CoA report on State Museums) and 21973 nr. 1 to nr. 5 (Memorandum on Privatization of Museums). We also conducted a large number of interviews with CoA officials involved, civil servants from the ministry and directors of museums. An extensive report of our research is in preparation.

[18] The explanation of this change, which can also be interpreted as the consequence of organizational learning processes, is outside the scope of this article.

[19] These developments too, can be analyzed as consequences of organizational learning processes. Again, however, this is outside the scope of this paper.