

Québec City
September 22, 1975

File: PLB 1365-06

"Measure for Measure"

"Money for Value"

or

"Keeping out the Elephants"

Efficiency and Program Effectiveness
in Government Programs

1. It is alleged that a new employee (who shall remain nameless) of the Auditor General's office in Ottawa encountered a known public servant leaning on a lamppost on the Sparks Street Mall. On enquiring as to why a public servant should be lounging around in such an obviously unproductive fashion, the Auditor General's representative found himself engaged in the following dialogue.

- . "I am keeping the elephants off the Mall".
- . "But there are no elephants around the Mall".
- . "Then I am running a pretty effective program, eh?"

The response of the Auditor General's office is not recorded.

2. I wish to talk to you today about the problem of demonstrating "value for money" in the public sector. More specifically, I want to assert the impossibility of any objective audit or conclusive evaluation of the effectiveness of government programs. I plan to spend about fifteen minutes sketching what I recognize to be

a somewhat negative position, and about ten minutes trying to salvage something from the wreckage. I hope to spend the remainder of the time getting your reaction to the dilemmas I have outlined.

3. I should also observe that nothing I say here is intended to conflict with or reflect on the essential principles expressed in the Report of the Independent Review Committee (the Wilson Committee). I have the highest regard for the work of the Committee, and generally support the views they have expressed. But I believe that their observations on the issue of "value for money", and evaluation of public sector programs, are open to misinterpretation, and indeed have been misinterpreted, in much of the discussion of their ideas.
4. Let me say first that I am delighted to be here and to have this opportunity to seek your comments on these matters. As one concerned with evaluation inside a government, I have probably developed a different point of view from your own, as legislative auditors. I probably also work at a somewhat different end of the analytical spectrum I am going to mention. But I expect in the end you will recognize the problems I describe, and perhaps you will have different resolutions of them.
5. My argument, in brief, is as follows:
 - (i) In government operations, no conclusive evaluation according to generally accepted practice is possible;

there is no analog to the famous "bottom line", and without it there is no demonstration of "value for money". I believe we delude ourselves in talking as if there were, or as if a determination that "value for money" had not been received could be made without the subjective weighing of conflicting social goals that is the unique responsibility of Ministers in our system of government.

- (ii) Moreover, the GAO in the US do not, contrary to some suggestions, do effectiveness evaluation as I understand the term; nor do any other external auditors I know about. Indeed, even the beloved bottom line is rather suspect and subjective when one comes to look at it (as you probably know better than I).
- (iii) On the other hand, every level in the hierarchy or accountability chain of any decentralized system requires some flow of information laying out the evidence as to the operation of the programs concerned.
- (iv) The House, and the public, are the ultimate levels in the chain of accountability for government programs, and require the most penetrating (though also most aggregate) information flow - and this need entails the full disclosure of all relevant information on the factual basis and consequences of government programs.

- (v) Within this system, a legislative auditor presumably has the responsibility either to provide such factual information, or to attest that the government has the capacity to do so and indeed is doing so in an objective manner.
6. In other words, my argument is that there cannot be any objective determination of "value for money" in the context of the public sector, but there must be greater disclosure of the factual evidence bearing on the performance of government programs. (For this purpose, I interpret "factual" to mean information which could be replicated by other analysts operating from the same data and following the same procedures.)
7. Before proceeding, I must emphasize that this argument is purely a personal view, the view of an outsider in the federal government. I do not know what the view of the federal government with respect to the role of the Auditor General may ultimately be, nor do I have the responsibility for advising on that matter. The views expressed here are individual views for which I alone am responsible.
8. Speaking very generally, there are at least three reasons for evaluation: first, to determine if existing policies and the instruments for their implementation are appropriate; second, to analyze the practicality and potential effect of new initiatives; and third, to

strengthen accountability relationships by reviewing the fulfillment of past commitments. The first two are intimately bound up in the planning and budgetary processes. Current measures can be assessed to determine if they are satisfactory. If they are not, alternative ways of achieving desired changes can be studied. When evaluation is undertaken for these purposes, consideration is given to the other side of the "value for money" coin, namely, how to expend "money for value". This kind of analysis is of particular concern to the Treasury Board because of its duties in the overall resource allocation function. Indeed, most of the evaluation now done at TBS is of a prospective nature - both in the study of existing programs with a view to making changes and in the analysis of proposed programs. Though they turn up information which is inevitably incomplete and imperfect, these evaluations provide useful supplementary information to Ministers or program managers in improving the design or composition of the government's programs.

9. Alternatively, when evaluation is done purely for reasons of accountability it will be in aid of the decision, after the fact, as to whether satisfactory results have been obtained for the resources committed. In this sense, the problem may be one of demonstrating, retrospectively, "value for money". Evaluation done to strengthen an accountability relationship between one party who has delegated, and another who has accepted, certain responsibilities can run the gambit of formality

from the casual question put by the boss "How is it going?" to the assignment of a third party, an auditor, as an independent monitor of progress. This latter choice is made to ensure objectivity and differs in substance from evaluations performed mainly for planning purposes. It will usually, I suppose, take as given the purpose and justification for the task that has been delegated and examine only the way in which, and the extent to which, it has been carried out. Normally, then, such an evaluation will not be prescriptive except possibly in matters of administrative implementation. That is, the report will not usually make recommendations for change, certainly not directly to the party whose performance is under review! It may, however, address the question of whether the purposes set out (e.g., in legislation) are reflected in the way programs and activities are mounted, and whether objectives are achieved.

10. Like Parliament, the Cabinet may wish to review results after certain decisions have been taken and responsibilities delegated. In this event, third-party evaluations of a retrospective nature ("value for money") would be undertaken on its behalf. Indeed, this has occurred to some extent. Such studies comprised the original mandate of the Effectiveness Evaluation Division at the TBS and have been conducted on occasion by interdepartmental committees. Moreover many departments have internal or operational audit capabilities.

11. As Cabinet is not solely concerned with prospective analysis, so Parliament is not only concerned with retrospective analysis. One could speculate that Parliament would find helpful, independent "money for value" assessments to aid in the examination of appropriation bills. There could be various ways of obtaining this service ... perhaps a planning and evaluation unit attached to Parliament? (The creation of the Congressional Budget Office in the U.S. is an interesting precedent in this respect.) What I'm getting at here is simply that it is important to identify the reason for any evaluation which is to be undertaken: whether it is to serve the enforcement of accountability in a relationship where responsibility has been delegated, or whether it is principally to enable more informed decision-making in the future.

Auditors, legislative or otherwise, have traditionally been engaged in ex post analysis and for this reason I will now focus mainly on retrospective evaluation.

12. As in all large organizations, there are many levels of delegation in Canada's federal public sector. Niceties of constitutional law and political science aside, we might say that the electorate assigns Parliament or provincial legislatures the role of guardian of a country's political and social values. Parliament collects revenues, votes monies and enacts legislation to discharge this function. Elected representatives in turn delegate to Cabinet and individual Ministers the prime responsibility for development of policy objectives for Canadian society, the design of instruments by

which to make progress towards these objectives, and, upon legislative approval, the administration of those instruments where implementation by the federal government is required. Cabinet as a whole delegates certain of its management functions to its committees and further to their secretariats (PCO and TBS). Individual ministers delegate responsibilities to the departments and agencies under them. At each level in the hierarchy I have outlined, there are certain sanctions by which lower levels are held to account for their performance. These include elections, votes of confidence in the House, shuffles of Order-in-Council appointments and so forth. The outcome of each is based upon judgements made in the light of performance reviews and prospects for desired change. Good public service, then, in the broadest sense of the term, requires that every level of the hierarchy from the top down ask informed questions of the lower levels and press for some evaluation to provide the information flow necessary to strengthen accountability and aid decision-making. Combined with strong direction from the top such evaluation will do much to ensure that government functions effectively.

13. But one has to recognize physical and philosophical realities: there is no point in expending considerable effort to develop a perpetual motion machine simply because it is badly needed in our energy-scarce world. Likewise there is no point in pretending to assess the "value for money" attained in government programs if that simply cannot be meaningfully done. I would earnestly wish to be able to present conclusive, objective

effectiveness evaluation which stands on its own feet, independent of subjective interpretation. But I believe that cannot be done in the public sector, and I would like to suggest why.

14. Note that the expression "value for money" as a criterion by which to judge government activities was chosen by the Wilson Committee over more technical terms because it appeared "to be the least ambiguous and the most comprehensive of the available alternatives". No serious attempt was made, however, to define this expression more precisely. As I have said, my experience has largely been with the reverse question, "money for value". Taking the words in that order, money is the easiest to understand, especially after expenditures have been made. (It is more difficult, of course, to estimate costs before the fact.) At least money can be counted in unambiguous units and, after disbursement, outlays can be scrutinized and factual financial statements prepared - according to generally accepted practice. (I want to return to this point later.)

15. The notion of "for" - as with all short connectives in any language - introduces a complexity. Money is expended "for" what? In the first instance, for inputs - purchased materials, labour services, discharge of contractual obligations associated with acquisition or maintenance of assets, and so on. Going beyond this first step, however, there is a second - not "for" what object, but "for" what objectives, what purpose? I must back off that question rather hastily for the

moment, however; it is too prickly for the present. A less thorny version is of the form "money" is expended "for" purposes of carrying out what activity? Here we recognize a relatively recent issue in management of public sector programs - the notion of a cost-accounting system.

16. This issue is, of course, at the heart of the reform in the preparation of the Estimates which saw resource allocations charged against specified activities within an accepted program-activity structure for each department. Two difficulties come to mind in this allocation procedure. The first is the vexed problem of allocating joint costs. This problem becomes more severe, the more precisely specified are the activities. The second follows, not from such a conceptual difficulty, but from the practical burden and inconvenience of the record-keeping that is required. The first is philosophically impossible, so may result in arbitrary choices, whereas the second may be merely prohibitively expensive. Most managers in the Public Service still appeal to the second reason.

And then we come to the problem of "value". More trees have been sacrificed to sustain the debate on this concept in philosophy and economics than on almost any other. As you are well aware, one man's "value" is another's extravagance; pornography for one reveals "redeeming social value" for another. In the economy at large we get along well enough with the problem of value because we are willing to let each pursue his own. Indeed, the concept of "money for value" or, after the fact, "value for money" is most meaningful to me when I place myself in the position of a consumer in

the market place. There I have a pretty good idea whether or not I've gotten my money's worth in the goods and services I have paid for. At least I am personally aware of whatever satisfaction a bottle of wine, for example, has brought to me or my guests.

17. But as a voter I would encounter a great deal of difficulty in deciding if I'm getting value for my tax dollars - more difficulty, probably, than in deciding whether I am enjoying good government overall. For one thing, I am not directly affected by many government programs or regulations. I haven't held a LIP job, nor received Unemployment Insurance, nor have I even visited a national museum for some time. Of course, I have gotten protection from the RCMP and the Armed Forces, but how much? And hasn't everyone else? Quite unlike my position in the private sector, I am purchasing certain goods and services with my tax dollars from which I can hope to derive benefits only indirectly - or indeed, through my expectations for my children's future.

18. The point is that government programs are generally mounted precisely because the notion of "results" is so fuzzy, or the divergence between individual valuations and social valuations of the consequences of programs so great, that decentralizing operations through a market mechanism - or any similar unambiguous mechanism for recognizing and rewarding performance - is impossible or unacceptable.

19. It is of course true that the "bottom line", even as information flowing to shareholders whose interests are exclusively (or dominantly) financial, is more subjective than usually recognized. and yet evaluation in the private sector is not abandoned. But the key discipline on the enterprise arising from the external valuation of its services is crucial. Two consequences thus attend the recognition that a good or service must be, or ought to be, produced in the public sector. The first is the loss of this continuing expression of external valuation of "outputs". The second, more important, is the introduction of multiple, conflicting, objectives for the activity in place of a single dominant objective on the basis of which unambiguous incentive rules can be developed, and operations decentralized.
20. I recognize that this sounds like the old defensive arguments, but I believe I am suggesting a slightly deeper argument. Consider, to illustrate my point, how in fact matters would look if the information provided by public accounting firms were suddenly deemed to be of such crucial importance to the continued functioning of the economy that all such firms were nationalized tomorrow. In place of services sold to clients, objective accounting information would be offered to the community as a state-provided service. How then would you assess "results" in the firms you know so well? Could you really monitor productivity and quality change in such a way as to report numerically or abstractly on these matters? Could you really assess whether variances between results achieved and resources expended were developing over a year? Or would not all these issues get caught up, as in the public service, in unanswerable questions of quality changes and appropriate levels of service, and so on?

21. It should be noted that there are other difficulties in the application of the "value for money" concept to the public sector which arise because of the nature of the electoral process. Elections do not entail a separate referendum for each policy that has been pursued or each piece of legislation enacted. Responsibility for taking these individual decisions has been delegated and rests with the elected representatives. Instead, voters cast a ballot for a composite of the status quo and diverse campaign promises. The choice is therefore only vaguely indicative and, with the vagaries of politics, is afforded at infrequent and irregular intervals. These features are in sharp contrast to a competitive market place where daily purchases of individual products constitute votes for the producer. As I see it, only one analogy does hold between buyers in the market place and voters at the ballot box on this question of "value for money" - the ultimate judgments are made by "consumers" in the aggregate, and the more information they can bring to bear the better. Accountability in this sense is a global concept, resting on overall performance, not on the conduct or management of individual programs.

22. Because individual members of the electorate are not themselves in a good position to assess whether they have received "value for money", it is necessary that Parliament assume this responsibility. Moreover, it is in Parliament's self-interest to ensure that the monies it votes are well-spent. More broadly speaking, they would as legislators presumably wish to see that the "instructions" which they have given to the executive,

or the actions they have enabled, have been carried out so that the desired effects are achieved. And one cannot logically limit performance review just to programs which require the expenditure of money; the effects of tax measures and other regulations are equally important, although this aspect has been neglected in most discussions about evaluation. These matters however, are not relevant to the concept of "value for money". Beyond wishing to see that all government programs are achieving what they are supposed to achieve, I presume, like the GAO in the U.S. and the Wilson Committee, that Parliament is particularly concerned with the manner in which programs are implemented - or more precisely, whether they are operated with due regard for economy and efficiency.

23. Parliament will not find ready evidence to turn to for answers. There are basically two reasons for this:

(1) Public accounts do not hold the key as financial statements do in the private sector. There is no equivalent to the "bottom-line" - the net profit figure which would enable a global conclusion to be drawn about economy, efficiency and effectiveness, for profit-making is rarely a consideration in government.

(2) Cabinet and its individual Ministers cannot realistically be expected to undertake, much less to offer up for the consideration of Parliament, objective assessments of their own performance or

that of departments and agencies they head. This is not a criticism, nor is it an expectation reserved exclusively for Cabinet. It applies equally to any level of the hierarchy where responsibility has been delegated. If not required to report and demonstrate good works, no one will unilaterally do so, unless perhaps, there is something to gain. Thus, if the evaluation is generated by other than a third-party and is forthcoming to the higher level, it will normally take the form described at the start of my talk - that is, it will be prospective in nature, designed to support some new initiative. It will form part of the planning and budgetary processes. And, of course, rewards for Ministers and Deputies will reflect the corresponding initiatives in policies and programs.

24. Parliament cannot easily ignore the lack of evidence either, relying perhaps on the presumption that competitive forces are at work to ensure economy, efficiency and effectiveness. They are not. Government holds a monopoly position for most of the things it supplies. It is not subjected to any market test, or to any analog of such a test. Of course, Parliament might rely simply on the managerial ability and parsimonious motives of public servants. While this approach may indeed be valid, it seems to me that trust must be earned and that there is no better way of doing that, than through some performance review which demonstrates reliability.

25. It can be concluded that if Parliament is truly concerned with economy, efficiency and effectiveness, it must either require that periodic evaluations of existing programs be undertaken by the parties responsible or it must assign this duty to a third party. Such a choice is obviously available to Cabinet in its interface with departments as well as to other managers in the accountability structure. My message to you, though, is simply that Parliament must expect evaluation to be done. Where is the dividing line between what should be done by legislative auditors (external to Cabinet and the bureaucracy) and that which should be done as part of the on-going management process? I can't judge that issue - nor indeed would it be proper for me to express an opinion at this time. This is obviously a matter still "before the courts". What I do know, on the basis of my own experience, is that no dividing line between economy and efficiency, or efficiency and effectiveness will adequately separate questions of administration or implementation from questions of policy or political judgment. I once believed that such a separation is possible - indeed I talked earnestly to the Wilson Committee about it. But now, after much further reflection, it seems to me that even in matters of economy, any decisions must turn on social and political judgments, often with respect to issues on which auditors or officials are not only unable to offer evaluations, but even unable to provide much hard evidence. What I want to do, therefore, is provide some note of caution (based on painful experience) on the expectations as to what's

possible. In the next sections I try to do that in two ways. First, a quick look at the composition of the budget to see which parts might not be amenable to a "value for money" evaluation. This will remind you of the initial scope for analytical review. Next, a comment on each of the three kinds of evaluation which comprise the "value for money" assessment: economy, efficiency and effectiveness, with some reference to the techniques normally used and the particular difficulties encountered with each type.

26. The Blue Book reveals that 12.6% of the 1975-76 budget will be paid out as interest on the public debt. What does it mean to obtain value for this money (\$3.6 billion)? Would a "value for money" audit extend to an analysis of the benefits derived from deficit financing by the government? An additional 9.3% will go toward fiscal transfer payments (tax points, equalization payments and the like). Federal and provincial auditors can vie for that portion! The dilemma is that federal dollars are involved but these cannot be traced to the particular provincial programs eventually funded. Where do we seek the value for this money? Another 9.2% of the budget is spent on internal overhead and general government services. To determine "value for money" requires a defensible method for allocating these expenditures among the programs and activities which are jointly served. (For example, what value is obtained from the salaries and operating costs of Parliament?) That's now over 30% of the budget where the question of "value for money" poses - it seems to me - insurmountable problems conceptually. A further 9.9% will be spent on defence - a fine example of what

economists call a public good - which usually translates to mean that benefits cannot be valued. Many programs in the fields of culture, recreation, education, and public health will have this characteristic as well. Of the rest of the budget, a large portion, in the fields of medical care, welfare, economic development, etc. goes into shared-cost programs. Here the difficulty would be in attributing any value to the federal dollars as distinct from the provincial.

The carving job I have just performed on the budget suggests that there is very little scope for attaching definite "value" to the monies which are spent. It does not imply, however, that partial evaluations are altogether intractable. Let's turn, then, to the components: economy, efficiency and effectiveness.

27. Broadly speaking evaluation of existing programs or activities will focus on any one or a combination of three features: inputs, outputs, and effects. One type of evaluation requires reference solely to inputs (personnel, property, space, time, materials and so forth), and attempts to determine if these are being acquired at the lowest possible price for the same quality. This is sometimes called operational economy. The study of operational efficiency, however, looks at the relationship between inputs and outputs, or in the jargon of economists, the production function. Here the relevant question is whether inputs are of such a kind, and are deployed in such a way, as to produce a given output for the least possible cost.

Almost totally divorced from input considerations is the analysis of what I term "operational effectiveness". This seeks to determine if program output objectives are being met. One might be compelled subsequently to ask "why" or "why not", in which case productivity and technical feasibility would be studied on the input side. Beyond having output targets set in terms of actual goods and services, government programs are usually meant to accomplish greater things. They are really instruments of policy. What I will call the evaluation of instrument effectiveness seeks to assess the efficacy of a particular program or regulation as a tool for the implementation of policy. Effectiveness in this sense is taken quite literally, for it is the effects which must be identified and measured. Other kinds of evaluation over and above instrument effects, could be added to this list, but these would take us into the realm of more abstract social and economic goals. One can imagine, for example, a study that might enquire into the extent to which the government's bilingualism policy contributes towards the realization of a more "just, tolerant society".

28. Turning again to the three basic types of evaluation mentioned earlier to see how they relate to the "value for money" audit suggests some particular problems posed by each. In the context of managing government programs one might best demonstrate inadequate "value for money" on the grounds of operational economy. Financial records are there to show what resources have been purchased and at what prices. When the government is not the only buyer it is possible to relate the prices paid to

those paid by others for the same inputs (e.g., pencils, office furniture, accommodation, electricity, computers, fleet cars, airplanes, etc.). When the government is the sole purchaser of a particular input - such as military weapons - one may still ascertain whether or not the buying procedure was satisfactory. Making a judgement as to whether economy has been achieved can be a bit of a problem though, for economizing is a relative matter and there may be justifiable reasons for not always shopping patiently in the bargain basement. If there is an effort to "buy Canadian" for example, higher prices might be paid than would otherwise appear strictly necessary. In fact, government procurement is at times clearly an instrument of policy and the pursuit of more general socioeconomic objectives very definitely precludes routine identification of cases where value for money has not been attained. Despite the fact that the measures involved are often fairly unambiguous, and this area is the least likely to require political judgement, it remains difficult. In fact, after two years of work with an interdepartmental committee I have been unable to come up with any definitive analytical procedure to determine when designated social or economic benefits might warrant paying higher prices in the procurement process. The issue remains one in which Ministerial judgement is essential.

29. A given output is produced efficiently if there is no apparent alternative choice or utilization of inputs which generates the same output at a lower cost. Turned around, a program, activity or process is inefficient if there is a cheaper way to "get the job done". The tip-off that a process may be inefficient can come from a systematic or selective review of operations and procedures, from a degenerating performance measure,

from recognized technological advances which have not been adopted, or from other sources (e.g. complaints from clientele). Further investigation might be relatively straightforward, requiring only a cursory analysis, or might be extensive, perhaps involving sophisticated computer-based modelling techniques. On the basis of evidence developed in such investigation it is often possible for those most closely concerned to make a judgement as to whether the process, activity or program under review was, or more generally, is, inefficient, and in the latter case, it may also be possible to determine what remedial action is most appropriate.

You will note that my treatment of efficiency assumes a fixed output, with no differences in quality, for example. I recognize that this is a restrictive definition because very frequently the output of government will be in the form of services where the quality is one characteristic which can be varied, often directly in proportion to cost. When this is the case it may be necessary to determine the value of alternative outputs and relate this to their respective costs. Herein lies the difficulty in efficiency evaluation, especially in the context of planning. For example, every spring Revenue Canada may pay overtime wages in order to get refund cheques into the mail within a certain length of time. Money might very well be saved if refunds were processed over a longer period,

thus eliminating the necessity for paying an overtime premium. However, the level of service provided by Revenue Canada, which is most certainly a component of their output, would degenerate if this extra cost were not incurred. What is the value to be attached to any particular turnaround time? Who should decide on the appropriate balance between cost and level of service? (Bureaucrats? External auditors? Or elected representatives?) I think you will agree that this example is not atypical; that the question is relevant for countless services provided by the government. Judgements must also be tempered by the realization that there may be objectives, other than those which primarily guide the program, activity or process, which may prohibit the fullest achievement of efficiency. For example, in the interest of permitting flexible hours of work phones may sometimes ring unanswered. Government decentralization policy may also constrain efficiency.

30. Further removed from financial considerations is what I have called the evaluation of "operational effectiveness" and what the GAO and the Wilson Committee refer to as a review of program results. What we are all getting at here, I know, is an analysis which would show whether programs and activities are achieving the objectives set for them. I want to be careful, though, to distinguish between targets which are specified or understood in terms of tangible outputs and those which relate to effects in terms of furthering policy objectives. This is

not just a semantic problem - there are very real differences in the methodology one would use. In the first case one is dealing with tangible goods and services. Provided that there are specified targets for output it will not be difficult to see how actual production measured up. It may not be so simple, by the way, to judge whether the program is at fault if output targets are not met. Consideration must be given to other factors which may have influenced the outcome. The Wilson Committee gives us an illustration in its report. I will quote from the section on "value for money": "If, for instance, a dock is built to accommodate 200 ships a year and the Auditor General finds out that only five ships have used it during the year this is certainly a case of ineffectiveness - an expenditure not achieving the anticipated results". Personally I don't agree that even this is a clear cut case. A concern to alleviate unemployment by building ahead of demand, or perhaps an intent to create a future demand, may lead to such results. Indeed, perhaps the only reason that so few ships arrived at the dock was that they were all at anchor in the harbour due to a prolonged grain handlers or longshoremen's strike. The following year 200 ships may indeed use the facilities.

This may seem at first glance a somewhat trivial example, but from our actual evaluation experience I could cite similar examples of program outputs that

may be dramatically affected by extenuating circumstances. A national park which was intended to accommodate 10,000 visitors in the season but only one-half of this total passes through the gate because, unexpectedly, the price of gasoline rises prohibitively and people simply aren't travelling; and housing programs that are intended to place so many subsidized units on the market by a given date but are plagued by unforeseen shortages of materials. To tell the whole story of program results will require, therefore, an examination of more than operational or tangible program outputs.

The difficulty, thus, is that many of these simple questions of efficiency feed back onto the issue of targets or objectives - questions of what is the "right job" - possible shifts in composition, or levels of service, or other sorts of quality change. Although questions of cost effectiveness might be handled in a "factual" manner, questions of benefit-cost analysis cannot be.

31. Whether or not a third-party evaluator gets into instrument effectiveness, of course, will depend upon the purpose of the report to be submitted. Suppose it is intended to strengthen an accountability relationship which exists between the person or group who approved and funded the program in the first place and those charged with its implementation. The judicious analyst would probably show that output targets were not met and offer probable reasons why, properly identifying those which are beyond the control of the managers, but would stop short of saying that the program wasn't an appropriate choice (as this would needlessly call into question the judgement of those who made the initial

decisions). Alternatively the report may be more for planning purposes. Those who approved and funded the program may not have been totally sold on it, or they may admit to not having had adequate information. Perhaps the program was purely experimental. In such cases the evaluation should encompass what I have called instrument effectiveness - that is, an examination of program effects.

32. How does one do this? Well, first it is necessary to identify the purpose of the program. There we encounter the first problem, for in reality most government programs serve several policies. Armed Forces bases, for example, which are frequently located in isolated communities form part of the defence effort but also serve to generate employment there. Further, the identification of objectives frequently requires an heroic assumption that somewhere they will be well-documented. A search would lead to the legislation which enables the program or to the mandate of the department which operates it or to the Blue Book (the Estimates). No one of these sources will reveal the purely political objectives of the program, of course, which means that such evaluation will be limited to technical effectiveness. Apart from that, these sources yield mostly vague "motherhood" statements about the objectives. Let me cite some examples: Indian and Eskimo Affairs programs have as an overall stated objective the achievement within Canadian society of the cultural, economic and social aspirations these

native groups hold. Fisheries and Marine programs at DOE are intended to improve the management and sustained economic utilization of the marine and aquatic renewable resources of the country, compatible with a concern for the quality of the environment. The Consumer Affairs programs seek to optimize the welfare of consumers and protect them against economic and accident hazards. Trade and Industrial programs are meant to achieve efficient and sustained growth in the production and trade of Canadian goods and services. These statements do not give much guidance about program purposes. Moreover other statements derived from these sources may vary in their language or degree of detail. Thus, the Agricultural Stabilization Act which governs programs in the Department of Agriculture may refer only to the objective of stable incomes for farmers, but according to the Blue Book particular programs seek to ensure that agricultural markets permit producers to realize, as well, a fair return for their labour and investment. This hints at a problem in any evaluation of program results: Parliament will have passed enabling legislation which states the purpose for a program (or a host of programs) in one way, but has also voted funds to the program based upon another interpretation or where other dimensions have been added.

33. Identification of the purposes of the programs must eventually culminate in the selection of criteria by which to measure their effectiveness as instruments of policy. What do we mean by the welfare of consumers,

economic utilization in the fishery, income stability for farmers, poverty reduction, etc.? What sort of numbers would properly convey achievement of these rather abstract objectives? From my experience the selection of criteria has always meant seeking a consensus, for nearly everyone involved in the evaluation or influenced by it will have his own view of what is implied. Where there are multiple objectives, these might have to be weighted in importance and some lower-ranked ones sacrificed in the measurement. There is a danger in this, for the criteria themselves are then likely to become enshrined and subsequently the programs might be directed solely to the accomplishment of change in certain of these indicators at the expense of others. For example, suppose we agree that one measure of a reduction in poverty is the number of families which have been moved above the so-called poverty line. (So-called effectiveness evaluations in the U.S. have sometimes used this measure.) Government support programs might thereafter focus their attention on those families whose income is very close to the poverty line so that they may be pushed over, but the very poorest families could be neglected as a result.

34. Evaluation must then proceed to measure the effects of the program, observed as changes in the criteria chosen. This is the most difficult task in the whole methodology of evaluation for we must take measurements that will make it possible to distinguish the program's

effects, if any, from the effects of all other forces working in a situation - to isolate what happened as a result of the program from what would probably have had happened anyway. The little story with which I began illustrates this point.

35. Analysts of effectiveness are always forced to wrestle with this question of what would have happened in the absence of the particular program under review (and it is never as easy as estimating the probability that an elephant will appear on the Sparks Street Mall). Some interesting techniques have been devised. One that we have made some use of in the Treasury Board Secretariat is the simulation model. For instance we were involved in an interdepartmental evaluation of the Local Initiatives Program set up to measure the effect of LIP ~~has~~ on ~~reducing~~ the rate of unemployment during the winter months. (That was considered the most important policy objective of the program.) To do this we developed a simulation model which attempted to estimate what the rate of unemployment would have been if no LIP jobs had been offered. It took account of the fact that only if participants would have been unemployed during that period would the potential unemployment rate be directly affected. For example, there would be no effect if people joined the labour force specifically to participate. In the case of workers who were employed before they joined or would have become employed, however, the unemployment rate might have been indirectly affected

because the jobs they left or would have occupied could then have been filled by others who might have been unemployed. Once the theoretical model was constructed to simulate the absence of LIP, a good deal of information was needed to estimate the parameters of it. Fortunately the program had been designed to obtain certain relevant information through a questionnaire completed by participants. Labour force data were also scoured in an attempt to predict what growth would have occurred in the absence of LIP. Administrative records were studied to try to determine if any of the projects, e.g. day care centers, might have been financed in any event by other governments or community organizations. Such was the range of information needed.

36. Regression analysis is another technique designed to isolate cause and effect relationships when adequate data can be assembled. Several regression equations may be combined into what is called an econometric model, which is another way of simulating change. Using this method, for example, an effort was made to estimate the volume of research and development activity undertaken by business in Canada in response to government R&D incentive grants.

37. Where government services or direct payments have been given to individuals with the hope of changing their behaviour or employment prospects or earning power, for example, a basic technique is to measure the change for the sample of recipients under the program with the change observed in a control group. This requires

careful scientific selection of samples and involves the tricky political and moral problem of giving benefits to certain members of the population while denying them to others who are equally in need. Once the program is operational, of course, it may be impossible for an evaluation to establish such a control group. It is usually something that has to be set up in the planning stage of the program. After the fact it may only be possible to interview or otherwise monitor recipients and especially where effects are expected to be long-term, maintain longitudinal information on them. I am sure you will agree that most of the efforts of the GAO which are truly concerned with evaluating program results involve these types of social programs. This technique has not been used in Canada as much as in the U.S., partly because the federal government here is not involved in as many programs of this kind as the U.S. government, (e.g. some of the most successful U.S. studies have been concerned with programs aimed at disadvantaged elementary and high school students). A few years ago though, DREE attempted to evaluate its incentive grants by follow-up interviews.

38. Benefit-cost analysis provides another method of evaluation, close in spirit to traditional investment analysis or project evaluation in the private sector, but based on the adjustment of market data to reflect more general social valuations. This method does provide a way for valuing some benefits not normally captured in analysis based on financial data alone, and the relevant costs

extend beyond government budgetary cost to include other costs to canadian society as a whole. Our experience with it in planning suggests that it can best be applied to projects of an investment nature, be it an investment in human or physical capital. It has been quite useful for analysing new transportation facilities, for example, and manpower training programs. However, when the policy objective is "soft" and not solely concerned with economic productivity or growth then the benefit/cost approach becomes less tenable. In cases even of physical capital projects it is apparent that some of the paramount objectives have to do with things like enhancing national prestige, increasing national unity, or redistributing income.

39. One acceptable measure of effectiveness, by the way, for many of our transfer programs is the impact they have on the distribution of income. Some of the most persuasive evaluation of government programs has to do with such features of "target efficiency". But whether or not the changes observed are in the desired magnitude or direction will depend ultimately on the political judgement of those in power at the time or, if the information is disseminated to it, on the electorate.

40. It is clear to me that many government programs have to be assessed in this way. We are just not able to attach a dollar value to program results where this includes recognition of effects as well as tangible goods and services. Instead program results must be

expressed in terms of meaningful indicators of achievement. Then the changes which programs generate can be valued by those whose role it is in the accountability structure of the government to do just that - make value-judgements! The best we can do is facilitate these ultimate decisions about effectiveness by offering the best information available. If it would be helpful to directly associate budgetary costs with the results achieved, we should strive to do so, but there are many hurdles. It may simply not be possible to disentangle the separate effects of closely related but independently funded and operated programs. In shared-cost programs it would be an awesome task to isolate the separate effects attributable to federal and provincial dollars.

41. Fundamental to every evaluation made of government programs and activities is the issue of what information is to be disclosed, to which level, and by whom. It would be improper for me to draw any conclusions on the last of these components - who should do the evaluations. I have tried to show that external auditors may be chosen where the reason for the evaluation is to strengthen accountability, but a third party might also be appointed to provide independent analysis to the planning and budgeting processes. Quite obviously, the mandate must be carefully specified for this will determine the extent to which the report is critical or offers recommendations for change. The question of what information is disclosed will be determined by other factors as well. I like the Wall proposal

regarding the disclosure of information - separating "facts" from judgement. It is tempting to argue simply on grounds of principle that administrative decisions and judgements must be made openly in public, not individually in private. But I recognize that an isolated fact may be just as misleading as an outright falsehood. Moreover, there are not too many hard facts unearthed in any evaluation. As I have shown, the objectives of government programs are very often vague. The criteria by which they are to be assessed, especially for effectiveness, then, becomes a matter of potential debate. To handle this, differences of opinion can be resolved before the results of the evaluation are disclosed, or the format of the report may be such as to give equal time to each point of view. It is very important that evaluations strive to be as objective as possible. Sometimes this still permits conclusions to be drawn below the level of Cabinet or Parliament. In the area of economy and efficiency, for example, there may be indisputable standards that can be applied for purposes of operational management. As I have shown, however, even in these areas conclusions must be tempered by policy considerations. In effectiveness evaluation there is bound to be the problem of unwritten political objectives so it is only proper that elected representatives (at the level of Cabinet or Parliament) and ultimately, the electorate, make the value judgements required.

CONCLUSIONS

A good evaluation report - unlike a good balance sheet - does not speak for itself. Evaluation of a government program represents an attempt to communicate a mood - impressions gained after intensive study based on pointed questions launched from a broad perspective, outside the program, perhaps, but within a general framework of accepted objectives. Successful evaluation entails identifying explicit objectives, translating these into appropriate indicators or measures of attainment, isolating the impact of individual programs on the relevant indicators, and estimating the resources allocated to each program in attaining such impacts. The result, with luck, may be an informative, though incomplete, narrative useful to Ministers or senior officials in weighing identifiable facets of a program against more general policy considerations. But it cannot represent an authoritative judgment as to the net value of the program under study.

We are, in this respect, badly misled by the rather novel - albeit presently widespread - organizational form represented by the limited liability company. Long before the advent of the corporation formed to carry on trade, communities developed the social organizations necessary to establish "peace, order, and good government". It is not at all clear that the management tricks feasible and appropriate in the one are necessarily suitable to the other. Indeed, we might do well to reverse the usual perspective, and view

the corporate form and market structure as very special mechanisms established to handle those cases of social decisions where goals are so concrete and easily recognized (or interests so single-minded and identifiable), and interpersonal side-effects so negligible, that decisions can be decentralized and delegated to individual market units under strictly defined and easily communicated guidelines or payoff rules. In these circumstances, evaluation for resource allocation purposes can indeed be authoritative and unambiguous.

Misunderstanding of the role of evaluation in government programs stems in part also from a tendency to discuss the subject abstractly, in a vacuum. There is constant reiteration of basic principles without any concrete referent. There is general acceptance of the economist's notion of a "public good" - one whose production must rest in the public sector because decisions as to the appropriate level or distribution of "output" cannot properly be delegated to a private market mechanism. But there is not much recognition of the fundamental consequences of this notion. In effect, what it means is a fundamentally different mode of decision in the public sector. It is this feature of government operations which on the one hand renders "value for money" audits impractical, and on the other makes essential the active role of the legislative auditor in assuring the free flow of information necessary to permit informed social decisions. (It is also this feature which accounts for the critical problem of information overload facing all governments

as they grapple with problems of ever-increasing complexity.) The "private-sector model" must not be allowed to drive the public sector reality.

Thus I argue that the notion of "value for money" audit is a red herring. There is no point in looking down the road to a better day when systems and methodology and sophisticated quantitative techniques will have improved to the point that authoritative evaluation of government programs can be undertaken according to generally accepted principles. Such a day will not come because objective evaluation of government programs is conceptually not possible. What is possible, and indeed crucial, is vastly increased disclosure of information on the consequences of government programs. Such information must flow to the House, and to the people. And the Auditor-General (at the federal level) can help to make that happen. His is a critical responsibility in assuring the information flow that leads to an informed House, an informed electorate, and the increased scrutiny of government operations that each will bring.

For all the reasons noted above, I am convinced that the continuing attempt to find a dividing line between audit of waste, economy, or efficiency (subjects alleged to be appropriate to the legislative auditor) and program effectiveness (appropriate to government spokesmen) can lead only to a continuing divisive squabble about what constitutes policy and what administration or implementation, what constitutes effectiveness, and what mere efficiency. The legislative auditor does not have the responsibility, nor

the mandate, nor the right, nor the competence to make judgments about whether programs - even procurement decisions - have achieved "value for money". What he does have, in my opinion, is the profound responsibility to report, wherever necessary, the facts relating to government programs in such a way as to enable Parliament itself to make such judgments.

My appeal to you, in other words, is to concentrate on the real issue, which is disclosure of sufficient information to enable the House (and the public) to be informed critics of the government charged with developing policies and implementing programs to carry out the wishes of Parliament. Such information, even in the simplest cases, cannot be authoritative or complete. It cannot carry the stamp of the auditor expressing a professional judgment that "value for money" has or has not been attained. There should be no pretense that audit in this sense is feasible. But there could from time to time be the observation that the management information systems maintained by the government, and the information provided by it to the House, do (or do not) accurately describe features of programs under review.

Accountability rests on disclosure. In the certification of financial statements, the legislative auditor can speak with the authority of a professional following recognized practice, and therefore can serve the objective of full disclosure by attesting that statements prepared by the government present fairly the condition of the departments or programs concerned. In the matter of evaluation or audit, no such authoritative view is possible.

But the legislative auditor might nevertheless point to the absence of adequate information relevant to government programs, or to the absence of systems adequate to enable the government properly to discharge its own obligation to make judgments as to economy, efficiency, and effectiveness. In the extreme, it may fall to the legislative auditor to raise questions as to costs or consequences of particular programs or decisions. No professional judgment - no audit of "value for money" - is necessary or proper at that point; no challenge to unproductive activities need be offered in addition. In the hands of an informed Parliament, such questions can speak for themselves.