

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: The Measurement of Economic and Social Performance

Volume Author/Editor: Milton Moss, ed.

Volume Publisher: UMI

Volume ISBN: 0-870-14259-3

Volume URL: <http://www.nber.org/books/moss73-1>

Publication Date: 1973

Chapter Title: Evaluating Performance in the Public Sector

Chapter Author: Mancur Olson

Chapter URL: <http://www.nber.org/chapters/c3618>

Chapter pages in book: (p. 353 - 410)

The Public Sector



Evaluating Performance in the Public Sector

MANCUR OLSON

UNIVERSITY OF MARYLAND

I

GOVERNMENTS are not famous for efficiency. Though no one can be certain of this until the matter has been properly researched, casual observation suggests that most economists probably believe that governments are less efficient than enterprises in the market sector of society. This observation seems to be made not only about governments of different levels in the United States, but also about governments of a variety of types in many different societies. Economists are certainly not alone in associating governments with red tape, cumbersome procedures, and bureaucratic inflexibility: whatever the balance of lay opinion may be, there are certainly large numbers of laymen of assorted ideologies and nationalities who claim that governments they have observed are inefficient, by which they presumably mean less efficient than other enterprises with which they are familiar. Whether the inefficiency attributed to existing governments is believed inherent in all government activity is more doubtful, since some of those who decry the ineptitude or waste of an existing government contend that these problems would be solved if the electoral victory or social revolution they prefer were to take place. Some of the complaints about the inefficiency of government may also be focused on existing legislation which is or is thought to be ill suited to the ends it is supposed to attain, rather than with the effectiveness of public administration itself.

The empirical evidence needed to determine what degree of truth, if any, there is in these beliefs is not now available. Are there any theoretical considerations that should lead us to expect that government operations are generally less efficient than those of private enterprises? There

NOTE: The author thanks the National Science Foundation and Resources for the Future for support of his research.

are a wide variety of ad hoc explanations and obiter dicta about the causes of the alleged government inefficiency in the literature of economics, and some seminal writings devoted directly or indirectly to this subject.¹ Yet it remains true that the reasons for the exceptional inefficiency that is alleged to occur in the public sector have not been codified, nor incorporated into the body of economic theory, nor even stated in a sufficiently clear and general fashion to bring this matter (important as it obviously is) into the elementary textbooks. Some of the discussions of the causes of inefficiency in the production of public outputs are in the literature of what is, in the United States, calling the Planning-Programming-Budgeting System, or in discussions of cost-benefit analysis. But others are in the welfare economics or public finance literature on what is coming to be called the theory of public expenditure. The PPB and cost-benefit discussions and the literature on the theory of public goods or expenditure have never been properly integrated—indeed, at times it seems as though they are developing independently. Perhaps this is in part because the students of cost-benefit analysis and especially PPB are often applied economists with only a limited interest in pure theory, whereas those who write about the theory of public expenditure are often theorists with only a sporadic interest in the day to day workings of government. There is a promising and relatively well integrated set of writings on what is sometimes called “nonmarket decision making” or “public choice,” but this literature has so far been mainly concerned with the constitutional, electoral, pressure group, and legislative processes that determine what public policies or purposes will be chosen, rather than with the efficiency or inefficiency with which these policies or purposes are carried out. The complaints about government inefficiency that are of interest here refer mainly to public production or administration, rather than to any stupidities or inequities in basic public policies.

II

It might seem at first glance that the explanation of the alleged government inefficiency is obvious: governments are inefficient because they do not or cannot in general use perfectly competitive markets. This is, on reflection, not an explanation at all. To be sure, if a perfectly competitive equilibrium exists, it will be efficient, in the sense that any

¹ See, for example, Charles J. Hitch and Roland N. McKean, *The Economics of Defense in the Nuclear Age*, Cambridge, Harvard University Press, 1960; and Charles Shultze, *The Politics and Economics of Public Spending*, Washington, D.C., Brookings, 1968.

rational objection to its performance must devolve into a preference for a different distribution of welfare. But it does not follow from this that governments can't also operate with Pareto-optimal efficiency. Moreover, markets in which no entity can affect the price of what it sells or buys are very much the exception, so that observations that governments are less efficient than the private sector in most cases mean that governments are supposed to be even less efficient than firms or markets with monopoly or monopsony power. What is needed is an explanation (or a refutation) of the contention that governments are even less efficient than the imperfect markets that are typical.

Another apparent explanation can likewise quickly be shown to be insufficient. Democratic systems can perhaps operationally be defined as those with an opposition party or parties that could at least in some circumstances defeat the party in power in a free election. If the party in power appoints a sufficient number of persons to public office in return for political support, or favors enough firms with government business for similar reasons, it might thereby have bought an electoral majority with public funds. This is a proximate (but as we shall later see, definitely not the ultimate) reason why some governments have civil service merit systems, rules requiring competitive bidding, and other constraints on official decision making. These constraints are understandably sometimes considered a source of government inflexibility and inefficiency. But the need to preserve the integrity of the democratic process cannot by itself explain government inefficiency, since it is only those personnel and procurement decisions which are inefficient from some national (as opposed to purely partisan) view that need to be prohibited. If we know that a government has served its announced conception of the public interest with complete efficiency, we know that it cannot at the same time have been guilty of political favoritism in its purchases of inputs. It is not in any event clear that effective controls on political favoritism in resource procurement are indispensable to continuance of the democratic process, at least when most economic activity is in the private sector. That seems to be the lesson from U.S. experience at all levels of government in the nineteenth century, and in some state and local governments (e.g., Pennsylvania) until at least very recently. Moreover, dictatorial governments that could not have this need to preserve democratic processes are often also deemed to be inefficient. The U.S. Strategic Bombing Survey and other studies have suggested that the Nazi administration of its war effort was far from efficient, and perhaps inferior to that of Great Britain

and the United States. And as anyone who has traveled in the Soviet Union as well as the West is aware, many of the services provided by the Soviet government seem to be relatively inefficient.

Another possible explanation grows out of Kenneth Arrow's justly famous paradox of voting.² Since it is clear from Arrow's work that in certain circumstances social choices will not be stable and transitive, it might be supposed that government behavior would be too erratic and unstable to be efficient. Yet behavior of democratic governments does not often appear to be nearly as erratic or unpredictable as a naive application of Arrow's analysis might suggest. The conditions that give rise to the Arrow paradox often don't exist, and even when they do, the complex network of institutions and checks and balances in democratic governments may keep policy from changing rapidly enough to be erratic or unstable. Interestingly enough, democratic governments are often more stable and predictable than those that are dictatorial or likely to be overthrown by violence. Even if democratic governments usually reversed themselves, that would not be sufficient to explain the observations of government inefficiency: it would lead instead to charges that governments cannot make up their minds. But it is not obvious that such charges are particularly common, and when they are heard it is sometimes because the left hand of a bureaucracy has worked against the right hand for reasons that have nothing to do with the Arrow paradox. Thus, without denying the extraordinary theoretical and normative importance of the Arrow paradox, it seems that the economist must look elsewhere for an explanation (or refutation) of the commonplace observation that governments produce goods and services inefficiently.

III

The most fundamental way to approach the question of government production inefficiency versus private production inefficiency is to distinguish the demands typically placed upon governments from those placed upon private firms. Though there are many exceptions, governments (and several other types of nonprofit organizations) are most often called upon to supply public or collective goods, or to control the production of those with externalities, whereas the private sector is almost never paid (except sometimes by government) to deal with such goods. Those goods or services that in almost every country are provided mainly by government (e.g., law and order, defense, pure research, pollution control) are certainly collective goods or goods with

² *Social Choice and Individual Values*, 2nd ed., New York, Wiley, 1963.

more than average externalities. This paper will argue that governments are in fact as well as by reputation usually inefficient, and that this is mainly because they deal with collective goods and externalities. Though it will be necessary to amend this definition later, a collective good for the moment can be defined as a good such that, if it goes to one individual in some group, it is not possible (or at any rate economical) to exclude others in the group from the enjoyment of that good; in other words, nonpurchasers cannot practically be excluded from the consumption of the good. An externality here (and for that matter in most other discussions) is distinguished by the same defining characteristic, though it is clearly also fairly common to call a good with the nonexclusion property an externality if it is jointly produced with a private good or if the group affected by it is very small, and otherwise to call it a public or collective good.

Governments are of course also distinguished from the private sector by the fact that only they can significantly redistribute income. Questions about the efficiency of a redistributive program can also arise, especially when the program affects the incentives in the productive system. Some redistribution of income can usefully be regarded as due to the external economy a redistribution of income can have; that is, it is desired by those who give up the income because they think the transfer is in some sense in their own best interest.³ Redistributions of this sort are covered by what will be said about collective goods and externalities. There remain some redistributions of income which are desired, or which even take place, but which are plainly not desired by the class from which the transfer is obtained. Any inefficiencies in these redistributions will here be left aside, partly on the ground that they would for the most part involve failures to anticipate the secondary effects of the redistribution in the fundamental legislation rather than any defect in public production or administration.

Though most economists no doubt realize that the problem of efficiency in government has something to do with public goods and externalities, the full implications of the relationship have not been explored. Some of the consequences of collective goods for governmental

³ Even such redistributions can require coercion, because individuals in the group from which the transfer would come might believe that the benefits to them of the transfer (e.g., reduced crime) would occur only if so much income were transferred that most or all of them would have to transfer income to the poor. In such a case the individuals in the donor class would be rational to vote to force themselves to transfer income to the poor through taxes, but not rational if they gave individual and voluntary contributions.

efficiency have not been correctly explained, and others are included in one strand of the literature and ignored in another. This in turn has meant that some steps that might be taken to increase the efficiency with which collective goods are provided have not received appropriate attention.

IV

One cause of inefficiency in the provision of collective goods is familiar from the theoretical writings in welfare economics and public finance, but rarely mentioned in the PPB or cost-benefit literature. That is the difficulty of getting consumers to reveal their preferences concerning a collective good or externality, and preferences must of course be known to determine how much it is optimal to provide. It has been clear at least since Erik Lindahl developed his famous "voluntary theory of public exchange" that there can be no assurance that individuals will agree, through voting or by other means, to provide a Pareto-optimal amount of a collective good unless each individual who benefits from the collective good shares its marginal costs in the same proportion in which he shares the benefits of additional units. Otherwise, those who pay more than their share of the marginal costs will vote for a lower level of provision of the collective good than is optimal, since their own marginal benefits will be equated with the tax price they pay for each additional unit before the marginal cost to the society equals the total social benefits from another unit of the good. Conversely, those who pay a less than proportional share of the marginal costs will have an interest in a supra-optimal provision of the collective good. Thus there can be no assurance that a Pareto-optimal level of provision of a collective good will occur in a democratic system⁴ unless all voters, or in any event those who cast the deciding votes, pay a tax price that is given by their marginal rate of substitution between the collective good and other goods. But if an individual's tax price will rise if he is known to place a high value on additional amounts of a collective good, he has (as Paul Samuelson has made clear) an incentive to conceal his true preferences. Since he cannot by the definition of a collective good be excluded from the consumption of it, however little he pays in taxes, he will in most cases get as much of the collective good whether he claims to want it or not. Since every consumer

⁴In a dictatorial system without any voting, inefficiency will be even more likely, since the dictator will have no direct information at all on individual preferences for the collective good.

of the collective good is in this position, it will with rational behavior normally be very difficult if not impossible to learn what preferences are and thus how much of the collective good to provide.⁵

Though the difficulty of getting preferences revealed is surely a significant part of the problem of public sector resource allocation, it is not a sufficient explanation of inefficiency in the public sector. It could by itself explain only why governments provide too little or too much in the way of public goods, not any failure of governments to attain *technical* efficiency. It is this difficulty of attaining technical efficiency—a state in which no more of one good can be produced without less of another, i.e., any position on the production possibility frontier—that this paper is about. This is probably a more important source of inefficiency than any tendency to choose the wrong mix of public and private goods. Though there are complaints in some quarters that governments provide an inefficiently large supply of public goods, there are complaints in other quarters about “public squalor midst private affluence.” Though the fact that the complaints contradict one another doesn’t prove anything, it may hint that the losses that result from the concealment of preferences are not of such a great magnitude that they overwhelm all other factors. Some voters will for some public goods face tax prices that are too low and vote for too much, and some voters will for some goods be in the opposite situation. Though there is nothing that would make these two errors regularly offset each other, they can often countervail each other to some extent. Then there is the fact that those politicians who succeed are likely to be those who spend a lot of time talking to constituents and have a good intuitive feel for their true preferences. These politicians also have an incentive to propose tax and expenditure packages that will command more support than those proposed by their opponents. Undoubtedly more efficient proposals, such as many of those that would reduce the extent to which tax prices are out of line with marginal rates of substitution, will often have a political advantage, for the simple reason that one or more possible distributions of the gains from a Pareto-superior policy will leave some one (if not everyone) better off and none worse off. Thus, important as the revelation of preferences problem is, there must be a good deal more to the problem of efficiency in government.

A second reason why inefficiency is particularly likely in the provision of collective goods has been neglected both in the literature of

⁵ Ingenious schemes are being worked out to get around this problem, but there is to my knowledge no easy and universally acceptable solution.

welfare economics and in the literature of PPB and cost-benefit analysis. Writings of the latter sort are above all filled with injunctions on the importance of estimating the output of public programs. There is no doubt that these injunctions are necessary. It is perhaps the principal achievement of the PPB system that it has sensitized governments to the need to evaluate the level of public outputs. Somewhat surprisingly, there is little in the way of a parallel discussion in the welfare economics literature on public goods. Neither is there a widespread awareness in the applied or PPB literature of why the estimation of program outputs should be particularly difficult in the public sector.

It might seem at first glance that the problem of measuring outputs in the public sector is nothing more than the problem of concealment of preferences for public goods that has already been discussed. But this is not in fact the case. The concealment of preferences relates to how much an individual would pay for an additional unit of output of a collective good. It may relate to the problem a political leader would have in determining how much the citizenry would be willing to pay for additional *output* of some collective good, but it does not explain the problem a government administrator has in determining what volume of output a particular program produces. In other words, it relates to the problem of how much the output is worth, not to the prior problem of finding out how much output there is. For an individual citizen to be able to tell a public official how much value he puts on an additional expenditure on pollution control, he first has to know how much the expenditure will reduce the pollution—he has to estimate the production function for the collective good. But the individual citizen or consumer is in no position to do this, hence the need for program analysts and evaluation of output in government.

The reason that there is a special problem of measuring outputs in government is that governments typically produce collective goods or control externalities, and one of the characteristics of a collective good or externality is that, if it goes to anyone, it goes to everyone in some group. In other words, the collective good or externality is normally indivisible, in that it cannot be divided up in such a way that only those who pay their share of the costs get some of the good. In short, *the very characteristics of a collective good that make it a kind from which nonpurchasers cannot be excluded, also make it a kind of which the output is not in the form of divisible units that can be readily counted.*

It might be said that this is also true of services sold in the private sector (and all commodities ultimately produce services); the firm that

sells janitorial services to individual customers also may not be able to measure its volume of output in physical terms. But this objection is insufficient, because when a good or service can be sold separately to different individuals, there is no need for a measure of the physical volume of output. All the janitorial firm needs is information on the marginal revenue it gets from providing additional service. The customer is in a position to estimate the value to himself of an additional supply of service from a firm because he can experiment with different levels of purchase. His estimate of the value of the good shows up in the marginal revenue the firm receives. By contrast, in the case of a collective good going simultaneously to many individuals, the individual consumer cannot take more or less to see how that affects his well being.

Accordingly, quite apart from the problem of concealment of preferences which makes it very difficult to know how much a given amount of a collective good is worth to those who receive it, there is the further problem that it is nearly impossible to measure what amount or volume of output a government program is in fact producing. Thus the government supplying collective goods or dealing with externalities has less of the information it needs to act efficiently than a private business (including a monopoly). A private business can in general get information about the marginal revenue it would receive if it changed its output mix or level and in most cases can measure its volume of output in physical units as well. This means that private corporations and other businesses, however much or little monopoly power they may possess, usually have a measure of physical output that allows them to calculate their technical efficiency, and often information that allows them to estimate marginal costs and marginal revenues. Governments dealing with collective goods and externalities are in a far different position. They know only their expenditures, and normally have next to no idea what impact or volume of output these expenditures bring about, if any, and if they did know what they produced, they would still have quite a problem to find out what it was worth to its consumers.

V

The inherent connection between the nonexcludability of nonpurchasers of most collective goods and the difficulty of measuring the volume of their outputs is probably even more important than the problem of concealment of preferences that is familiar from welfare economics. This is partly because it in large part explains the weakness of the forces tending to weed out inefficiency and reward efficiency in organizations

providing collective goods. This weakness can best be explained by comparing a typical government agency providing a collective good with a monopoly selling an individual or private good. To make the comparison more compelling, assume that in both cases entry into the respective industry or area of responsibility of the firm or agency is forbidden. The monopoly firm without fear of new entrants still has a powerful incentive to be technically efficient. Though the monopoly may be able to survive even if it is not as efficient as it could be, it will always be able to increase its profits if it can lower its total cost curve. This will of course be true even if it has no incentive to increase output and total revenue, though it will in fact do this whenever marginal costs in the relevant range also fall. The government agency *might* also gain if it were *known* that it had reduced costs while maintaining or increasing output. But this is seldom known to have happened even when it has happened. Since the volume of output of the collective good is so difficult to measure, the agency and agency head may gain as much from *seeming* to be efficient as from being efficient. If a government administrator can persuade the powers that be that his predecessor has sent ice boxes to the Eskimo, but that he has put a stop to this waste, he may be in a strong position, even if the total volume of output he produced from a given budget in fact declined. As Charles Schultze and others have emphasized, the efficiency of an institution providing a collective good is often judged to an excessive extent by the presence or absence of obvious "mistakes," which need not be an accurate guide to its over-all degree of efficiency. It is also sometimes judged by the fashionableness of the procedures it uses (does it recruit people with advanced degrees or use a big computer), or by whether the administrator's style is in accord with current doctrine (does he run a "taut ship" or use "consultative management"), or by whether it is adept at praising itself in the right places (does it have a good press or friends on the Hill). The importance of an appearance of efficiency and usefulness naturally encourages, in spite of regulations to the contrary, considerable efforts at public relations, which from a social perspective are mainly a waste of resources.

The critic may say that this emphasis on the lack of knowledge of the physical volume of output of the relevant public goods misses the point: if it were known with absolute certainty that a given agency had increased efficiency it still might be the case that the agency head and the agency as a whole might go unrewarded, or even end up with a smaller appropriation, so the real trouble is the lack of an appropriate

incentive system in government. But this criticism in turn neglects the fundamental point that an adequate incentive system cannot be established when there is no way of measuring the volume or value of output.

Whether an institution produces a divisible private good or an indivisible public good affects not only information about its over-all performance and that of its manager, but also its internal processes. The more accurate the information about the output of an institution, the greater the possibility of estimating the usefulness of its component parts. If the changes in the dependent variable are not known, there is not even the hope of estimating the contributions each dependent variable has made to its movements. But with a firm producing private goods at least the over-all result is known, and there is then often the possibility of roughly apportioning the responsibility for it among the different parts of the firm. That is not only because of the rough inferences that may sometimes be drawn from time series data or experiments, but also (and more importantly) because the contribution of each part of a firm can often be measured directly. The multiproduct or geographically dispersed firm can often set up what the accountant calls separate "profit centers" to facilitate evaluations of subordinate divisions and managers. If the firm's output is at all homogeneous, it can simply count the units or components produced or handled by each factory or work group, and relate the totals to the resources that were used. Indeed, it is often possible to tell which nut tighteners have tightened the most nuts, and to pay them accordingly—that is, to use a "piece rate" or "commission" system. The fact that a piece rate reward system is often used by firms providing private goods, but can virtually never (except at earlier stages of the productive process) be used when a collective good is being provided, is entirely consistent with the argument that is being made here, and suggests that it can have profound implications for the procedures and incentives within an institution.

The lack of readily countable outputs in jurisdictions or agencies that produce collective goods, and the resulting lack of objective bases for judging performance, has led to complicated, cumbersome, and expensive restrictions on public management. Civil service or "merit" systems, requirements for competitive bidding even in situations where that is evidently uneconomical, separate resource constraints for personnel, money, and certain classes of supplies, and "red tape" in general are the natural concomitant of unmeasured outputs. Presumably no one would care very much whether the police commissioner gave a good job to his

nephew, or bought uniforms from his brother, if the police department were *demonstrably* as efficient as it could be. But since we do not know whether the police department is giving good value or not we know that the commissioner could get away with hiring his nephew, however incompetent, or buying his brother's uniforms, however costly, if we did not outlaw it directly, so we do. The civil service rules, competitive purchase provisions, and red tape requirements no doubt prevent some of the chicanery that the lack of measured output allows, since such regulations limit the extent to which resources can be diverted from public purposes; they may even sometimes increase output. But they also impose enormous costs. In some cases the labyrinth of restrictions that limits the freedom of action of public administrators is so elaborate that it prevents timely action in certain areas, at which point there is often a tendency to rely on contracts with outside suppliers (sometimes non-profit firms) to get the job done. This can be only a temporary solution, since the lack of measured outputs allows inefficiency and corruption on the part of the contractors, and thereby leads to regulations that hem them in as well.

The fact that it is the lack of an adequately measured output, rather than any other factor, that mainly accounts for the cumbersome regulations characteristic of government bureaucracy is also demonstrated by the tendency of nationalized industries producing private goods to have fewer regulations (at least in some countries) than civil service departments providing collective goods. This in turn illustrates a point that must be emphasized: it is not that governments are inherently inefficient, but rather that the provision of collective goods and the handling of externalities, whether by government or other institution, is typically inefficient.

The difficulty of weeding out the inefficient and rewarding the efficient in institutions providing collective goods creates some problems that tend to get worse over time. The criteria for promotion, when the volume and value of output are not known, cannot by definition include productivity, so other considerations must determine who rises in the organization. These other considerations may be very diverse. Whatever they are, it seems likely that talented people will be more likely to rise than stupid ones, since they will be able to find out sooner what has to be done to get promoted and be able to do it better. One consideration that apparently often affects promotion in organizations producing collective goods is loyalty to the organization, its leadership, its ideology, and its established way of doing things. Bureaucrats with such loyalty

will not "rock the boat." In military organizations, police forces, fire departments, foreign ministries, and churches, at least, it often seems to be the case that it is the consummate organization man who rises to the top. Such men, however able, tend to be conservative, in the sense that they are convinced that the organization's hierarchy and practices are basically sound. Where it is not productivity, but able and unwavering adherence to the established policy, that leads to promotion, an organization is likely to become increasingly inefficient as time goes on. The more conditions change, the farther its initial policies will be from the optimal ones.

The dynamic process just described can perhaps be illustrated most aptly by considering a well-known type of institution that provides only collective goods—the bureaucracies that handle foreign affairs, and particularly the U.S. State Department. The foreign ministries of many countries, such as France and Great Britain, have traditionally been able to recruit their professionals from the most able university graduates. Though the U.S. State Department may not be able to be quite so selective, it is still able to skim the cream. As a glance at a typical graduating class of, say, an Ivy League institution will reveal, a great many of the C and low B students go on to the business schools and then to the great corporations. But the handful that have won admission to the Foreign Service will have marks that are surpassed only by those who go on to the academic world, and in addition personality traits that have impressed references and interviewers.

The State Department, like some foreign ministries of other countries, does not, however, have a reputation for efficiency that is at all proportional to the apparent quality of its recruits, nor even as good as some of the great corporations that have had to be satisfied with the gentlemanly C students. The modest character of the State Department's reputation is illustrated by the fact that on the same day two of the best known Washington-based editorial columns carried tales of the Presidential dissatisfaction with the bureaucratic ineptitude in the State Department. The Evans-Novak column had this to say:

At the first staff meeting of the high-level State Department officials following Henry Kissinger's secret flight to Peking last month, Secretary of State William P. Rogers announced somewhat defensively he wanted everyone to know that "the State Department has been in on this from the start."

In fact, except for Rogers himself and possibly one or two of his top aids . . . State Department Far East experts were as much in the dark as everyone else.

Likewise, when President Nixon made the spectacular turnabout that killed his tattered economic game plan and floated the dollar, not one State Department economic expert had been clued in. . . .

In countless other matters involving less dramatic policy decisions, President Nixon's White House has circumvented the State Department bureaucracy, leaning not on career Foreign Service officers so often scorned by so many White House staffs in the past, but on non-career advisers, particularly the President's own national security staff in the White House.

Under Mr. Nixon, however, this effort to suffocate the professionals has been carried much farther than in the Kennedy or Johnson administration, which built the foundation of the now institutionalized foreign policy apparatus in the White House. For example, the skilled technicians on Dr. Henry Kissinger's national security staff now number about twice as many as the staffs of McGeorge Bundy or Walt Rostow in the 1960s.

. . . Rogers is under rising pressure from President Nixon to give him "non-careerist" State Department judgments on new policies. . . .

This White House aversion to careerist judgments is not new. The cautious, self-protective bureaucrat often finds 20 reasons for being against a bold shift in policy. . . .⁶

In the same newspapers Joseph Kraft pointed out that it was dangerous to exclude all those with extensive experience in foreign policy from the circle of Presidential advisers on international economic policy.⁷

Some participants in interagency meetings have compared State Department representation at these meetings unfavorably with that coming from civilian officials in the Defense Department. Yet the Defense Department has never been able to attract college graduates of comparable promise into its civilian bureaucracy. Of course, the State Department and comparable departments in some other countries may not deserve a reputation for striped pants conservatism and inflexibility. We do not and cannot if this paper is correct hope to *know* whether they are especially inefficient.

Yet the paradoxical relationship between the reputations of the new recruits of the Foreign Service and the reputation of the State Department as a bureaucracy is suggestive, and exactly what should be expected from the fact that they produce a collective good that is perhaps uniquely difficult to measure. State Department procedures and promotions have to be determined on some bases other than measured productivity. It is difficult to say what these bases are, but even critics with other perspec-

⁶ Rowland Evans and Robert Novak, "Smothering the State Dept.," *Washington Post*, August 26, 1971, p. A-17.

⁷ Joseph Kraft, "Dollar Diplomacy," *Washington Post*, August 26, 1971, p. A-17.

tives have argued that they are bureaucratic caution and fidelity to a quintessential State Department style.

The Defense Department also produces an unmeasured collective good, and it has its own well-earned reputation for waste of public funds. Yet it is possibly significant that the higher ranking *civilians* in the Defense Department, unlike most of their counterparts in the State Department, are mainly men who have been appointed from the outside in mid-career. They are often businessmen who have made money, scientists who have made discoveries, lawyers who have won cases, scholars whose books have won reputations, or politicians who have picked winners. It may be that on the average the high-ranking outsiders have proven themselves in areas where output is less badly measured. And at the very least, they know from experience that things are not everywhere done the way the Defense Department is in the habit of doing them.

It might seem that all of the above difficulties that have been ascribed to the lack of measured output are explained instead by the problem of concealment of preferences. But this is not so. As long as an organization has some measure of the volume or level of its outputs, it can produce whatever amount of each output it chooses to produce with as much efficiency as it could have attained had it known what each output was worth to its clients (i.e., had preferences been revealed). That is, it can obtain maximum technical efficiency. To be sure, an institution cannot know what mix or level of output to produce until it knows how much the outputs are worth to those who want them. But that is another matter. The problems of weeding out inefficiency, rewarding productivity, and maintaining effectiveness over time stem not so much from the familiar revelation of preferences problem as from the fact that collective goods by their very nature make it difficult to get a measure of the volume of output.

VI

How can the problem of inefficiency in the provision of collective goods be solved? Unfortunately, it cannot be completely solved, since it is an inherent consequence of a defining characteristic of collective goods. The hopes of the most optimistic advocates of cost-benefit analysis and the Planning-Programming-Budgeting System will never be realized.

But the problem can be ameliorated. Indeed, some possibilities for improving governmental efficiency grow directly out of the argument that it is the lack of a measure of the volume or level of the relevant public

goods that is the largest part of the problem. In some cases it is almost certainly possible to devise arrangements that would markedly increase efficiency, and in most others possible to propose steps that would lead to modest but still significant increases in efficiency.

The possibilities of improvement are greatly augmented by the fact that the most fragmentary and approximate information on the volume of collective goods produced could bring about substantial improvements in efficiency. This is true, in any event, if the organization or official whose performance is to be judged cannot usually determine or predict which fragments of information are to be found. The fragments can of course then be assumed to be a random sample from the mass of unknown information on the output of an agency. Even if the sample of information that is assumed to be a random sample is not in fact either random or representative of the "population" of information about which inferences are made, it can nonetheless lead to substantial improvements in efficiency. This is because, if it is used to influence the incentives placed before organizations and officials, and they cannot control or predict which fragments of information will be found, it can (even when it is rather inaccurate) give agencies and bureaucrats a reason to be as efficient as they can be. It may be argued that rewards and punishments based on fragmentary information on output cannot affect incentives in more than a marginal way, since the organizations and officials being judged know that it is highly probable that their output at any given moment will be unobserved. But this is not correct. If there is neither risk aversion nor risk preference, much the same behavior can be obtained by a "one out of ten" output measurement as by a "one out of one" system if the rewards and punishments are multiplied by a factor of ten.

Societies use such systems regularly; there is statistical evidence that only a fraction of the burglaries committed are "solved," much less lead to punishment for the burglar. Nor are the burglars punished probably a random sample of the population of burglars: they are perhaps on average less intelligent and fleet of foot. Yet people often dare leave their houses, and this is partly because a convicted burglar tends to get a sentence that has a negative value several times larger than the gain from an average burglary. Nine out of ten people who park overtime can go off scot-free without loss to society, provided only that the policeman chooses the parking places to check in a way drivers cannot predict and the fine is ten times as great as it would be if every offender were caught. There are, to be sure, limits to the "economies of evaluation"

that can be gained in this way; a lifetime sentence for a burglar without a prior arrest or a hundred dollar fine for a single parking violation are regarded as unjust when at the same time other offenders go unpunished (if the punishment were certain and this were known *ex ante* on all sides, there would usually be no offenses, except for those so rewarding that they left the offender with a net gain, and thus no injustice—except for those who “needed” to be burglars or park overtime, and that would reflect some *other* injustice). There may also be ethical objections if the rewards that are given or denied to agencies and officials are too great in relation to the intrinsic importance of the fragments of information about them. But it is surely possible to multiply rewards for fragments of favorable information to some extent, and thus to create a large incentive for efficiency with a small amount of information. This is particularly important in one important class of cases, but to identify this class it is necessary first to consider the definition of collective goods again.

The characteristic that has been used to distinguish collective goods and externalities in this paper so far is that of “nonexclusion.” Though the infeasibility of excluding nonpurchasers from the consumption of some goods was sufficient to pinpoint what this paper argues is the largest source of inefficiency in the public sector, it is not sufficient for other purposes. In particular, it conceals an area in which substantial gains in government efficiency are possible.

As John Head has pointed out, there is a second condition that is also a sufficient (but again not a necessary) condition for a collective good. This is the characteristic that Paul Samuelson has emphasized in his writings on public goods, and which Head has labeled “jointness.” Jointness is dramatically evident when, as Samuelson put it, increased consumption by one individual does not reduce consumption of anyone else. If one more person turns on his television set to a given channel, that does not affect the reception of those who were already enjoying the broadcast; if the minimum size of bridge which is technically feasible is such that it is never congested, an additional person can cross it without cost to anyone. In other words, there is what Head calls jointness whenever the marginal social cost of providing a collective good to another person, once it is already provided to someone, is zero. But when the marginal social cost of a good is zero, the only price for that good that is consistent with complete or Pareto-optimal efficiency is of course also zero. And if that is the socially optimal price, it follows that there is no profit to the entrepreneur who would provide that good at that price, so (at least if the number of people who would enjoy that good is large)

the collective good can normally only be provided optimally by governments or other organizations with the power to pay for the good while charging a zero price for its use.

Some economists are fond of pointing out that a pure, Samuelsonian collective good is relatively rare; if new consumers are added, or some old ones take more, there will usually be at least a little reduction in the consumption of the rest. The waves from the television tower may not ever be used up, but the bridge will usually suffer some degree of congestion.

This contention illustrates the importance of making a link between two phenomena that are often considered independently, but that ought in the interests of theoretical parsimony and policy insights to be labeled with a single concept. Our profession has long been aware of the problem of the industry such that decreasing average costs make it impossible to have marginal cost pricing without running at a loss. The marginal cost pricing that Pareto-optimal efficiency demands will therefore normally require government ownership or price discrimination.

The Samuelsonian pure public good is nothing more and nothing less than a good with a positive average cost and no marginal cost—it is the extreme case of the decreasing cost industry.⁸ In any industry such that, at the socially optimum level of output, the marginal cost is less than average cost, there is market failure, and thus a *prima facie* case for asking whether some form of government intervention could make things more efficient. All of these cases of market failure through decreasing costs, whether they involve the Samuelsonian extreme of zero marginal costs or not, should be classified together. Thus here we consider all such goods as having jointness—one of the two conditions, either of which is sufficient (but neither of which is necessary) for a collective good or externality.

Though Head has distinguished the nonexcludability and jointness characteristics of collective goods, many writers have failed to do so, perhaps because these two characteristics often go together. In the case of the classic public good of defense, for example, it is probably true both that excluding nonpaying beneficiaries would be prohibitively expensive and also that additional persons can be defended by a given military force at much less than the per capita cost for the entire population (an exception to this latter point would occur if a larger population

⁸ That is, where the level of output is defined in terms of the number of consumers who receive a given level of service, rather than in terms of the level of service.

made a country a more appealing prize for an enemy). There is, however, a real need to distinguish the two properties that can make a good collective, because many goods have the one property and not the other, and because the two properties create very different situations for output measurement. These differences have, unfortunately, been seriously neglected, especially in much of the applied literature on output measurement in government.

When a collective good has the jointness or decreasing cost property, but the costs of excluding nonpurchasers are not great, there are promising, unexploited opportunities for output evaluation that are not present when exclusion is exceptionally costly. Though, at least where there are many customers involved, *laissez faire* will not provide an optimal supply of such goods, it does not follow that the government should *never* charge a price. If the marginal cost price of the good is positive, and the transactions cost of charging that price is not prohibitive, it follows that a government can determine the value of the collective good at the margin simply by charging the marginal cost price. This policy will of course be twice blessed, in that charging such a price will also lead to an optimal level of use of the service. The government activity will of course have to run at a loss (or use discriminatory pricing), but if that were not true, there would in this sort of situation be no point in government intervention in any case.

Even when the marginal cost price of a collective good is precisely zero, it is often feasible to get market evaluation of output, as long as exclusion is feasible. Admittedly, there is a fundamental conflict between the desirability of charging a zero price (which is necessary to give consumers an optimal incentive to use the good) and the desirability of a positive price (which is necessary to obtain information about what the output of the good is and what it is worth to those who bought it). But the need for output measurement and evaluation does not require that a positive price be charged to everyone all the time, or even most people most of the time. As a practical matter, it requires only price exclusion on an experimental or sampling basis. Where large numbers of people enjoy some good, we can be confident that the information about the output and valuation of a good provided by a random sample of the users will give essentially the same information as could be obtained from the entire population. And even if we cannot in practice get a sample that is random, strong incentives based on information from an imperfect sample may still induce a far more energetic, flexible, and efficient administration.

Consider the example of a bridge or highway. The prevailing practice is either to make the use of roads and bridges free at all times, or alternatively to charge tolls that are large enough (or more than large enough) to pay for the cost of the road or bridge plus the interest on the loan that was floated to build it. The former policy can be rational in those cases (which may with present practice be fairly numerous) in which the transactions cost of collecting a toll would always exceed the value of the information and the incentive provided through the toll (the main cost here is probably not the toll taker, but rather the time and other loss to the motorist of the extra stops and starts). But, as William Vickrey has argued in discussions of particular cases, it would be easy with present computer and electromagnetic technology to develop toll systems that would have only negligible transaction costs, at least when applied to certain cases (such as controlled access roads during commuter hours). The latter policy (average cost pricing) can be rational if (as J. de V. Graaff has in essence argued) average cost pricing is in some sense better than the alternative forms of taxation. This may be true in particular cases, but it is hardly conceivable that it could always be true.

Naturally, in those road and bridge cases in which marginal cost is significant (at least during some hours) but less than average cost (at least during some hours), marginal cost pricing should be considered. This might in practice mean charging tolls only during rush hours, or a tax on all day parking in the central city, or a special tax on automobile commuters who have regular working hours. If there is a lack of public understanding of the utility of such policies, or if no easy way of keeping transaction costs within bounds can be found, it is tempting to rule out the use of prices. But it must be remembered that marginal cost pricing will not only improve incentives to the user, but usually also provide better information on which to base governmental decisionmaking.

When marginal cost pricing of a general kind is for any reason infeasible for a good with a marginal cost that is less than average but positive, then experimental pricing applied to a sample of potential users can be useful. The commuters in a given representative sample of firms or residential blocks can be charged marginal cost prices and the results blown up for the population as a whole. This need not lead to injustice to those who happen to be chosen for the experiment, since they can be given a lump sum payment as large as the amount they would be required to pay in tolls. This payment can even be significantly larger than the

cost of the tolls to the experimental subjects, so long as it is not so much larger that it has a substantial "income effect."

VII

The approach proposed here can more readily be illustrated by considering the important case of television. Since an additional receiver in no way damages television reception for others, and since "scramblers" that can exclude nonpurchasers already exist, this is an example which has the property that a zero marginal cost is socially optimal and the further property that exclusion of nonpayers costs very little. It is, of course, clear that television paid for by advertisers who use part of their time to advertise their products is not consistent with Pareto-optimality. If the marginal social cost of another viewer is zero, but the viewer must in fact watch commercials for X minutes out of every hour, the viewer obviously has to pay a price for watching television, and this is, of course, inefficient. The magnitude of this inefficiency is enormous. Since we know, at least for those cases in which people are able to choose how many hours of work they do, that leisure must be valued at least at the going wage rate, the value of the time lost watching television advertisements since television was introduced must amount to tens, or more probably hundreds, of billions of dollars for the United States alone. Something must be subtracted from this sum to take account of the informational and entertainment value advertisements have to consumers, and something added to take account of the positive annoyance, loss of dramatic effect, and misinformation that television advertisements sometimes cause.

Pay television or "fee vee" may well be more efficient than television paid for by sponsor advertising. Even those who would not in any event buy the advertised good must waste the time they sit through the commercial, and this is a waste. By contrast, with competitive pay television, the viewers need pay no more than the full average cost of the programs they receive. The fact that prices can readily vary from program to program and channel to channel also makes it easier for pay television to adjust to minority tastes and to obtain full rights to televise expensive events (like the right to show special sporting events on television even in the same community in which the event is held). Of course, pay television isn't optimal either, for the reason that it charges a positive price for something that has no marginal social cost.

It would seem that the best arrangement, given the considerations that

have been discussed so far, is a government television system which meets its cost out of tax funds and provides programming free. This arrangement certainly can be Pareto-optimal. But in practice it may have disadvantages. One disadvantage is that a *single* government agency or corporation which has the benefit of a legal monopoly has some dangers a pluralistic arrangement need not have. The dependence of the state owned French television monopoly on the government in power has led to the use of that system for nearly naked government propaganda. The British experience with BBC has been happier, but (at least in the days before competitive private television) there were charges of some degree of subservience to the government (e.g., during the general strike). Nor is it likely that a federally owned monopoly system in the United States would have interviewed Daniel Ellsberg in a secret location about the purloined top secret papers at a time when he was alleged to have taken them, and was apparently in hiding from the authorities, but CBS did. A publicly subsidized television system in a democracy should therefore not only have great autonomy vis-à-vis the current government, but also many independent, competitive stations or networks, some of which would be willing to defy the government, or air a risky idea, even when the others would not.

Given that free, tax supported television is Pareto-optimal, and therefore such that it would be possible to compensate losers for changing to such a system and still leave everyone else better off, we must ask why free public television has not tended to sweep the field, or at least expand vis-à-vis private advertising television. In the United States, none of the scores of public service or educational television stations has obtained audiences that are in the same league with the commercial stations. Why hasn't even one of the public service stations been able to get a sufficient appropriation from some government to compete with the commercial stations, even, say, in the area of news? The tendency to introduce advertising in some of the continental television systems is puzzling in the same way. Most significant of all, why did the United Kingdom add commercial television after BBC television was already established, and protected by a measure of national pride and the ideology of most of the intellectual elite? And why does independent television in Britain have so many viewers, when the viewers must pay the substantial time cost of the advertisements, when they can watch equally light entertainment on BBC television for free?

One explanation is that a single state owned television or radio system, in the absence of competition from independent systems (be they public

or private), may not be very effective in satisfying consumer demands. It would take a careful study, or in any event wider observation than this writer has managed, to be sure about this. But many people with relevant experience as observers have asserted that this is true. One thinks particularly of the English critic who said of the "leg shows" on BBC, in the days before independent television, that they looked as though they were put on by amateur groups from local convents.

Part of the explanation for this alleged ineptitude in meeting popular demands is surely that publicly sponsored broadcasting is not perfectly responsive to the wishes of the individuals in the electorate, at least in the short run. If it were it would be totally subservient to the politicians in office. In practice certain elites are likely to have a disproportionate influence, and certain minorities almost none; an intellectual elite may raise the intellectual level, an ecclesiastical or moralistic lobby may constrain the amount of pornography, and so on. But this surely is not the whole story. Democracy in the successful democracies surely can't be so ineffective and unrepresentative that this is a total explanation for the long run.

Probably part of the explanation is that a single, state owned television station tends to be run by a bureaucracy that has only a limited incentive to meet consumer demands efficiently. The individuals in the bureaucracy are more likely to be promoted if they fit in with the prevailing professional ethics than if they please customers. If the consumers can readily relieve them by voting, it will become the organ of the party in power. And since the consumers can't be expected to know the production function for television, they can't know whether the state system is as effective as it could be or not. In other words, a single state system will have the problems this paper has argued are characteristic when public goods are provided and there is no measure of the quantity of output.

There is surely no perfect arrangement but one measure that could be helpful in some cases would be to take advantage of the fact that exclusion would be possible on a sampling basis. A sample of potential viewers could be given a flat sum of money and in turn forced to pay for the program they watched. If there are several alternative channels, each with independent management, this system will reveal a good deal more information about what the output of each television channel is, in terms relevant to its users, than if there is only one channel. The rewards to the different program directors and channels or networks should then be related to their success in satisfying the sample consumers. Other objec-

tives could also be pursued at the same time, by making some part of the reward to program directors and networks dependent on some "expert" evaluation of their intellectual level, artistic quality, moral standards, and so on. This approach could be used to improve noncommercial or educational stations in a country with a mainly commercial system and also to increase the efficiency of a wholly public television system.

VIII

There are several other areas in which exclusion on a sample basis (or exclusion of all nonpurchasers, if that is economic) can increase efficiency. It would take at least a substantial volume to analyze all of these appropriately, but it is possible to indicate the general nature of some of the possibilities from a brief and superficial discussion.

Though a great many, if not most, collective goods have the property that it is not feasible or economic to exclude nonpurchasers, there is in the case of every collective good nonetheless a considerable scope for the use of the exclusion principle. This is because collective goods, like other goods, are of course made with the aid of intermediate or producer goods, many of which are private goods. Thus by using markets in all but the very last stage of production of a collective good, and making this stage as small as possible, the area in which the information needed for efficiency is lacking can often be made quite narrow. This is done now whenever the private sector, rather than government factories, supplies weapons, uniforms, and the like to the military, or when a government office is built by a private contractor.

The scope for this can be expanded by close attention to the question of identifying the stage of the productive process that has outputs from which nonpurchasers can be excluded and where markets (even markets with considerable degrees of monopoly power) are possible. It may well be possible to have market-oriented enterprises (which might sometimes be government owned) provide that part of training for military pilots that imparts basic flying skills, especially those that are also in demand in the private sector. It may be the case that private companies can provide rented cars more cheaply than a government motor pool. There are a vast number of other possibilities of this sort.

Some public goods are themselves intermediate goods, from the police services that help to protect merchants from theft to the water for irrigation that comes from a reservoir created by a government. As others have pointed out before, in this case the concealment of preferences presents no great problem, because the worth of the collective good to

the recipient must be given by what it adds to his profits, and this can in principle be estimated without too much difficulty. As Musgrave has emphasized, it is in cases of this sort—especially irrigation projects—that cost-benefit analysis has come closest to being successful. This is to be expected. If, however, the intermediate good is a pure collective good in the sense that nonpurchases cannot ever be excluded, the hairy problem of estimating the volume or quantity of output will still be present.

Where private goods have been gratuitously collectivized, the exclusion principle can very easily be used to increase efficiency. If there is neither any inherent difficulty of excluding nonpurchasers, nor decreasing costs such that the socially optimal output cannot be produced at a profit, then all that is needed is to replace the bureaucratic mechanisms with market mechanisms. The area of public housing is one that, in several countries, probably could be made more efficient in this way. If the argument of this paper is correct, it is the provision of collective goods rather than government per se that breeds inefficiency. Thus the point is not that government enterprise (or of course income redistribution or externality-internalizing subsidies) should necessarily have no role in housing, but rather that any government owned productive undertakings should be fully subject to the discipline of competitive markets.

For many collective goods, exclusion is not altogether impossible in a physical sense—just prohibitively costly. But costs that would be prohibitive if every nonpurchaser were excluded need not be prohibitive if borne only for a small sample of the relevant consumers. In other words, some consumers can be denied a collective good unless they pay for it, and given prior lump sum payments of similar value to avoid any injustice. The exclusion costs in these cases (unlike the television case considered earlier) will be substantial in relation to the value the good has to those excluded, but they may still be small in relation to the gain in efficiency that could result for the mass of consumers who consume the collective good but are not in the sample from which the needed information is to be obtained.

It is sometimes also feasible to exclude nonpurchasers from one level or aspect of a public good even when it is not feasible to exclude on an all or nothing basis. In the case of fire protection, for example, it may be difficult to protect a given house from burning down without putting out any fires that start next door. But in some cases, with enough fire fighting resources a given set of properties can be protected from fire without at the same time protecting all others, at least to the same degree. Thus, that degree of fire protection that is desired for property

and convenience only, and is beyond that needed to protect life and widespread conflagration, can be denied to a small sample of property owners. This has been done: firemen in Missouri not long ago watched from fire trucks as the property of a noncontributor burned down across the street, having assured themselves that no lives were at stake. This particular solution seems a bit drastic, but the principle need not be applied in any drastic way. It would be perfectly possible for a sample of individuals and small neighborhoods to decide whether they would rather have a marginal decrease in fire protection and a tax rebate, or stay with things as they are. It would similarly be possible for individuals and small neighborhoods to be given the choice of more policemen passing by if they would pay the price, or fewer if they would prefer to save the money instead. Quite apart from the more general arguments for permitting consumer choice in the level of collective goods supplied, this approach would, even if applied to only small samples, provide information on the true output and value to consumers of certain collective goods.

Some collective goods or external economies are joint products with private goods. If it is impossible or very costly to change the proportions in which the private good and the externality are produced, efficiency can sometimes be increased by using some assessment of the output of the private goods as a proxy measure for the output of the externality. The ability to read and write, for example, is simultaneously a private investment good that may increase future earning and satisfactions, and also a collective good in countries with democratic procedures partly dependent on the understanding of written arguments. The elementary and secondary educational systems that attempt to teach or improve reading and writing skill are widely thought to be inefficient, at least in big central cities. The big city public school systems certainly often do have the traits that this paper argues are typical of organizations without a measure of the volume of output. They often promote on bases other than observed productivity (e.g., seniority, bureaucratic conformity, and possession of degrees), and rigidly adhere to time-honored rules and procedures.

One way to introduce a greater orientation to output and efficiency is to take advantage of the fact that the private good aspect of basic education is, like all private goods, such that nonpurchasers can be excluded. This means that a market system can often be used to supply basic education. The collective good value of education and distributional objectives can be met by giving parents vouchers that can be used only to pay

for basic education. There is, to be sure, nothing new in this suggestion. It has come from many quarters, and has been particularly attractive to many advocates of *laissez faire*. The approach here, though, would lead to a slightly different emphasis than that which is evident in some *laissez faire* arguments for voucher systems. It would recommend special controls or subsidies designed to insure that the collective good part of the joint product was not slighted; it would insure that the schools at which vouchers could be spent would, so to speak, give "civics" sufficient attention in comparison with "business correspondence," racial integration in comparison with sports, and so on. This is because the parents would of course have no incentive to favor schools with a sufficient intensity of external economies.

IX

Where the exclusion of nonpurchasers is impossible even on a sample basis, it becomes again particularly important to distinguish the revelation of preferences problem from the lack of a measure of the volume or level of a collective good. When exclusion is impossible, there is no completely satisfactory way of getting consumers individually to reveal how much a further unit of a collective good is worth to them. So the preferences revealed by the political process must be used. But this process will work better, and the government agency providing the collective good will be more likely to attain technical efficiency, if there is a measure of the volume of each collective good.

The volume of many of the important collective goods—law and order, clean air and water, public health—is roughly measurable. The probability that a representative citizen will fall victim to a crime, the levels of sulphur dioxide in the air, the biological oxygen demand in the rivers, the risk of contagious disease, and the expectancy of healthy life can all be measured, and sometimes have been. The measurements sometimes cannot be at all accurate, but then neither is a businessman's estimate of the elasticity of the demand curve he faces, or even the exact level of his marginal costs, yet the private sector often works fairly well with this approximate information.

It is often possible, in other words, to get "social indicators." A social indicator, at least as this writer defines it, is a nonmonetary measure of social output or performance—a measure of welfare or illfare to which no price has yet been attached. It can be used to supplement the information about welfare in the national income and product accounts. A measure of the volume of a public good that is of direct normative interest to a

society would then have to be a social indicator. Since preferences will often be concealed, and will in any event differ for people with different preference orderings or value judgments, it is utopian to expect consensus about the monetary value to be placed upon each social indicator. Rational public decisions about resource allocation will of course require that the politicians in power put some value or price on the alternative outcomes that could be obtained by using the same public funds in different ways, but there will rarely be a consensus that they have used the right values. But, to repeat, there is no reason in principle (outside the defense and international relations area) why tolerable physical or social indicator measurements cannot be obtained.

This paper has argued that volume or physical measurements of the output of public programs are all that are required to attain technical efficiency, and that this failure to reach the production possibility frontier is probably the largest source of inefficiency in government. It is also true that the art of measuring the output of public programs is in its infancy, and that there hasn't been a great deal of progress thus far. Thus the mind leaps to the possibility that the *relatively* tractable task of collecting social indicators could open up a smoother path to output measurement in government. This is a possibility which, in this writer's opinion, deserves more attention than it has so far received. It has probably been neglected in part because the PPB system got its form in the Defense Department, and it happens that this is one area where direct social indicator measurement can contribute least (it is not feasible to measure "national security," partly because it depends on a small number of other countries whose behavior is not readily predictable, either individually or in the mass; so output evaluation in defense has been based on engineers' experiments on how weapons will perform, etc.).

The problem here, as with all conceivable approaches to output measurement in government when exclusion is impossible, is that so many different variables influence the social states we are interested in. Crime rates depend not only upon policing, but also on family patterns, street lighting, income levels, the courts, the degree of urbanization, and what not; health depends not only on the public health service, but also on diet, smoking, exercise, the drug industry, nervous tension, and no doubt other factors that haven't even been specified yet. And it is presumably as complicated in other areas.

If the exclusion and even the sample exclusion possibilities set out above are infeasible, there are only two further conceivable ways in which the outputs of government programs can be measured: experi-

mentally or statistically. We must either set up controlled experiments which keep all independent variables except the control variable essentially constant, and then calculate the effect of the control variable on output in those conditions; or alternatively we must collect a good deal of disaggregated data on the level of the variables we ultimately want to change (the social indicators) and use them along with data on resources used, and with well-specified statistical models, to estimate social production functions.

Neither approach promises striking results. Experiments are often infeasible or morally offensive. Child abuse and unhappy homes may or may not be a cause of crime, but it is out of the question to do the experiments needed to find out. Longitudinal statistical studies of criminals and noncriminals might give the answer, but for confidence they would require data we are not likely to have. To find out the value of a certain level of a collective good, someone has got to be excluded from that level of the collective good, which may be politically or morally objectionable even on an experimental basis. But "nature" may not have provided enough variety in levels of the collective good to allow statistical estimates either. In practice, cost-benefit analyses and PPB studies have often had to deal with the intermediate goods, or avoided outright efforts at output measurement, and it is easy to see why.

On the other hand, there is no telling what might be done to increase governmental efficiency if more social indicators could be collected and better experiments could be conducted. And there are many relatively obvious ways in which institutional improvements could be made. It would help, if the argument of this paper is at all correct, to prohibit inbreeding in bureaucracies: to insist that in any bureaucracy people can come in at any level from the outside, and thus sometimes from areas where productivity is better measured and where the conventional wisdom is different. If pension rights are fully vested so insiders can compete for outside jobs without handicap there is no need for this to lead to injustice. There is also, as others have pointed out, the need to be wary of the influence of organizations representing inputs used in government on the allocations chosen by the political and administrative process.

It follows from what has been said here about the lack of output measurement in government that administrators and politicians are often more likely to be judged by the extent to which they please politically organized suppliers of resources to the government than by their output. But whether through experiments, statistical estimations, or changes in

institutional arrangements, general efforts to improve efficiency in government can never realize their full potential unless they start with the source of the problem: the defining characteristics of collective goods.

COMMENT

CHARLES L. SCHULTZE, The Brookings Institution

Mancur Olson's basic thesis is that government is fundamentally and incurably inefficient in much of what it undertakes. His argument has four steps: (a) government produces collective goods; (b) no particular individual can be excluded from the use or enjoyment of such goods—if they are available for one, they are available for all; (c) the output of collective goods, therefore, does not take the form of divisible physical units that can be counted; (d) since output cannot be measured, productive efficiency cannot be measured, and hence it cannot be achieved. He also argues that problems of inefficiency in the conduct of government programs do not stem from the fact of "concealed preference"—i.e., from our inability to put a value on the physical output of government. Rather they stem from the impossibility of measuring physical output itself.

In evaluating Olson's proposition it will be useful to start with a formal definition of efficiency. In the case of *private* goods there are three aspects of efficiency: first, efficiency in exchange—all marginal rates of substitution (*MRS*) are equal among individuals, an equality that is secured by rational consumers reacting to market prices ($p_1/p_2 = MRS_i = MRS_j$; etc.); second, efficiency in production—the ratios of the marginal products of productive factors are equal in all lines of activity and that equality is secured by rational producers minimizing costs, given factor prices ($MPL/MPK^a = MPL/MPK^b = W/p_k$); third, joint efficiency in production and exchange—marginal rates of substitution are equal to marginal rates of transformation (*MRT*), and that equality is given by producers choosing a volume of output such that prices are equal to marginal costs ($MRS = MRT$; $p = MC$). Because of the non-excludability property of public goods, the first and third equalities collapse into one: marginal rates of transformation are set equal to the sum of marginal rates of substitution ($\Sigma MRS = MC$). The first and third equalities have to do with efficiency in resource allocation; the second is a requirement for efficiency in production. While Olson does not explicitly develop these various aspects of efficiency, he is clearly con-

cerned with efficiency in production, rather than efficiency in allocation.

I shall argue (1) that our inability to measure output is indeed an important reason for the difficulty of measuring and securing efficiency in the execution of government programs; (2) that the concept of *physical* output, however, has little relevance—the major problem in measuring output is the difficulty, or indeed impossibility, of attaching *values* to various aspects of governmental activities; (3) and that another major reason for inefficiency in government program execution stems from the incentive problem which necessarily plagues bureaucracies.

In both the public and private sector we are not basically interested in measuring output in the purely physical sense. Private output is measured in terms of the values which arise in the market place. A refrigerator is five times the output of a portable TV set, not because it is five times as large or five times as heavy or has metal five times as durable or because of any other set of physical characteristics, but because consumers value it at five times a portable TV set. The problem in the public sector is not our inability to measure physical output—we can often do so. But the purely physical aspect of output is economically meaningless. The real problem of measuring performance or efficiency in the public sector is that we do not have market values to place on the goods or services produced by the government.

Let us take tanks and automobiles as examples. An “automobile,” not otherwise described, is a meaningless measure of output for personal consumption, just as a tank per se is a meaningless measure of national defense output. An automobile is two tons of steel, other metals, and textiles, so assembled that it moves under its own power when fueled. But a Cadillac Eldorado with a built-in bar and all the accessories is five times the output of a Volkswagen. There are myriad things which contribute to the satisfaction of consumer desires for transportation, convenience, and status provided by an automobile. How do we measure the “output” represented by disc brakes versus shoe brakes; by hand-tooled leather upholstery versus cheap textiles; by radial steel-belted tires versus conventional tires; by electric window openers versus manual openers; by two-speed versus one-speed windshield wipers; by fuel injection versus a carburetor; etc. Because of these and many other kinds of differences, an increase in “automobile” output represented by one Cadillac has twice the weight (as measured by relative prices) of an increase in output represented by one Ford.

The relationship of the physical characteristics of automobiles to consumer satisfaction is not basically different from the relationship

of the physical characteristics of a "tank" to national defense. We can count cars and we can count tanks. We can go further and provide physical measures of a host of characteristics of a car and a tank. But in the case of cars, the exceedingly complex collection of physical characteristics can be made commensurable and their contribution to output measured on the basis of the price consumers are willing to pay. And in the private sector, productive efficiency comes from minimizing the cost of producing particular collections of physical things called a car. But the "output" whose costs are minimized is meaningful only in terms of relative values assigned by consumers. In the case of a tank, however, we have no measure of value to associate with the tank. As a consequence, we must attempt to impute some kind of value to various potential physical characteristics of a tank on the basis of their presumed relationship to national defense. In any event, the problem is not the physical output represented by a car or by a tank. The problem of measuring output in the case of the tank is that we have no value to attach to it. The problem is not that it is impossible to count tanks.

Education is another example. This is a private good with public externalities, produced in part by the public sector and in part by the private sector. The externalities are most usually joint products with the private good. It is impossible to exclude particular individuals from the enjoyment of the public externalities and therefore impossible to ration this aspect of the good through the price system. But even viewed purely as a private good, a measure of the physical output of education is meaningless. The quality aspects of education are most complex and difficult to measure from an analytic standpoint. Do we measure output by the number of students educated? Are Princeton and Podunk Junior College to be counted the same? To the extent that we wanted to treat education as a private good and to accept the values which arise in the market place, the ultimate output measure of education would depend on consumer valuation. In that case, we would agree, for example, to treat an additional year at Princeton as, say, three times the year at Podunk. But the fact that one year at Princeton is three times a year at Podunk would arise from our acceptance of market values. Whether we should accept market values or not, whether education is a good which should be produced publicly because we don't want to accept market values is not at issue here. What is at stake is that in the case of education as both a private and a public good, output can only be defined in value terms. The concept of *physical* output is literally meaningless.

It might be argued that even though the concept of physical output has no economic meaning in measuring the *level* of output, physical output does take on meaning when we are measuring changes. We can construct an output index by designing indexes of each of the specific physical characteristics of a particular output, weighting each characteristic by some value. But on reflection, two things become evident. First, value weights are needed to obtain our index. The output increase represented by an additional Cadillac is not the same as the increase represented by an additional Ford; nor is one added student in the third grade the same as one added student at Harvard medical school. A measure of efficiency which assumed that the two changes were equal would surely provide misleading advice to decisionmakers. Second, only in those cases in which the nature of the physical characteristics of the good in question does not change can we unambiguously construct a component index to be weighted. To measure the economic meaning of a change in the physical nature of a product we must again rely on valuation. This is the nature of the linking procedure through which we handle changes in the physical nature of output. And, of course, precisely the same problems arise when output indexes are constructed by deflating expenditure data. In this case the quality problem rears its head in the construction of the price index where in the other it is in the construction of the output index.

In short, identifying and measuring the physical characteristics of output is neither more nor less difficult on the average in the public than in the private sector. In either case, the physical characteristics of almost all outputs are exceedingly complex. There are a few cases—usually of intermediate output—in which simple physical measures are meaningful, e.g., the Treasury check-writing process and the Veterans Administration processing of insurance premiums. But in almost all important situations, the physical characteristics of output are very complex, and the achievement of efficiency in production depends in significant part on assigning the proper weight to each of the many characteristics of output. Without such weights the attempt to minimize cost for some few arbitrary physical characteristics inevitably leads to distortion.

There is the classic case of the early Soviet nail factory in which quotas were set in terms of the *weight* of nails produced. In response to this implicit incentive a small number of very large nails was produced. When an attempt was made to correct this obvious distortion by measuring output in terms of the *number* of nails, quotas were met

and Stakhanovite awards were won by producing a very large number of very small nails. A similar situation arises at the present time in attempting to measure the performance of those insurance companies who contract with the Social Security Administration as intermediaries in the payment and regulation of claims for hospital services under Medicare. One might think that the ratio of administrative costs to total benefits paid or to total business handled would be a good measure of the efficiency of "output" of these insurance intermediaries. It turns out, however, that insurance firms can easily minimize their administrative costs by doing practically no monitoring of claims, no weeding out of excessive charges, no checking on overutilization of services, and the like. As a consequence the measure of output in this case turns out again to be exceedingly complicated, and the real problem comes in assigning weights to the various possible aspects of output. There is no way of securing superior performance without carefully specifying the output to be desired, which in turn cannot be done without assigning, either explicitly or implicitly, some sort of weights to each of the aspects.

In three different places in his paper, Olson implicitly recognizes that output cannot really be measured in the absence of knowledge about relative values. In discussing the problem of output measures in the service sector, he notes that private firms producing janitorial services need not measure physical output—they simply observe the marginal revenue from supplying janitorial services and act in the light of that knowledge to minimize the cost of supplying such services. He is precisely correct. But janitorial services are not really different from automobiles or toys. The output meaning of the latter does not arise from the fact they can be touched, or seen, or broken. Rather their output meaning is imputed from the services they render. Indeed, it is impossible even to define an automobile as a physical output in anything less than a 20,000 page book of detailed specifications.

In another case Olson talks about the usefulness to decisionmakers of pricing at marginal costs the services of a bridge which has positive marginal costs. As he correctly points out, the efficiency gains from this practice would extend beyond the simple welfare rules of setting price equal to marginal costs. Pricing practices provide "better information on which to base government decisionmaking." I agree. But if one asks what kinds of information are provided to governmental decisionmakers through the price rationing device, the answer is that information is obtained about demand, and hence about values. Charging a price for a bridge-crossing, with the price moving up the slope of the marginal

cost curve as congestion increases, provides information that helps to make better investment decisions about building a new bridge. It is the information on values provided by the pricing mechanism which is important. More generally, Olson's suggestions about pricing government services go not to the securing of information about physical outputs so much as to securing information about values—indeed how could we price anything unless we already had a measure of the units to be priced!

Olson also discusses the desirability of contracting out certain operations of the Defense Department. He uses as an example contracting out basic pilot training. But to do this, the Air Force would clearly have to spell out the output it wanted from the contractor. In turn, defining output could not be done simply in terms of the number of pilots turned out. The proficiency of pilots is relevant, as is the time span of the training and the auxiliary skills acquired. Somehow these and other important aspects of pilot training would have to be taken into account and weighted by Air Force "values" in order to draw up a contract.

More generally in the case of the Defense Department, there is much inefficiency in the procurement of physical weapons systems produced under private contract. This inefficiency stems from many reasons but two are most important: (1) the inability of the Department to settle on the "value" to it of the myriad physical characteristics going into a weapons system, as a consequence of which design changes, production interruptions, and redesign problems are legion and (2) the inability to devise real incentives for procurement officers and higher decisionmakers to care enough about unit costs.

In summary then, Olson is partly correct when he points out that inefficiency on the part of government is fundamental and in many cases incurable. He is also correct when he points out that these problems stem from an inability to measure output. But he is wrong, I believe, to suggest that this is not a problem of concealed preferences and therefore value, but rather a problem of measuring physical output. Considered from an economic standpoint, physical output is a meaningless concept in both the public and the private sector. Finally I would note that even when output is measured or measurable, after taking relative values into account, major efficiency problems remain because the incentives of governmental decisionmakers are often quite different from those which would tend to maximize output for a given volume of resources or minimize resource input for a given output. The molding and maintenance of a political consensus, high risk aversion to sharp changes in hiring, firing, and location policies, and the

difficulty of translating analytical output measures into politically appealing terms are but a few of the reasons why the measurement of output—important as it is—represents only one aspect of the problem of improving governmental efficiency. With these last remarks I am sure Olson would agree.

GENERAL PROBLEMS OF MEASURING PERFORMANCE AND OUTPUT IN THE PUBLIC SECTOR

Performance Measures Versus Output Measures

Performance measures and output measures are not the same. In general, questions of performance go well beyond the measurement of the actual outcome of production decisions. For example, if the performance of the auto industry were at issue, a number of questions might be raised going well beyond the accuracy of our current measures of automobile output:

1. We might inquire into the status of the industry with respect to its competitive situation. From a resource allocation standpoint, the industry may be producing the wrong level of output very efficiently.
2. Even if an industry is competitive, we might ask how dynamic it is in adopting technological change. A perfectly competitive industry may still be a stagnant one.
3. We may wish to go behind consumer evaluation and look at the performance of automobiles in terms of their safety, their durability, or their economy of operation. In this case, our performance measure would implicitly question consumer evaluation on grounds of lack of information or Galbraithian manipulation of preferences.

In the context of the national accounts, it is output, not performance in the above sense, that we wish to measure. In other words, we want to measure actual outcomes which reflect collective decisions and evaluations without looking behind those decisions. The kinds of data and analytic structures which are useful for performance measurement may not be the same, and will almost always be much more complex and detailed than the data and analysis needed for output measurement.

Measurement of Public Output in the National Income Accounts

Even without raising some of the very fundamental questions about the nature and the structure of the accounts which have been brought forward in this conference, it is possible to point to several areas in which modest improvements could probably add to the usefulness and

in some sense the accuracy of the accounts. One of these has to do with the way in which state and local government expenditures are deflated in the measurement of constant dollar GNP. In fiscal year 1970, state and local governments spent \$132 billion. Of that amount \$70 billion was compensation of employees. The deflation of this component of the state and local government sector is done by exceedingly aggregate techniques. Compensation of employees is divided into two parts: (i) the compensation of employees in education; and (ii) all other compensation. In each of these two components, changes in real output are measured by changes in employment. This is equivalent to deflating each of the two components by an index which is simply the aggregate payroll in that component divided by the aggregate number of employees. As a consequence, a shift in activity within either the "education" or the "all other" sector from a function involving low wage employees to a function involving high wage employees is washed out in the deflation. A \$30,000 per year doctor and a \$5,000 per year garbage worker are considered equivalent in terms of the output value of their services. A shift in employment from one to the other does not show up as a change in output. What is needed is a compensation index for state and local employees which is a true index. It should *not* simply be a measure of average compensation per man reflecting both wage increases and shifts among occupational categories. The construction of such an index would require breaking state and local employment into a number of different components, constructing wage indexes for each component, and then combining them with base period employment weights. There are several ways of accomplishing this. Edward Denison has suggested "specification pricing." Demographic, educational, and other characteristics of employees would be specified for a number of representative categories, e.g., a forty-year-old college engineering graduate. Wage and salary data would be collected for each category, weighted, and combined into an index. Another alternative would be the use of occupational specifications. Which of the two alternatives is preferable would depend principally upon data availability. In any event, the resulting wage index used as a deflator would at least allow the measure of this major component of state and local output to vary as employment was shifted from low wage to high wage categories of employees. It may turn out that in any given period the difference in the two measures (i.e., the current measure and the proposed measure) would be minimal.¹ Nevertheless,

¹ Subsequent to the conference, Charles Waite of the Office of Business Economics informed me by letter that he has calculated an alternative series of

the effort is worth pursuing because we have no a priori reason to believe that shifts in the mix of output will remain unimportant.

Other Possible Improvements in Measuring Public Sector Output

There are several major components of the public sector whose level of output and rate of change in output are probably systematically misstated in the national income accounts.

The accounts make a number of necessarily arbitrary distinctions between activities which are economic in nature and those which are not. The former are included in, the latter excluded from, gross national product. More precisely, distinctions of this nature generally turn on where and how activities are carried out rather than on the nature of the activities themselves: the household versus the business enterprise is the chief distinction. Education in the home is "non-economic" while education carried on in a public or private institutional setting is economic; preparation of food in a restaurant is economic while preparation of food in the home is noneconomic. This distinction leads to some anomalies. The unpaid volunteer time of mothers in a cooperative kindergarten outside the home is not valued in the accounts while the wages of the maid who prepares food at home are counted. But even these anomalies are roughly consistent with concepts generally employed in the accounts. The volunteer mothers, for example, most generally are diverting time from household activities not covered in the accounts and by accounting definition their opportunity costs are zero.

But there are two major areas in the public sector where the valuations of output, by the conceptual usages of the accounts themselves, are significantly understated. One is the military budget where the wages paid draftees rather than their foregone earnings are used to measure output. The other is education where student wages foregone are excluded from the measure of output value.

deflated state and local educational output, using data on wage rate changes for five categories of employees. The increase in deflated output between 1961 and 1970 as shown by this alternative measure is only slightly different from the increase shown by the method of aggregate deflation currently in use—61.3 per cent for the former versus 60.6 per cent for the latter. This result occurred, despite the shift in employment from elementary and secondary to higher education, because the number of relatively low paid noninstructional personnel in higher education is increasing nearly as rapidly as instructional personnel. As a consequence, the difference in *average* wages paid in elementary and secondary education on the one hand and higher education on the other is nowhere near as large as the difference between the wages of teachers in the two systems.

Let us start with education. Unlike the case of the volunteer mothers in the kindergarten, the time of many students, particularly in higher education, does represent a diversion from activities otherwise included in the accounts. If the combination of public and private decisions about education roughly balances marginal costs and marginal benefits, the value of education is substantially understated in the national income accounts. The marginal product represented by foregone wages is a diversion from output which the accounts otherwise measure. More importantly, the *raison d'être* for including foregone earnings in measuring the value of educational output stems from the hypothesis that costs can be a surrogate for value only if all costs are included.

The current practice in constructing the accounts implies that part of the value of education is economic in nature and part is not and that the division of the value of the product into economic and noneconomic components is precisely equal to the division of costs into monetary and imputed components. But there is no warrant in the nature of education as an output to make such a distinction. None of the analysis of the impact of education on economic growth, for example, makes this distinction. This is not to suggest that the existing main structure of the accounts necessarily be altered. But a supplementary measure of educational output which included foregone earnings and which classified education as an investment outlay would be exceedingly useful for analyses of economic growth, savings, and investment. Only if we assume that education is substantially overproduced—that is, output is pushed to the point where marginal benefits are substantially below marginal costs—does the current measure of educational output in the accounts, which leaves out foregone earnings, provide an appropriate evaluation of that output.

The personnel component of the military budget provides another clear example of the same problem. The wages paid draftees have been until recently substantially less than their foregone earnings. As a consequence the resource cost of the defense budget is significantly understated. While the pay raises which have accompanied the move toward a "volunteer army" have eliminated most of the differential, it looms very large in the historical data.

In the case of education and military spending, both of which are explicitly defined to be economic activities subject to measurement in the national income accounts, a transfer of manpower resources (students or draftees) from other sectors of the economy to these activities automatically tends to reduce national output. The reduction does not

stem from any explicit hypothesis that the value of output per man is necessarily less in these areas than in the sectors from which the manpower has been withdrawn. Nor does it arise because activities are being transferred from market to nonmarket sectors. Rather the reduction stems simply from the failure to impute foregone earnings to the military or educational activity in question.

Imputation of foregone earnings in the measurement of output would also require that the receipts side of the accounts be appropriately modified in order to maintain equality between receipts and expenditures in the particular activities concerned. In the case of military spending the excess of foregone earnings over the military wages paid draftees could be treated as a tax levied upon the draftees. In the case of education, the problem is not so simple. It could be handled, as is imputed rent, by treating the individual as a firm investing in education and the value of foregone earnings would be an imputed cost paid by the person as consumer to himself wearing another hat as educational entrepreneur.

ERNEST W. GROVE, U.S. Department of Agriculture

The principal theme of some papers at this conference has been a *non sequitur*: There is a tremendous need for elaboration of the national accounts in various directions; therefore it can and should be done. Olson's paper is in this by now highly respected tradition. He recognizes many obstacles in the way of his various proposals for measuring government output or the demand for government services, but he remains an incorrigible optimist nevertheless. His analysis of the problem of efficiency in government is good as far as it goes, but his suggested solutions are just another stab in the continued dark.

The much and justifiably abused PPB system provides Olson's starting point, but he concludes that "it is perhaps the principal achievement of the PPB system that it has sensitized governments to the need to evaluate the level of public outputs"—a thoroughly backhanded compliment at this late date. Defense Department origins of PPB are noted—but not the most notorious fiasco in attempting to measure "government output," the infamous "body count" in Vietnam. When first instituted, the "body count" system doubtless had much logic (of a sort) in back of it. What better measure could there be of a soldier's work performance?

But inevitably, after a while, it no longer mattered what "bodies" were being counted, and the GI and his officers were given a bad time if bodies were not produced.

This is quite typical of the way the PPB system has worked since it was forcibly and inflexibly foisted on all government departments in 1965. The result was a triumph of technique over purpose, and a tremendous burgeoning of new jobs and unnecessary work. It is impossible to disagree with the basic principles underlying the PPB system; the problem arises in their practical application, as all kinds of cooked-up statistics, fudged or invented data, and other necessary subterfuges are likely to arise when basically sound principles are indiscriminately applied to all areas of government activity. And George Jaszi was right when he said that impractical theory has to be bad theory.

Olson argues, without any substantial evidence, that even "fragmentary and approximate information" on government output could bring about substantial improvements in efficiency. And he states that such "fragments of information," even though "rather inaccurate," would "give agencies and bureaucrats a reason to be as efficient as they can be." But how would this approach differ from the present PPB system, now applied throughout government and very largely a failure? We already have things like "Operations Evaluation and Improvement Divisions" and "Offices of Management Improvement," but they are little more than window dressing and completely under the thumbs of management types now running the government.

The question must be raised, therefore, as to who would collect this fragmentary and probably inaccurate information. It would be available in many cases only in field offices of the agencies, so it should logically be collected by the agencies themselves—as is now the case! Does Olson believe that agency managements would be unable or unwilling to insure that only favorable information came to light? Or, on the other hand, does he believe that, if such information were to be collected by an outside and independent agency, there would be no danger of creating a Gestapo, USA?

Emphasizing the carrot instead of the stick, Olson further states that "it is surely possible to multiply rewards for fragments of favorable information to some extent, and thus to create a large incentive for efficiency with a small amount of information." This is not at all certain; in fact, it is most unlikely. The present incentive awards system is close to bankruptcy because of favoritism and image-building. Published dollar totals of "savings" as a result of accepted suggestions are considerably

less than honest. And if Olson would choose to administer his "rewards" from somewhere outside the agency, the problems would be different but no less serious.

The proposal for "price exclusion on an experimental or sampling basis" is interesting but quite unrealistic. How long would it take to persuade the Congress to approve that kind of operation? The development of "social indicators" is a more practical proposal, but the inherent imprecision of such indicators would take administrators off the hook and negate the whole purpose. And the measurement of government output through controlled experiments or statistical model building is not considered promising even by Olson. So he winds up with the suggestion that we "prohibit inbreeding in bureaucracy!" This is a good proposal, of course, but it is comparable to support for home and mother. Much more to the point would be Ralph Nader's proposal that professional employees in government develop "an ethic of whistle-blowing" so as to keep their wrongheaded superiors in line.¹

It is just about impossible to measure most aspects of government output, and this reckons without consideration of the many negative aspects of such output. Production may not be what was intended, but something totally different—and quite undesirable. It follows that, although measurement of government output would enable us to solve some problems, government inefficiency is actually based on something much more fundamental—the bureaucratic organization itself.

There is a growing body of opinion which argues that our burgeoning technical and organizational structures have become the main source of evil in our society, and that, to correct this situation, ordinary rank-and-file employees should be given—or should simply take on their own initiative—many more rights and much more responsibility than they now have. The whole purpose of organizational structures in government is to reduce—or even eliminate—the human element in all relations with the citizenry. So we have universal specialization, divided responsibilities—even thoroughly fragmented in some cases—an elaborate hierarchy of supervision and rigidly specified and pyramidal organization charts. The result has been that both general inefficiency in government agencies and grossly negative outputs, which are all too common, "have less to do with the malice of individuals than with unexamined and unquestioned institutional practices"—and with the universal Eichmann syndrome "wherein individual motives, consciences, or

¹ See Philip M. Boffey, "Nader and the Scientists: A Call for Responsibility," *Science*, Vol. 171, 1971, pp. 549-551.

goals become irrelevant in the context of organizational behaviors.”²

Giving ordinary government employees more responsibility would not be likely to improve the measurement of government output, but it would almost surely guarantee that some of the more grossly negative outputs would not continue indefinitely, as they do now, and it might even improve routine efficiency as a result of greater involvement by rank-and-file employees. As Nader has pointed out, “it is difficult for outsiders to monitor” large bureaucracies, so that any significant improvement is probably dependent on insiders accepting some responsibility. Nader feels that such insiders, especially the professional employee, should feel a “duty to dissent,” preferably with this duty “protected by an organization of his peers, by his professional society, and by law that requires due process and substantive justice.”³

This is the general direction in which improvements in government efficiency must be sought. Little if any progress can be expected if efforts are confined to the scattered and narrowly technical tracks laid out by Olson.

REPLY BY OLSON

I

It is clear, despite his gentle and disarming style, that Charles Schultze thinks that one of the fundamental conceptual points in my paper is wrong. Though we agree on a great many questions, and seem to see the applications and practical issues in much the same way, he contends that my initial and basic argument—that there is another systematic cause of inefficiency in the provision of public goods distinct from the concealment of preferences problem—is erroneous.

I have found Schultze’s comment extraordinarily helpful. It has certainly stimulated and advanced my thinking on this subject. In addition, it has allayed my initial uneasiness that my basic conceptual point, even if neglected in the literature, was already familiar to most leading specialists, and my concern that my initial point was so simple as to be obvious. If such a leader in this area of inquiry as Charles

² See Matthew P. Dumont, “Down the Bureaucracy!” *Trans-Action*, October 1970, pp. 10–14. Dumont goes on to state that inefficiency and inhumanity “can be seen in pure culture in . . . the men and women who buzz out their lives in the spaces defined by the United States government” (p. 10).

³ Boffey, *op. cit.*, p. 549.

Schultze thinks my basic conceptual point is wrong, it is hardly likely that that point is generally taken for granted in the profession. Nor can the point be as obvious as I once thought: it is simple, certainly, but that is not the same as being obvious.

Useful and intelligent as Schultze's comment is, it is not technically correct. Nor is its critical conclusion valid. To show this, however, it is necessary to go into some important questions which I possibly should not have taken for granted in my paper.

Since this is an area in which questions have to be posed very precisely, it will be well to make clear one point that is *not* in dispute, even though that requires repetition of a point already stated in my paper, namely, that the lack of revealed preferences is indeed part of the problem of attaining efficiency in the provision of collective goods. The paper distributed to the conference said this quite explicitly, among other places, at the beginning of section IV. Since this passage (and others like it) apparently lack the prominence they should have had in the original paper, I quote it again here:

One cause of inefficiency in the provision of collective goods is familiar from the theoretical writings in welfare economics and public finance, but rarely mentioned in the PPB or cost-benefit literature. That is the difficulty of getting consumers to reveal their preferences concerning a collective good or externality, and preferences must of course be known to determine how much it is optimal to provide.

Since the problem of the revelation of preferences will normally be present only when nonpurchasers cannot be excluded, and such non-exclusion is also necessary to the additional cause of inefficiency that my paper introduced, it is often true that when preferences are revealed, both problems tend to be cleared up. As I have only lately come to realize, the fact that the possibility of exclusion will usually clear up not only the revelation of preferences problem but also the further source of inefficiency my paper explained makes that further source of inefficiency more difficult to understand. Yet, as we shall see, my further source of inefficiency can be solved in the production of certain goods when the preferences of consumers are not known, and can also remain unsolved when they are, so that the two sources of inefficiency are demonstrably distinct.

Charles Schultze and I are also in agreement, I think, about the importance of incentives—no amount of information about outputs or anything else will be sufficient to bring efficiency if the relevant actors

do not have incentives to be efficient. At the same time, I see no reason to change the view expressed in my paper, that it is impossible to set up incentive systems that will reward efficiency unless there is enough information about output as well as inputs to distinguish efficiency from inefficiency; you can't reward efficiency until you can find out who deserves the rewards.

II

Part of the difficulty arises because too little attention is usually given to distinguishing the different stages of production and their different relationships to the choices of consumers or governments. Charles Schultze is probably rather typical in comparing the tank and the car as though the former were a public good and the latter a private good: The car, related to consumer satisfaction, plays the same role as the tank related to national defense. "We can count cars and we can count tanks. We can go further and provide physical measures of a host of other characteristics of a car and a tank." But in the case of the car we get a measure of output from the price consumers are willing to pay for this complex collection of characteristics.

In fact, both the tank and the car are of course private goods. There is nothing about a tank which makes it impossible, or even expensive, to exclude those who do not purchase the tank from getting it. And there is of course an international market in tanks and other armaments, albeit one in which most of the buyers and some of the sellers are governments. Defense is a collective good, but it is one which is normally produced with the aid of a wide variety of intermediate goods, most of which are private goods. In this respect, defense is like most other public goods: police departments purchase automobiles and patrolmen; public health agencies purchase doctors and vaccines, and so on. Public goods are of course sometimes also intermediate goods used in the production of private goods, as when government levees protect factories from flooding or public police protect factories from theft.

What has just been said is obvious, and does not of course tell Schultze or any other skilled economist anything he did not know before. The tendency to treat a tank as a public good is no doubt the result of a minor oversight or a casual mode of expression. There would be no need to touch upon any such oversight, but for the fact that in this case this inadvertence is intimately tied up with a far more important difficulty.

III

This more important difficulty arises from the neglect of the distinction between feasibility of prices and markets in general and the special problem of the concealment of preferences. My paper argued that there was a special problem of output measurement with public goods distinct from that due to any tendency for preference to be concealed, but never denied that prices were needed even to attain technical efficiency—which, to repeat the paper, is conventionally defined as a situation such that no more of one output can be attained without getting less of another. Prices are indeed very important to the attainment of technical efficiency. Schultze implies at diverse points in his comment that the problem of the revelation of preference is a sufficient general explanation of the difficulty of pricing public goods (and private intermediate goods used in the production of public goods). Consider the following passage from his comment:

Olson also discusses the desirability of contracting out certain operations of the Defense Department. He uses as an example contracting out basic pilot training. But to do this, the Air Force would clearly have to spell out the output it wanted from the contractor. In turn, defining output could not be done simply in terms of the number of pilots turned out. The proficiency of pilots is relevant, as is the time span of the training and the auxiliary skills acquired. Somehow these and other important aspects of pilot training would have to be taken into account and weighted by Air Force "values" in order to draw up a contract.

More generally in the case of the Defense Department, there is much inefficiency in the procurement of physical weapons systems produced under private contract. This inefficiency stems from many reasons but two are most important: (1) the inability of the Department to settle on the "value" to it of the myriad physical characteristics going into a weapons system, as a consequence of which design changes, production interruptions, and redesign problems are legion and (2) the inability to devise real incentives for procurement officers and higher decisionmakers to care enough about unit costs.

Much the same lumping together of the problem of revelation of preferences and the pricing problem was evident in Schultze's previous comment, paraphrased earlier, about tanks and cars.

Schultze's emphasis on the complexity of goods like pilots, tanks, and automobiles is extremely useful. A great many goods obviously have many different characteristics, and thus cannot adequately be measured in terms of any single physical quantity. But Schultze's conviction that

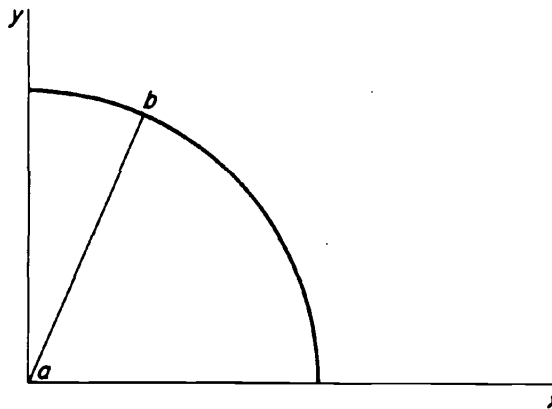
the existence of such complexity negates or limits my argument seems to grow out of his tendency to derive the difficulty of getting market prices entirely from the concealment of consumer preferences for public goods.

In fact, the lack of a market price for output is easily distinguishable from the problems caused by the concealment of preferences: first, because it is often possible and useful to have market prices for many outputs even when consumer preferences are not known, and second, because, as my paper argued, it is sometimes impossible to get a market value or other measure of output of a good for reasons that do not derive from any concealment of preferences.

The first reason why the concealment of preferences problem and the lack of market prices problem are distinct can be demonstrated by taking a situation in which consumer preferences are not known, and demonstrating that prices could still be useful, in that they could put the individuals in the economy on a higher point on their (unknown) indifference maps. Assume an economy such as is depicted in Figure 1, which produces two private goods, x and y , subject to the constraint imposed by the production possibility frontier that is shown. Though the indifference maps are not known, I assume that they are smooth, convex to the origin, and do not intersect. Both x and y are always goods to everyone in the society: it is always better to have more of x and y than less.

Now assume a dictator in this economy who does not care about consumer preferences and thus does not bother to establish the markets

FIGURE 1



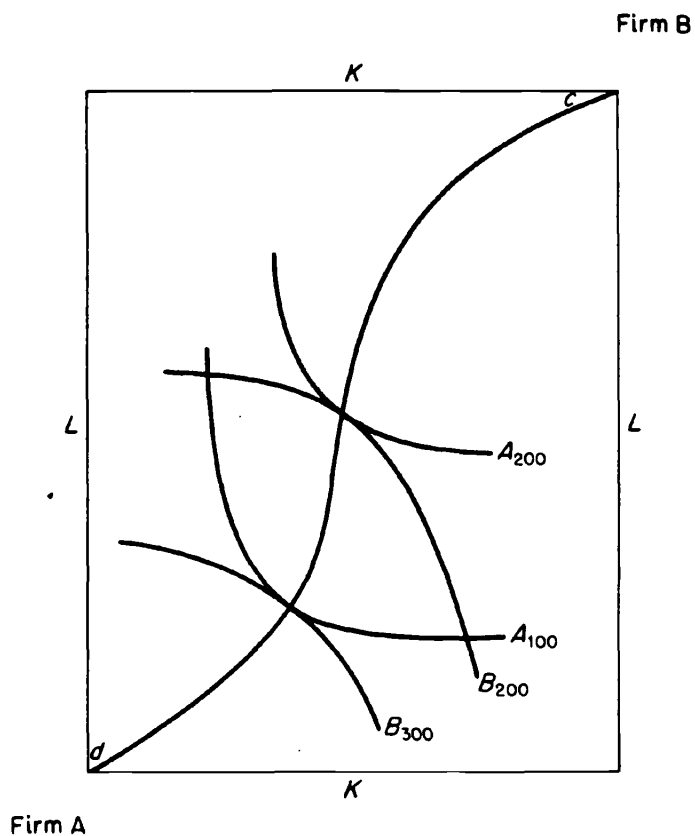
in consumer goods x and y that would reveal information about these preferences. He randomly chooses some mix of goods x and y which all consumers receive for free and are not allowed to sell or supplement. The dictator, however, shares the view that x and y are both goods and strives to organize the economy in such a way as to produce and distribute as large an output as possible. That is, he strives to attain the technical efficiency that my paper dealt with. If the capricious dictator chooses a mix of goods involving, say, two units of y for every unit of x , consumers would necessarily consume some bundle of goods along the line ab , and the dictator because of his belief that more of this mix of goods was better than less would strive to get the economy to produce at point b .

One thing that would have to be done to reach point b is that, in the absence of external effects, every firm or government enterprise that produced both x and y would have to have the same marginal private rate of transformation of x and y . The proof of this is simply that, if more x could be obtained by giving up a unit of y in one firm than in another, more output could be obtained by shifting the production of x toward the first firm and the production of y toward the second. This marginal private rate of transformation is given by the slope of the production possibility frontier at point b , and any firm which could transform x into y , but produces only x , must be able to produce more x per unit of y than this, whereas the converse must be true for enterprises that produce only y .

Another necessary condition for reaching point b , or technical efficiency, is given in Figure 2. This familiar figure assumes fixed endowments of factors of production K and L , given respectively by the lengths of the horizontal and vertical sides of the box in the figure, and the production functions for firms A and B are given by the isoquants arranged in the familiar Edgeworth box fashion. It is immediately obvious that production cannot be maximized except along the locus of points cd . This reminds us that, as even the more elementary textbooks explain, the marginal technical rate of substitution of any pair of factors must be the same for every producer who uses both factors if there is to be technical efficiency.

As long as the allocation of the bundles of x and y among the individuals in the society (i.e., the distribution of income) is not changed, and all of the x and y produced is distributed, we can be sure that welfare is higher at point b than at any point beneath it on line ab . This is true, notwithstanding the fact that consumer tastes are not revealed and did

FIGURE 2



not influence the bundle of consumer goods chosen, because the assumption that indifference curves portray rational consumers for whom x and y are goods, and the assumption that the distribution of income is not changed, insure that every consumer is better off at point b than at any lower point along line ab .

The gains in welfare that come from attaining technical efficiency in this situation could in principle be got by using prices to guide the allocation of production and inputs among enterprises. Indeed, if a bureaucratic mechanism could somehow attain technical efficiency, it could naturally do so only by duplicating what an ideal price system for producers and inputs for this situation would do. An xy price ratio for

producers given by the slope of the production possibility curve at point *b* could allocate production efficiently among maximizing enterprises and insure that they would in the aggregate produce the (arbitrarily chosen) output mix. The appropriate price for inputs could insure that they were allocated among enterprises producing *x* and *y* in such a way as to maximize output, because maximizing enterprises in the absence of external effects and monopoly will, as Schultze's comment points out, satisfy the condition that $MP_K/MP_L = W_K/W_L$. In short, the lack of revelation of preferences assumed in the foregoing model does not make the use of prices inappropriate or impossible: ¹ prices of producer goods can fulfill a useful function in a society producing private goods, even when consumer preferences are not revealed.

The relevance of this to Schultze's comment can be demonstrated by considering Schultze's passages about the lack of prices for the complex of characteristics embodied in a privately trained military pilot or a tank, and then introducing such intermediate goods explicitly into a situation of the sort depicted in the model developed above. Schultze was mistaken in supposing that the difficulty of putting an appropriate price on different pilots or tanks produced by contractors for the Department of Defense flowed entirely or even mainly from the concealment of voters' preferences about defense. As the foregoing model makes clear, there is nothing intrinsic in ignorance of consumer preferences which makes it inappropriate to use prices to organize production. Schultze is of course correct in saying that some goods have a variety of characteristics that make it impossible to describe them adequately in terms of one physical characteristic. But this does *not* mean that the concealment of consumer preferences is necessarily the problem; inputs or producer goods, however complex they might be, can often usefully be priced in markets, even if consumer tastes for the final goods they are used to produce are not known. If there are two alternative intermediate goods with many different features, they can be traded in free markets for such producer goods, where they will command prices that depend particularly upon their respective marginal productivities and marginal costs (and also to some degree on the mix of final goods that we assume was arbitrarily chosen in the absence of knowledge of con-

¹ Critics may reasonably object that my oversimplified model leaves out the heroic administrative task of keeping consumers from buying and selling *x* and *y*, and reflects the uncertainty about the level of supply of inputs when the relation of consumer income to input supply is not specified, and abstracts from many other problems. But these difficulties have nothing to do with the point at issue with Schultze.

sumer tastes). There is nothing inherent in an ignorance of consumer preferences that makes it impossible to attain *technical* efficiency with markets in which complex intermediate goods fetch different prices. There is nothing, for example, in the Soviet neglect of the preferences of final consumers that requires that the nail factories Schultze talked about must not produce that producer good for markets in which construction enterprises would be the buyers. Such markets could readily deal with the problem of the factory that makes nails of a type that no builder can use.

We can now turn to the second difficulty with Schultze's lack of a distinction between the feasibility of prices and the concealment of preferences. This arises from the theoretical possibility that preferences could be honestly revealed, even when outputs would not be known and market prices would not be useful for the allocation of resources. This second difficulty can most simply be seen by imagining the development of a truth serum that could be used to insure that every citizen would provide honest and complete information about his marginal rate of substitution between "national security"—say, to simplify, the probability of being victim of an enemy attack—and other goods. Even with this utopian information about citizen preferences, it would not be possible to determine an optimal defense policy or defense budget, because of the argument set out in my paper for the conference. Defense is a public good and we don't know what quantity of defense results from a given defense budget; we don't know whether a larger defense budget will deter an attack or provoke an arms race, whether antiballistic missiles will be effective or not, whether it is better to use tanks or Maginot lines, to carry a big stick or to sign a nonaggression pact. We do not, in short, know the production function for the collective good of national security, so that beliefs about that production function rest as much upon ideology and fashion as upon empirical estimation. Because the production function for the public good of national security is not known, it will not, as my paper pointed out, even be possible to rely entirely on markets and contracts. When the quantity of output is not known, there is no way to determine when a contractor has appropriately fulfilled his contract, or to measure the marginal productivity of a particular input or intermediate good.

IV

The value of any good to the consumer obviously depends on two things: how much or how many units of the good he consumes; and the value or

price he puts on each unit. It is not possible to say that the total value of the good to the consumer equals price times quantity, since he may value intramarginal units more than the marginal unit, but the total value of a good is plainly a function of the worth he puts on each of the units of the good he consumes and on the number of units.

Paul Samuelson argued in his famous articles on public goods that it followed from the very nature of these goods that those who received or consumed them could have an incentive to conceal the true value or price they would attach to marginal units of the good. Since the government would then be ignorant of even the relevant portions of the consumers' or citizens' utility functions, and thus of the value or price it should attribute to additional units of the collective good, it could not know how much of the collective good it was optimal to provide.

My paper for this conference accepted Samuelson's point that the government often cannot determine what *price* consumers would be able to pay for additional units of a collective good, and went on to say that it also followed from a defining characteristic of public goods that the government also could not straightforwardly determine what *quantity* or volume of output of the collective good it was supplying. The indivisibility of most collective goods—the fact that they are consumed by groups—entails that there cannot be so many experiments to test the consequences for consumers (the final output) with these goods as with private goods. This suggests that the government is not only, as Samuelson pointed out, going to know less about the utility function of consumers of public goods, but it is also going to know less about the production function for these goods than it would if it were producing private goods. This means it is more difficult to attain technical efficiency in the production of public goods.

Schultze argues in his comment that I was wrong in thinking there is any systematic tendency of the sort I put forth that makes it more difficult to measure the quantity of output of public than of private goods, and he holds that the concealment of preferences for public goods is sufficient to explain the lack of information about the output of public goods that I was writing about.

One way to test whether Samuelson's concealment of preferences is sufficient to explain the lack of information about the output of public goods is to consider real types of cases of public good production in which preferences are not concealed. If the ignorance about the level of output that I allege occurs is not then evident, Schultze is proved right and I am proved wrong, but if it is, Schultze is wrong and I am right.

Happily, the case of public intermediate goods offers a test. These are goods that facilitate production or lower costs for business firms, but which at the same time have the indivisibility or nonpurchasers-can't-be-excluded defining characteristic that I have emphasized. As Richard Musgrave and others have pointed out, the government can find out what the preferences for a public intermediate good are, because its worth to the recipients is obviously given by its effect on profits. Though there is no operational way to ascertain the consumer's utility function if he has no incentive to reveal his preferences, the estimation of the production function poses no comparable epistemological problem. The value of a given reduction in larceny to a warehouse company, for example, can readily be estimated from the dollar value of the given reduction in loss from the thefts.

Is there any special problem in determining the output of public intermediate goods that have the indivisibility property? Indeed there is. Take, for example, the value of police services to business firms. Whereas what a given reduction in the incidence of specified crimes would be worth to a business can be estimated relatively accurately, the effect of a given change in the policies or expenditures of a police department on the level of crime cannot be determined straightforwardly, if it can be determined at all. The crime rate depends not only on police policy and expenditure, but also on housing patterns, street lighting, family stability, income distribution, unemployment rates, educational policies, racial conflict, correctional policies, community attitudes, geographical migration, age distribution, and even the weather, not to mention other variables we cannot now specify with any confidence. I know of no single study that has accurately estimated the marginal productivity of police expenditures in the production function for crime control.² Many other examples of situations where the concealment of preferences problem does not exist, but where there are horrendous problems of output measurement, can readily be found.

The reason there is a special problem in measuring the output of public goods, even when preferences are not concealed, can best be seen by contrasting the information a consumer can get about public goods with the information he can get when deciding whether to buy a private good. The rational consumer or citizen who must decide whether to support or oppose candidates advocating a larger or smaller expenditure on a given public good must make some guess about the production

² See Belton M. Fleisher, *The Economics of Delinquency*, Chicago, Quadrangle, 1966, to get an impression of the difficulty of the econometric problem at issue.

function by which that public good is produced; he must estimate the impact more police or defense expenditures, for example, will have on the likelihood he will be a victim of a crime or that the country will be conquered (and then go on to decide what price he attaches to that). Though I have emphasized that very little is known about the production function for most collective goods, the citizen at issue (or those experts or officials on whose judgment he may choose to rely) has that slight degree of information that comes from comparing the history of expenditures on a given collective good with the apparent state of affairs—he can compare the crime rate or national security of high spending periods with that of low spending periods. He can also make comparisons with foreign governments or jurisdictions with different allocations, always somehow trying to allow for the many other variables that also determine the level of national security or crime rates or whatever. From this sort of judgment something may be learned (as, for example, that *extremely* militaristic nations usually perish by the sword). A government can also increase or decrease its level of spending on a collective good in the hope of getting more information in this way (e.g., in defense try either pacifism or a militarist solution), though experiments of this kind can of course be profoundly costly.

The consumer of a private service can also get more information by taking more or less of the private service over time and also by drawing on the experience of others. He can try to hire a housekeeper for a fewer or a greater number of hours to see what difference it makes and he can also get references from friends; he can try the restaurant or read what the food critic or other “consumer report” has to say. The situation at first glance seems much the same as for the public good.

But there is one profound difference affecting the information about the production function. Since a collective good by definition goes to a whole nation or group, all of whose members must take more or less of the good together, there can be only one experiment for each nation or community or catchment area for a public good. If, for example, an inoculation against a contagious disease is at issue, the question may arise whether 25, 50, or 90 per cent of the community has to be inoculated to eliminate the disease. If there is no international travel and migration, each comparable country may have a different policy, and some experimental evidence can conceivably be obtained in this way, but not much. The number of possible experiments is fewer, and the cost of each of them is greater, because the good is a public good.

But each individual can by definition take more or less of a private

Evaluating Performance in the Public Sector 409

good or service to see what output he is in fact getting. For a nationwide service such as is offered by Avis, Hilton, or MacDonal'd's, there are literally millions of experiments to determine what these service firms really offer. It may even pay a "consumer reports" firm to sell the results of its experiments to the consumers. Because collective goods go by definition to a group, and because private goods can by definition be taken in greater or lesser amounts by individuals, societies systematically have more information about the quantity or volume of output of the latter than the former. This is what Charles Schultze has denied; so his comment (though unusually suggestive and helpful) is in this most fundamental respect altogether mistaken.

