A THEORY OF NONMARKET FAILURE: FRAMEWORK FOR IMPLEMENTATION ANALYSIS*

CHARLES WOLF, JR.
The Rand Corporation

I. THE INADEQUACIES OF MARKETS

The principal rationale for public policy intervention lies in the inadequacies of market outcomes. Yet this rationale is really only a necessary, not a sufficient, condition for policy formulation.\(^1\) Policy formulation properly requires that the realized inadequacies of market outcomes be compared with the potential inadequacies of nonmarket efforts to ameliorate them. The "anatomy" of market failure provides only limited help in prescribing therapies for government success.\(^2\)

That markets may fail to produce either economically optimal or socially desirable outcomes has been elaborated in a well-known and voluminous literature.\(^3\) Although the last word has not been written, the essential points

---

* Many of the central ideas of this paper have developed from discussions I have had with Graham Allison over the past six years. Indeed, the question of whether the paper we agreed should be written would be written jointly or separately, and if the latter then by whom, was almost as frequent a topic of these discussions as the content of the paper. Fortunately, a six-month stay at Oxford in 1976, for whose support I am indebted to The Rand Corporation and the Ford Foundation, provided the answer to this question. I am also indebted to Pat Crecine, George Eads, Gene Fisher, John Flemming, Robert Klitgaard, Nathan Leites, John Martin, Joseph Newhouse, Robert Roll, and Harry Rowen for comments on an earlier draft.

\(^1\) The point is the same as Sidgwick's familiar comment: "It does not follow that whenever laissez faire falls short government interference is expedient; since the inevitable drawbacks of the latter may, in any particular case, be worse than the shortcomings of private enterprise." Henry Sidgwick, Principles of Political Economy 414 (1887). See also Alexander Cairncross, The Market and the State, in Essays in Honour of Adam Smith (Thomas Wilson & Andrew S. Skinner eds. 1976).

\(^2\) For some cogent observations closely similar to this line of argument, see Roland N. McKean, Divergence between Individual and Total Costs within Government, 54 Am. Econ. Rev. 243 (Papers & Proceedings, May 1964).

in the accepted theory are worth summarizing as background for the subsequent discussion of nonmarket failures.4

There are four sources or types of market inadequacies:5

A. Externalities and Public Goods

Where economic activities create "spillovers," whether benefits or costs, that are not, respectively, appropriable by or collectible from the producer, then market outcomes will not be (Pareto) efficient. Since these external benefits or costs do not enter the calculations upon which production decisions are based, too little output will tend to be produced where the externalities are (net) benefits, and too much where they are (net) costs, compared with socially efficient output levels. Education is an example of putatively positive externalities (benefits), which provide a rationale for government intervention—through subsidy or direct public sector production—to compensate for the market's tendency toward insufficient output. Chemical and noise emissions from aircraft or other industrial activities are examples of negative externalities (costs), which provide a rationale for government intervention—through taxing or direct regulation—to compensate for the market's tendency toward excessive output.

A distinction can be made between private goods with externalities and public goods: the former applies where most of the benefits or costs associated with output are, respectively, collected or paid by the producer, although some are not; and the latter applies where most of an activity's consequences comprise nonappropriable benefits (for example, national security) or noncollectible costs (for example, crime, an archetypical public "bad").6 Externalities and public goods are thus one condition—though neither necessary nor sufficient—for government intervention.

---

4 As Arrow observes: "The clarification of these concepts [relating to market failure] is a long historical process, not yet concluded." Kenneth J. Arrow, supra note 3, at 13.

5 I use the term "market inadequacies" and "market failures" interchangeably, although strictly speaking the former is more inclusive. Most economists would confine "market failure" to departures from Pareto-efficient outcomes, thereby excluding distributional issues except to the extent that distribution affects efficiency (see pp. 110–11 infra). By way of contrast, some noneconomists argue that distribution has, or should have, priority over efficiency (for example, Rawls's second principle of a just society), and they fault the market precisely because of its failure to accord this priority. See John Rawls, A Theory of Justice (1971). As will be clear in the text, I am including distributional considerations within market "inadequacies."

6 Externalities are thus a more general concept than public goods. Stated another way, a public good is the limiting case of a "private" good with externalities: "private" benefits approach zero, and the external benefits remain. More precisely, if \( v_{ij} \) is the valuation placed, or price paid, by the \( i \)th person for the \( j \)th unit of a good \( s \), and \( mc_j \) is its marginal cost of production, then the condition for an optimum (efficient) level of output for a private good with externalities is:

\[
mc_j = v_{ij} + \sum_{m=i+1}^{k} v_{mj},
\]
B. Increasing Returns

Where economic activities are subject to increasing returns and declining marginal costs, the market mechanism will also fail to generate an efficient outcome. Under conditions of decreasing costs, the lowest cost mode of production is by a single producer. In a free market, the result will therefore be monopoly, and, assuming single-part pricing, the outcome will be inefficient in both static and dynamic terms: statically, because output will be less than is efficient; dynamically, although more arguably (vide Schumpeter), because incentives for innovation will be weaker than would likely prevail under a more competitive regime.

Where increasing returns exist, various types of government intervention may be justified to correct the market outcome: (1) by directly regulating a "natural" monopoly (for example, public utilities) or by setting prices or allowable rates of return on capital; (2) by legal protection to prevent a single-firm takeover and to encourage competition (for example, through antitrust legislation). The various types of intervention admittedly depart from a theoretically efficient outcome, although they seek to approach it.\(^7\)

where \(v_{ij}\) is the price paid by \(i\), and the \(\sum\omega_{mj}\) are externalities (experienced by all other \(k\) individuals as a result of \(i\)'s consumption of the \(j\)th unit of \(s\), positive if the externalities are benefits and negative if costs.

For a pure public good, \(v_{ij} = 0\). Consumption is collective and no single unit is purchased by anybody. The optimum condition then is

\[
m\omega_{j} = \sum_{m=1}^{k} v_{mj}.
\]

Compare E. J. Mishan, supra note 3.

Total demand for public goods is determined by vertical summation of individual demand curves, rather than horizontal summation as in the case of private goods. (The point is sometimes misstated as equivalent to a zero marginal cost of production. For example, the marginal cost of national defense in, say, the United States or NATO is not zero, although non-taxpayers, as well as citizens of other countries, receive the benefits of such defense.)

The generalized explanation for the existence of externalities and public goods is that markets do not exist for capturing some benefits or levying some costs. Nonexistence of markets in these cases is explained by (1) the high costs or inability of excluding beneficiaries (for example, from benefits of national defense or police expenditures), or of establishing property rights as a basis for claiming liability when they are infringed (for example, noise emissions in airport vicinities); and (2) the lack of information required for market transactions to be concluded (for example, ascertaining what the "true" \(v_{ij}\) are in the previous discussion), in part at least because of the free-rider problem associated with (1).

7 Some discussions of market failure include increasing returns (for example, Francis M. Bator, supra note 3), while others exclude it. Arrow, for example, contrasts increasing returns ("essentially a technological phenomenon") with market failure (which relates to "the mode of economic organization"). Kenneth J. Arrow, supra note 3. I think this causation does not always hold. Improvements in technology can eliminate or at least reduce externalities by resolving the exclusion problem; for example, electronic warning and protection devices may be an efficient means of lowering the risk of theft for households purchasing them. One can imagine acoustical and air-filtration devices that would reduce the injury inflicted by the emissions or identify their source as a basis for imposing and collecting costs. Conversely, the "technological" phenomenon
C. Market Imperfections

Where the price, information, and mobility characteristics of "perfect" markets depart significantly from the realities, market outcomes will not be efficient, again providing a rationale for government intervention. Where prices and interest rates, for one reason or another, do not indicate relative scarcities and opportunity costs, where consumers do not have equal access to information about products and markets, where information about market opportunities and production technology is not equally available to all producers, or where factors of production are restricted in their ability to move in response to such information, market forces will not allocate efficiently and the economy will produce below its capacity. In such cases, which apply to some extent in all markets and to a greater extent in some, the implication for public policy is to reduce, if not remove, these imperfections: to facilitate availability of information, to lower barriers to entry and mobility, and so on.

However, where many of the conditions required for efficient functioning of markets do not exist, improving some of these conditions will not necessarily improve the efficiency of the market as a whole. Consequently, the policy implications of market imperfections may be ambiguous. And in some cases, public intervention may be justified even where it seems to add to these imperfections. For example, patent regulations, which are intended to restrict access to technological information, reduce the efficiency of resource use in the short run in the interest of long-run efficiency.

D. Distributional Inequity

Most economists exclude distributional effects from market failure strictly defined. That the distributional results of well-functioning markets may not accord with society's preferences is acknowledged, as is the plausible trade-

---

of increasing returns can be reconciled with efficient pricing and output by suitable modes of economic organization, for example, through multipart pricing. For a discussion of various pricing and market devices to reconcile increasing returns with efficient operation, see Charles Wolf, Jr., William R. Harris, Robert E. Klitgaard, J. Richard Nelson, John P. Stein, & Mario Baeza, Pricing and Recoupment Policies for Commercially Useful Technology Resulting from NASA Programs (Rand Corp. No. R-1671-NASA, Jan. 1975). For increasing returns are a source of market inefficiency only as long as markets do not exist for separate units of the same good. Allowing for enough subscribing, in the Arrow-Debreu sense, and hence separability of commodities, increasing returns are theoretically compatible with competitive equilibrium as are externalities.

8 This is, of course, the essential message of second-best theory (Richard G. Lipsey & Kelvin Lancaster, supra note 3). For example, changing a tariff that has applied equally to imports from all countries, so that it applies instead only to a few countries, may reduce efficiency. Trade will be diverted as well as created, and the loss from the former may exceed the gains from the latter. See Jacob Viner, The Customs Union Issue (Carnegie Endowment for International Peace, 1950).
off between efficiency and equity. In welfare economics the trade-off is usually dealt with by considering the relative efficiencies of various redistributive measures (for example, income taxes, excises, subsidies, unemployment relief, and income transfers), for achieving a specified redistribution (that is, minimizing the allocative distortions resulting from the income and substitution effects of redistribution). The term “market failure” is usually confined to departures from competitive equilibrium and Pareto-efficient outcomes, and excludes departures from distributional equity.

Nevertheless, from one perspective, it is theoretically correct to consider distributional inequity as an example of market failure. From this perspective, income distribution is a particular type of public good. An “equitable” redistribution does not result from freely functioning markets because philanthropy and charity yield benefits that are not appropriable by donors. Left to its own devices, the market outcome will entail no redistribution or too little, because of the usual free-rider problem associated with public goods and incomplete markets.

There is also a different perspective for viewing distributional equity, quite unrelated to market failure in the strict sense. From this perspective, the equilibrium redistribution previously described may be inequitable in terms of one or another ethical norm. Even if the market could surmount the “failure” discussed above, its distributional outcome might still be socially and ethically inadequate.

On these grounds, many people criticize the distributional outcomes of even perfectly functioning markets. Furthermore, most public policy deci-


10 Compare Walter Nicholson, Microeconomic Theory (1972).

11 The point can be formulated more precisely. Individual demand functions for redistribution can be defined in the same notional sense they can be defined for defense or for law and order. For example, the demand for redistribution can be expressed as the desired change in current distribution (as measured, say, by the Gini coefficient) with demand declining as the required amount of voluntary individual philanthropy per dollar of earned income rises. Presumably, individual willingness to pay for redistribution declines as its price rises. A cost function for redistribution can also be defined in terms of the same two variables. In principle, individual demands would be summed (vertically), and the social equilibrium level of redistribution would be that for which the marginal optimization condition is satisfied (see note 6 supra). This equilibrium redistribution is not achieved because there is either no market or an incomplete market for philanthropy, just as there is an incomplete market for defense. In both cases, voluntary donations (if unmotivated by special tax incentives) would be lacking for the usual nonappropriability, nonexcludability reasons.

12 In this sense, distributional inequity is a market “inadequacy” rather than “failure.”

13 John Rawls, supra note 5, is probably the most cogent recent effort to distinguish equity from efficiency and, accordingly, to fault market outcomes and the Pareto-efficient criterion of competitive equilibrium. Sometimes Rawls's points about equity or fairness, in contrast to efficiency, seem to me to have peculiar implications. For example, one of his central ideas, the
sions are usually at least as concerned with distributional issues (namely, who gets the benefits and who pays the costs) as with efficiency issues (namely, how large are the benefits and costs). Since my principal aim is to compare market inadequacies with the inadequacies of nonmarket remedies, I include distributional inequity among the offenders.

II. A Theory of Nonmarket Failure

Demand and Supply Characteristics

Market failure provides the rationale for attempted nonmarket (that is, government) remedies. Yet the remedies may themselves fail, for reasons similar to those accounting for market failure. In both cases, incentives influencing individual organizations ("firms," in the one case, and entities acting for or constituting "government," in the other) may lead to outcomes that diverge substantially from what is socially preferable. The basis for the market/nonmarket distinction is that market organizations derive their revenues from prices charged for output sold in markets where buyers can choose what to buy as well as whether to buy, while nonmarket organizations receive their revenues from taxes, donations, or other nonprice sources. Just as the absence of particular markets accounts for market failure, so nonmarket failures are due to the absence of nonmarket mechanisms for reconciling calculations by decision makers of their private and organizational costs and benefits with total costs and benefits. Nor, for reasons we will suggest, are prospects for invention of suitably compensatory nonmarket mechanisms to avoid nonmarket failure notably brighter than for creating suitable markets where their absence leads to market failures.

"difference principle," is intended to provide a tightly constrained basis for permissible differences in income and status, where the constraints derive from initial premises relating to justice or equity in contrast to efficiency. In elaborating the principle, Rawls argues that extra benefits received by the advantaged are just (fair) if and only if directly linked with some benefits realized by the least advantaged. According to the difference principle, a distribution is "just" even if benefits are forgone that have this property, as long as the original distribution retains it.

Some curious consequences follow from this position. Suppose a particular program (say, subsidized loans to new entrepreneurs from disadvantaged minority groups) provides tangible benefits to the disadvantaged, and even more substantial gains as a by-product to already advantaged groups. As I understand him, Rawls would term the distribution resulting from this program "just" or fair ("but not the best just arrangement"), even though it forgoes an alternative program that might entail small extra benefits for the already advantaged, as well as huge extra benefits for disadvantaged groups. "Dog-in-the-manger" behavior and spite become "just" in this curious formulation.

14 Although government is the principal nonmarket organization, there are also others: foundations, state-supported universities, churches, PTAs, and the Boy Scouts. The theory and types of nonmarket failure to be developed here apply to the performance shortfalls of these other nonmarket organizations, as well as governments.

15 This is essentially the same criterion used by Robert Bacon & Walter Eltis, Britain's Economic Problem: Too Few Producers (1976).
Where the market’s “hidden hand” does not turn “private vices into public virtues,” it may be hard to construct visible hands that effectively turn nonmarket vices into public virtues.

Public policies intended to compensate for market inadequacies generally take the form of legislative or administrative assignment of certain functions to a government agency in order thereby to produce certain outputs, which are expected to redress the shortcomings of the market. These outputs are of four types: (1) regulatory services (environmental regulation, radio and television licensing, interstate commerce regulation, pure food and drug control); (2) “pure” public goods (national defense, space research and development; (3) quasi-public goods (education, postal services, health research); or (4) administering transfer payments (federal, state, and local welfare programs, social security, and so on). The “value” of these outputs is expressed in national accounts as exactly equal to the cost of inputs used in producing them. But this accounting convenience does not tell us anything about the efficiency or social value of the outputs. Nor does it tell us why producing these outputs is likely to result in specific types of nonmarket failure. To explain this prospect, we need to examine the distinctive supply and demand characteristics that differentiate nonmarket outputs from market outputs.

On the supply side, there are several such characteristics:

(a) Nonmarket outputs are usually hard to define in principle, ill-defined in practice, and extremely difficult to measure independently of the inputs which produce them. They are generally intermediate products which are, at best, only remote proxies for the “real” or final intended output: for example, environmental impact precautions enforced by the Environmental Protection Agency; licenses issued or rejected by the Federal Commerce Commission; forces and equipment developed and deployed by the military services; students taught at different levels by the public school system; research projects funded by the National Institutes of Health; cases processed and payments disbursed by social welfare agencies. Units for measuring final product are usually nonexistent, and it is often hard even to distinguish “more” from “less.” Consider, for example, the difficulty of measuring military “worth,” specifying “quantities” of national defense, or education, or even regulatory services, in terms that are separate from the inputs used in producing them. Measuring outputs by their inputs becomes accepted because measuring outputs directly is so difficult.

(b) Evidence of output quality is also elusive, in part because the information that would in the market be transmitted by consumer behavior is missing. Consider, for example, the difficulty of determining whether the “quality” of education or welfare programs or environmental regulation is “better” or “worse” now than two or three years ago. Moreover, such signaling as
may be provided concerning "consumer" (that is, citizen) reactions tends to be too little and possibly nonrepresentative (for example, letter writers may be cranks, but the nonwriters are not thereby implying approval), or too gross and too late (for example, through congressional hearings or the ballot box) to be an effective means of monitoring output quality. To monitor output quality requires precise, representative, and regularized feedback which is hard to realize for nonmarket output.16 Congressional committees, the Congressional Budget Office, ombudsmen, consumer groups, voter and consumer surveys, and other "watchdog" devices help, but their separate and collective effectiveness in monitoring output quality inspires only limited confidence.

(c) Nonmarket outputs are usually produced by a single agency whose exclusive cognizance in a particular field is legislatively mandated, administratively accepted, or both (for example, the regulatory agencies, the public school system, NASA's role in space). It is rare that this exclusivity is contested. Where it is (for example, between the Air Force and the Army in providing battlefield air support), resolution is frequently on grounds unrelated to output quality. In sum, the absence of sustained competition is another factor contributing to the difficulty of evaluating output quality.

(d) Finally, nonmarket output is generally not connected with any "bottom line," comparable to the profit-and-loss statement of market output, for evaluating performance. Nor is there a reliable mechanism for terminating nonmarket activities if they are unsuccessful. Perhaps the closest analogy to a market "test" in the case of nonmarket output is military performance in war. Because it faces competition in war, the military does have special incentives to produce quality output. Yet even in this case, the effectiveness of these incentives is diminished by a paradox. The more successful potential military performance is, the more likely is military conflict to be deterred; and the more effective deterrence is, the less seriously is the risk of war likely to be taken and, hence, the weaker it becomes as an incentive to motivate high performance.

There are also distinctive characteristics that apply to the demand for nonmarket activities and to the process by which these demands become effective.

(a) As a result of the activity, perhaps hyperactivity, of information media, environmental groups, and consumer organizations, there has in the past few decades been an enormous expansion in public awareness of the

16 Hirschman's notions of "exit" and "voice" satisfy some of these criteria, but their effectiveness as monitoring and signaling devices is limited because they apply only to "insiders," not to consumers. Albert O. Hirschman, Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States (1970).
shortcomings of market outcomes. Increased awareness of monopolies, oligopolies, imperfect markets, negative externalities (for example, pollution), and distributional inequities, has resulted in intensified and politically effective demands for remedial action by government.

(b) In the political process, which mediates these demands, rewards often accrue to legislators or executives from articulating and publicizing problems, and legislating proposed solutions rather than assuming responsibility for implementing them.17

(c) In part as a consequence of this reward structure, the rate of time discount of political actors tends to be higher than that of “society.” Furthermore, there is often an appreciable disjuncture between the time horizons of political actors and the time required to analyze, experiment, and understand a particular problem (namely, a market inadequacy) in order to see whether a practical remedy exists at all.

The result of these demand characteristics is often a premature, but politically effective, demand to establish public programs for producing some nonmarketed output, as a symbolic response to the originally perceived market inadequacy. The equal opportunity and model cities programs of the 1960s and the decision in the early 1970s to emphasize “targeted” cancer research are examples. In these cases, as in others, the political effectiveness of public demands can lead to nonmarket activities with infeasible objectives and redundant costs.18

The supply and demand characteristics of the nonmarket sector are fundamental to the theory of nonmarket failure. They provide a basis for formulating a typology of nonmarket failure analogous to that which already exists for market failure. In both cases, the “failures”—whether market or nonmarket—are evaluated against the same criteria of success: allocative Pareto efficiency19 and distributional equity judged according to some explicit social or ethical norm. Nonmarket remedies “fail” to the extent they, too, result in outcomes that depart from the efficiency or distributional goals by which market outcomes are judged to fail. Although the touchstones of success are similar, the ways in which nonmarket solutions fail differ from those in which market outcomes fail.

17 Anthony Downs, Inside Bureaucracy (1967).
18 See pp. 124–26 infra.
19 Hence, in both cases other efficiency criteria are neglected, namely, dynamic efficiency, x-efficiency, and technological efficiency. Except for the later treatment of one type of nonmarket failure (namely, redundant and rising costs, pp. 124–26 infra), these other sources of efficiency are omitted from the discussion. This omission does not gainsay the argument that the additional types of efficiencies may be larger in their collective impact on economic performance (productivity) than is allocative efficiency. See Charles Wolf, Jr., Efficient Performance with Inefficient Pricing: A Puzzle for Economists Who Believe in the Free Market (Rand Corp. No. P-5915-1, July 1978).
Types and Sources of Nonmarket Failure

There are four types of nonmarket failure resulting from one or more of the distinctive demand and supply characteristics of nonmarket output.

A. Internalities and Private Goals

All operating agencies require, to conduct their activities, certain explicit standards. The requirement does not principally arise from an agency's need to justify its activities externally but rather from the practical problems associated with internal, day-to-day management and operations: evaluating personnel; determining salaries, promotions, and perquisites; comparing subunits within the agency in order to help in allocating budgets, offices, parking space, and so on.\(^{20}\) Lacking the direct-performance indicators available to market organizations from consumer behavior and the profit-and-loss bottom line, public agencies must develop their own standards. These standards are what I will call "internalities": *The goals that apply within nonmarket organizations to guide, regulate, and evaluate agency performance and the performance of agency personnel.* I refer to these internalities synonymously as "private" goals because they, rather than, or at least in addition to, the "public" purposes stipulated in the agency's assigned responsibilities, provide the motivation behind individual and collective behavior within the agency. This structure of rewards and penalties constitutes what Arrow refers to as "an internal version of the price system."\(^{21}\)

It is, of course, true that market organizations also must develop their own internal standards in order to regulate the same quotidian functions required for the management of any organization. But there is an important difference. The internal standards of market organizations are generally related, even if indirectly, to meeting a market test, to responding to or anticipating consumer behavior, to contributing to the firm's bottom line. Sales, revenues, and costs materially affect the internal standards of market organizations. For market organizations, the internal version of the price system must be connected to the external price system. If the two are disconnected, the survival of a market organization will be jeopardized by the response of consumers, competitors, or stockholders, even in imperfect markets.


The situation of nonmarket organizations is different because the supply and demand characteristics associated with their output are different. Because measures of output are often so hard to define, because feedback and signaling from "consumers" are lacking or are unreliable, internal standards for nonmarket organizations cannot be derived from these sources. Furthermore, because there are usually no competing producers, the incentive to devise internal standards that will control costs is weakened. Under these circumstances, nonmarket agencies often develop internalities that do not bear a very clear or reliable connection with the ostensible public purpose which the agencies were intended to serve.

In formal terms, internalities or private goals become arguments in the utility functions that agency personnel seek to maximize. Hence, internalities affect the results of nonmarket activities, as predictably and appreciably as externalities affect the results of market activities, in both cases causing divergences between actual outcomes and socially preferable ones. The existence of externalities means that some social costs and benefits are not included in the calculus of private decision makers. The existence of internalities means that "private" or organizational costs and benefits are included in the calculus of social decision makers. Whereas externalities are central to the theory of market failure, what goes on within public bureaucracies—the "internalities" that motivate their action and performance—are central to the theory of nonmarket failure.

In the market context, externalities result in social demand curves higher or lower than market demand curves, depending on whether the externalities are, respectively, positive or negative. And the levels of market output that result will be, respectively, below or above the socially efficient ones; hence, there is market failure.\(^{22}\) In the nonmarket context, "internalities" boost agency supply curves above technically feasible ones, resulting in redundant total costs, higher unit costs, and lower levels of real nonmarket output than the socially efficient ones; hence there is nonmarket failure.\(^{23}\)

Whether the nonmarket failure associated with internalities is greater or less than the market failure associated with externalities is an analytically interesting, and operationally crucial, question. Unfortunately, the answer is, in general, indeterminate. The nonmarket sector in principle allows for

\(^{22}\) Recalling the optimum condition noted earlier (note 6 supra), if the \(\sum v_{m_i}\) are positive, the \(j\) units produced under market conditions will be less than is socially optimal; where the \(\sum v_{m_i}\) are negative, the \(j\) units produced will exceed the social optimum.

\(^{23}\) If the optimal condition were complied with, producing \(j\) units of output would be less than is socially optimal absent internalities, because \(mc_j\) is inflated by the internalities of the nonmarket producers. See note 41 infra.
externalities in determining social demand, and hence comes closer on this count to an efficient level of output. But it does so at a likely cost in terms of internalities arising on the supply side. These are reflected in inflated total costs, which push the nonmarket sector away from a socially efficient level, as well as mode, of output. Which failure is the greater, nonmarket or market, depends on whether the supply distortions created by internalities in nonmarket output are larger or smaller than the demand distortions created by externalities in market output.

What determines the specific internalities developed by particular nonmarket organizations? Three different hypotheses suggest possible answers.

One hypothesis is that internal standards are based on norms that, when an organization was started, appeared to be reasonable proxies for the elusive final output it was intended to produce. Thereafter, they may become formalized as organizational routines or standard operating procedures which are accepted as a principal measure of the organization's performance. While market organizations also establish standard operating procedures, these must generally meet a market test. If the costs of adhering to them exceed those connected with changing them, they will be altered. The standard operating procedures of nonmarket organizations must stand up to a different test. Generally, a congressional hearing or scandal of some sort is required for change; and these may or may not be related to agency performance.

A second hypothesis is that those internalities are selected which maximize the income (and nonincome perquisites) of agency members.

The third hypothesis is that specific internalities arise because they tend to

---

24 The \( \Sigma w_m \)'s are, in principle, included in determining output decisions.

25 For example, a budget-maximizing internality (see infra) may arise at the time nonmarket organizations are first established because new organizations have to hire staff and acquire facilities to handle their assigned responsibilities. Through a simple, inertial process, the proxy variable (increased staff and budget), which was essential for a particular nonmarket agency to get started, becomes accepted and retained as a convenient indicator of agency performance. McFadden's attempt to infer what an agency (namely, California State Highway Division) is trying to maximize by observing its prior behavior (for example, with respect to project and route selection compared to optimal choices) is in the spirit of this hypothesis. See Daniel McFadden, The Revealed Preferences of a Government Bureaucracy: Theory, 6 Bell J. Econ. 401 (1975).

26 For example, larger budgets generally mean larger numbers of supergrade jobs, and the anti-new-technology internality of primary schools (see p. 122 infra) protects skills, positions, and income of senior members. This hypothesis is close to the view taken in Roland N. McKean, supra note 2; William A. Niskanen Jr., Bureaucracy and Representative Government (1971); and Jack A. Stockfish, Analysis of Bureaucratic Behavior: The Ill-Defined Production Process (Rand Corp. No. P-5591, Jan. 1976). In some cases, the first and second hypotheses lead to similar predictions, for example, the budget maximand. In others, the predictions probably differ, for example, the information-acquisition maximand (see pp. 122-23 infra) is hard to reconcile with the first hypothesis.
increase the benefits received by a constituency group which has succeeded in co-opting a particular nonmarket organization. Often, the co-optation is by a constituency that the nonmarket agency has been set up to regulate.\textsuperscript{27}

What are some of the specific internalities that often accompany nonmarket activities, and lead to nonmarket failures?

1. \textit{Budget growth ("more is better")}. Lacking profit as a standard for motivating and evaluating performance, a nonmarket agency may adopt the agency's budget as its maximand, or at least as an important argument, in the agency's utility function. Performance of the agency's personnel and subunits is then evaluated in terms of their contribution to expanding its budget or protecting it from cuts. Incentives within the agency will develop to reward participants for "justifying costs rather than reducing them,"\textsuperscript{28} a characterization that has been applied to the Defense Department and the military services, but surely is not confined to them.

The following instruction from a former Chief of Naval Operations to subordinate commands shows how government budgetary procedures may be translated into internal agency pressures to spend rather than save resources:

Fiscal Year 1972 outlay targets promulgated . . . as part of the President's budget for FY 1973 are over \$400 million above targets in the earlier FY 1972 budget . . . Difficulty of achieving these targets during remaining months of 1972 fully appreciated, but importance of avoiding shortfall in meeting newly established FY72 targets to avoid resultant adverse effects on anticipated FY73 outlay ceiling \textit{dictates need for top management attention. Anticipate any shortfall in FY72 outlay target could be translated into program loss under FY 1973 outlay ceiling.}\textsuperscript{29}

Stripped of bureaucratic jargon, the commander is advising his subordinates to find ways to spend funds quickly and plainly implying his intention to evaluate their performance in terms of how well they succeed! As one observer, commenting on the motivations behind actions of the military services, notes: "The welfare of a service is measured by its budget."\textsuperscript{30} The result of a budget internality is likely to be a distortion in the level of agency activity; in other words, a nonmarket failure to produce a socially optimal outcome.\textsuperscript{31}

\textsuperscript{27} This hypothesis is favored in George J. Stigler, \textit{The Theory of Economic Regulation}, 2 Bell J. Econ. \& Management Sci. 3 (1971), and applied empirically to transportation and professional licensing. The list in note 51 \textit{infra} of regulatory agencies, and of the constituencies they affect most directly, suggests other examples of this hypothesis.


\textsuperscript{29} Quoted in Arthur E. Fitzgerald, \textit{The High Priests of Waste} 385-86 (1972).

\textsuperscript{30} Nancy Nimitz, \textit{infra} note 28, at 1.

\textsuperscript{31} Using plausible demand and cost functions, William A. Niskanen, Jr., \textit{infra} note 26, shows how the budget internality will lead to an output level above the socially efficient one.
Variants of the budget maximand can lead to similar nonmarket failures. For example, managers of the West German public television and telephone system reportedly have asserted that their primary objective is to raise rates and sales so as to maximize gross revenues. This, they explained, was necessary to “finance their further growth.”\(^{32}\) If revenue maximization is the internal performance standard, output will rise as long as marginal revenue is positive, again resulting in nonmarket failure to produce a socially efficient outcome.

When an American businessman was asked in 1972 to assume management responsibility for the postal service, he found that its serious financial predicament was due, in part, to its system of determining pay scales for postmasters: “Postmasters were actually paid [based] on how many employees they had, how many branch offices they had, or how many trucks . . . Can you imagine a greater disincentive?”\(^ {33}\)

A variant of the budget internality is the agency’s employment level. A public agency, eschewing or precluded from profit maximization as its objective, may attempt to maximize the size of its staff. For example, British Rail, a nationalized industry and one of the half-dozen largest employers in Britain, operates under acute pressure from trade unions and government to maintain high employment levels and avoid “redundancies.” Operating under such incentives, featherbedding by managers and foremen becomes a rewarded practice. High employment per unit of service, the reciprocal of high labor productivity, is aspired to, resulting once again in nonmarket failure.

2. Technological advance ("new and complex is better"). Often compatible with the budget internality is one relating to “advanced,” “modern,” “sophisticated,” or “high” technology.\(^{34}\) Nonmarket agencies, whose activities may be justified in the first instance by one or more of the acknowledged sources of market failure, may establish technical “quality” as a goal to be sought in agency operations. In medicine, a bias toward “Cadillac”-quality health care may result, and in the military a sometimes compulsive tendency toward development of the “next generation” of more sophisticated equipment. Explicit consideration of whether these advances are worth their extra costs is regarded as inappropriate because the operating agencies either are not intended to maximize net revenues (in the case of hospitals) or earn no

\(^{32}\) I am indebted to James Rosse for this example.
\(^{34}\) This is not the place to attempt to define precisely what is meant by “high technology,” a subject richly clothed in confusion in both popular and professional discussion. To consider whether the term does, or should, refer to products or processes, novelty or efficiency, costs and/or effectiveness would take us too far afield. For present purposes, I will conveniently assume that high technology, like a camel, is easy to recognize if difficult to describe.
revenue since they are producing a public good (in the case of military services).\textsuperscript{35}

An example is provided by the purchase of disposable syringes by the British national health service in the late 1960s when these gadgets were invented. Their novelty suggested merit. Only later was it demonstrated that repeated use of durable syringes had, in fact, been accompanied by equal or lower rates of attributable infection, and at lower cost.\textsuperscript{36}

Perhaps especially in the military services is the development of systems embodying the latest technology taken to be an organizational imperative. As one practitioner has observed: "In the Air Force, advancing technology has become a part of the professional ethic."\textsuperscript{37} The technological ethic is not confined to the Air Force. Organizational pressures toward sophistication, complexity, and technological novelty play a powerful role in the acquisition process of other services as well.\textsuperscript{38} Nuclear-powered supercarriers are no less an illustration than the FB-111 or the F-15 aircraft.

The American space program is pervaded by a similar, indeed legislatively encouraged, imperative. From NASA's legislative mandate for "the

\textsuperscript{35} Newhouse has shown formally how the addition of a "quality" argument in the maximands of nonprofit hospitals tends toward misallocation of resources in the health care industry. A nonmarket failure results because managers trade off quality against quantity, producing a different product from that which consumers would choose if they were spending the resources that nonprofit hospitals receive from public or philanthropic sources. Joseph P. Newhouse, Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital, 60 Am. Econ. Rev. 64 (1970). In the Newhouse model misallocation is reduced because a nonprofit hospital's choice of high quality is assumed to shift consumer demand upward, thereby adding to the market value of outputs. However, this may not occur. As long as the nonprofit hospital draws a subsidy (from government or philanthropy) based on the presumed market failure (for example, externality) which the subsidy is intended to correct, the hospital can price its output below cost, while indulging its practitioners' taste for quality. The original source of market failure is not thereby redressed.

\textsuperscript{36} Martin S. Feldstein, Economic Analysis for Health Services Efficiency: Econometric Studies of the British National Health Service (1968).

\textsuperscript{37} Richard G. Head & Ervin J. Rokke, American Defense Policy (3rd ed. 1973). The particular attraction in the U.S. Air Force of technological advance as an organizational internality is well known. The process of its adoption is probably an example of the hypothesis concerning initially valid proxies whose validity may have diminished after the proxy had already become accepted and engrained in agency operating routines. For example, when the Air Force was established as a separate service in 1947, two circumstances impelled it toward emphasizing technological advance as an organizational internality: (a) the two decades of struggle with the U.S. Army to win acceptance of the new aviation technology, independent of artillery and infantry; and (b) the major technological advances achieved during World War II (for example, in radar and nuclear weapons) and the resulting belief that the outcome of a future war "would be determined solely by the technological power of weapons that adversaries could bring to bear in its first moments." Harvey Sapolsky, The Polaris System Development: Bureaucratic and Programmatic Success in Government 77 (1972).

preservation of the role of the United States as a leader in aeronautical and space science and technology,"\textsuperscript{39} it has been a short step to formalize the development of novel and complex technology as an internal agency norm, whether or not it seems likely to be efficient.

The technological internality can have perverse consequences, not only in excessive zeal for what is complex and novel, but in mindless opposition to what is simple and familiar. In the Vietnam War, use of a modified propeller-driven cargo aircraft, with long loiter time and a slow stalling speed as a platform for delivering guided munitions as well as airborne artillery, was by far the most efficacious source of American firepower. Yet turning this "gunship" idea into an operating system was delayed five years, largely because of service opposition to what was viewed as a technologically retrograde step!

A bias against new technology can, of course, equally lead to nonmarket failure. Parts of the American educational system, for example, resist even the experimental use of such new technology as videotaping for presentations to large classes, computer-aided instruction, and performance contracting, all of which might reduce the demand for teachers. Indeed, the education industry's behavior often suggests the opposite of the maxim that "new and complex is better." While a maxim that "familiar and simple is better" may be generally preferable, rigid application of it can have equally perverse effects on performance. Resistance by the education sector to technological advance is similar in quality, although opposite in direction, to the military's frequently uncritical enthusiasm for technology. In both cases, a private organizational goal, an "internality," contributes to nonmarket failure.

3. Information acquisition and control ("knowing what others don't know is better"). Another element in the utility functions of some nonmarket organizations is information. Frequently in nonmarket, as well as in market, organizations, information is readily translated into influence and power.\textsuperscript{40} Consequently, information becomes valued in its own right—an internality for guiding and evaluating the performance of agency members.

Acquisition and control of information may be particularly important as a goal for agencies involved in foreign policy, because existing constraints already limit such other possible internalities as budget growth or technological advance. An example is Kissinger's use of the National Security Council framework and the Committee of 40 as means of acquiring exclusive information and, hence, of increasing influence for the National Security Council in the 1968-73 period. The careers of council staff members came to depend

\textsuperscript{39} National Aeronautics and Space Act of 1958, Public Law 85-568, Section 102(c)(5).

\textsuperscript{40} For a more general treatment of the importance of informational access and control in "post-industrial" society see Daniel Bell, The Coming of Post-Industrial Society (1973).
on their ability to understand and adjust to the incentives created by this particular internality. Staff members succeeded by demonstrating their ability to collect and protect new information, which Kissinger's organizational and procedural rearrangements made possible, for the "private" use of the council. Information available only to the National Security Council seemed to become an end in itself, an internal standard motivating staff behavior. The effect of this internality on the conduct of foreign policy, and more particularly on the effectiveness of the State Department, in contrast with the council, is surely debatable. That the informational internality will lead to nonmarket miscarriages is likely, since it connects in no obvious way with the final and elusive outcome sought, the successful conduct of foreign policy.

In associating these specific types of internalities with nonmarket activity, I do not imply their absence from market activity. For the usual reasons pertaining to more or less imperfectly competitive markets—which, of course, are the only markets that exist—these characteristics also apply, to some extent, to market activity. But this extent is likely to be more limited. Price competition among firms and products, as well as competition within firms among managers seeking promotions, generally limits the extent of cost-inflating internalities in market activities, as compared with nonmarket activities.

What can be said to summarize the difference between internalities associated with nonmarket output and externalities associated with market output? Whereas externalities in the market sector are costs and benefits realized by the public but not collectible from or by producers, the internalities associated with nonmarket output are usually benefits perceived as such by producers and paid for by the public as part of the costs of producing the nonmarket output. Consequently, internalities tend to raise costs and supply functions. These shifts, moreover, are likely to increase over time as nonmarket agencies succeed in building special constituencies within the Congress and the public that are more immediately concerned than is the broader taxpaying public over whom the costs are spread.

Internalities are thus elements of the private goals of producers: private in the sense that their role is primarily that of satisfying interests of nonmarket producers rather than contributing to the public sector's intended final output. Such internalities and private goals, often quite remote from an elusive final product, are as frequent and important in nonmarket activities as externalities are in market activities.  

---

41 The existence of internalities in organizations producing nonmarket outputs can be related to the condition for determining an optimal (efficient) level of output. Recalling the notation used earlier (note 6 supra) the condition is:
B. Redundant and Rising Costs

Whether policy takes the form of regulation, administering transfer payments, or direct production of public goods, there is a tendency for these nonmarket activities to exhibit redundant costs ("x-inefficiency")—that is, for production to take place within production possibility frontiers—and for cost functions to rise over time. If technological possibilities exist for lowering cost functions, raising productivity, or realizing economies of scale, these opportunities are likely to be ignored or exploited less fully by nonmarket than by market activities. Nonmarket failure, in the form of technically inefficient production and redundant costs, is the result. Moreover, these redundancies may well rise over time.42

The sources of these nonmarket failures lie in the demand and supply characteristics associated with nonmarket output. As public awareness of the inadequacies of market outcomes grows, demands for remedial action intensify. Dissatisfaction with existing circumstances may result in misperceiving the cause as a market failure, rather than something more intractable like genetics, physical laws, or resistant sociology. With rewards frequently accruing in the political arena to publicizing the problems and initiating action labeled as a remedy, nonmarket activities may be authorized which have quite infeasible objectives. Objectives may be internally inconsistent: for example, bringing all students' reading scores up to the mean; or minimizing the time individuals are unemployed while maximizing their earnings; or providing foreign aid to accord with "need" but also to encourage better development. Or objectives may be specified for which no known technology exists; for example, providing "dignified" work for people with low IQs, or training people with IQs of 70 to be draftsmen, or achieving a

\[ mc_j^f + \sum_{p=1}^{n} mc_{p_j^f} = \sum_{m=1}^{k} v_{m_j^f}, \]

where \( mc_{p_j^f} \) is the marginal cost of the \( p^{th} \) internality associated with production of the \( j^{th} \) unit of the \( s^{th} \) public good.

This specification is closely related to Stigler's "positive" theory of regulation: a benefit of some outside constituency becomes an agency goal and an argument in the agency maximand. I think Stigler errs, however, in denying what seems to me a generally valid proposition about public policy intervention: even though co-optation of a regulatory agency frequently occurs after it gets underway, nonmarket activity is rarely undertaken without a case being first made on normative grounds, based on market failure or distributional equity. George J. Stigler, supra note 27; see also Richard A. Posner, Theories of Economic Regulation, 5 Bell J. Econ. & Management Sci. 335 (1974).

42 The term "redundancies" has a different meaning here from that referred to at pp. 119–20 supra. Clearly, maintaining low productivity to avoid employment redundancies, as in the case of British Rail cited earlier, is one source of cost redundancies.
cure for cancer by 1980. Redundant costs may result at any positive level of nonmarket output. Redundant costs may also result from the difficulty of measuring output, and the resulting need, as well as latitude, to establish agency goals—the internalities that become accepted as proxies for nonmarket output. The cost-inflating effect of internalities may endure because nonmarket activity is conducted without competition. Or redundant costs may rise over time because of the absence of a reliable termination mechanism for nonmarket output, thereby allowing agency managers to move toward higher levels of internal goals.

Those responsible for market activities usually have an incentive to expand production and to lower costs over time, because of actual or potential competition or because of opportunities for additional profits. By contrast, those responsible for nonmarket production may be spurred to increase costs (for example, staff), or to increase output even if its incremental value is less than incremental costs (for example, the German TV case cited earlier), resulting in redundant costs that rise over time. These tendencies toward supportable if those receiving the benefits, \( \int_0^{q^*} D(q)\,dq \), are politically more effective (even though they pay nothing or at least pay less than the benefits) than those who pay the full costs, \( \int_0^{q^*} S(Q)\,dq \), or at least pay the difference between the full costs and the amount paid by the first group. To avoid tautology requires that the ingredients of “political effectiveness” (for example, organizational skill, media pressure) for the gainers and the losers can be evaluated independently of the nonmarket undertaking.

Hence, cost functions for nonmarket activity are likely to drift upward because of private goals (internalities). This upward drift is what I mean by “rising costs.” By “redundant costs,” I

---

43 In the words of one observer, whose comment is all the more insightful because it preceded his own not inconsiderable role in providing evidence in its support: “[N]ew agencies, from which better things might be hoped, are put under unremitting pressure to produce glamorous new programs—before the necessary analysis has been performed.” James R. Schlesinger, Systems Analysis and the Political Process, 11 J. Law & Econ. 281, 293 (1968).

44 In effect, demand and supply functions may not intersect, yet the demand for nonmarket activity may still be politically “effective.” Nonmarket output of at least \( q^* \) will be politically
redundant and rising costs were described by a departing director of the United Nation's Food and Agriculture Organization with reference to his own organization:

Eighty percent of its budget is destined to pay for a gigantic centralized bureaucracy in Rome, 11 percent to put out publications that no one reads, and the remaining 9 percent to holding meetings and for travel expenses that are largely unnecessary. 46

The details of this example may be extreme, but the general picture probably has wide applicability to nonmarket agencies and activities.

C. Derived Externalities

Government intervention to correct market failure may generate unanticipated side effects, often in areas remote from that in which the public policy was intended to operate. Indeed, there is a high likelihood of such derived externalities, because government tends to operate through large organizations using blunt instruments whose consequences are both far-reaching and difficult to forecast. In the Russian proverb, "When elephants run, other animals tremble."

The likelihood of externalities is further enhanced by both demand and supply characteristics associated with nonmarket output. Strong political pressure for nonmarket intervention may create an effective demand for action before there is adequate knowledge or time to consider potential side effects. Furthermore, derived externalities are generally more likely to occur later than sooner. Hence, the short time horizon and high time discounts of political actors predispose them to overlook potential externalities. And, finally, the frequently ill-defined nature of both quantity and quality of nonmarket outputs limits the motivation, as well as the means, for thinking seriously about their potential unintended side effects.

mean the tendency of nonmarket activities to be carried on inside, rather than on, the production possibility frontier at any given time. The two tendencies thus relate to dynamic efficiency and x-efficiency, respectively.

It should be possible to test the hypothesis advanced here that (a) rising costs and (b) redundant costs tend to be associated with nonmarket activities compared with market activities. One might use for this purpose cost data in sectors where production has been carried on in both a market and a nonmarket mode (for example, education, fire protection, housing) within a given country or in comparisons between market and nonmarket modes in different countries (for example, health care in the United Kingdom and in the United States). The few empirical studies already done of production by market and nonmarket organizations (for example, private versus governmental production in fire protection and in refuse collection) suggest that the former tends to be more efficient and that redundant costs tend to be associated with nonmarket organizations. See Roger S. Ahlbrandt, Municipal Fire Protection Services: Comparison of Alternative Organizational Forms (1973); and Robert M. Spann, Public versus Private Provision of Government Services in Budgets and Bureaucrats: The Sources of Governmental Growth 71 (Thomas E. Borcherding ed. 1977).

46 International Herald Tribune, April 26, 1976.
Of course, cost-benefit analysis tries to internalize such externalities, for example, by calculating the benefits of hydroelectric projects to include flood control, irrigation, and "feeder industries," as well as electric power. But the limitations of such analyses are numerous and well known, resulting in part from the unanticipated nature of some of the side effects.\(^{47}\)

Derived externalities are hard to anticipate because the consequences of public policies may be far removed from the target. For example, when standards for noise emissions were established by the Environmental Protection Agency to compensate for the market's failure to allow for these externalities, it was unanticipated that one result would be strains (that is, costs) in American foreign policy relations with the French and British over the Concorde. That an embargo in soybean exports to Japan in 1973 would affect U.S. military-base negotiations in that country was also not anticipated (although perhaps it should have been). That long-standing "Buy America" and other trade restrictions—once again, presumably based on a need for public-policy intervention to compensate for market inadequacies—would make more difficult a move toward standardization and rationalization of weapons systems and forces in NATO was also difficult to forecast.

Another instance of derived externalities is provided by public regulation of utilities. Permissible profits are typically calculated on the basis of return on capital, with the intention of holding prices closer to marginal cost, thereby overcoming one source of market failure. But a derived externality often results as an unintended consequence. The regulated utilities may respond by inefficient substitution of capital for labor to raise the allowable profit base.\(^{48}\) The resulting nonmarket failure may equal or exceed the market failure that regulation was intended to remedy.

Of course, derived externalities may be positive rather than negative. Construction of a North Sea barrier in the Veere inlet, for the safety of the Zeeland population in the Netherlands, meant the loss of mussel and oyster beds but also the start of trout raising, the end of ocean-going boating but the beginning of a recreational industry based on smaller vessels in the new Veere Lake, none of which was anticipated when the Veere barrier was originally decided upon.

\(^{47}\) A detailed attempt to internalize such externalities, as well as a candid acknowledgment of the limitations of cost-benefit analysis which tries to do so, is contained in Jack Hirshleifer, James C. DeHaven, & Jerome W. Milliman, Water Supply: Economics, Technology, and Policy (1960). Hirschman, in his notion of the "hiding hand," emphasizes the benefits, rather than the costs, of unanticipated consequences from selected development projects undertaken by governments. Of course, whether the hand principally hides benefits or costs depends on which development projects are selected for retrospective examination. Albert O. Hirschman, Development Projects Observed (1967).

\(^{48}\) Harvey Averch & Leland L. Johnson, Behavior of the Firm under Regulatory Constraint, 52 Am. Econ. Rev. 1052 (1962).
All of these examples represent a type of nonmarket failure: externalities, whether negative or positive, deriving from a public policy intended to compensate for an existing market failure. They have in common, also, the characteristic of not having been foreseen at the time the policy was initiated. Clearly, policy choice would be improved if such derived externalities could be taken into account when policy analysis and choice are under way.49

D. Distributional Inequity

Nonmarket activities, whether intended to overcome the distributional inequities of market outcomes or to remedy other inadequacies in the market’s performance, may themselves generate distributional inequities. The resulting inequities are often indexed on power rather than income or wealth.

Public policy measures—whether intended to correct distributional inequities, or to regulate industry (because of externalities or increasing returns), or to produce public goods, or to redress market imperfections—place authority in the hands of some to be exercised over others. Whether the authority is exercised by the social worker, the welfare-case administrator, the tariff commissioner, the utilities regulator, the securities examiner, or the bank investigator, power is intentionally and inescapably lodged with some and denied to others. The power may be exercised with scruple, compassion, and competence. It may be subject to checks and balances, depending on the law, on administrative procedures, on the information media, and on other political and social institutions. Nevertheless, such redistribution of power provides opportunities for inequity and abuse. Corrupt practices are one type of abuse; for example, government contracts obtained through bribery, perhaps illustrated by the case of Lockheed’s F-104 sales abroad; import licenses or preferential exchange rates conferred on the relatives, friends, or associates of officials and politicians who exercise discretionary authority. Less conspicuous inequities can result from the decisions of welfare authorities in classifying cases and conferring or withholding aid to fatherless families with dependent children, or to potential recipients of aid for the aged. Anecdotes reflecting the vagaries, perversities, and inequities associated with welfare programs are too numerous to recount, as well as too inexact to yield precise conclusions.

49 To the extent that better analysis can anticipate and calibrate the derived externalities associated with nonmarket output, they become analytically identical to the externalities associated with market output. Hence, the optimum condition for nonmarket output with derived externalities is the same as that specified above for market output with externalities. See note 6 supra. However, determining the $\Sigma v_m$/ex ante may be even harder for nonmarket “derived” externalities because of the bluntness of nonmarket instruments and the frequent remoteness of their effects both in time and place.
In the specific case of public policies intended to redistribute income, a frictionless, impersonal, and automated redistributive mechanism might avoid the inequitable distribution of power that can result from discretionary authority. But even a sharply progressive tax system—which is intended to serve this purpose—reserves considerable room for auditors to exercise judgment and hence power. The same applies to the redistributive expenditure programs mentioned above. One need not ascribe to those who administer public programs less humane motives than the average to contend that some distributional inequities may result from efforts to rectify other inequities, as well as from efforts intended to remedy still other market inadequacies. And, of course, there is still a presumption that the distributional inequities created by progressive taxes or by redistributive expenditure programs are smaller than the original inequities which such measures relieve.

Nonmarket activities may also result in distributional inequities indexed on income rather than power. It is truistic that any public policy will benefit some and take from others. Indeed, this will ensue whether or not the particular market inadequacy, which gave rise to a nonmarket intervention in the first place, was explicitly distributional in character. Public policy measures will increase the demand for some factors, skills, services, and products, and levy costs on others. Those who are specialized in the former will benefit at the expense of those in the latter, by comparison with the previously prevailing situation. If public expenditures are increased for defense or education, because these are instances of public goods in the one case or private goods with large externalities in the other, organizations and individuals specialized in producing one or the other output will realize increases in their real income.50

Consequently, groups that are potentially benefited by a public policy measure intended to compensate for market failure can be expected to urge, and very likely believe, that more compensation is needed to bring about a socially optimal outcome than would otherwise be estimated. Educators, accepting the argument that some government subsidy is necessary to take account of positive externalities ignored by the market, are likely to argue that these externalities are greater than was originally allowed for and hence warrant a larger subsidy. A similar point applies to the professional and business community concerned with aerospace technology and research and development.51 The result is likely to be nonmarket failure in the form of a

---

50 Imposition of nondistorting lump-sum taxation to capture these economic rents is arguable in theory and unrealistic in practice.

51 A recent paper by an executive of the General Electric Company displays the following suggestive matching between certain government organizations and policy areas, on the one hand, and their business and professional "constituencies," on the other.
larger public subsidy or a more protective regulatory policy for the benefit of "constituencies" that are well organized. Hence, a distributional inequity from the standpoint of nonbenefiting groups occurs, even though they may have acknowledged the existence of a market failure and the legitimacy of nonmarket intervention in the first place.\(^5^2\)

The role of nonmarket activities in producing distributional inequities, whether these are reflected in maldistribution of power or of income, derives from specific demand and supply characteristics associated with nonmarket output.

On the demand side, the principal causal characteristic is heightened public awareness of the inequities generated by the market and the resulting clamor for redistributive programs, often without prior consideration of the inequities that may be generated by these programs themselves.

On the supply side, distributional inequities result from the typical monopoly of nonmarket output in a particular field and the related absence of a reliable feedback process to monitor agency performance. In the absence of competing producers, those who feel adversely affected, whether as victims of arbitrary administrative authority or as general taxpayers, have notably less direct and less effective means of expressing their dissatisfaction than is available to consumers of marketed output who can withhold purchases or shift them to other producers. By contrast, those who realize special distributive benefits from particular nonmarket activities are likely to have, or to create more direct and more effective means for expressing their support, through organized lobbying and advocacy, than is available to consumers in the marketplace.

This does not imply that the inequities of the market are less than those of

<table>
<thead>
<tr>
<th>Government Organizations</th>
<th>Related Business Organizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Department of Defense, NASA</td>
<td>Defense-space contractors</td>
</tr>
<tr>
<td>Department of Agriculture</td>
<td>Farmers; dairy, meat processors</td>
</tr>
<tr>
<td>Environmental Protection Agency</td>
<td>Auto manufacturers; electric utilities</td>
</tr>
<tr>
<td>Securities and Exchange Commission</td>
<td>Brokers; underwriters; issuers</td>
</tr>
<tr>
<td>Interstate Commerce Commission</td>
<td>Railroads; truckers</td>
</tr>
<tr>
<td>Federal Communications Commission</td>
<td>Radio &amp; TV stations &amp; networks; cable and pay TV</td>
</tr>
<tr>
<td>Tariff Commission</td>
<td>Trade unions; business subject to import competition</td>
</tr>
<tr>
<td>Food and Drug Administration</td>
<td>Drug industry; food &amp; beverage industry</td>
</tr>
<tr>
<td>Federal Power Commission</td>
<td>Electric utilities; natural gas producers</td>
</tr>
<tr>
<td>Nuclear Regulatory Commission</td>
<td>Atomic energy equipment builders</td>
</tr>
</tbody>
</table>


\(^5^2\) The distributional type of nonmarket failure is the core of Stigler's theory of economic regulation. Stigler finds empirical evidence to support this hypothesis in interstate variations in trucking regulation and in occupational licensing. George J. Stigler, *supra* note 27.
the nonmarket, but it does suggest there is an identifiable process by which inequities can result from nonmarket activities, as they result from market activities.

* * * *

In summary, nonmarket inadequacies can be tabulated in comparison with the inadequacies of the market.

**Market and Nonmarket Failures**

<table>
<thead>
<tr>
<th>Market</th>
<th>Nonmarket</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Externalities and public goods</td>
<td>1. Internalities and private goals</td>
</tr>
<tr>
<td>2. Increasing returns</td>
<td>2. Redundant and rising costs</td>
</tr>
<tr>
<td>3. Market imperfections</td>
<td>3. Derived externalities</td>
</tr>
<tr>
<td>4. Distributional inequity</td>
<td>4. Distributional inequity</td>
</tr>
<tr>
<td>(income and wealth)</td>
<td>(influence and power)</td>
</tr>
</tbody>
</table>

These parallel categories should not be misunderstood. Nonmarket inadequacies are not the "duals" of those associated with market activities. The externalities on the market side are qualitatively related to the internalities on the nonmarket side only in the sense that each is a major source of failure in the market and nonmarket contexts, respectively. (Indeed, externalities in the market sector are conceptually closer to derived externalities in the nonmarket sector than either is to its horizontal neighbor in the two lists.)

However, two points apply to both lists.

(1) For the several types of market and nonmarket failures, it is much easier to estimate signs than magnitudes. Estimating magnitudes requires detailed empirical work in specific cases and contexts. Moreover, it is no easier to determine the magnitude of, say, the (negative) national security externalities associated with U.S. reliance on Middle Eastern oil than to determine, say, the magnitude of internalities that affect the behavior of the U.S. Air Force. Or, to take a more tractable example, it is probably no more difficult to estimate the derived externalities (negative as well as positive) resulting from environmental regulation than it is to estimate the (negative) externalities resulting from unregulated strip mining, or from noise emissions near metropolitan airports.

(2) The types and sources of market failure indicate the circumstances in which government intervention is worth contemplating, and in which alternative policies are worth analyzing as possible remedies. Similarly, the types and sources of nonmarket failure indicate the circumstances in which government intervention may itself misfire, and in which potential correctives are worth analyzing as possible remedies for likely shortcomings of government intervention.
The existing theory of market failure provides a useful corrective to the theory of perfectly functioning markets. In a similar sense, the theory of nonmarket failure outlined in the preceding pages is intended as a corrective for the implicit theory of perfectly functioning governments. Just as market failures or inadequacies have provided the theoretical underpinning for applied policy analysis, so nonmarket failures may provide the theoretical underpinning for implementation analysis. The analysis of how specific nonmarket activities (public policies) can be expected to operate and to depart in predictable ways from their costs and consequences as originally estimated.

III. Implementation Analysis

Even the most sophisticated policy analysis usually neglects implementation issues. Policy studies rarely raise, and almost never answer such questions as who would have to do what, and when, and with what foreseeable resistance, modifications, and compromises if alternative A were chosen, or B, or C? As far as implementation is concerned, the Napoleonic dictum, "On s'engage et puis on voit," is tacitly accepted. Analysts implicitly assume that the costs and benefits, as modeled in the analysis, will not be altered by implementation.

One numerical estimate suggests how badly policy studies turn out when prior estimates are compared with later results. Cost-estimating relationships, based on several dozen weapons systems developed in the United States, showed that on average costs for these systems (holding constant system differences in technology, performance, size, and so on) rose by a factor of three between the time development was begun and delivery was completed.\(^{53}\)

The question is whether policy studies can better deal with implementation matters so that the implementation chapter of policy studies will not remain "missing."\(^{54}\) In recent years, interest in the question of implementation has increased substantially, as reflected in the new public policy journals (Policy Analysis, Policy Sciences), several recent books and case stud-


ies, and the curricula of graduate schools of policy analysis.\textsuperscript{55} Most discussion has emphasized the typically large gaps between programs as designed and as executed, the lack of appropriate methods for anticipating these gaps and taking them into account, and consequently the failure of virtually all policy analysis to address implementation issues systematically.

To move from these justifiable criticisms to the systematic analysis of implementation issues requires an acceptable paradigm. In the following discussion, I will suggest that the preceding treatment of nonmarket failures provides this paradigm: a method of analyzing how public policy (that is, nonmarket) efforts to compensate for market failures may themselves fail in predictable ways. Anticipating nonmarket failures can be invaluable for trying to avoid them, or for developing mixed market and nonmarket alternatives that will diminish the more undesirable consequences of each.

Policy analysis can deal more effectively with implementation issues by linking the theory of nonmarket failure with explicit consideration of the "who-what-when" questions mentioned above. The reasons for implementation shortfalls—for costs to rise and effectiveness to fall ex post—in public policies intended to correct inadequacies of the market lie in the predictable inadequacies of nonmarket activities themselves. Hence, implementation analysis, as a regular segment of policy analysis, should proceed by applying the theory of nonmarket failures outlined above.

The principal connections among these stages and processes are summarized in Figure I. Emphasis in the preceding and following discussion is on the connections marked with solid lines. The roman numerals refer to the sections of this paper where the indicated connections have been discussed. Section I has focused on the connection between market output and market failures, Section II on those between market failures, nonmarket output and nonmarket failures; and Section III emphasizes the connections between nonmarket failures, implementation analysis, and policy analysis. The connections shown with dotted lines are alluded to, but not treated in detail.

The aim of this section is thus to incorporate implementation analysis systematically into policy studies. As a first step it is useful to distinguish two parts: (A) descriptive and (B) normative-inventive.

(A) The descriptive part of implementation analysis should use the structure of nonmarket failure outlined above as a check list for comparing each policy alternative under consideration. More specifically, the following implementation questions would be addressed as part of the analysis of policy alternatives and prior to choosing among them:

\textsuperscript{55} The term "missing chapter" was first used in print by Graham T. Allison, \textit{supra} note 56, although it was originated by Andrew Marshall and me five years earlier, to describe the usual neglect of implementation analysis in policy studies.
1. If policy A (or B, or C) were to be adopted, what government departments, agencies, or bureaus would have to be assigned what precise responsibilities?

2. To the extent these designated agencies are already in existence, rather than new agencies to be created, what are the internalities and private goals that now motivate them, and how is behavior affected as a result? (If one looks at how these agencies really operate, especially at their criteria for recruiting, evaluating, and promoting personnel, how is output actually measured, and how are success and effectiveness in producing it assessed? Are staff members rewarded for adding to or justifying costs, or for reducing them; for generating new technology, or opposing it, or objectively evaluating different types; for connecting the agency with new information sources and restricting access to them, or for facilitating informational flows to and from other agencies?)

If the policies under consideration call for creating new administering agencies, can the corresponding internalities, and the way in which they will influence agency behavior, be anticipated (for example, by the evident connection between particular policies and the interest groups advocating them)? Can these prospective agency internalities be modified by program redesign, and with what expected effects on agency behavior?
3. What externalities may result from the alternative policies—over what time period and with what prior probabilities attached to them—in consideration of the content, scale, and impact, perhaps in policy areas remote from the target area of the programs under consideration? From the analyst's point of view, allowing for unforeseen externalities is thus an exercise in trying hard to remember what he would otherwise overlook.

4. Based on the track records of the agencies involved, on scrutiny of the alternative policies for the possible existence of inconsistent or otherwise infeasible program objectives, and on considerations covered in (2) above, can estimates be made of the prospective occurrence of redundant and rising costs associated with the agency responsibilities identified in (1) above? Can cost-estimating relationships be calculated (as in the system acquisition example referred to earlier) expressing the upward drift in cost functions to be expected over time?

5. Finally, in accord with the way in which each of the policies or programs would be expected to operate, how much discretionary authority is allowed, and to whom? What changes would ensue in distribution, not only income distribution but also in the distribution of power that may be wielded by some over others?

It is evident that many, indeed most, of the foregoing questions are not answerable in precise terms. Answers are likely to be judgments and opinions, hence subject to some disagreement by reasonable people even after substantial empirical work to attain "objective" information. Nevertheless, even "soft" answers, which display divergent opinions, should be valuable for the normative dimensions of implementation analysis.

(B) There are three normative purposes to be served by the previous descriptive treatment of alternative policies and programs. One purpose is simply to facilitate evaluation of the specified alternatives with respect to the ease or difficulty of implementing them: of translating "what is good to be done" into an estimate of what actually would get done. In effect, this would amount to an ex post adjustment in the costs and benefits as modeled before implementation considerations were brought into the analysis.

A second purpose is to facilitate comparison between the actual inadequacies associated with the market and the potential inadequacies associated with implementing the nonmarket remedies under consideration. Juxtaposing the market failures to be remedied, and the nonmarket failures to be anticipated from the intended remedies, would permit an assessment that has

56 "If to do were as easy as to know what were good to do, chapels had been churches, and poor men's cottages princes' palaces. ... I can easier teach twenty what were good to be done, than be one of the twenty to follow mine own teaching." William Shakespeare, The Merchant of Venice 224 (Classics Club Library ed. 1937).
been rare in previous policy studies and should become difficult to avoid in future ones.\textsuperscript{57} 

The third purpose of applying the paradigm of nonmarket failure is to stimulate invention: new ideas for policies and programs, or combinations among those under consideration, or of parts of them, or of entirely different approaches to the problem. A systematic comparison between market failures and nonmarket failures in a particular problem area (the second purpose mentioned above), and among the potential nonmarket failures associated with various alternative policies (the first purpose), should contribute to a result Dr. Johnson associated with the prospect of being hanged: namely, “to concentrate the mind wonderfully.” Invention of new options, or discovery of ways to improve existing options, may result. If nonmarket solutions have been needed as countermeasures against market failures, we now need to develop countermeasures against nonmarket failures (hence, “counter-countermeasures” against market failures).

Besides evaluation of the existing set of options, the normative-inventive part of implementation analysis should focus on the following set of questions, which are as important, and as formidable, as the previous set:

1. Are there relatively simple and easily administered “fixes” in the operation of markets which would sufficiently alleviate the acknowledged market failure to provide an acceptable solution?\textsuperscript{58}

2. Can policies be invented that, while recognizing the need for nonmarket interventions because the market’s inadequacies are so great (for example, in the case of public goods or of private goods with major externalities), nevertheless try to retain certain valuable characteristics of market solutions (for example, competition by several producers, tangible and public performance measures, a bottom-line incentive structure in operating agencies)? In particular, can mechanisms be devised for the “reprivatization” of certain public services, for example, using publicly funded vouchers for the “purchase” of education or open bidding on private contracts for waste disposal or postal services.\textsuperscript{59}

\textsuperscript{57} This comparison is similar to what has been referred to as “zero-based budgeting” in discussions of planning, programming, and budgeting systems (PPBS). Arthur Smithies, The Budgetary Process in the United States (1955). The inadequacies of a particular market outcome, with little or no public intervention (a “zero” budget), may be preferable to the inadequacies of the nonmarket remedy.

\textsuperscript{58} Some possible examples are: (a) estimating the separate effect of noise emissions on property values in airport vicinities and obliging airlines to compensate property owners accordingly, while leaving to the airlines the choice of aircraft power plant, acoustical damping, or other measures to reduce noise; (b) using foreign trade policy as an adjunct or alternative to antitrust policy in maintaining competitive pressures in monopolistic industries; (c) reducing market imperfections (for example, by removing or lowering barriers to entry or providing adjustment assistance to facilitate factor mobility).

\textsuperscript{59} Peter F. Drucker, The Age of Discontinuity: Guidelines to Our Changing Society (1969);
3. Can improved measures for nonmarket output be devised, so that those nonmarket failures resulting from the lack of a suitable metric can be reduced? Can tests be made of the connections, or lack thereof, between the intermediate outputs that are often reflected in agency internalities and the final outputs that are intended?

4. Can the internalities (standards, goals) that provide the incentives for individual and agency behavior be revised so as to be more closely connected with the final intended output?

5. Can improved information, feedback, and evaluation systems be built into new policies and programs in order to reduce the risks of co-optation by a "client" group and to publicize it if it occurs?

The normative questions of implementation analysis are no less formidable than those relating to the descriptive aspects discussed earlier. At best, attempts to respond systematically to the implementation questions raised by the nonmarket failure paradigm are likely to result in uncertain answers. Yet even without firm or complete answers, or indeed even without answers at all, there is considerable merit in the exercise. Addressing the questions in specific policy contexts requires that they be reformulated with precise reference to those contexts. For each policy alternative, the cardinal implementation issues ("who has to do what, when, how?") cannot be avoided. What has been omitted from virtually all policy studies, and what has significantly contributed to the failure of many implemented policies, must then be given explicit attention.

Another type of criticism can also be advanced. If, in fact, these formidable implementation questions can be answered in some fashion, why can't the answers simply be fed back into conventional policy analyses in accord with the standard methodology described earlier? Why can't the descriptive aspects of implementation analysis be made part of the analytical models and included in the usual cost-effectiveness calculations? And why can't the normative-inventive aspects simply be added to the policy alternatives to be run through the analytical models? These questions imply a direction of development that policy analysis should take: incorporation of implementation considerations within the existing "chapters" of the standard analyses, rather than as a separate chapter. However, we just don't know enough yet.


60 Such revisions are apt to involve consideration of agency personnel practices, and in this respect would move implementation analysis in a direction taken by management consulting.
to be able to do this. Raising the implementation questions as issues for consideration can perhaps elicit enough of a response to fill an important gap in existing policy studies. But the response is unlikely to be sufficiently rigorous for formal inclusion in analytical models at this stage. We need first to consider in a rough and qualitative way what has been largely ignored in policy research as a step toward more systematic inclusion in analytic methodology in the future.

IV. CONCLUSIONS

The foregoing argument can be summarized in several propositions:

1. The essential rationale for public policy measures lies in specific failures of the market of itself to produce efficient or otherwise socially preferred outcomes.

2. However, this rationale provides only a necessary, not a sufficient, justification for public policy interventions. Sufficiency requires that specifically identified market failures be compared with potential nonmarket failures associated with the implementation of public policies. Such a comparison is needed to arrive at a balanced assessment of whether, as well as what kind of, policy intervention will come closer to a socially preferred outcome.

3. There are four sources of the market's failure to produce socially preferred outcomes: externalities and public goods, increasing returns, market imperfections, and distributional inequity. The most general explanation for these failures is that markets don't exist, and perhaps can't be created, that will suitably capture the full social benefits or levy the full social costs of market activity. (This general explanation can be extended to include in part, though not completely, the last of the four market failures, distributional inequity.)

4. Similarly, as a result of distinctive demand and supply characteristics associated with nonmarket output, there are four sources and types of nonmarket failure: "internalities" and private goals (relating, for example, to agency budgets, technology, and information acquisition and control); redundant and rising costs; derived externalities; and distributional inequity (indexed on power, as well as on income or wealth). The distinctive demand and supply characteristics that give rise to these nonmarket failures include the following: premature but politically effective demands for government action; difficulties of defining and measuring output; lack of a "bottom line" for evaluating performance; absence of competition; and lack of an effective termination mechanism. Where there is nonmarket failure, there is an absence of nonmarket incentives that reconcile the calculations of costs and benefits by government decision makers with total costs and benefits.
5. In order to make more reliable comparisons among alternative public policies, as well as between them and market outcomes, policy analysis should explicitly consider how particular policy alternatives will be implemented. Implementation analysis, as a regular component of policy analysis, should link the formal modeling and cost-effectiveness comparisons among alternative policies with consideration of how policies are likely to be altered if implemented. It is intended to explain and anticipate the frequent tendency of implemented policies to result in higher costs and lower benefits, as well as different consequences, from those calculated in conventional policy studies.

6. There are two parts to implementation analysis: (a) a descriptive part, in which the previously mentioned sources of nonmarket failure are systematically traced for each policy alternative, and (b) a normative-inventive part, in which the costs and consequences of each alternative—as modified by the description of implementation realities—are compared with one another, as well as with the market outcome and its attendant shortcomings. This part of implementation analysis should explicitly consider changes and inventions in the policy options under consideration—changes that can be expected to reduce either the failures of nonmarket activities (for example, by devising improved measures of final output and translating them into agency internalities, or by reprivatizing public services) or the failures of market solutions (for example, by measures that reduce market imperfections).