Exit, Voice, and Loyalty

Responses to Decline in Firms, Organizations, and States

Albert O. Hirschman

Harvard University Press
Cambridge, Massachusetts
and London, England

Library of Congress Catalog Card Number: 77-99517
ISBN 0-674-27660-4
Printed in the United States of America

Contents

Exit, Voice, and Loyalty

1
Introduction and Doctrinal Background

Under any economic, social, or political system, individuals, business firms, and organizations in general are subject to lapses from efficient, rational, law-abiding, virtuous, or otherwise functional behavior. No matter how well a society's basic institutions are devised, failures of some actors to live up to the behavior which is expected of them are bound to occur, if only for all kinds of accidental reasons. Each society learns to live with a certain amount of such dysfunctional or mis-behavior; but lest the misbehavior feed on itself and lead to general decay, society must be able to marshal from within itself forces which will make as many of the faltering actors as possible revert to the behavior required for its proper functioning. This book undertakes initially a reconnaissance of these forces as they operate in the economy; the concepts to be developed will, however, be found to be applicable not only to economic operators such as business firms, but to a wide variety of noneconomic organizations and situations.

While moralists and political scientists have been much concerned with rescuing individuals from immoral behavior, societies from corruption, and governments from decay, economists have paid little attention to *repairable lapses* of economic actors. There are two reasons for this neglect. First, in economics one assumes either fully and undeviatingly rational behavior or, at the very least, an *unchanging level* of rationality on the part of the economic actors. Deterioration of a firm's performance may result from an adverse shift in supply and demand conditions while the willingness and ability of the firm to maximize profits (or growth rate or whatever) are unimpaired; but it could also reflect some "loss of maximizing aptitude or energy" with supply and demand factors being un-

changed. The latter interpretation would immediately raise the question how the firm's maximizing energy can be brought back up to par. But the usual interpretation is the former one; and in that case, the reversibility of changes in objective supply and demand conditions is much more in doubt. In other words, economists have typically assumed that a firm that falls behind (or gets ahead) does so *"for a good reason"*; the concept—central to this book—of a random and more or less easily "repairable lapse" has been alien to their reasoning.

The second cause of the economist's unconcern about lapses is related to the first. In the traditional model of the competitive economy, recovery from any lapse is not really essential. As one firm loses out in the competitive struggle, its market share is taken up and its factors are hired by others, including newcomers; in the upshot, total resources may well be better allocated. With this picture in mind, the economist can afford to watch lapses of any one of *his* patients (such as business firms) with far greater equanimity than either the moralist who is convinced of the intrinsic worth of every one of *his* patients (individuals) or the political scientist whose patient (the state) is unique and irreplaceable.

Having accounted for the economist's unconcern we can immediately question its justification: for the image of the economy as a fully competitive system where changes in the fortunes of individual firms are exclusively caused by basic shifts of comparative advantage is surely a defective representation of the real world. In the first place, there are the well-known, large realms of monopoly, oligopoly, and monopolistic competition: deterioration in performance of firms operating in that part of the economy could result in more or less permanent *pockets* of inefficiency and neglect; it must obviously be viewed with an alarm approaching that of the political scientist who sees his polity's integrity being threatened by strife, corrup-

tion, or boredom. But even where vigorous competition prevails, unconcern with the possibility of restoring temporarily laggard firms to vigor is hardly justified. Precisely in sectors where there are large numbers of firms competing with one another in similar conditions, declines in the fortunes of individual firms are just as likely to be due to random, subjective factors that are reversible or remediable as to permanent adverse shifts in cost and demand conditions. In these circumstances, mechanisms of recuperation would play a most useful role in avoiding social losses as well as human hardship.

At this point, it will be interjected that such a mechanism of recuperation is readily available through competition itself. Is not competition supposed to keep a firm "on its toes"? And if the firm has already slipped, isn't it the experience of declining revenue and the threat of extinction through competition that will cause its managers to make a major effort to bring performance back up to where it should be?

There can be no doubt that competition is one major mechanism of recuperation. It will here be argued, however (1) that the implications of this particular function of competition have not been adequately spelled out and (2) that a major alternative mechanism can come into play either when the competitive mechanism is unavailable or as a complement to it.

Enter "Exit" and "Voice"

The argument to be presented starts with the firm producing saleable outputs for customers; but it will be found to be largely—and, at times, principally—applicable to organizations (such as voluntary associations, trade unions, or political parties) that provide services to their members without direct monetary counterpart. The per-

formance of a firm or an organization is assumed to be subject to deterioration for unspecified, random causes which are neither so compelling nor so durable as to prevent a return to previous performance levels, provided managers direct their attention and energy to that task. The deterioration in performance is reflected most typically and generally, that is, for both firms and other organizations, in an absolute or comparative deterioration of the *quality* of the product or service provided.[1] Management then finds out about its failings via two alternative routes:

(1) Some customers stop buying the firm's products or some members leave the organization: this is the *exit option.* As a result, revenues drop, membership declines, and management is impelled to search for ways and means to correct whatever faults have led to exit.

(2) The firm's customers or the organization's members express their dissatisfaction directly to management or to some other authority to which management is subordinate or through general protest addressed to anyone who cares to listen: this is the *voice option.* As a result, management once again engages in a search for the causes and possible cures of customers' and members' dissatisfaction.

The remainder of this book is largely devoted to the

1. For business firms operating in situations of monopoly or monopolistic competition, performance deterioriation can also be reflected in cost and resulting price increases or in a combination of quality drops and price increases. On the other hand, changes in either price or quality are ruled out when both are rigidly dictated by a perfectly competitive market; in this admittedly unrealistic situation, deterioration can manifest itself only via increases in cost which, with price and quality unchanged, will lead straightaway to a decline in net revenue. Under perfect competition, then, managers learn about their failings directly and exclusively from financial evidence generated within the firm, without any intermediation on the part of the customers who remain totally unaware of the firm's troubles. It is perhaps because the whole range of phenomena here described has no place in the perfectly competitive model that it has not been paid attention to by economists.

comparative analysis of these two options and to their interplay. I will investigate questions such as: Under what conditions will the exit option prevail over the voice option and vice versa? What is the comparative efficiency of the two options as mechanisms of recuperation? In what situations do both options come into play jointly? What institutions could serve to perfect each of the two options as mechanisms of recuperation? Are institutions perfecting the exit option compatible with those designed to improve the working of the voice option?

Latitude for Deterioration, and Slack in Economic Thought

Before setting out to answer some of these questions, I shall now step back briefly and indicate how I conceive the subject of this book to be related to economic and social science thought around us.

Talking with students of animal behavior (at the Center for Advanced Study in the Behavioral Sciences) about the social organization of primates I learnt about the smoothness and efficiency with which leadership succession, a problem human societies have found so intractable, was handled in certain baboon bands. Here is how the process is described for a typical band of Hamadryas baboons lorded over by one male leader:

> Sub-adult males steal very young females from their mothers and attend them with every semblance of solicitous maternal care. The young female is rigorously controlled, and repeated retrieval trains her not to go away . . . At this stage there is no sexual behaviour, the female being yet two to three years from child bearing . . . As these young interlopers mature and the overlord ages, the younger animal starts initiating group movements although the direction of eventual movement is dependent upon the older animal's choice. A highly complex relation-

ship develops between the two animals which, by paying close attention to one another and by reciprocal "notification," cooperate in governing group movement. Old males retain command of group direction but gradually relinquish sexual control over their females to the younger male animal . . . It seems that eventually old males resign entirely from their original reproduction units but retain great influence within the band as a whole, and young males refer to them continuously particularly before developing the direction of march.[2]

Compare this marvel of gradualness and continuity with the violent ups and downs to which human societies have always been subject as "bad" government followed upon "good," and as strong or wise or good leaders were succeeded by weaklings, fools, or criminals.

The reason for which humans have failed to develop a finely built social process assuring continuity and steady quality in leadership is probably that they did not have to. Most human societies are marked by the existence of a surplus above subsistence. The counterpart of this surplus is society's ability to take considerable deterioration in its stride. A lower level of performance, which would mean disaster for baboons, merely causes discomfort, at least initially, to humans.

The wide latitude human societies have for deterioration is the inevitable counterpart of man's increasing productivity and control over his environment. Occasional decline as well as prolonged mediocrity—in relation to achievable performance levels—must be counted among the many penalties of progress. A priori it would seem futile, therefore, to look for social arrangements that

2. John Hurrell Crook, "The Socio-Ecology of Primates," in J. H. Crook, ed., *Social Behaviour in Animals and Man* (to be published by Academic Press, London). The passage quoted summarizes research by Hans Kummer, "Social Organization of Hamadryas Baboons," *Bibliotheca Primatologica,* no. 6 (Basle: S. Karger, 1968).

would wholly eliminate any sort of deterioration of polities and of their various constituent entities. Because of the surplus and the resulting latitude, any homeostatic controls with which human societies might be equipped are bound to be rough.

Recognition of this unpleasant truth has been impeded by a recurring utopian dream: that economic progress, while increasing the surplus above subsistence, will also bring with it disciplines and sanctions of such severity as to rule out any backsliding that may be due, for example, to faulty political processes. In the eighteenth century the expansion of commerce and of industry was sometimes hailed not so much because of the increase in well-being that it would make possible, but because it would bring with it powerful restraints on the willfulness of the prince and thereby reduce and perhaps eliminate the system's latitude for deterioration. One characteristic passage from Sir James Steuart's *Inquiry into the Principles of Political Oeconomy* (1767) will suffice to make the point:

> How hurtful soever the natural and immediate effects of political revolutions may have been formerly, when the mechanism of government was more simple than at present, they are now brought under such restrictions, by the complicated system of modern oeconomy, that the evil which might otherwise result from them may be guarded against with ease . . .
>
> The power of a modern prince, let it be by the constitution of his kingdom ever so absolute, immediately becomes limited so soon as he establishes the plan of oeconomy . . . If his authority formerly resembled the solidity and force of the wedge (which may indifferently be made use of, for splitting of timber, stones and other hard bodies, and which may be thrown aside and taken up again at pleasure), it will at length come to resemble the delicacy of the watch, which is good for no other purpose than to mark the progression of time, and which is immediately destroyed, if put to any other use, or touched with any but

the gentlest hand . . . modern oeconomy, therefore, is the most effectual bridle ever was invented against the folly of despotism.[3]

This noble hope echoes nearly two hundred years later in the writings of a Latin American intellectual similarly predicting, against all likelihood, that economic progress and latitude for deterioration will be negatively, rather than positively, correlated:

[In the pre-coffee era, policy makers] are lyrical and romantic because they cannot yet defer to a product whose output is constantly on the increase. It is a time of childhood and play. Coffee will bring maturity and seriousness. It will not permit Colombians to continue playing fast and loose with the national economy. The ideological absolutism will disappear and the epoch of moderation and sobriety will dawn . . . Coffee is incompatible with anarchy.[4]

History has cruelly disappointed the expectations of both Sir James Steuart and Nieto Arteta that economic growth and technical progress would erect secure barriers against "despotism," "anarchy," and irresponsible behavior in general. Yet their line of thought is hardly extinct. It is, in fact, not unrelated to today's widespread belief that a major war is unthinkable and therefore impossible in the nuclear age.

The common assumption of these constructs is simply stated: while technical progress increases society's surplus above subsistence it also introduces a mechanism of the utmost complexity and delicacy, so that certain types of social misbehavior which previously had unfortunate

3. (Chicago: University of Chicago Press, 1966), I, 277, 278–279.

4. Luis Eduardo Nieto Arteta, *El café en la sociedad colombiana* (Bogotá: Breviarios de orientación colombiana, 1958), pp. 34–35. This posthumously published essay was written in 1947, only a year before the outbreak of the sanguinary civil disturbances known as *la violencia*, just as Sir James Steuart wrote about the definitive conquest of despotism not long before the rise of Napoleon.

but tolerable consequences would now be so clearly disastrous that they will be more securely barred than before.

As a result society is, and then again it is not, in a surplus situation: it is producing a surplus, but is not at liberty *not* to produce it or to produce less of it than is possible; in effect, social behavior is as simply and as rigidly prescribed and constrained as it is in a no-surplus, bare subsistence situation.

The economist cannot fail to note the similarity of the situation with the model of perfect competition. For this model contains the same basic paradox: society as a whole produces a comfortable and perhaps steadily increasing surplus, but every individual firm considered in isolation is barely getting by, so that a single false step will be its undoing. As a result, everyone is constantly made to perform at the top of his form and society as a whole is operating on its—forever expanding—"production frontier," with economically useful resources fully occupied. This image of a relentlessly *taut economy* has held a privileged place in economic analysis, even when perfect competition was recognized as a purely theoretical construct with little reality-content.

These various observations add up to a syndrome, namely, to man's fundamentally ambivalent attitude toward his ability to produce a surplus: he likes surplus but is fearful of paying its price. While unwilling to give up progress he hankers after the simple rigid constraints on behavior that governed him when he, like all other creatures, was totally absorbed by the need to satisfy his most basic drives. Who knows but that this hankering is at the root of the paradise myth! It seems plausible, indeed, that the *rise* of man above the narrowly constrained condition of all other living creatures was frequently sensed, though it can hardly ever have been avowed, as a *fall;* and a radical but basically simple act of the imagination may well have metamorphosed this condition which one was really

yearning for into its exact opposite, the Garden of Eden.[5]

But we must leave paradise and return to social thought, for there is another side to our story. The simple idea that the ability to produce a surplus above subsistence makes it possible and indeed likely that occasionally less than the maximum producible surplus will be produced has not gone wholly unnoticed. In fact, next to the traditional model of the permanently *taut* economy, elements of a theory of the *slack* economy begin to be available. I am not referring now to unemployment and depression economics —the slack associated with these phenomena results from malfunctions at the macroeconomic level which frustrate firms and individuals in their supposedly undiminished zeal to maximize profits and satisfaction. Nor is the question of slack involved in the dispute about what it is that business firms, and particularly the large corporations, really do maximize: profits, growth, market shares, community goodwill, or some composite functions of such objectives. The assumption underlying this dispute is that, whatever it is that firms do, they do it the best they can even though the criterion for "best" performance is becoming rather murky. Finally I am not concerned with the large body of writings showing that the actions of conscientiously maximizing private producers and consumers may fail to produce a *social* optimum, because of the existence of monopolistic elements and externalities.

5. Samuel Johnson intimated this thought in his fable about the Happy Valley of Abyssinia. When Prince Rasselas first analyzes the discontent he feels in the paradiselike valley, he compares his condition to that of some grazing goats in the following terms: "What makes the difference between man and all the rest of the animal creation? Every beast that strays beside me has the same corporal necessities with myself; he is hungry and crops the grass, he is thirsty and drinks the stream, his thirst and hunger are appeased, he is satisfied and sleeps; he rises again and is hungry, he is again fed and is at rest. I am hungry or thirsty like him, but when thirst or hunger cease, I am not at rest; I am like him pained with want, but am not, like him, satisfied with fulness." (Samuel Johnson, *Rasselas*, II.)

Here again the difference between actual and potential output is not due to some "failure of nerve" at the microeconomic level. But of late there has been increasing attention to the possibility of just such a failure.

A seminal contribution in this area was H. A. Simon's suggestion that firms are normally aiming at no more than a "satisfactory" rather than at the highest possible rate of profits.[6] This notion was given considerable underpinning in 1963 by Richard Cyert and James March, who in their book *A Behavioral Theory of the Firm*[7] introduced the concept of "organizational slack." At about the same time, Gary Becker showed that some of the basic and empirically well-tested microeconomic theorems (for example, that market demand curves for individual commodities are negatively inclined) are consistent with a wide range of irrational and inefficient behavior on the part of consumers and producers even though these theorems had originally been derived on the assumption of undeviating rationality.[8] The importance of slack was later affirmed in a particularly sweeping form by Harvey Leibenstein.[9] Finally, in a widely discussed polemical essay, Professor M. M. Postan has recently contended that Britain's economic ailments are better understood by focusing on microeconomic slack than on any mistaken macroeconomic policies. He writes:

6. H. A. Simon, "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69:98–118 (1952). An early, completely forgotten empirical work with a related theme has the significant title *The Triumph of Mediocrity in Business*, by Horace Secrist, published in 1933 by the Bureau of Business Research, Northwestern University. The book contains an elaborate statistical demonstration that, over a period of time, initially high-performing firms will on the average show deterioration while the initial low performers will exhibit improvement.

7. Richard M. Cyert and James G. March, *Behavioral Theory of the Firm* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1963).

8. Gary S. Becker, "Irrational Behavior and Economic Theory," *Journal of Political Economy*, 52:1–13 (February 1962).

9. Harvey Leibenstein, "Allocative Efficiency versus X-Efficiency," *American Economic Review*, 56:392–415 (June 1966).

For many (perhaps most) of these . . . ailments the morbid causes will be found not in the malfunctioning of the life processes in the body economic, such as the low rate of savings, or the high level of prices, or the insufficient allocation of national resources to research and development, but in specific failures of its individual cells—management, design, salesmanship, or the behavior of groups of labor.[10]

I feel considerable kinship with this group of writings for I had adopted a similar position in dealing with the problem of development. The basic proposition of *The Strategy of Economic Development* (1958) was that "development depends not so much on finding optimal combinations for given resources and factors of production as on calling forth and enlisting for development purposes resources and abilities that are hidden, scattered or badly utilized." [11] And the term slack actually came under my pen when I summarized later on the essential argument of that book in an article co-authored with C. E. Lindblom:

At any one point of time, an economy's resources are not to be considered as rigidly fixed in amount, and more resources or factors of production will come into play if development is marked by sectoral imbalances that galvanize private entrepreneurs or public authorities into action . . . The crucial, but plausible, assumption here is that there is some "slack" in the economy; and that additional investment, hours of work, productivity, and decision making can be squeezed out of it by pressure mechanisms.[12]

Various reasons have been invoked for explaining slack. Leibenstein's emphasis is on the uncertainties surrounding the production function and on the nonmarketability

10. M. M. Postan, "A Plague of Economists?" *Encounter* (January 1968), p. 44.

11. (New Haven: Yale University Press, 1958), p. 5.

12. "Economic Development, Research and Development, Policy Making: Some Converging Views," *Behavioral Science*, 7:211–212 (April 1962).

of managerial and other skills. Cyert and March refer primarily to the bargaining process that takes place among the various parties whose (shaky) coalition is required for factors to be hired and for output to be produced and marketed. I stressed rather similarly the existence of obstacles to entrepreneurial and cooperative behavior needed for the making of development decisions.

Those who have found that the individual economic operators and, as a result, the economy are ordinarily far from doing as well as they might, can be expected to react to their shocking discovery along two principal lines. The immediate and most obvious reaction is a determined search for ways and means to take up the slack, to retrieve the ideal of the taut economy. As long as the pressures of competition do not seem to be sufficient, the pressures of adversity will be invoked.[13] Frequent changes in the environment, forcing the firm to be "on its toes," will be identified as one way of inducing performance closer to the firm's potential.[14] Insofar as innovation is concerned, the inducing and focusing virtues of strikes and war have been stressed.[15] My own search concentrated on pressure mechanisms such as intersectoral and intrasectoral imbalances and on production processes that exact high penalties for poor performance or do not tolerate it at all.[16] Finally, the advocates of social revolution have contributed to this line of thought: one of their most seductive arguments has long been that only revolutionary changes can tap and liberate the abundant but dormant, repressed, or alienated energies of the people.[17]

13. See Leibenstein, "Allocative Efficiency versus X-Efficiency."

14. Charles P. Bonini, "Simulation of Information and Decision Systems in the Firm" (unpub. diss. Carnegie Institute of Technology, 1962).

15. Nathan Rosenberg, "The Direction of Technological Change: Inducement Mechanisms and Focusing Devices," *Economic Development and Cultural Change*, 18 (October 1969).

16. Hirschman, *Strategy*, chs. 5–8.

17. See, for example, Paul Baran, *The Political Economy of Growth* (New York: Monthly Review Press, 1957).

Quite a different reaction to the discovery of slack occurs when the discoverer asks himself, after having got over his initial shock, whether slack may not after all be a good thing, a blessing in disguise. The idea that slack fulfills some important, if unintended or latent, functions was put forward by Cyert and March, who point out that it permits firms to ride out adverse market or other developments. During such bad times slack acts like a reserve that can be called upon: excess costs will be cut, innovations that were already within one's grasp will at last be introduced, more aggressive sales behavior that had been shunned will now be engaged in, and so on. Slack in the political system has been rationalized in a very similar manner. The discovery that citizens do not normally use more than a fraction of their political resources came originally as a surprise and disappointment to political scientists who had been brought up to believe that democracy requires for its functioning the fullest possible participation of all citizens. But soon enough a degree of apathy was found to have some compensating advantages in as much as it contributes to the stability and flexibility of a political system and provides for "reserves" of political resources which can be thrown into the battle in crisis situations.[18]

The immediate response to the discovery of slack has thus been either to assert the rationality of a certain level of slack or to look for ways of extirpating excessive levels by invoking exogenous forces such as adversity, imbalances, revolution, and so on. Both these approaches look at slack as a gap of a given magnitude between actual and potential performance of individuals, firms, and organizations. This book takes a further, more radical step in recognizing the importance and pervasiveness of slack. It assumes not only that slack has somehow come into the

18. See below, pp. 31–32.

world and exists in given amounts, but that it is *continuously being generated* as a result of some sort of entropy characteristic of human, surplus-producing societies. "There's a slacker born every minute," could be its motto. Firms and other organizations are conceived to be permanently and randomly subject to decline and decay, that is, to a gradual loss of rationality, efficiency, and surplus-producing energy, no matter how well the institutional framework within which they function is designed.

This radical pessimism, which views decay as an ever-present force constantly on the attack, generates its own cure: for as long as decay, while always conspicuous in some areas, is hardly in undisputed command everywhere and at all times, it is likely that the very process of decline activates certain counterforces.

Exit and Voice as Impersonations of Economics and Politics

In examining the nature and strength of these endogenous forces of recovery, our inquiry bifurcates, as already explained. Its breakup into the two contrasting, though not mutually exclusive, categories of exit and voice would be suspiciously neat if it did not faithfully reflect a more fundamental schism: that between economics and politics. Exit belongs to the former realm, voice to the latter. The customer who, dissatisfied with the product of one firm, shifts to that of another, uses the market to defend his welfare or to improve his position; and he also sets in motion market forces which may induce recovery on the part of the firm that has declined in comparative performance. This is the sort of mechanism economics thrives on. It is neat—one either exits or one does not; it is impersonal —any face-to-face confrontation between customer and

firm with its imponderable and unpredictable elements is avoided and success and failure of the organization are communicated to it by a set of statistics; and it is indirect —any recovery on the part of the declining firm comes by courtesy of the Invisible Hand, as an unintended by-product of the customer's decision to shift. In all these respects, voice is just the opposite of exit. It is a far more "messy" concept because it can be graduated, all the way from faint grumbling to violent protest; it implies articulation of one's critical opinions rather than a private, "secret" vote in the anonymity of a supermarket; and finally, it is direct and straightforward rather than roundabout. Voice is political action par excellence.

The economist tends naturally to think that his mechanism is far more efficient and is in fact the only one to be taken seriously. A particularly good illustration of this bias appears in a well-known essay by Milton Friedman which advocates the introduction of the market mechanism into public education. The essence of the Friedman proposal is the distribution of special-purpose vouchers to parents of school-age children; with these vouchers the parents could buy educational services that would be supplied in competition by private enterprise. In justifying this scheme he says:

> Parents could express their views about schools *directly,* by withdrawing their children from one school and sending them to another, to a much greater extent than is now possible. In general they can now take this step only by changing their place of residence. *For the rest, they can express their views only through cumbrous political channels.*[19]

19. "The Role of Government in Education," in Robert A. Solo, ed., *Economics and the Public Interest* (New Brunswick, N.J.: Rutgers University Press, 1955), p. 129. A revised form of this essay was included in Friedman's *Capitalism and Freedom* (Chicago: University of Chicago Press, 1962) as ch. 6 and the cited passage appears unchanged on p. 91. The italics are mine.

I am not interested here in discussing the merits of the Friedman proposal.[20] Rather, I am citing the above passage as a near perfect example of the economist's bias in favor of exit and against voice. In the first place, Friedman considers withdrawal or exit as the "direct" way of expressing one's unfavorable views of an organization. A person less well trained in economics might naïvely suggest that the direct way of expressing views is to express them! Secondly, the decision to voice one's views and efforts to make them prevail are contemptuously referred to by Friedman as a resort to "cumbrous political channels." But what else is the political, and indeed the democratic, process than the digging, the use, and hopefully the slow improvement of these very channels?

In a whole gamut of human institutions, from the state to the family, voice, however "cumbrous," is all their members normally have to work with. Significantly, one major, if problem-plagued, effort presently underway toward better public schools in the large cities is to make them more responsive to their members: decentralization has been advocated and undertaken as a means of making the channels of communication between members and management in the public school systems less "cumbrous" than heretofore.

But the economist is by no means alone in having a blindspot, a "trained incapacity" (as Veblen called it) for perceiving the usefulness of one of our two mechanisms. In fact, in the political realm exit has fared much worse than has voice in the realm of economics. Rather than as merely ineffective or "cumbrous," exit has often been branded as *criminal,* for it has been labeled desertion, defection, and treason.

Clearly, passions and preconceptions must be reduced

20. For a good discussion see Henry M. Levin, "The Failure of the Public Schools and the Free Market Remedy," *The Urban Review,* 2:32–37 (June 1968).

on both sides if advantage is to be taken of an exceptional opportunity to observe how a typical market mechanism and a typical nonmarket, political mechanism work side by side, possibly in harmony and mutual support, possibly also in such a fashion that one gets into the other's way and undercuts its effectiveness.

A close look at this interplay between market and nonmarket forces will reveal the usefulness of certain tools of economic analysis for the understanding of political phenomena, *and vice versa.* Even more important, the analysis of this interplay will lead to a more complete understanding of social processes than can be afforded by economic or political analysis in isolation. From this point of view, this book can be viewed as the application to a new field of an argument on which much of *The Strategy of Economic Development* was based:

> Tradition seems to require that economists argue forever about the question whether, in any disequilibrium situation, *market forces acting alone* are likely to restore equilibrium. Now this is certainly an interesting question. But as social scientists we surely must address ourselves also to the broader question: is the disequilibrium situation likely to be corrected at all, by market or nonmarket forces, or by both acting jointly? *It is our contention that nonmarket forces are not necessarily less "automatic" than market forces.*[21]

I was concerned here with disturbances of equilibrium and the return to it. Kenneth Arrow has argued along very similar lines for movements from less-than-optimal to optimal states:

> I propose here the view that, when the market fails to achieve an optimal state, society will, to some extent at least, recognize the gap, and nonmarket social institutions

21. Hirschman, *Strategy*, p. 63. Italics in the original.

will arise attempting to bridge it . . . this process is not necessarily conscious.[22]

These views do not imply, as both Arrow and I immediately hastened to add, that any disequilibrium or nonoptimal state whatever will be eliminated by some combination of market and nonmarket forces. Nor do they exclude the possibility that the two sets of forces could work at cross-purposes. But they leave room for a conjunction—which could quite possibly be inadequate—of these two forces, whereas both laissez-faire and interventionist doctrines have looked at market and nonmarket forces in a strictly Manichaean way, it being understood that the laissez-faire advocate's forces of good are the interventionist's forces of evil and vice versa.

A final point. Exit and voice, that is, market and nonmarket forces, that is, economic and political mechanisms, have been introduced as two principal actors of strictly equal rank and importance. In developing my play on that basis I hope to demonstrate to political scientists the usefulness of economic concepts *and to economists the usefulness of political concepts.* This reciprocity has been lacking in recent interdisciplinary work as economists have claimed that concepts developed for the purpose of analyzing phenomena of scarcity and resource allocation can be successfully used for explaining political phenomena as diverse as power, democracy, and nationalism. They have thus succeeded in occupying large portions of the neighboring discipline while political scientists—whose inferiority complex vis-à-vis the tool-rich economist is equaled only by that of the economist vis-à-vis the physicist—have shown themselves quite eager to be colonized and have often actively joined the invaders. Perhaps it

22. "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review,* 53:947 (December 1963).

takes an economist to reawaken feelings of identity and pride among our oppressed colleagues and to give them a sense of confidence that their concepts too have not only *grandeur*, but *rayonnement* as well? I like to think that this could be a by-product of the present essay.

2
Exit

The availability to consumers of the exit option, and their frequent resort to it, are characteristic of "normal" (non-perfect) competition, where the firm has competitors but enjoys some latitude as both price-maker and quality-maker—and therefore, in the latter capacity, also as a quality-spoiler. As already mentioned, the exit option is widely held to be uniquely powerful: by inflicting revenue losses on delinquent management, exit is expected to induce that "wonderful concentration of the mind" akin to the one Samuel Johnson attributed to the prospect of being hanged.

Nevertheless the precise modus operandi of the exit option has not received much attention, to judge from a determined though inevitably fragmentary search of the vast literature on competition.[1] Most authors are content with general references to its "pressures" and "disciplines."

Insofar as the apologetic literature is concerned, this neglect of what could be considered one of the principal virtues of the "free enterprise system" may be particularly surprising; but some of the reasons for it have already been suggested. Those who celebrate the invigorating qualities of competition are loath to concede that the system could fail for even a single moment to make everybody perform at his peak form; should such a failure nevertheless occur in the case of some firm, that firm must *ipso facto* be assumed to be mortally sick and to be ready to leave the stage while some vigorous newcomer is presumably waiting in the wings to take its place. This "view of the American economy . . . as a biological process in which the old and the senile are continually being replaced by the young and the vigorous," as Galbraith puts it mock-

1. Which was carried out by David S. French.

ingly,[2] does not leave room for showing how competition helps to cure the temporary and remediable lapses whose importance is stressed here. It would seem that the apologists of competitive enterprise have missed, in their eagerness to stake extravagant claims for their system, one of the more substantial points to be made in its favor.

The technical economic literature, on the other hand, has been very largely concerned with discussing the conditions under which competitive market structures result or fail to result in an efficient allocation of resources within a static framework. One nonstatic aspect of competition has also been amply, if rather inconclusively, scrutinized, namely, its aptitude to generate innovation and growth. But, as far as I have been able to ascertain, no study, systematic or casual, theoretical or empirical, has been made of the related topic of competition's ability to lead firms back to "normal" efficiency, performance, and growth standards after they have lapsed from them.[3]

How the Exit Option Works

The conceptual elements needed for such an exploration are straightforward. The first one is a variant of the

2. John Kenneth Galbraith, *American Capitalism: The Concept of Countervailing Power* (Boston: Houghton Mifflin Co., 1956), p. 36.

3. John Maurice Clark, who had a most lively sense of the multiplicity of functions competition is expected to perform, does mention that "another thing desired is that competition should keep firms vigilant to eliminate inefficiencies of process or product, before losses have so depleted their resources as to make rehabilitation difficult or impossible." *Competition as a Dynamic Process* (Washington: Brookings Institution, 1961), p. 81. In ch. 4, "What Do We Want Competition to Do for Us?" Clark dealt at some length with what he considered to be the ten principal functions of competition. Strangely, the rescue of faltering firms is not among them; the sentence cited is found, almost as an afterthought, at the end of a section entitled "Elimination of Inefficient Elements," which deals primarily with the "unpleasant services demanded of competition" in seeing to it that faltering firms are liquidated rather than restored to health.

familiar demand function, with the difference that quantity bought is made to depend on changes in quality rather than on price. Just as quality is normally assumed to remain unchanged when the effect of price changes on demand are considered, so it is now convenient to assume that price does not change when quality drops. Costs also remain constant, for by definition the quality decline results from a random lapse in efficiency rather than from a calculated attempt, on the part of the firm, to reduce costs by skimping on quality. Under these conditions, *any* exit whatever of consumers in response to quality decline will result in revenue losses; and, of course, the more massive the exit the greater the losses following upon any given quality drop. Whereas an increase in price can result in an increase in the firm's total revenue in spite of some customer exit, revenue can at best remain unchanged and will normally decline steadily as quality drops.[4]

Secondly, there exists a management reaction function which relates quality improvement to the loss in sales—upon finding out about customer desertion, management undertakes to repair its failings. Perhaps the simplest way to visualize such a relationship is as a discontinuous three-value function. No reaction occurs for a small drop in rev-

4. The response of demand and revenue to quality changes can be graphically represented by means of a demand curve with the familiar downward slope if the vertical axis of the traditional diagram is made to measure quality deterioration rather than price increase. This is done in Appendix A, figure 2, which also shows, in its lower portion, the effect of quality decline on revenue. This diagram makes clear that the effect on total revenue of a decline of demand caused by quality decline is much simpler—and more damaging—than that caused by price rises. In the former case, total revenue declines whenever the *quality*-elasticity of demand is greater than zero, whereas in the case of price increases total revenue falls of course only if *price*-elasticity of demand is greater than unity. (Unit elasticity of demand has no precise meaning in the case of quality-elasticity. When the concept of "quality-elasticity of demand" is put together in analogy to price-elasticity, two different scales—some measure of quality and money—are divided into one another. Hence, any numerical measure other than zero and infinity is the result of arbitrary scaling.)

enue, full recovery follows upon a drop of intermediate size; and, then again, if the revenue decline exceeds a certain large percentage of normal sales volume, no recuperation ensues—beyond a certain point, losses will weaken the firm so badly that bankruptcy will occur before any remedial measures can take effect.[5]

The interaction between the exit function and the reaction function can now be described. If there is to be a drop in quality it is desirable that it be of the size which leads to recuperation. Evidently if demand is highly inelastic with respect to quality change, revenue losses will be quite small and the firm will not get the message that something is amiss. But if demand is very elastic, the recuperation process will not take place either, this time because the firm will be wiped out before it will have had time to find out what hit it, much less to do something about it. This is a case of "too much, too soon." For the recuperation potential of the firm to come into play, it is therefore desirable that quality elasticity of demand be neither very large nor very small. This proposition, which is intuitively evident, can also be phrased as follows: For competition (exit) to work as a mechanism of recuperation from performance lapses, it is generally best for a firm to have a mixture of *alert* and *inert* customers. The alert customers provide the firm with a feedback mechanism which starts the effort at recuperation while the inert customers provide it with the time and dollar cushion needed for this effort to come to fruition. According to traditional notions,

5. It would be easy to think instead of a continuous reaction curve. Remedial action would be small with small sales losses and would then increase and later decline. It is even conceivable that, as a result of the reaction, the firm would come to produce at qualities superior to the ones at which it started out—to that extent one might speak of a point of "optimal deterioration" in quality. At a later point, beyond a certain loss in sales, the reaction would turn into reinforcement as demoralization and other results of financial stringency would compound quality deterioration and thus hasten the firm's downfall. Such a shape of the reaction function would not change materially the points that will be made in the text.

of course, the more alert the customers the better for the functioning of competitive markets. Consideration of competition as a recuperation mechanism reveals that, although exit of some customers is essential for bringing the mechanism into play, it is important that other customers remain unaware of, or unperturbed by, quality decline: if all were assiduous readers of *Consumer Reports,* or determined comparison shoppers, disastrous instability might result and firms would miss out on chances to recover from their occasional lapses.

As has already been noted, in perfect competition (which includes perfect consumer knowledge as one of its many exacting assumptions) the firm is not deprived of an effective correction mechanism because performance deterioration, which cannot possibly affect either quality or price, is reflected directly in a decline in revenue (due to increasing costs). But assume now a small departure from the perfectly competitive model so that the firm has some latitude in varying quality; then performance deterioration *can* (and is perhaps likely to) take the form of quality decline and if the market in which the firm sells is highly competitive, that is, full of highly knowledgeable buyers, the firm will be competed out of existence in very short order. In other words, while the perfectly competitive world is a feasible one from the point of view of an effective recuperation mechanism, the world of quasi-perfect competition is not. If one gives up, as he must in most real cases, the concept of a firm with no latitude as to quality whatever, then the optimal arrangement is not one as close as possible to that of perfect competition, but one rather far removed from it; and incremental moves in the direction of perfect competition are not necessarily improvements—the argument of the second best applies here in full force.

Competition as Collusive Behavior

No matter what the quality elasticity of demand, exit could fail to cause any revenue loss to the individual firms *if the firm acquired new customers as it loses the old ones.* But why would a firm whose output deteriorates in quality attract any new customers at all? One can actually think of a situation in which this seemingly quite unlikely event would come to pass: when a uniform quality decline hits simultaneously all firms of an industry, each firm would garner in some of the disgruntled customers of the other firms while losing some of its previous customers to its competitors. In these circumstances the exit option is ineffective in alerting management to its failings, and a merger of all firms would appear to be socially desirable —that is, monopoly would replace competition to advantage, for customer dissatisfaction would then be vented directly and perhaps to some effect in attempts at improving the monopoly's management whereas under competition dissatisfaction takes the form of ineffective flitting back and forth of groups of consumers from one deteriorating firm to another without any firm getting a signal that something has gone awry.

While a simultaneous and uniform deterioration of all firms in a certain type of business is of course highly unlikely, a slight modification of the previous situation serves to endow it with greater realism and relevance. A competitively produced new product might reveal only through use some of its faults and noxious side-effects. In this case the claims of the various competing producers are likely to make for prolonged experimenting of consumers with alternate brands, all equally faulty, and hence for delay in bringing pressure on manufacturers for effective improvements in the product. Competition in this situation is a considerable convenience to the manufacturers

because it keeps consumers from complaining; it diverts their energy to the hunting for the inexistent improved products that might possibly have been turned out by the competition. Under these circumstances, the manufacturers have a common interest in the maintenance rather than in the abridgement of competition—and may conceivably resort to collusive behavior to that end.[6]

The argument presented so far maintained the premise that the unsatisfactory features of the product turned out by the various competing firms could be eliminated as a result of pressures and a resultant search for solutions. But even if this premise is dropped, the competitive solution may again be inferior to one in which a single firm is the sole producer. For the presence of a number of competing firms fosters in this case the perpetual illusion that "the grass is always greener on the other side of the fence," that is, that an escape from defectiveness is possible through purchase of the competitor's product. Under monopoly, consumers would learn to live with inevitable imperfection and would seek happiness elsewhere than in the frantic search for the inexistent "improved" product.

The reader can judge whether elements of the foregoing situations can be detected in the economic and commercial life around us.[7] A few comments may be in order, however,

6. This is even more the case if the most determined comparison shoppers are those who would make most trouble for the manufacturers if there were no possibility of exit. The competitive mechanism then rids management of its potentially most troublesome customers. This argument is explained more fully below.

7. To help him judge I should like to provide some sample passages from letters recently fired off, by irate owners of "lemons," (*a*) to the Ford Motor Company: ". . . You can be assured that I absolutely will not purchase another Ford of any kind no matter what your usual form letter to me will say . . ." ". . . Needless to say my Falcon is the last of any Ford product I would consider to purchase. I am a young girl of 25, reasonably attractive, who has depleted her bank account buying Falcon transmissions, when there are other things in this world where the money could be put to much better use . . ."; and (*b*) to the General Motors Corporation: ". . . At home we have a Chevrolet bus and a Chevrolet van. You may be

on the relevance of the preceding notions for organizations other than business firms. The basic point is that competition may result merely in the mutual luring over of each others' customers on the part of a group of competing firms; and that to this extent competition and product diversification is wasteful and diversionary especially when, in its absence, consumers would either be able to bring more effective pressures upon management toward product improvement or would stop using up their energies in a futile search for the "ideal" product. It will be immediately evident that competitive political systems have frequently been portrayed in just these terms. Radical critics of societies with stable party systems have often denounced the competition of the dominant parties as offering "no real choice." It is of course a very open question whether, in the absence of the competitive party system, citizens would be better able to achieve fundamental social and political changes (assuming, for the sake of the argument, that such changes are desirable). Nevertheless the radical critique is correct in pointing out that competitive political systems have a considerable capacity to divert what might otherwise be a revolutionary ground swell into tame discontent with the governing party. Although this capacity may normally be an asset, one can surely conceive of circumstances under which it would turn into a liability.

A less speculative illustration of the issue under discussion can be drawn from the history of the trade union movement in this country. A preliminary step to the CIO-

---

sure that after all of this trouble and inconvenience and wasted time I shall never own a General Motors product again . . ." ". . . I have had a G.M. auto and wagon for many years now but maybe FORD has a better idea. I'll try to put up with this LEMON till the '70 models come out, but you can be sure there will be no G.M. product of any kind on my driveway . . ."

Copies of the letters from which these excerpts are taken were mailed by their authors to Ralph Nader who has kindly made them available to me.

AFL merger of 1955 was the No-Raiding Agreement which was concluded between the two organizations two years earlier. The text of this agreement referred to a statistical study of all petitions over a two-year period, addressed by CIO-AFL unions to the National Labor Relations Board for certification as the official bargaining agents in industrial plants. It was found that most petitions were unsuccessful and that those which were granted were about equally divided between CIO petitions to displace an AFL union and AFL petitions to displace a CIO union. These results, so the report says, "compel the conclusion that raids between AFL and CIO unions are destructive of the best interests of the unions immediately involved and also of the entire trade union movement." [8] As reasons for this conclusion, the document cites unrest and disunity created among the workers as a result of the raids, successful or not, and the desirability of devoting the energies of the trade union movement to the organization of unaffiliated workers rather than to raiding. Implicit in this conclusion is the judgment that the disadvantages of exit-competition outweighed in this case its possible efficiency-inducing advantages and *perhaps* the assumption that these advantages can be better secured via the alternative mechanism—voice—which must now be examined more closely.

8. American Federation of Labor and Congress of Industrial Organizations, *Constitution of the AFL-CIO* (Washington, D.C., January 1956), AFL-CIO Publication no. 2, p. 36. I am indebted to John Dunlop for the reference and for discussing this point with me.

3
Voice

If the exit option has not been investigated in detail by economists, its existence and effect on performance—generally presumed to be wholesome—nevertheless underlies many judgments and attitudes toward economic institutions. Nothing remotely similar can be said about the voice option. The very idea that this is another "recuperation mechanism" which can come into play alongside, or in lieu of, the exit option is likely to be met with a mixture of incredulity and raised eyebrows. Yet, in this age of protest, it has become quite apparent that dissatisfied consumers (or members of an organization), rather than just go over to the competition, can "kick up a fuss" and thereby force improved quality or service upon delinquent management. It is therefore both legitimate and timely to examine the conditions under which the voice option is likely to make an effective appearance, either as a complement to exit or as a substitute for it.

To resort to voice, rather than exit, is for the customer or member to make an attempt at changing the practices, policies, and outputs of the firm from which one buys or of the organization to which one belongs. Voice is here defined as any attempt at all to change, rather than to escape from, an objectionable state of affairs, whether through individual or collective petition to the management directly in charge, through appeal to a higher authority with the intention of forcing a change in management, or through various types of actions and protests, including those that are meant to mobilize public opinion.

It is becoming clear, as was already pointed out in the introductory chapter, that voice is nothing but a basic portion and function of any political system, known sometimes also as "interest articulation." [1] Political scientists

1. For a recent treatment in a comparative perspective see G. A. Almond and G. B. Powell, Jr., *Comparative Politics: A Developmental Approach* (Boston: Little, Brown and Co., 1966), ch. 4.

have long dealt systematically with this function and its various manifestations. But in doing so they have ordinarily confined their attention to situations in which the only alternative to articulation is aquiescence or indifference (rather than exit), while economists have refused to consider that the discontented consumer might be anything but either dumbly faithful or outright traitorous (to the firm he used to do business with). A niche thus exists for this book, which affirms that the choice is often between articulation and "desertion"—voice and exit, in our neutral terminology.

First a few remarks on the working of voice in isolation, as compared to that of exit. As before, the initial assumption is a decline in the performance of a firm or organization which is remediable provided the attention of management is sufficiently focused on the task. If conditions are such that the decline leads to voice rather than to exit on the part of the discontented member-customers, then the effectiveness of voice will increase, up to a certain point, with its volume. But voice is like exit in that it can be overdone: the discontented customers or members could become so harassing that their protests would at some point hinder rather than help whatever efforts at recovery are undertaken. For reasons that will become clear this is most unlikely to happen in relations between customers and business firms; but in the realm of politics—the more characteristic province of voice—the possibility of negative returns to voice making their appearance at some point is by no means to be excluded.

An interesting parallel appears here between economics and exit, on the one hand, and politics and voice, on the other. Just as in economics it had long been thought that the more elastic demand is (that is, the more rapidly exit ensues whenever deterioration occurs) the better for the functioning of the economic system, so it has long been an article of faith of political theory that the proper function-

ing of democracy requires a maximally alert, active, and vocal public. In the United States, this belief was shaken by empirical studies of voting and political behavior which demonstrated the existence of considerable political apathy on the part of large sections of the public, for long periods of time.[2] Since the democratic system appeared to survive this apathy rather well, it became clear that the relations between political activism of the citizens and stable democracy are considerably more complex than had once been thought. As in the case of exit, a mixture of alert and inert citizens, or even an alternation of involvement and withdrawal, may actually serve democracy better than either total, permanent activism or total apathy. One reason, stressed by Robert Dahl, is that the ordinary failure, on the part of most citizens, to use their potential political resources to the full makes it possible for them to react with unexpected vigor—by using normally unused reserves of political power and influence—whenever their vital interests are directly threatened.[3] According to another line of reasoning, the democratic political system requires "blending of apparent contradictions": on the one hand, the citizen must express his point of view so that the political elites know and can be responsive to what he wants, but, on the other, these elites must be allowed to make decisions. The citizen must thus be in turn influential and deferential.[4]

2. See Robert A. Dahl, *Modern Political Analysis* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1966), ch. 6 for data and principal sources.

3. Robert A. Dahl, *Who Governs?* (New Haven: Yale University Press, 1961), pp. 309–310. This point is remarkably similar to the one made by March and Cyert about the virtues of "organizational slack" in the economic system. See *The Behavioral Theory of the Firm* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1963), pp. 36–38.

4. Gabriel A. Almond and Sidney Verba, *The Civic Culture: Political Attitudes and Democracy in Five Nations* (Boston: Little, Brown and Co., 1965), pp. 338–344. A similar thought is expressed by Robert Lane who shows that, in certain respects, "one can assign different political roles to the political activists and the indifferents and that a balance between the two can achieve beneficient results." *Political Life* (New York: Free Press of Glencoe, Inc., 1959), p. 345.

The essential reasoning behind this thesis is quite similar to the argument made earlier on the need for exit to stay within certain bounds. Voice has the function of alerting a firm or organization to its failings, but it must then give management, old or new, some time to respond to the pressures that have been brought to bear on it.

Finally, then, the relation between voice and improvement in an organization's efficiency has considerable similarity with the modus operandi of exit. This does not mean, however, that exit and voice will always both have positive effects at first and destructive ones at a later stage. In the case of any one particular firm or organization and its deterioration, either exit or voice will ordinarily have the role of the *dominant* reaction mode. The subsidiary mode is then likely to show up in such limited volume that it will never become destructive for the simple reason that, if deterioration proceeds, the job of destruction is accomplished single-handedly by the dominant mode. In the case of normally competitive business firms, for example, exit is clearly the dominant reaction to deterioration and voice is a badly underdeveloped mechanism; it is difficult to conceive of a situation in which there would be too much of it.

Voice as a Residual of Exit

The voice option is the only way in which dissatisfied customers or members can react whenever the exit option is unavailable. This is very nearly the situation in such basic social organizations as the family, the state, or the church. In the economic sphere, the theoretical construct of pure monopoly would spell a no-exit situation, but the mixture of monopolistic and competitive elements characteristic of most real market situations should make it possible to observe the voice option in its interaction with the exit option.

We return to the simple relationship between deterioration of a product and declining sales, but look now at those who continue as customers. While they are not yet ready to desert the firm, they are likely to experience different degrees of unhappiness about the quality decline. Being presumably endowed with some capacity to articulate this discontent, these nonexiting customers are therefore the source of the voice option. The other determinant of voice is of course the degree of discontent of the nonexiting customer which depends roughly on the degree of deterioration.

In a first approximation, then, voice can be viewed as a residual. Whoever does not exit is a candidate for voice and voice depends, like exit, on the quality elasticity of demand. But the direction of the relationship is turned around: with a given potential for articulation, the actual level of voice feeds on *in*elastic demand, or on the lack of opportunity for exit.[5]

In this view, the role of voice would increase as the opportunities for exit decline, up to the point where, with exit wholly unavailable, voice must carry the entire burden of alerting management to its failings. That such a see-saw relationship between exit and voice exists in fact to some extent is illustrated by the many complaints about quality and service that have been prominently published for years in the Soviet press. With exit-competition playing a much smaller role in the Soviet economy than in the market economies of the West, it was found necessary to give voice a more prominent role.

Similarly, voice is in a much more commanding position in less developed countries where one simply cannot choose between as many commodities, nor between as many varieties of the same good, nor between as many ways of traveling from one point of the country to another, as in an

5. The relationship between the volumes of exit and voice that is indicated here is spelled out in more formal terms in Appendix A.

advanced economy. Therefore, the atmosphere in the former countries is more suffused with loud, often politically colored protests against poor quality of goods or services than it is in the advanced countries where dissatisfaction is more likely to take the form of silent exit.

Turning now to the reaction function, that is, to the effect of voice on recuperation of efficiency on the part of voice-exposed management, we shall assume that exit is the dominant reaction mode. In a preliminary appraisal of the combined effect of exit and voice, the possibility of voice having a destructive rather than constructive effect may therefore be excluded. Obviously sales losses and complaints or protests of those who remain members are not easily added to derive an aggregate recuperative effect.[6] Both the propensity to protest and the effectiveness of complaints vary widely from one firm-customer complex to another. But three general statements can be made:

(1) In the simple model presented up to now, voice functions as a complement to exit, not as a substitute for it. Whatever voice is forthcoming under those conditions is a net gain from the point of view of the recuperation mechanism.[7]

6. Voice may cause direct monetary losses to the firm, as, for example, when dissatisfied consumers are able to turn in defective merchandise. If voice appears exclusively in this particular incarnation, then its likely effectiveness in making an impression on profit-conscious managers can be precisely measured against that of exit. See Appendix A.

7. Voice could usefully complement competition also in a more familiar context. Economists who have hopefully eyed competition's ability to allocate resources efficiently have generally concluded that the most serious impediment to the hope's fulfillment is the existence of external diseconomies in production and consumption (pollution, littering of beaches with beer cans, and so forth). Obviously, these diseconomies could be contained or prevented through effective articulation of protests on the part of those who suffer from them. In other words, the voice of the nonconsumer on whom the diseconomies are inflicted could become a valuable adjunct to the competitive mechanism. Once this is realized it is perhaps less surprising that the voice of the *consumer* too has a role to play in complementing the mechanism.

(2) The more effective voice is (the effectiveness of exit being given), the more quality-inelastic can demand be without the chances for recuperation stemming from exit and voice *combined* being impaired.

(3) Considering that beyond a certain point, exit has a destructive rather than salutary effect, the optimal pattern from the point of view of maximizing the combined effectiveness of exit and voice over the whole process of deterioration may be an elastic response of demand to the first stages of deterioration and an inelastic one for the later stages. This pattern has long been held to be characteristic of consumer responses to price increases for certain commodities which are vitally needed in limited quantities even at high prices, but whose consumption will easily expand beyond this point if prices drop. It may similarly apply to quality elasticity of demand, especially if the only alternative available for a deteriorating product is a higher-priced substitute. Eventually, of course, as quality becomes abominable, demand will vanish (just as it does, because of the budget constraint, when price increases indefinitely), but there may well be a number of goods and services whose demand will move from quality-elastic to quality-inelastic for a wide range of quality declines. The reason for which even such a pattern may be too much weighted by exit will be commented on at some length in Chapter 4.

Voice as an Alternative to Exit[8]

Up to now, the treatment of voice has suffered from a certain timidity: the new concept has been viewed as wholly subordinated to exit. In judging the volume of voice to be determined by the quality elasticity of demand, one implicitly assumes that customers who are faced by

8. See Appendix B for a more technical discussion of the topics treated in this section.

a decline in quality first decide whether to shift to another firm or product regardless of their ability to influence the behavior of the firm from which they usually buy; only if they do not shift, does it possibly occur to them to make a fuss. If the matter is put in this way it is immediately evident, however, that the decision whether to exit will often be taken *in the light of the prospects for the effective use of voice.* If customers are sufficiently convinced that voice will be effective, then they may well *postpone* exit. Hence, quality-elasticity of demand, and therefore exit, can also be viewed as depending on the ability and willingness of the customers to take up the voice option. It may, in fact, be more appropriate to put matters in this way, for if deterioration is a process unfolding in stages over a period of time, the voice option is more likely to be taken at an early stage. Once you have exited, you have lost the opportunity to use voice, but not vice versa; in some situations, exit will therefore be a reaction of *last resort* after voice has failed.

It appears, therefore, that voice can be a substitute for exit, as well as a complement to it. What are the conditions, then, under which voice will be preferred to exit? The question can be formulated more precisely as follows: If a competing or substitute product B is available at the same price as the normally bought product A and if, because of the deterioration of A, B is now clearly superior from the point of view of A's customers, under what conditions will a customer of A *fail* to go over to B?

Once voice is viewed as a substitute for exit, an important component of the voice option consists in this decision to continue as a customer of the deteriorating and now inferior product (or as a member of the deteriorating organization), for it will presumably be taken only by those who wish for and expect A to recover its original superiority over B, and not necessarily by all of them. Ordinarily, a customer or member will undergo the sac-

rifice of staying with *A* because he feels that he wants and is able to "do something" about *A* and because only by remaining a customer or member will he be able to exert this influence. Nevertheless, the decision not to exit in the face of a clearly better buy (or organization) could also be taken by customers (or members) who expect the complaints and protests of *others,* combined with their own faithfulness, to be successful. Others may not care to switch to *B* when they feel that they would soon want to switch back, because of the costs that may be involved. Finally there are those who stay with *A* out of "loyalty," that is, in a less rational, though far from wholly irrational, fashion.[9] Many of these "loyalists" will actively participate in actions designed to change *A*'s policies and practices, but some may simply refuse to exit and suffer in silence, confident that things will soon get better. Thus the voice option includes vastly different degrees of activity and leadership in the attempt to achieve change "from within." But it always involves the decision to "stick" with the deteriorating firm or organization and this decision is in turn based on:

(1) an evaluation of the chances of getting the firm or organization producing *A* "back on the track," through one's own action or through that of others; and

(2) a judgment that it is worthwhile, for a variety of reasons, to trade the certainty of *B* which is available here and now against these chances.

This view of the matter shows the substitutability of *B* for *A* as an important element in the decision to resort to voice, but as only one of several elements. Naturally, the consumer will resort to voice if *A*'s original margin of superiority over *B* was wide enough to make it worthwhile for him to forego a *B* that is superior right here and now. That will hardly ever be the case if *A* and *B* are very close substitutes. But given a minimum of nonsubstituta-

9. See ch. 7, below.

bility, voice will depend also on the willingness to take the chances of the voice option as against the certainty of the exit option and on the probability with which a consumer expects improvements to occur as a result of actions to be taken by himself or by others with him or just by others.

It is useful to compare this formulation with the related one provided by Edward Banfield in his study of political influence: "The effort an interested party makes to put its case before the decisionmaker will be in proportion *to the advantage to be gained from a favorable outcome multiplied by the probability of influencing the decision.*[10]

Banfield derived this rule from his study of public policy decisions in a large American city and of the participation of various groups and individuals in the decisionmaking process. He, like most political scientists looking at the "articulation-of-interests" function, was analyzing situations in which individuals or groups had essentially the choice between passivity and involvement. The present model is more complicated because it allows for exit, as a result of the availability of a substitute product. Banfield's formulation correctly states the benefits of the voice option,[11] but for our purposes there is need to introduce cost which so far has been identified as the foregoing of the exit option. In fact, in addition to this opportunity cost, account must be taken of the direct cost of voice which is incurred as buyers of a product or members of an organization spend time and money in the attempt to achieve changes in the policies and practices of the firm from which they buy or of the organization to which they belong. Not nearly so high a cost is likely to be attached to the exercise of the exit option in the case of products

10. Edward C. Banfield, *Political Influence* (New York: Free Press of Glencoe, 1961), p. 333. Italics in the original.

11. It should be noted that our concept of voice, as defined at the beginning of this chapter, is much wider than Banfield's "influence," which appears to exclude any expression of opinion or discontent that is not addressed directly to the officeholding decisionmaker.

bought in the market—although some allowance should be made for the possible loss of loyalty discounts and for the cost of obtaining information about substitute products to which one intends to switch.[12]

Hence, in comparison to the exit option, voice is costly and conditioned on the influence and bargaining power customers and members can bring to bear within the firm from which they buy or the organizations to which they belong. These two characteristics point to roughly similar areas of economic and social life in which voice is likely to play an important role and to hold exit at bay, at least for a time. As voice tends to be costly in comparison to exit, the consumer will become less able to afford voice as the number of goods and services over which he spreads his purchases increases—the cost of devoting even a modicum of his time to correcting the faults of any one of the entities he is involved with is likely to exceed his estimate of the expected benefits for a large number of them. This is also one of the reasons for which voice plays a more important role with respect to *organizations* of which an individual is a member than with respect to *firms* whose products he buys: the former are far less numerous than the latter. In addition, of course, the proliferation of products tends to increase cross-elasticities of demand and to that extent it would increase the probability of exit for a given deterioration in quality of any one product picked at random. For these reasons, voice is likely to be an active mechanism primarily with respect to the more substantial purchases and organizations in which buyers and members are involved.

Similar conclusions with respect to the *locus* of the voice option are reached when one focuses on the other characteristic which distinguishes voice from exit, namely, the requirement that a customer must expect that he himself

12. When loyalty is present, however, the cost of exit may be substantial. The point will be discussed in ch. 7, below.

or other member-customers will be able to marshal some influence or bargaining power. Obviously, this is not the case in atomistic markets. Voice is most likely to function as an important mechanism in markets with few buyers or where a few buyers account for an important proportion of total sales, both because it is easier for few buyers than for many to combine for collective action and simply because each one may have much at stake and wield considerable power even in isolation.[13] Again, it is more common to encounter influential members of an organization than buyers with a great deal of influence on the policies of firms from which they buy,[14] and the voice option will therefore be observed more frequently among organizations than among business firms.

Certain types of purchases may nevertheless lend themselves particularly to the voice option, even though many buyers are involved. When the consumer has been dissatisfied with an inexpensive, nondurable good, he will most probably go over to a different variety without making a fuss. But if he is stuck with an expensive durable good such as an automobile which disappoints him day-in and day-out, he is much less likely to remain silent. And his complaints will be of some concern to the firm or dealer whose product he has bought both because he remains a potential customer in one, three, or five years' time and because adverse word-of-mouth propaganda is powerful in the case of standardized goods.

The upshot of this discussion for the comparative roles of voice and exit at various stages of economic development is two-edged: the sheer number of available goods and varieties in an advanced economy favors exit over voice, but the increasing importance in such an economy

13. See Mancur Olson, Jr., *The Logic of Collective Action* (Cambridge, Mass.: Harvard University Press, 1965).

14. See, however, the description of the influential buyer in John Kenneth Galbraith, *American Capitalism: The Concept of Countervailing Power* (Boston: Houghton Mifflin Co., 1956), pp. 117–123.

of standardized durable consumer goods requiring large outlays works in the opposite direction.

Although the foregoing remarks restrict the domain in which the voice option is likely to be deployed, especially as a substitute for exit, the territory left to it remains both considerable and somewhat ill-defined. Moreover, once voice is recognized as a mechanism with considerable usefulness for maintaining performance, institutions can be designed in such a way that the cost of individual and collective action would be decreased. Or, in some situations, the rewards for *successful* action might be increased for those who had initiated it.

Often it is possible to create entirely new channels of communication for groups, such as consumers, which have had notorious difficulties in making their voice heard, in comparison to other interest groups. Consumers have, in fact, made such progress in this regard that there is now talk of a "consumer revolution" as part of the general "participation explosion." The former phrase does not refer to the long established and still quite useful consumer research organizations, but to the more militant actions by or on behalf of consumers that have been taken recently, the most spectacular and resourceful being the campaigns of Ralph Nader, who has established himself as a sort of self-appointed consumer ombudsman.[15] The appointment since 1964 of a consumer adviser to the President has been a response to this emergence of the consumer voice which was quite unexpected in an economy where competition-exit is supposed to solve most of the "sovereign" consumer's problems. As a result of these developments, it looks as though consumer voice will be institutionalized at three levels: through independent entrepreneurship à la Nader, through revitalization of

15. The broad range of Nader's work, with respect to both products and action, is brought out in his article "The Great American Gyp," *The New York Review of Books*, November 21, 1968.

official regulatory agencies, and through stepped-up preventive activities on the part of the more important firms selling to the public.[16]

The creation of effective new channels through which consumers can communicate their dissatisfaction holds one important lesson. While structural constraints (availability of close substitutes, number of buyers, durability and standardization of the article, and so forth) are of undoubted importance in determining the balance of exit and voice for individual commodities, the propensity to resort to the voice option depends also on the general readiness of a population to complain and on the *invention* of such institutions and mechanisms as can communicate complaints cheaply and effectively. Recent experience even raises some doubts whether the structural constraints deserve to be called "basic" when they can suddenly be overcome by a single individual such as Ralph Nader.[17]

Thus, while exit requires nothing but a clearcut either-or decision, voice is essentially an *art* constantly evolving in new directions. This situation makes for an important bias in favor of exit when both options are present: customer-members will ordinarily base their decision on *past* experience with the cost and effectiveness of voice even though the possible *discovery* of lower cost and greater effectiveness is of the very essence of voice. The presence of the exit alternative can therefore tend to *atrophy the development of the art of voice.* This is a central point of this book which will be argued from a different angle in the next chapter.

16. Traditionally such firms have been engaged in considerable "auscultation" of voice through market surveys.

17. For another, most vivid case in point, within the context of community action in Venezuela, see Lisa Redfield Peattie, *The View from the Barrio* (Ann Arbor, Mich.: University of Michigan Press, 1968), ch. 7; the "art" of eliciting voice, this time in low-income neighborhoods of American cities, is also the subject of her article "Reflections on Advocacy Planning," *Journal of the American Institute of Planners* (March 1968), pp. 80–88.

4
A Special Difficulty in Combining Exit and Voice

The groundwork has now been laid for telling the reader about the empirical observation that was mentioned in the Preface as the origin of this essay. In a recent book, I tried to explain why the Nigerian railways had performed so poorly in the face of competition from trucks, even for such long-haul, bulky cargo as peanuts (which are grown in Northern Nigeria, some eight hundred miles from the ports of Lagos and Port d'Harcourt). Specific economic, socio-political, and organizational reasons could be found for the exceptional ability of the trucks to get the better of the railroads in the Nigerian environment; but having done so I still had to account for the prolonged incapacity of the railroad administration to correct some of its more glaring inefficiencies, *in spite of active competition,* and proposed the following explanation:

> The presence of a ready alternative to rail transport makes it less, rather than more, likely that the weaknesses of the railways will be fought rather than indulged. With truck and bus transportation available, a deterioration in rail service is not nearly so serious a matter as if the railways held a monopoly for long-distance transport—it can be lived with for a long time without arousing strong public pressures for the basic and politically difficult or even explosive reforms in administration and management that would be required. This may be the reason public enterprise, not only in Nigeria but in many other countries, has strangely been at its weakest in sectors such as transportation and education where it is subjected to competition: instead of stimulating improved or top performance, the presence of a ready and satisfactory substitute for the services public enterprise offers merely deprives it of a precious feedback mechanism that operates at its best when the customers are securely locked in. For the management

of public enterprise, always fairly confident that it will not be let down by the national treasury, may be less sensitive to the loss of revenue due to the switch of customers to a competing mode than to the protests of an aroused public that has a vital stake in the service, has no alternative, and will therefore "raise hell." [1]

In Nigeria, then, I had encountered a situation where the combination of exit and voice was particularly noxious for any recovery: exit did not have its usual attention-focusing effect because the loss of revenue was not a matter of the utmost gravity for management, while voice did not work as long as the most aroused and therefore the potentially most vocal customers were the first ones to abandon the railroads for the trucks. It is particularly this last phenomenon that must be looked at more closely, for if it has any generality, then the chances that voice will ever act in conjunction with exit would be poor and voice would be an effective recuperation mechanism only in conditions of full monopoly "when the customers are securely locked in."

As a preliminary to generalizing about this sort of situation, another example, closer to home, may be helpful. If public and private schools somewhere in the United States are substituted in the story for the railroads and lorries of Nigeria, a rather similar result follows. Suppose at some point, for whatever reason, the public schools deteriorate. Thereupon, increasing numbers of quality-education-conscious parents will send their children to private schools.[2] This "exit" may occasion some impulse toward an improvement of the public schools; but here again this im-

1. *Development Projects Observed* (Washington: Brookings Institution, 1967), pp. 146–147.

2. Private schools being costly and income distribution unequal, the public schools will of course be deserted primarily by the wealthier parents. Nevertheless, the willingness to make a financial sacrifice for the sake of improving the children's education differs widely within a given income class, especially at intermediate levels of income. In its pure form, the phenomenon here described is best visualized for a school district with many middle-class parents for whom

pulse is far less significant than the loss to the public schools of those member-customers who would be most motivated and determined to put up a fight against the deterioration if they did not have the alternative of the private schools.

In the preceding examples, insensitivity to exit is exhibited by public agencies that can draw on a variety of financial resources outside and independent of sales revenue. But situations in which exit is the predominant reaction to decline while voice might be more efficacious in arresting it can also be observed in the sphere of private business enterprise. The relation between corporate management and the stockholders is a case in point. When the management of a corporation deteriorates, the first reaction of the best-informed stockholders is to look around for the stock of better-managed companies. In thus orienting themselves toward exit, rather than toward voice, investors are said to follow the Wall Street rule that "if you do not like the management you should sell your stock." According to a well-known manual this rule "results in perpetuating bad management and bad policies." Naturally it is not so much the Wall Street rule that is at fault as the ready availability of alternative investment opportunities in the stock market which makes any resort to voice rather than to exit unthinkable for any but the most committed stockholder.[3]

---

the decision to send the children to private school is a significant, yet tolerable burden.

3. The passages in quotes are from B. Graham and D. L. Dodd, *Security Analysis*, 3d ed. (New York: McGraw-Hill, 1951), p. 616. The argument is spelled out in some detail in ch. 50, "Stockholder-Management Controversies." In the fourth edition of this work (1962), the authors return only briefly to this argument, and seem to be aware that the institutional odds are heavily stacked against any substantial success of their exhortations: "In quixotic fashion perhaps," they say wistfully, "we wanted to combat the traditional but harmful notion that if a stockholder doesn't like the way his company is run he should sell his shares, no matter how low their price may be" (p. 674).

While it is most clearly revealed in the private-public school case, one characteristic is crucial in all of the foregoing situations: those customers who care *most* about the quality of the product and who, therefore, are those who would be the most active, reliable, and creative agents of voice are for that very reason also those who are apparently likely to exit first in case of deterioration.

One interest of this observation is that it could define a whole class of economic structures where a tight monopoly would be preferable, within the framework of the "slack" or "fallible" economy, to competition. But before jumping to this conclusion, we must take a closer look at the observation by translating it into the ordinary language of economic analysis.

In terms of that language, the situations just described have more than a faint odor of paradox. We all know that when the price of a commodity goes up, it is the *marginal* customer, the one with the smallest consumer surplus, the one, that is, who cares *least,* who drops out first. How is it then that with a decline in quality the opposite seems quite plausible: *Is it possible that the consumers who drop out first as price increases are not the same as those who exit first when quality declines?* [4] If this question were to be answered in the affirmative, it would be easier to understand why combining exit and voice is so troublesome in some situations.

The basic reason for our paradox lies in the still insufficiently explored role of quality (as contrasted with price) in economic life. Traditional demand analysis is overwhelmingly in terms of price and quantity, categories which have the immense advantage of being recorded, measurable, and finely divisible. Quality changes have usually been dealt with by economists and statisticians

4. Appendix C refers to this possibility as the "reversal phenomenon." The discussion in the following pages should be read in conjunction with Appendixes C and D by those who find diagrams clearer than language.

through the concept of the *equivalent* price or quantity change. An article of poor quality can often be considered to be simply less in quantity than the same article of standard quality; this is the case, for example, of the automobile tire which lasts on the average only half as long (in terms of mileage) as a high quality tire. Alternatively, poor quality can often be translated into higher costs and prices; for example, increased pilferage in the rendering of railroad freight service will result in higher insurance premiums. In the latter case, a large part of the quality deterioration can be described by the statement: "now everybody really pays more for the same railroad service than before." To the extent that this statement is correct, there would be no reason to expect the effect of quality deterioration on demand (that is, for who gets out first) to be any different from the effect of a uniform rise in price. In other words, if a quality decline can be fully expressed as an equivalent rise in price that is *uniform for all buyers* of the article, the effects on customer exit of the quality decline and of the equivalent rise in price would be identical.

The crucial point can now be made. For any one individual, a quality change can be translated into equivalent price change. But this equivalence *is frequently different for different customers of the article because appreciation of quality differs widely among them.* This is so to some extent even in the just mentioned case of automobile tires and of increased pilferage of freight sent by rail. Appreciation of the longer life of quality tires will depend on the time discount of each individual buyer. In the case of rail freight, the increase in the insurance premium fully offsets only the increase in average direct monetary costs which is occasioned to the shipper by the worsening in service. For some shippers this may be all they care about, but there will surely be others for whom the lessened reliability of rail service represents costs (in inconvenience,

reputation of their own reliability, and so forth) that cannot be fully made good through an insurance scheme. That appreciation of quality—of wine, cheese, or of education for one's children—differs widely among different groups of people is surely no great discovery. It implies, however, that a given deterioration in quality will inflict very different losses (that is, different equivalent price increases) on different customers; someone who had a very high consumer surplus before deterioration precisely because he is a connoisseur and would be willing to pay, say, twice the actual price of the article at its original quality, may drop out as a customer as soon as quality deteriorates, provided a nondeteriorated competing product is available, be it at a much higher price.

Here, then, is the rationale for our observation: in the case of "connoisseur goods"—and, as the example of education indicates, this category is by no means limited to quality wines—the consumers who drop out when quality declines are not necessarily the marginal consumers who would drop out if price increased, but may be intramarginal consumers with considerable consumer surplus; or, put more simply, the consumer who is rather insensitive to price increases is often likely to be highly sensitive to quality declines.

At the same time, consumers with a high consumer surplus are, for that very reason, those who have most to lose through a deterioration of the product's quality. Therefore, they are the ones who are most likely to make a fuss in case of deterioration until such time as they do exit. "You can actively flee, then, and you can actively stay put." This phrase of Erik Erikson [5] applies with full force to the choice that is typically made by the quality-conscious consumer or the member who cares deeply about the policies pursued by the organization to which he be-

5. *Insight and Responsibility* (New York: W. W. Norton & Co., Inc., 1964), p. 86.

longs. To make that kind of consumer and member "actively stay put" for a while should be a matter of considerable concern for many firms and organizations, and particularly for those, of course, that respond more readily to voice than to exit.

Before the varieties of consumer behavior in the case of connoisseur goods are further explored, a brief homage to the hoary concept of consumer surplus is in order, for it appears to have the useful property of measuring the potential for the exercise of influence on the part of different consumers. This potential is the counterpart of the concept's traditional content. Consumer surplus measures the gain to the consumer of being able to buy a product at its market price: the larger that gain the more likely is it that the consumer will be motivated to "do something" to have that gain safeguarded or restored. In this way it is possible to derive the chances for political action from a concept that has dwelt so far exclusively in the realm of economic theory.[6]

Evidently the nature of the available substitute has something to do with the question whether or not connoisseur goods will be rapidly forsaken, in case of deterioration, by the more quality-conscious customers. In the discussion of the exit and voice options in Chapter 3, it was assumed that the only available competing or substitute good was initially of inferior quality, but carried the same price tag. Usually, of course, many other combinations of price and quality exist: in particular, consumers may often have had some hesitation between the good they actually bought, a better-quality substitute with a higher price, and a poorer-quality substitute with a lower price. Suppose now that only the former type of substitute exists

6. For a similar transformation of a time-honored economic concept, the gain from trade, into a political category, namely the influence a trading partner may acquire in the gain-receiving country, see my *National Power and the Structure of Foreign Trade* (Berkeley: University of California Press, 1945, rev. ed. 1969), ch. 2.

and that the quality of the connoisseur good normally bought by a group of consumers deteriorates. In this case it is immediately plausible that the consumers who valued the deteriorating good most will be the first ones to decide that it is worth their while to go over to the higher-quality, higher-price substitute. If only a lower-price, lower-quality good is available, on the other hand, these highly quality-conscious consumers, even though they suffer greatly as a result of quality deterioration, will stick with it longer than their less quality-conscious colleagues. These and similar propositions can be easily proved by indifference curve analysis.[7]

Hence the rapid exit of the highly quality-conscious customers—a situation which paralyzes voice by depriving it of its principal agents—is tied to the availability of better-quality substitutes at higher prices. Such a situation has, for example, been observed in the field of housing. When general conditions in a neighborhood deteriorate, those who value most highly neighborhood qualities such as safety, cleanliness, good schools, and so forth will be the first to move out; they will search for housing in somewhat more expensive neighborhoods or in the suburbs and will be lost to the citizens' groups and community action programs that would attempt to stem and reverse the tide of deterioration. Reverting to the public-private school case, it now appears that the "lower-priced" public schools have several strikes against them in their competition with private schools: first, if and when there is a deterioration in the quality of public school education these schools will lose the children of those highly quality-conscious parents who might otherwise have fought deterioration; second, if, thereafter, quality comes to decline in the private schools, then this type of parents will keep their children there for much longer than was the case

7. See Appendix D, which also discusses in more technical terms a number of other points made in this section.

when the public schools deteriorated. Hence, when public and private schools coexist, with the quality of education in the latter being higher, deterioration will be more strenuously fought "from within" in the case of the private than in that of the public schools. And because exit is not a particularly powerful recuperation mechanism in the case of public schools—it is far more so in that of private schools which have to make ends meet—the failure of one of our two mechanisms is here compounded by the inefficiency of the other.

The relevance of the foregoing observation is greatest in certain important discontinuous choices and decisions, such as between two kinds of educational institutions or two modes of transportation.[8] If one assumes a complete and continuous array of varieties, from cheap and poor-quality to expensive and high-quality, then deterioration of any but the top and bottom variety will rapidly lead to a combination of exits: the quality-conscious consumers move to the higher-price, higher-quality products and the price-conscious ones go over to the lower-price, lower-quality varieties; the former will still tend to get out first

8. In Appendix D it is shown that the reversal phenomenon can occur only when there are at least three goods: the intermediate variety which is the one that deteriorates or whose price increases, another variety that is higher-priced and higher-quality, and a third with the opposite characteristics. In this constellation the less demanding consumers will exit first (toward the lower-priced, lower-quality good) when the *price* of the intermediate good increases, whereas the quality-conscious consumer will exit first (toward the higher-priced, higher-quality good) when *quality* decreases. Even though in the above example only two goods are made explicit, namely public and private school education, the required third alternative on the "other side" of the normally bought good would be present if there were a price increase for public education, namely, informal education at home. This would no doubt be the alternative chosen by many of the less demanding consumers if public schools ceased being free. Hence the presence of the reversal phenomenon cannot be ruled out in this case. A similar reasoning applies to other seemingly dichotomous choices: upon looking more closely, it is usually found that a third alternative exists; some inferior commodity can be found in case the price of the usually bought good increases.

when it is quality that declines rather than price that rises, but the latter will not be far behind.

The proposition that voice is likely to play a more important role in opposing deterioration of high-quality products than of lower-quality products can nevertheless be maintained for the case of a good with many varieties, if these varieties can be assumed not to be spread with equal *density* over the whole quality range. If only because of economies of scale, it is plausible that density is lower in the upper ranges of quality than in the lower and middle ranges. If this is so then deterioration of a product in the upper quality ranges has to be fairly substantial before the quality-conscious will exit and switch to the next better variety. Hence the scope for, and resort to, the voice option will be greatest in these ranges; it will be comparatively slight in the medium- and low-quality ranges.

This finding permits two inferences. First, it can be related to the discussion of education which suggested that the role of voice in fending off deterioration is particularly important for a number of essential services largely defining what has come to be called the "quality of life." Hence, a disconcerting, though far from unrealistic, conclusion emerges: since, in the case of these services, resistance to deterioration requires voice and since voice will be forthcoming more readily at the upper than at the lower quality ranges, the cleavage between the quality of life at the top and at the middle or lower levels will tend to become more marked. This would be particularly the case in societies with upward social mobility. In societies which inhibit passage from one social stratum to another, resort to the voice option is automatically strengthened: everyone has a strong motivation to defend the quality of life at his own station. That cleavages between the upper and lower classes tend to widen and to become more rigid in upwardly mobile societies has become increasingly obvious;

but it has not been an easy observation to make in a culture in which it had long been taken for granted that equality of opportunity combined with upward social mobility would assure both efficiency and social justice.[9]

A rather different inference results if the assumption of a progressive thinning out of varieties at the upper end of the quality scale is brought into contact with the plausible notion that a combination of exit and voice is needed for best results. If this notion is accepted, then the recuperation mechanism may rely too much on exit at the lower end of the quality scale, *but suffer from a deficiency of exit at the upper end.* An illustration of the latter proposition will be found toward the end of the book.

9. The fallacies of this belief were laid bare in Michael Young's incisive satire *The Rise of Meritocracy* (1958, Penguin Edition 1968). See also below, pp. 108–112.

5

How Monopoly Can be Comforted by Competition *

The realization that a tight monopoly is preferable under certain circumstances to a looser arrangement in which competition is present comes hard to a Western economist. Nonetheless, the preceding argument compels recognition that a no-exit situation will be superior to a situation with some limited exit on two conditions:

(1) if exit is ineffective as a recuperation mechanism, but does succeed in draining from the firm or organization its more quality-conscious, alert, and potentially activist customer or members; and

(2) if voice could be made into an effective mechanism once these customers or members are securely locked in.

There are doubtless many situations in which the first condition applies—some additional examples will be given in this and later chapters. The second condition is a very large subject indeed: as was already pointed out, to develop "voice" within an organization is synonymous with the history of democratic control through the articulation and aggregation of opinions and interests.

By itself, the fact that the members or customers are locked in cannot therefore ensure that an effective volume of voice will be forthcoming. As will be argued below, one important way of bringing influence to bear on an organization is to threaten exit to the rival organization. But this threat cannot be made when there is no rival, so that voice is not only handicapped when exit is possible, but also, though in a quite different way, when it is not. Neverthe-

*In writing this chapter I inexcusably failed to refer to John Hicks's celebrated statement of 35 years ago: "The best of all monopoly profits is a quiet life." Had I remembered it, I would have been rather less critical about the economist's neglect of the "lazy monopolist." At the same time, I would have been able to express even more sharply the principal point of the chapter: On certain assumptions about the existence and intensity of voice, competition can afford an even quieter life than does monopoly.—A.O.H., September 30, 1971.

less, it is often possible to make probabilistic statements such as: considering the authority structure and responsiveness of organizations in a given society, and the general readiness of individuals and groups to assert their interests, it is likely that in this or that particular case, voice is going to do a more creditable job of maintaining efficiency when the customers or members are locked in than when some exit is available.[1]

Perhaps the best way of looking at the matter is to recognize that we face here a choice of two evils. Next to the traditional full-fledged monopoly whose dangers and possible abuses have long been exposed, attention should also be paid to those organizations whose monopoly powers are less complete, but who are characterized by sturdy, if undistinguished survival after exit of the more alert customers or members. Often there will be a real question which one of these two institutional varieties is the more unsatisfactory.

The point of view here adopted contrasts with the spirit

1. One may note an interesting symmetry here with the case of perfect competition. As pointed out in ch. 1, n. 1, the firm which produces for a perfectly competitive market finds out about its failings directly through increases in its costs rather than indirectly through customers' reactions because it cannot change either the price or the quality of its product. It will experience losses which will depend on the size of its lapse from efficiency. If the lapse is small, small too will be the losses and the firm will have an opportunity to recover. If one moves just a small step away from perfect competition, to a situation, that is, where the firm has some market power as a price- and quality-maker while demand remains very elastic, then one lands in a very different situation: a small lapse can produce a slightly deteriorated product which will lead to so large a loss of revenue that the firm immediately succumbs. It is now suggested that a similar situation may prevail at the other end of the spectrum. In some situations, a full monopoly may be preferable, from the point of view of the effectiveness of our recuperation mechanisms, to a monopoly just slightly hampered by competition. For this limited competition may result in revenue losses too small to alert management to its failings while it could weaken voice decisively by drawing away from the firm its most vocal customers. At both extremes of perfect competition and pure monopoly the recuperative mechanism may therefore work better than if only a *small step* were made from these extremes in the direction of market power and competitive structure, respectively.

that has long animated the concern over monopoly and the struggle against it. The monopolist has traditionally been expected to utilize to the utmost his ability to exploit the consumer and to maximize profits by restricting production. Public policies have been based primarily on this expectation. Even Galbraith, ordinarily so ready to repudiate the "conventional wisdom," takes this exploitative behavior to be the prime and perhaps only danger which must be guarded against. In his *American Capitalism* he merely pointed out that competition has become an unrealistic alternative to the monopolistic tendencies of advanced capitalist economies and extolled an alternative, already existing remedy, to wit, "countervailing power." But what if we have to worry, not only about the profit-maximizing exertions and exactions of the monopolist, but about his proneness to inefficiency, decay, and flabbiness? This may be, in the end, the more frequent danger: the monopolist sets a high price for his products not to amass super-profits, but because he is unable to keep his costs down; or, more typically, he allows the quality of the product or service he sells to deteriorate without gaining any pecuniary advantage in the process.[2]

In view of the spectacular nature of such phenomena as exploitation and profiteering, the nearly opposite failings which monopoly and market power allow, namely, laziness, flabbiness, and decay have come in for much less scrutiny. To find these problems recognized as public policy issues one has to look beyond the "Anglo-Saxon" world where economic thinking is usually carried on in terms of some maximizing or "taut economy" model. When, a few years ago, a prestigious French economic official put forward proposals for various public controls of business, he did single out incompetence and "abandon"

2. Compare the following remark of a student of Brazilian society: "The large Brazilian landholding is an evil not because it is inhuman and brutal, but because it is inefficient." Jacques Lambert, *Os dois Brasis* (Rio de Janeiro: INEP-Ministerio da Educação e Cultura, 1963), p. 120.

on the part of faltering corporate management as an important problem.[3]

Political power is very much like market power in that it permits the powerholder to indulge either his brutality or his flaccidity. But here again the dangers of abuse of power, of invasion of individuals' rights have—for very good reasons—stood in the center of attention, rather than those of maladministration and bureaucratic ineptitude. Accordingly, the original purpose of the now so widely discussed office of ombudsman was to help redress citizens' grievances against officials who had exceeded the constitutional limits of their power. Later, however, the institution experienced a "shift in its main purpose" which today "has become promotion of better administration," the correction of malpractices and the like.[4] This presumably means that the institution is now also used to correct and reprimand official *indolence* though it was originally devised for the purpose of stemming abuses of power on the part of overactive and overbearing officials.

Such versatility is admirable, but cannot be expected to be the rule. It would be surprising if every one of the safeguards against a monopolist's single-minded pursuit of profits turned out to do double duty as a cure of his propensity toward flabbiness and distraction. Exit-competition is a case in point. While of undoubted benefit in the case of the exploitative, profit-maximizing monopolist, the

3. François Bloch-Lainé, *Pour une réforme de l'entreprise* (Paris: Editions du Seuil, 1963), pp. 54–57, 76–77. "Anglo-Saxon" literature, particularly on trade unions, has paid some attention to the possible existence of "sleepy" or "lazy" monopolies. See, for example, Richard A. Lester, *As Unions Mature* (Princeton: Princeton University Press, 1958), pp. 56–60, and Lloyd G. Reynolds and Cynthia H. Taft, *The Evolution of Wage Structure* (New Haven: Yale University Press, 1956), p. 190. But the exploitative potential of the monopoly has always stood in the center of the discussion and it has been the exclusive motive for regulation and antitrust legislation.

4. Hing Yong Cheng, "The Emergence and Spread of the Ombudsman Institution," *The Annals*, special issue on "The Ombudsman or Citizen's Defender" (May 1968), p. 23.

presence of competition could do more harm than good when the main concern is to counteract the monopolist's tendency toward flaccidity and mediocrity. For, in that case, exit-competition could just fatally weaken voice along the lines of the preceding section, without creating a serious threat to the organization's survival. This was so for the Nigerian Railway Corporation because of the ease with which it could dip into the public treasury in case of deficit. But there are many other cases where competition does not restrain monopoly as it is supposed to, but *comforts and bolsters* it by unburdening it of its more troublesome customers. As a result, one can define an important and too little noticed type of monopoly-tyranny: a limited type, an oppression of the weak by the incompetent and an exploitation of the poor by the lazy which is the more durable and stifling as it is both *unambitious and escapable.* The contrast is stark indeed with totalitarian, expansionist tyrannies or the profit-maximizing, accumulation-minded monopolies which may have captured a disproportionate share of our attention.

In the economic sphere such "lazy" monopolies which "welcome competition" as a release from effort and criticism are frequently encountered when monopoly power rests on location and when mobility differs strongly from one group of local customers to another. If, as is likely, the mobile customers are those who are most sensitive to quality, their exit, caused by the poor performance of the local monopolist, permits him to persist in his comfortable mediocrity. This applies, for example, to small city or "ghetto" stores which lose their quality-conscious patrons to better stores elsewhere as well as to sluggish electric power utilities in developing countries whose more demanding customers will decide at some point that they can no longer afford the periodic breakdowns and will move out or install their own independent power supply.

The United States Post Office can serve as another

example of the lazy monopolist who thrives on the limited exit possibilities existing for its most fastidious and well-to-do customers. The availability of fast and reliable communications via telegraph and telephone makes the shortcomings of the mail service more tolerable; it also permits the Post Office to tyrannize the better over those of its customers who find exit to other communication modes impractical or too expensive.

Those who hold power in the lazy monopoly may actually have an interest in *creating* some limited opportunities for exit on the part of those whose voice might be uncomfortable. Here is a good illustration of the contrast between the profit-maximizing and the lazy monopolist: the former would engage, if he could, in discriminatory pricing so as to extract maximum revenue from its most avid customers, while the lazy monopolist would much rather price these customers out of the market entirely so as to be able to give up the strenuous and tiresome quest for excellence. For the most avid customers are not only willing to pay the highest prices, but are also likely to be most demanding and querulous, in case of any lowering of standards.[5]

Instances of such topsy-turvy (from the point of view of profit maximization) discrimination are not easy to document in economic life, in part perhaps because we have not looked for them very hard and in part simply because price discrimination in general is not easily practiced. But a closely analogous situation is familiar from politics. Latin American powerholders have long encouraged their political enemies and potential critics to remove themselves from the scene through voluntary exile. The

5. There is another way in which the lazy monopolist may be able to rid himself of the voice of these customers: he can extend *just to them* especially high-quality, "gold-plated" service. This would be discrimination with respect to quality rather than to price. The purpose, once again, is not to extract maximum revenue, but to buy "freedom to deteriorate."

right of asylum, so generously practiced by all Latin American republics, could almost be considered as a "conspiracy in restraint of voice." An even more straightforward illustration is supplied by a Colombian law that provided for paying former presidents as many U.S. dollars if they resided abroad as they would receive in Colombian pesos if they lived in their own country. With the U.S. dollar being worth from five to ten pesos while the law was in effect, the officially arranged incentive toward exit of these potential "trouble makers" was considerable.

Even without such special incentives, exit for disgruntled or defeated politicians has always been easier in some countries than in others. The following comparison between politics in Japan and in Latin America supplies another illustration of the corroding influence exit can have on vigorous and constructive political processes via voice:

> The isolation of Japan set rigid boundaries to the possibilities of political opposition. The absence of easy opportunities for tolerable exile was a powerful teacher of the virtues of compromise. The Argentinean newspaper editor in danger of arrest or assassination could slip across the river to Montevideo and still find himself a home, amid familiar sounds and faces and familiar books, easily able to find friends and a new job. (Nowadays, perhaps, he would arrange a refuge in one of the mushrooming international organizations beforehand.) But to all but a tiny fraction of Japanese only one place has ever been home.[6]

In this view, Japan gained an advantage from being a "no-exit" polity while the ever-beckoning opportunity to exit that was characteristic of Hispano-American societies contributed perhaps as much to the factionalism and *personalismo* typical of their politics as the Spanish national character, the *machismo* cult, and similar conventionally given reasons.

6. R. P. Dore, "Latin America and Japan Compared," in John J. Johnson, ed., *Continuity and Change in Latin America* (Stanford: Stanford University Press, 1964), p. 238.