The MIT Press

The Anatomy of Market Failure
Author(s): Francis M. Bator
Source: *The Quarterly Journal of Economics*, Vol. 72, No. 3 (Aug., 1958), pp. 351-379
Published by: The MIT Press
Stable URL: http://www.jstor.org/stable/1882231
Accessed: 15/10/2010 03:17

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=mitpress.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

The MIT Press is collaborating with JSTOR to digitize, preserve and extend access to *The Quarterly Journal of Economics.*

http://www.jstor.org

THE ANATOMY OF MARKET FAILURE

By FRANCIS M. BATOR*

Introduction, 351. — I. The conditions of market efficiency, 353. — II. Neoclassical external economies: a digression, 356. — III. Statical externalities: an ordering, 363. — IV. Comments, 371. — V. Efficiency, markets and choice of institutions, 377.

What is it we mean by "market failure"? Typically, at least in allocation theory, we mean the failure of a more or less idealized system of price-market institutions to sustain "desirable" activities or to estop "undesirable" activities.[1] The desirability of an activity, in turn, is evaluated relative to the solution values of some explicit or implied maximum-welfare problem.

It is the central theorem of modern welfare economics that under certain strong assumptions about technology, tastes, and producers' motivations, the equilibrium conditions which characterize a system of competitive markets will exactly correspond to the requirements of Paretian efficiency.[2] Further, if competitively imputed incomes are continuously redistributed in costless lump-sum fashion so as to achieve the income-distribution implied by a social welfare function, then the competitive market solution will correspond to the one electronically calculated Pareto-efficient solution which maximizes, subject only to tastes, technology and initial endowments, that particular welfare function.[3]

* Center for International Studies and Department of Economics, Massachusetts Institute of Technology.

I am much indebted to R. S. Eckaus and R. M. Solow for detailed comment and discussion.

1. "Activities" broadly defined, to cover consumption as well as production.

2. I.e., to the conditions which define the attainable frontier of maximal utility combinations with given preference functions, resource endowments and technology. A community is on its Paretian frontier if it is impossible to make anyone better off (in terms of his own ordinal preference function) without making someone else worse off. Associated with the utility possibility frontier, in turn, is a production possibility frontier denoting maximal alternative output combinations. (Cf. my "Simple Analytics of Welfare Maximization," *American Economic Review*, XLVII (Mar. 1957), 22–59, and references therein.)

3. In other words, given the "right" lump-sum taxes, markets will match the allocation called for by the point of tangency of the relevant W-function with the utility-possibility frontier, i.e., by the "bliss point." The W-function need not, of course, be explicit — it could be implicit in the political power-configuration which characterizes a community. On the other hand, it cannot be just any kind of function. It has to have some special characteristics which reflect a number of ethic-loaded restrictions, e.g., that individuals' preference functions are to count, and to count positively (cf., *ibid.*, and Section V below).

Many things in the real world violate such correspondence: imperfect information, inertia and resistance to change, the infeasibility of costless lump-sum taxes, businessmen's desire for a "quiet life," uncertainty and inconsistent expectations, the vagaries of aggregate demand, etc. With most of these I am not here concerned: they have to do with the efficiency of "real life" market institutions operated by "real life" people in a nonstationary world of uncertainty, miscalculation, etc.

What follows is an attempt, rather, to explore and order those phenomena which cause even errorless profit- and preference-maximizing calculation in a stationary context of perfect (though limited) information and foresight to fail to sustain Pareto-efficient allocation. I am concerned, in other words, with the decentralizing efficiency of that regime of signals, rules and built-in sanctions which defines a price-market system.[4]

Specifically, Section I sets out the necessary conditions for efficiency of decentralized price-profit calculations both in a "laissez-faire" and in a "socialist" setting of Lange-Lerner civil servants. Section II is a brief digression on an often discussed mode of failure in these conditions: neoclassical external economies. It is concluded that the modern formulation of the doctrine, in terms of "direct interaction," begs more questions than it answers; further, that the usual emphasis on "divorce of scarcity from effective ownership" is misplaced. Section III, then, suggests a comprehensive ordering of types of market failure, with generalized indivisibility, public goods, and, last and least, nonappropriability as the villains of the piece. Section IV consists of some comments on the Meade and Scitovsky classifications of external economies; on the analytical link between indivisibility and public goods; on the significance of "exclusion"; on organizational arrangements designed to offset externality; and on blends of the various types of market failure. Section V concludes with some cautionary notes on the relevance of market-efficiency for choice of institutions.

4. In most of what follows, I shall assume that individual preferences, though not necessarily sensitive only to own-consumption, are representable by strictly convex indifference surfaces (i.e., by an ordering (one for each individual) such that all points on a straight line connecting two equivalent points x and y are preferred to x (hence to y)). But convexity is too restrictive. It excludes not only such characteristics of man's psyche as violate the "usual" regularities — these I do want to exclude — but also such physical and topographical facts as lumpy consumption-goods. Rather than attempt a specification of preferences with convex-like properties where choice must be made among discrete bundles, I dodge the problem by attributing lumpiness only to inputs (including, however, inputs that are intermediate outputs).

I. The Conditions of Market Efficiency

The central theorem of modern welfare economics, the so-called *duality theorem*, asserts a correspondence between Pareto efficiency and market performance. Its analytical essence lies in the remarkable fact that with all-round convexity, independence of tastes, etc., the technocratically formulated, institutionally neutral, Paretian maximum-of-welfare problem has embedded within it a set of constants: "duals," Lagrangean multipliers, shadow-prices, which have all the analytical characteristics of prices, wages, rents, interest rates.[5] Correspondence between Pareto-efficiency and market performance implies, at the least, that decentralized decisions in response to these "prices" by atomistic profit- and satisfaction-maximizers sustain just that constellation of inputs, outputs and commodity-distribution, that the maximum of the specified social welfare function calls for. It implies, in other words, that decentralized market calculations correctly account for all "economic" costs and benefits to which the relevant W-function is sensitive.[6]

Duality can fail in many ways. Specifically, and in a statical and "laissez-faire" context:[7]

(1) Duality will fail unless the Pareto-efficient (a) input-output points (production) and (b) associated commodity distribution points (exchange) which associate with the maximum of the welfare function in hand are characterized by a complete set of marginal-rate-of-substitution (*MRS*) equalities (or limiting inequalities) which, in

5. The theorem holds for the statical steady-state flow model of the Walrasian sort where the solution values are stationary time-rates; it holds, also, for dynamical systems involving capital formation (given, still, convexity throughout). For these last, the solution values are time paths of inputs, outputs, prices, etc. (A set of points is convex if, and only if, the straight lines connecting all possible pairs do not anywhere pass outside the set. The set of feasible output points bounded by a production possibility curve is convex, for instance, if the curve itself is concave-to-the-origin or a straight line. On all this, see Section V of "Simple Analytics," *ibid.*)

6. Given, again, optimal lump-sum redistribution of as-imputed incomes. While I make use of the lump-sum transfer device throughout this paper to abstract from the income distribution problem and permit exclusive attention to Pareto efficiency, it is well to note that this involves a measure of sleight-of-hand. No decentralized price-market type "game" can reveal the pattern of taxes and transfers that would maximize a particular welfare function. "Central" calculation — implicit if not explicit — is unavoidable. Moreover, since distribution (hence correct redistribution) of numeraire-incomes interdepends with allocation in production and exchange, the supposedly automatic, nonpolitical character of market mediation is a myth on the strictest neoclassical assumptions. This is not to say, even on our stratospheric levels of abstraction, that markets are "useless." Where they do compute well we are saved an awful lot of calculation.

7. With optimal redistribution.

turn, yield a set of price-like constants. Where no such constants exist, reference will be to *failure of existence.*[8]

(2) Should such an associated set of Lagrangean parameters exist, duality would nevertheless fail, specifically in production, unless the bliss configuration of inputs and outputs, evaluated in terms of these price parameters, will yield: (a) a local profit-maximum position for each producer, rather than, as possible, a profit minimum; (b) non-negative profits for all producers from whom production is required; (c) maximum profits-in-the-large for each producer. Failure on counts (a) and (c) will be labeled *failure by signal,* that on count (b) *failure by incentive.*[9]

(3) Even if all efficient production configurations, or the one which maximizes a particular welfare-function, coincide with points of maximum and non-negative producers' profits, market mediation may fail in production. If prices are determined by market forces, they will not correspond to a Paretian maximum unless self-policing perfect competition obtains in all markets. Self-policing competition requires "very many" producers in every market.[1] If, then, for whatever reason, some markets are saturated by a few firms of "efficient" scale, the full welfare-maximum solution of inputs, outputs *and prices* will not be sustained. There will be *failure by structure.*

(4) Finally, even if all above is satisfied, market performance could still fail, and fail in a statical sense, due to arbitrary legal and organizational "imperfections," or feasibility limitations on "keeping book," such as leave some inputs or outputs "hidden," or preclude their explicit allocation or capture by market processes (e.g., the restriction, unless I go into baseball, on the sale of the capitalized value of my lifetime services). Failure is *by enforcement.*

8. We could consider, instead, the configuration which associates with the initial pattern of ownership of endowment. Or we could play it safe and extend the conditions to cover each and every Pareto efficient configuration. But this would be overly strict, since many efficient situations have no relevance either to any interesting W-functions or in terms of the initial distribution of scarcities.

It may be worth noting, incidentally, that "existence," as used above, is not the same as existence in the sense of, e.g., Arrow and Debreu (in "Existence of an Equilibrium for a Competitive Economy," *Econometrica,* Vol. 22 (July 1954), pp. 265–90). They use the term to denote the complete set of conditions which defines competitive equilibrium, and this includes, in addition to all that is implied by (1) above, conditions akin to my conditions (2), and some analogous conditions on consumers.

9. This is slightly misleading: as we shall see, failure on count (c) leads both to signaling and to incentive troubles. Anyway, the labels are only for expository convenience.

1. Or at least the potentiality of very many producers, ready and able to "enter the fray" instantaneously. This may be sufficient in the constant-cost case, where the equilibrium number of firms per industry is indeterminate.

All the above are germane to duality in its usual sense, to the statical Pareto-efficiency of laissez-faire markets with genuine profit- and satisfaction-seekers.[2] Conditions (1), (2) and (4) are relevant, also, to the decentralizing efficiency of a Lange-Lerner type organizational scheme. In its "capitalist" version, with profit-motivated operation of privately-owned means of production where it is simply an anti-monopoly device to assure parametric take-prices-as-given behavior, conditions (1), (2) and (4) are all necessary for efficiency. Of course condition (3): self-policing competition, no longer matters.

In its true socialist version, a Lange-Lerner system can afford to "fail" also "by incentive," (2b). Socialist civil servants, under injunction to maximize profit (in the small) in terms of fixed centrally-quoted prices, care or should care not at all about absolute profitability. By assumption the scheme can dispense with the built-in incentive of positive profit: the lure of bureaucratic advancement, the image of Siberia, or the old school tie presumably substitute for the urge to get rich. But if prices and the injunction to maximize profit are to be used to decentralize, condition (1): existence, and (2a) and (2c): correct and unambiguous signals, remain crucial.[3] So does condition (4): the solution of quantities and prices need not be profitable and self-enforcing, but it does have to be enforceable. If the nectar in apple blossoms is scarce and carries a positive shadow price, it must be possible to make every beekeeper pay for his charges' meals.

It warrants repetition that this has to do with whether a decentralized price-market game will or will not *sustain* a Pareto-efficient configuration. The word sustain is critical. There exists a host of further considerations which bear on dynamical questions of adjustment, of "how the system gets there." (E.g., will some "natural" price-market type computational routine of price-quantity responses with a meaningful institutional counterpart tend to track the solution?) These are not here at issue. We shall be concerned only with the prior problem of whether a price-market system which finds itself

2. The mathematically minded will object that (3) and (4), at least, do not really violate "duality" in its strict mathematical sense; the dual minimum problem still yields Lagrangean constants. True, yet I think it suggestive to use "duality" rather more loosely as a label for the general welfare theorem, particularly as this does not lead, in this context, to any ambiguity.

3. It is tempting, but wrong, to suggest that in a true Lange-Lerner world totals do not matter and only margins count. It is true that the non-negativeness of profits is immaterial. Where there is any sharing of shadow-price sets by two or more production points, however, totals necessarily become a part of the signaling system and if 2(c) does not hold they may lead down the garden path.

at the maximum-welfare point will or will not tend to remain there.[4]

The relevant literature is rich but confusing. It abounds in mutually reinforcing and overlapping descriptions and explanations of market failure: external economies, indivisibility, nonappropriability, direct interaction, public goods, atmosphere, etc. In a sense, our problem is simply to sort out the relations among these. In doing so, it is appropriate and useful to begin with a brief review of the neoclassical doctrine of external economies and of its modern formulation in terms of "direct interaction."

II. Neoclassical External Economies: A Digression

By Way of Some History

Marshall, as has often been pointed out, proposed the external economy argument to explain, without resort to dynamics, the phenomenon of a negatively sloped ("forward falling") long-run industry supply curve in terms consistent with a horizontal or rising marginal cost curve (MC) in the "representative" firm. The device permits — in logic, if not in fact — long-run competitive equilibrium of many firms within an industry, each producing at its profit-maximum price-equal-to-a-rising-MC position, without foreclosing the possibility of a falling supply price with rising industry output.[5]

The mechanism is simple. It is postulated that an expansion in the output of the industry as a whole brings into play economies which cause a downward shift of the cost curves of all the component

4. More precisely, whether the point of maximum welfare is or is not a point of self-policing and "enforceable" market equilibrium, where, following common usage, equilibrium is defined to subsume both the first-order and the second-order inequalities for a maximum. A firm, for instance, is taken to be in equilibrium only at a point of maximum profit. This way of defining equilibrium does bring in issues of stability, hence some implicit dynamics. In particular, the word "sustain" is taken to imply some scanning or reconnaissance by producers and consumers at least in the neighborhood of equilibrium. But I do not think it does any harm to subsume this much stability in the equilibrium notion. The possibility of a firm in *unstable* "equilibrium," i.e., in equilibrium at a point of minimum profit, is hardly likely to be of import.

On the other hand, correspondence between Pareto-efficiency and the equilibrium state of perfectly competitive markets is not sufficient to insure market efficiency. It is the burden of "failure by structure" that markets may fail to be competitive, and of "failure by enforcement" that legal or institutional constraints may prevent competitive markets from allocating efficiently, even though there does exist a competitive equilibrium for each Pareto-efficient configuration. "Existence" in the sense of Arrow and Debreu (*op. cit.*) is necessary but not sufficient for market-efficiency in the present context.

5. This refers to a so-called Marshallian supply curve. It has nothing whatever to do with the Walrasian "maximum quantity supplied at a given price" type schedule.

firms. These economies, however, are not subject to exploitation by any one of the myriad of tiny atomized firms. Their own MC curves, at $p = MC$, rise both before and after the shift, due, presumably, to internal diseconomies associated with the entrepreneurial function which defines the firm. Even the modern formulation is not entirely without ambiguity — institutional ambiguity is intrinsic to the device of parametrization: how many firms does it take for the demand curve of each to be perfectly horizontal? — but it does provide a means for "saving" the competitive model, of ducking the monopoly problem.

Marshall, and also Professor Pigou, "preferred," as it were, the other horn of what they perhaps saw as a dilemma. The external economy device, while saving competition, implies a flaw in the efficacy of the "invisible hand" in guiding production.[6] "Price equal to MC" is saved, but wrong. Market forces, they argued, will not give enough output by industries enjoying external economies and will cause industries with rising supply curves to overexpand. Hence the Marshall-Pigou prescription: to harmonize private production decisions with public welfare, tax the latter set of industries and subsidize the former.

It took the better part of thirty years, and the cumulative powers of Allyn Young, and Messrs. Robertson, Knight, Sraffa, and Viner, to unravel the threads of truth and error which run through the Marshall-Pigou argument.[7] The crucial distinction, which provides the key to it all, is between what Viner labeled technological external economies, on the one hand, and pecuniary external economies on the other. The latter, if dominant, cause the long-run supply curve of an industry, say A, to decline because the price of an input, B, falls in response to an increase in A's demand for it. The technological variety, on the other hand, though also a reversible function of industry output, consists in organizational or other improvements in efficiency which do not show up in input prices.[8]

6. That there are difficulties also with income distribution was by that time generally recognized.

7. The strategic articles, with the exception of Young's ("Pigou's *Wealth and Welfare*," this *Journal*, XXVII (1913), 672–86), as well as Ellis and Fellner's 1943 treatment, have all been reprinted in American Economic Association, *Readings in Price Theory*, ed. Stigler & Boulding. For an excellent modern discussion, see R. L. Bishop, *Economic Theory* (to appear).

8. Note, however, that there need be nothing about an organizational improvement to make it obvious in advance whether it will turn out to be technological or, through "internalization," pecuniary. Many trade-association type services which are justified by the scale of an industry could as well be provided commercially, and vice versa.

As regards pecuniary external economies, Robertson and Sraffa made it clear that in a sense both the Marshall-Pigou conclusions were wrong. For one thing, no subsidy is called for. The implied gains in efficiency are adequately signaled by the input price, and profit-maximizing output levels by the A-firms are socially efficient. Second, monopoly troubles may be with us, via, as it were, the back door. For what causes the price of B to drop in response to increased demand? We are back where we started: a declining long-run supply curve.

In the end, then, if *internal* technological economies of scale are ruled out, we are left with only *technological* external economies. All pecuniary external economies must be due to technological economies somewhere in the system.[9] It is true — and this is what remains of the original Marshall-Pigou proposition — that technological externalities are not correctly accounted for by prices, that they violate the efficiency of decentralized market calculation.

The Modern Formulation[1]

In its modern version, the notion of external economies — external economies proper that is: Viner's technological variety — belongs to a more general doctrine of "direct interaction." Such interaction, whether it involves producer-producer, consumer-consumer, producer-consumer, or employer-employee relations, consists in interdependences that are external to the price system, hence unaccounted for by market valuations. Analytically, it implies the nonindependence of various preference and production functions. Its effect is to cause divergence between private and social cost-benefit calculation.

That this is so, is easily demonstrated by means of a simplified variant of a production model suggested by J. E. Meade.[2] Assume a world of all-round perfect competition where a single purchasable and inelastically supplied input, labor ($\overline{L}$), is used to produce two homogeneous and divisible goods, apples (A) and honey (H), at nonincreasing returns to scale. But while the output of A is dependent only on $L_A : A = A(L_A)$, honey production is sensitive also to the level of apple output: $H = H(L_H, A(L_A))$. (Professor Meade

9. Pecuniary diseconomies, in contrast, need have no technological counterpart. Finite-elastic supplies of unproduced inputs are a sufficient cause. Recall, incidentally, that only narrowly statical reversible phenomena are admissible here.

1. While this section makes some slight use of elementary calculus, the reader uninterested in technicalities may avoid, without loss of continuity, all but some simple notation.

2. *Economic Journal*, LXII (Mar. 1952). Meade uses a two factor model and, while he does not explicitly solve the Paretian maximum problem, shows that market imputed rates of remuneration will not match marginal social product.

makes pleasurable the thought of apple blossoms making for honey abundance.)[3]

By solving the usual constrained maximum problem for the production-possibility curve, it can be shown that Paretian production efficiency implies

$$p_H \frac{\partial H}{\partial L_H} = w \tag{1}$$

$$p_A \frac{dA}{dL_A} + p_H \frac{\partial H}{\partial A} \frac{dA}{dL_A} = w \tag{2}$$

where p_H, p_A, and w represent the prices, respectively, of honey, apples and labor.[4] Equation (1) is familiar enough and consistent with profit maximizing. Each competitive honey producer will do for profit what he must for efficiency: hire labor until the value of its social as well as private marginal product equals the wage rate. Not so the apple producers; unless $\frac{\partial H}{\partial A} = 0$ — unless the cross effect of apples on honey is zero — their profit-maximizing production decisions will be nonefficient. Specifically, if apples have a positive external effect on honey output, market-determined L_A will be less than is socially desirable.[5]

A different way to see this is to examine the relations of private to social marginal cost. The marginal money cost of apples to the com-

3. Both functions are assumed homogeneous of degree one. Moreover, apple blossoms (or the nectar therein) are exhaustible, rationable "private" goods: more nectar to one bee means less to another. On the need for this assumption, see Section III-3 below.

4. Assuming internal tangencies and all-round convexity (the last is implicit in constant returns to L: the A-effect on H reinforces convexity), as well as nonsatiation and nonredundancy ($\bar{L} = L_A + L_H$), the maximization of $p_A A + p_H H$, subject to the production functions and the supply of labor, is equivalent to finding a critical value for the Lagrangean expression, $F = p_A A(L_A) + p_H H[L_H; A(L_A)] + w(\bar{L} - L_A - L_H)$. To do so, differentiate F with respect to L_A and L_H, treating p_A, p_H and w as arbitrary constants and set the resulting first order partial derivatives equal to zero. This will give exactly (1) and (2). (Needless to say, the value weights can be varied at will, or taken as given.)

5. To see this, rewrite (2) to read $\frac{dA}{dL_A} = \frac{w}{p_A + p_H \frac{\partial H}{\partial A}}$ and match it against the profit-maximizing rule, $\frac{dA}{dL_A} = \frac{w}{p_A}$. Clearly, $\frac{\partial H}{\partial A} \lesseqgtr 0 \longrightarrow$

$$\left(\frac{dA}{dL_A}\right)_{\text{Private}} \lesseqgtr \left(\frac{dA}{dL_A}\right)_{\text{Social}}$$

petitive apple producer is $\frac{w}{dA/dL_A}$; that of honey to the beekeeper, $\frac{w}{\partial H/\partial L_H}$. It is the ratio of the two: $\frac{\partial H/\partial L_H}{dA/dL_A}$, that competitive market-mediation brings into equality with the equilibrating configuration of relative prices. Markets will be efficient if, and only if, this *private* marginal cost ratio reflects the true marginal cost to society of an extra apple in terms of foregone honey: the marginal rate of transformation between H and A.

What is MRT in the model? Differentiating (totally) the two production functions and dividing the value of one derivative into the other, we get, in absolute (cost) terms:

$$MRT \equiv \left| \frac{dH}{dA} \right| = \frac{\partial H/\partial L_H}{dA/dL_A} - \frac{\partial H}{\partial A}.$$

If, then, $\frac{\partial H}{\partial A} > 0$, the true marginal *social* cost of an "extra" apple, in terms of honey foregone, is less than the market-indicated private cost. It is less precisely by the amount of positive "feedback" on honey output due the "extra" apple.

By combining (1) and (2), eliminating w, and dividing through by p_H and $\frac{dA}{dL_A}$, we get the condition for Pareto efficiency in terms of private MC's:

$$\frac{\partial H/\partial L_H}{dA/dL_A} = \frac{p_A}{p_H} + \frac{\partial H}{\partial A}.$$

Clearly, price equal to private marginal cost will not do. Further, if prices are market-determined, they will diverge from true, *social* marginal cost.

Any number of variations on the model suggest themselves. As Meade pointed out, interactions can be mutual and need not be associated with the outputs. Even in the above case, it is perhaps more suggestive to think of L_A as producing some social value-product both in the A industry and the H industry. In the most general formulation, one can simply think of each production function as containing all the other variables of the system, some perhaps with zero weight. Moreover, by introducing two or more nonproduced inputs one can, as Meade does, work out the consequences for income distribution and input proportions.[6]

6. The question of whether technological external economies involve shifts of each other's production functions, or mutually induced movements along such functions, is purely definitional. If one chooses so to define each producer's

Some Queries

The modern formulation of the doctrine of external economies, in terms of direct interaction, is not only internally consistent: it also yields insight. Yet one may well retain about it some dissatisfaction. There is no doubt that the Robertson-Sraffa-Viner distinction between the technological and the pecuniary sort gets to the nub of what is the matter with the original Marshallian analysis. It cuts right through the confusion which led Marshall and Pigou to conclude that the price mechanism is faulty in situations where in truth it is at its best: in allocating inputs in less than infinitely elastic supply between alternative productive uses. It also facilitates unambiguous formulation of the more difficult "falling supply price" case. But in a sense it only begs the fundamental question: what is it that gives rise to "direct interaction," to short circuit, as it were, of the signaling system?

Most modern writers have let matters rest with the Ellis-Fellner type explanation: "the divorce of scarcity from effective ownership."[7] Does nonappropriability then explain all direct interaction? In a sense it does, yet by directing attention to institutional and feasibility considerations which make it impracticable for "real life" market-institutions to mimic a price-profit-preference computation, it diverts attention from some deeper issues. Surely the word "ownership" serves to illuminate but poorly the phenomenon of a temperance leaguer's reaction to a hard-drinking neighbor's (sound insulated and solitary) Saturday night, or the reason why a price system, if efficient, will not permit full "compensation," in an age of electronic scramblers, for an advertisement-less radio program, or for the "services" of a bridge.[8]

function as to give axes only to inputs and outputs that are purchased and sold, or at least "controlled," and the effects of everything else impinging on production (e.g., of humidity, apple blossoms, etc.) are built into the curvature of the function, then it follows that externalities will consist in shifts of some functions in response to movements along others. On the other hand, if, as in our apple-honey case, it seems useful to think of the production function for H as having an A-axis, then, clearly, induced movement along the function is a signal of externality.

7. *Op. cit.*

8. Moreover, in the one sense in which nonappropriability fits all cases of direct interaction, it explains none. If all it denotes is the failure of a price-market game properly to account for (to appropriate) all relevant costs and benefits, then it is simply a synonym for market failure (for generalized externality), and cannot be used to explain what causes any particular instance of such failure. I use it in a much narrower sense, to mean the inability of a producer of a good or service physically to exclude users, or to control the rationing of his produce among them. In my sense not only bridges but also, say, television programs are fully appropriable: it is always possible to use scramblers.

It may be argued, of course, that at least the two latter examples are out of order, that radio programs and bridges do not involve "direct," i.e., non-price, interaction. But is this really so? Does not the introduction of a new program directly affect my and your consumption possibilities, in ways other than by a change in relative prices? Does not a bridge, or a road, have a direct effect on the production possibilities of neighboring producers, in precisely the sense in which apples affect the possibilities of beekeepers?[9]

True, perhaps bridges and roads are unfair: they violate the neoclassical assumption of perfect divisibility and nonincreasing returns to scale. But they surely do involve non-price interaction. In fact, lumpiness and increasing returns are perhaps the most important causes of such interaction. Are they to be denied status as externalities? More generally, are we to exclude from the class of externalities any direct interaction not due to difficulties with "effective ownership," any failures other than "by enforcement"?

It would be, of course, perfectly legitimate to do so — tastes are various. But I think it more natural and useful to broaden rather than restrict, to let "externality" denote any situation where some Paretian costs and benefits remain *external to* decentralized cost-revenue calculations in terms of prices.[1] If, however, we do so, then clearly nonappropriability[2] will not do as a complete explanation. Its concern with the inability of decentralized markets to sustain the solution-prices and quantities called for by a price-profit-preference type calculation, as computed by a team of mathematicians working with IBM machines, tends to mask the possibility that such machine-

9. It is possible, of course, to interpret these examples as involving very large changes in price: from infinity to zero. But it does not help to do so. The shared characteristic of bridges and programs is that there is no price which will efficiently mediate both supply and demand.

I have puzzled over ways of limiting the notion of "direct interaction" to something less than all instances where there is some interaction not adequately signaled by price. Robert Solow has suggested to me that this might be done by distinguishing situations where something is not subject to a market test at all from instances where no single price constitutes a correct test for both sides of a transaction (e.g., where the correct ration price for the services of an expensive facility is zero). I am inclined, rather, to drop the attempt to use "direct interaction" as an explanation of market failure; it is best used, if at all, as yet another synonym for such failure.

1. Recall that it is the existence of such "externality," of residue, at the bliss-point, of Pigouvian "uncompensated services" and "incidental uncharged disservices" that defines market failure. It may be objected that to generalize the externality notion in this way is to rob it of all but descriptive significance. But surely there is not much to rob; even in its strictest neoclassical formulation it begs more than it answers. In its generalized sense it at least has the virtue of suggesting the right questions.

2. As defined in fn. 8, p. 361 above.

calculated solution q's may well be nonefficient.[3] It explains failure "by enforcement," but leaves hidden the empirically more important phenomena which cause failure by "nonexistence," "signal," and "incentive." Section III is designed to bring these deeper causes of generalized externality into the foreground.

III. Statical Externalities: An Ordering

If nonappropriability is, by itself, too flimsy a base for a doctrine of generalized (statical) externality, what broader foundation is there? Section I's hierarchy of possible modes of market failure suggests a fivefold classification. If, however, one looks for an organizing principle not to modes of failure but to causes, there appear to be three polar types: (1) Ownership Externalities, (2) Technical Externalities,[4] and (3) Public Good Externalities. These are not mutually exclusive: most externality phenomena are in fact blends. Yet there emerges a sufficient three-cornered clustering to warrant consolidation.[5]

Type (1): Ownership Externalities

Imagine a world which exhibits generalized technological and taste convexity, where the electronically calculated solution of a Paretian maximum-of-welfare problem yields not only a unique set of inputs, outputs and commodity-distribution, but where initial endowments plus lump-sum transfers render income distribution optimal in terms of the community's social welfare function. Assume, further, that everything that matters is divisible, conventionally rationable, and either available in inelastic total supply,[6] or producible at constant returns to scale; also that tastes are sensitive only to own-consumption. We know, then, from the duality theorem, that

3. Or that the algorism may break down for lack of a consistent set of p's.

4. I should much prefer "technological," but since this would necessarily confuse my Type (2) with Professor Viner's "technological" I fixed on "technical."

5. In effect, we end up with a five-by-three ordering of types of "failure": five "modes" vs. three "causes." Its relation to Meade's categories (*op. cit.*) and to Tibor Scitovsky's classification (in "Two Concepts of External Economies," *Journal of Political Economy*, LXII, April 1954) is discussed in Section IV below. I have had the benefit of reading, also, William Fellner's "Individual Investment Projects in Growing Economies," *Investment Criteria and Economic Growth* (Proceedings of a Conference, Center for International Studies, Massachusetts Institute of Technology, 1955) and an unpublished paper by Svend Laursen, "External Economies and Economic Growth."

6. The supply of such nonproduced scarcities need not, of course, remain constant. On the other hand, their ownership distribution must not be so concentrated as to preclude competitive rationing. There must exist no "indivisible" lake full of fish, etc., such as might be subject to monopolization, but thousands of lakes, all perfect substitutes.

the bliss point implies a unique[7] set of prices, wages and rents, such as would cause atomistic profit- and preference-maximizers to do exactly what is necessary for bliss. In particular, all required production points give maximum and non-negative producer's profits.

This is an Adam Smith dream world. Yet it is possible that due to more or less arbitrary and accidental circumstances of institutions, laws, customs, or feasibility, competitive markets would not be Pareto-efficient. Take, for instance, the Meade example of apples and honey. Apple blossoms are "produced" at constant returns to scale and are (we assumed) an ordinary, private, exhaustible good: the more nectar for one bee, the less for another. It is easy to show that if apple blossoms have a positive effect on honey production (and abstracting from possible satiation and redundancy) a maximum-of-welfare solution, or any Pareto-efficient solution, will associate with apple blossoms a positive Lagrangean shadow-price.[8] If, then, apple producers are unable to protect their equity in apple-nectar and markets do not impute to apple blossoms their correct shadow value, profit-maximizing decisions will fail correctly to allocate resources (e.g., L) at the margin. There will be failure "by enforcement."

This is what I would call an *ownership* externality. It is essentially Meade's "unpaid factor" case. Nonappropriation, divorce of scarcity from effective ownership, is *the* binding consideration. Certain "goods" (or "bads") with determinate non-zero shadow-values are simply not attributed. It is irrelevant here whether this is because the lake where people fish happens to be in the public domain, or because "keeping book" on who produces, and who gets what, may be impossible, clumsy, or costly in terms of resources.[9] For whatever legal or feasibility reasons, certain variables which have positive or negative shadow value are not "assigned" axes. The beekeeper thinks only in terms of labor, the orchard-owner only in terms of apples.

The important point is that the difficulties reside in institutional arrangements, the feasibility of keeping tab, etc. The scarcities at issue are rationable and finely divisible and there are no difficulties with "total conditions": at the bliss-configuration every activity would pay for itself. Apple nectar has a positive shadow price, which

7. Or, where there are corners, only inessentially indeterminate.

8. Set up a variant of the Apple-Honey model of Part II, introducing apple blossoms, B, explicitly. Add a production function, $B = B(L_A)$, and substitute $B(L_A)$ for $A(L_A)$ as the second input in honey production. The solution will give out a positive Lagrangean shadow price for B, and profit-maximizing producers of the joint products: A and B, will push L_A to the socially desirable margin.

9. Though on this last, see Section IV, first paragraph.

would, if only payment were enforceable, cause nectar production in precisely the right amount and even distribution would be correctly rationed. The difficulty is due exclusively to the difficulty of keeping accounts on the nectar-take of Capulet bees as against Montague bees.[1]

Many of the few examples of interproducer external economies of the reversible technological variety are of this type: "shared deposits" of fish, water, etc.[2] Much more important, so are certain irreversible dynamical examples associated with investment. For instance, many of Pigou's first category of externalities: those that arise in connection with owner-tenant relationships where durable investments are involved, have a primarily organizational quality.[3] Perhaps the most important instance is the training of nonslave labor to skills — as distinct from education in a broader sense (which partakes more of Type (3)). In the end, however, and in particular if restricted to reversible statical cases, it is not easy to think of many significant "ownership externalities" pure and simple. Yet it turns out that only this type of externality is really due to nonappropriability.

Type (2): Technical Externalities

Assume, again, that all goods and services are rationable, exhaustible, scarcities, that individual ordinal indifference maps are convex and sensitive only to own-consumption and that there exist no ownership "defects" of Type (1). If, then, the technology exhibits indivisibility or smooth increasing returns to scale in the relevant range of output, these give rise to a second and much more important type of market failure: "technical externality."[4]

1. More generally, it could as well be due to difficulty in knowing who "produced" the "benefit" — oil wells drawing on the same pool are an example. The owner cannot protect his own; in fact it is difficult to know what one means by "his own." Moreover, in the case of *dis*economies, at least, it may be that both the source and the recipient of the "bad" are identified: one factory producing soot and nothing but one laundry in the neighborhood, yet it is difficult to see how a price can be brought to bear on the situation. Presumably the laundry can pay for negative units of smoke.

2. Though indivisibility elements enter into some of these. Why can't somebody "own" part of a lakeful of fish?

3. When not simply due, in a world of uncertainty, to inconsistent expectations.

4. Again, this is not the same as Viner's "technological." Note, incidentally, that the above formulation unabashedly begs the question of whether smooth increasing returns to scale could or could not arise without indivisibility somewhere. The issue is entirely definitional: it is conceptually impossible to disprove either view by reference to empirical evidence. (Cf. "Simple Analytics," *loc. cit.*, fn. 37 and references.) (*Continued on page 366.*)

The essential analytical consequence of indivisibility,[5] whether in inputs, outputs or processes, as well as of smooth increasing returns to scale, is to render the set of feasible points in production (input-output space) nonconvex. A connecting straight line between some pairs of feasible points will pass outside the feasible set. Nonconvexity, in turn, has a devastating effect on duality.[6]

In situations of pure "technical externality" there does, of course, still exist a maximal production possibility frontier (*FF*); and with a Samuelson-type social indifference map (*SS*) — i.e., a map "corrected" for income distribution which provides a ranking for the community as a whole of all conceivable output combinations[7] — it is possible, in concept, to define a bliss point(s).[8] Also, where indivisibility is exhibited by outputs, and only outputs, or, stronger, where smoothly increasing returns to scale is the only variety of nonconvexity — isoquants for one, are properly convex — the locus of efficient output combinations can be defined in terms of conditions on marginal-rates-of-input-substitution.[9] Moreover, bliss could possibly occur at a point where *SS* is internally tangent to *FF*, perhaps to a convex *FF*. But even in the least "pathological," most neoclassi-

The pioneer work on decreasing cost situations is Jules Dupuit's remarkable 1844 essay, "On the Measurement of Utility of Public Works," translated in *International Economic Papers*, No. 2, ed. A. T. Peacock, *et al.* Harold Hotelling's "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," in the July 1938 issue of *Econometrica*, is the originating modern formulation. Cf., also, references to work by R. Frisch, J. E. Meade, W. A. Lewis and others in Nancy Ruggles' excellent survey articles on marginal cost pricing (*Review of Economic Studies*, XVII (1949–50), 29–46, and 107–26).

5. Indivisibility means lumpiness "in scale" and not the kind of indivisibility-in-time we call durability. (Durability, as such, does not violate convexity.) Lumpiness has to do with the impossibility to vary continuously, e.g., the capacity service-yield per unit time of such things as bridges.

6. The best known and perhaps most important variety of nonconvexity occurs where isoquants are properly convex, but returns to scale are increasing, hence the full set of feasible input-output points is nonconvex. (In a two-input, one-output situation, slices by (vertical) planes through the origin perpendicular to the input plane will cut the production surface in such a way as to give a nonconvex boundary.) A production point lying in an "increasing returns" region of a production function implies that (1) the associated average cost curve (*AC*) is downward sloping at that level of output; (2) the associated marginal cost curve (*MC*), while it may be rising, could as well be falling and will certainly lie below *AC;* and (3) the production possibility curve of the community may be nonconvex. On all this, see Part V of "Simple Analytics," *loc. cit.*

7. Cf. P. A. Samuelson, "Social Indifference Curves," this *Journal*, LXX (Feb. 1956), 1–22. Such a function presumes that *numeraire*-incomes are continuously redistributed so as to maximize in utility space over the community's operative social welfare function.

8. This is saying very little, of course, except on the level of metaphysics.

9. Inequalities due to kinks and corners are as good as equalities where all is smooth.

cally well-behaved case, where there exists a meaningfully defined set of shadow prices associated with the bliss point, genuinely profit-seeking competitive producers, responding to that set of prices, would fail to sustain optimal production. At best, even if at the bliss-configuration all MC's are rising, some producers would have to make continuing losses, hence would go out of business; market calculations would necessarily fail "by incentive." If, in turn, prices are not centrally quoted but permitted to set themselves, monopoly behavior will result. There will be failure "by structure."

Further, bliss may require production at levels of output where losses are not only positive, but at a constrained maximum;[1] $p = MC$ may be correct, though MC at that point is falling. If so, the embedded Lagrangean constants may still retain meaning as marginal rates of transformation, but they will fail to sustain efficient production even by Lange-Lerner civil servants who care only about margins and not about absolute totals. There will be failure "by signal": producers under injunction to maximize profit (in the small) will not remain where they ought to be.

If, moreover, we drop the assumption of smooth increasing returns to scale and permit indivisibilities such as give scallop-like effects and kinks in cost curves and in the production-possibility curve, things get even more complicated. Bliss could require production at points of positive but locally minimum profit, where MC exceeds AC but is falling. Worse, even if bliss should occur at points where production functions are locally convex and MC (greater than AC) is rising, prequoted prices may still not sustain the solution unless production functions are in fact convex throughout. Though positive and at a local maximum, profits may not be at their maximum-maximorum: other hills with higher peaks may induce producers with vision at a distance to rush away from bliss. Alternatively, if prices are not administered, competition may not be self-policing and markets could fail "by structure."[2]

1. Subject to the requirement that total cost for that level of output be a minimum, i.e., that each producer be on his least-cost expansion path.

2. Where sharp indivisibility gives a nonconvex production possibility curve with corners and kinks, duality may fail even if there exists a price vector in terms of which decentralized producer-calculations would sustain the bliss-point output mix. The existence of such a vector does not assure that it will coincide with the price-vector which would efficiently ration that bill of goods among consumers. The point is that there may not exist a *single* set of prices which will at the same time keep both consumers and producers from rushing away from where they ought to be. The prices which will effectively mediate production may cause consumers' calculations to go wrong and vice versa.

It should be noted, incidentally, that none of the above takes space and distance considerations into account. For some interesting effects of plant-indivisi-

On the other hand, given our assumptions, the Paretian contract locus of maximal (ordinal) utility combinations which is associated with any one particular output point is defined, as in the trouble-free neoclassical model, by the usual subjective, taste-determined, marginal-rate-of-substitution equalities (or, at corners, inequalities). These *MRS* equalities, in turn, imply a set of shadow-prices which, if centrally quoted, would efficiently ration among consumers the associated (fixed) totals of goods. In the sphere of exchange, then, a decentralized price system works without flaw.

In what sense do these Type (2) situations exhibit "externality"? In the (generalized) sense that some social costs and benefits remain external to decentralized profitability calculations. With Type (1) externalities, though it is not feasible to police the bliss values of all quantities and prices, there exists embedded in the solution a set of prices whose use for purposes of decentralized signaling would sustain, if only appropriation or exclusion were feasible, both itself and the maximum welfare configuration of inputs, outputs, and distribution. This is not the case here. In Type (1) situations, at the bliss point there is complete correspondence between social and private pay-off, both at the margin and in totals.[3] Profits are at their maxima and non-negative throughout. Here there is no such correspondence; there may well be divergence, either at the margin: bliss-profits may be at a "minimum," or in *totals*. The private totals in terms of which producers in an (idealized) market calculate — total revenue minus total cost — will not reliably signal the social costs and benefits implied by the relevant social indifference curves.[4] Hence at the set of prices which would correctly ration the bliss point bill of goods, that bill of goods may not be produced by profit seekers, or even by Lange-Lerner civil servants.[5]

bility where there are interplant flows and transport takes resources, see T. C. Koopmans and M. Beckmann, "Assignment Problems and the Location of Economic Activities," *Econometrica*, Vol. 25 (Jan. 1957).

3. More correctly, there would be such correspondence, if only the p's could be policed.

4. This is particularly awkward since the very nonconvexities which cause a divergence between private and social total conditions render output-mix calculations based on margins alone wholly inadequate. Even if bliss gives all local profit maxima, there may be several such open to any one producer, hence he must make total calculations in order to choose.

5. There is one qualification to be made to the above. It may be that the bliss configuration gives unique and positive profit maxima throughout, though some production functions exhibit nonconvexities at a distance. It was to exclude this case that we assumed that increasing returns or indivisibility obtain in the "relevant ranges." Should this happen, no "externality" divergence of social and private calculation will occur, at least in a statical context. But unless all

A point to note, in all this, is that in relation to "technical externalities" the nonappropriability notion, as generally conceived, tends to miss the point. Strictly speaking, it is, of course, true that price mediation, if efficient, cannot be counted on to "appropriate" the full social benefits of activities showing increasing returns to scale or othe types of indivisibility to those engaged in them. But the existence of such "uncompensated services" has in this case nothing whatever to do with "divorce of scarcity from ownership," with feasibility limitations on "exclusion." It is entirely feasible to own a bridge and profitably ration crossings; indeed, a private owner would do so. The point is, rather, that such profitable rationing, such "compensation" for services rendered, would inefficiently misallocate the "output" of bridge crossings. If in terms of scarce resource inputs the marginal cost of an additional crossing is zero, any positive toll will, in general, have the usual monopolistic effect: the resulting output configuration will not be efficient.[6]

This, incidentally, is where most pecuniary external economies lead: a supplier is required to produce in a range of declining AC due to internal technological economies of scale and hence cannot make "ends meet" at the socially correct price. The crucial associated difficulty at the level of social organization is monopoly.

Can we leave matters at that? Not quite. There is a third kind of externality, recently emphasized by Professor Samuelson, caused by so-called "public goods."

Type (3): Public Good Externalities

In some recent writings on public expenditure theory, Samuelson has reintroduced the notion of the collective or public good. The defining quality of a pure public good is that "each individual's consumption of such a good leads to no subtractions from any other individual's consumption of that good . . .",[7] hence, "it differs from a private consumption good in that each man's consumption of it, X_2^1 and X_2^2 respectively, is related to the total X_2 by a condition of

is convex throughout, the existence of such a locally stable tangency cannot be taken as evidence that the point is in fact the bliss-point — a difficulty of considerable significance for dynamical efficiency.

6. Of course, if at bliss the bridge were to be used "to capacity," it is possible that the Lagrangean ration price (now positive) would make commercial operation profitable. If so, an administered price setup would efficiently mediate the demand and supply of crossings. But while a Lange-Lerner system would work fine, laissez-faire markets would fail "by structure."

7. P. A. Samuelson, *Review of Economics and Statistics*, XXXVI (Nov. 1954), 387.

equality rather than of summation. Thus, by definition, $X_2^1 = X_2$ and $X_2^2 = X_2$."[8]

As Samuelson has shown, the form of the marginal rate of substitution conditions which define the Pareto-efficient utility possibility frontier in a world where such public goods exist, or at least where there are outputs with important "public" qualities, renders any kind of price-market routine virtually useless for the computation of output-mix and of distribution, hence, also, for organizational decentralization. Where some restraints in the maximum problem take the form: total production of X *equals* consumption by Crusoe of X *equals* consumption of X by Friday, Pareto efficiency requires that the marginal rate of transformation in production between X and Y equal not the (equalized) MRS of each separate consumer, but rather the algebraic *sum* of such MRS's. This holds, of course, in what in other respects is a conventionally neoclassical world: preference and production functions are of well-behaved curvature, all is convex.

If, then, at the bliss point, with Y as numeraire, Px is equated to the marginal Y-cost of X in production (as is required to get optimal production), and X is offered for sale at that p_x, preference-maximizing consumers adjusting their purchases so as to equate their individual MRS's to p_x will necessarily under-use X. Moreover, a pricing game will not induce consumers truthfully to reveal their preferences. It pays each consumer to understate his desire for X relative to Y, since his enjoyment of X is a function only of total X, rather than, as is true of a pure private good, just of that fraction of X he pays for.

The two Samuelson articles[9] explore both the analytics and the general implications of "public goods." Here the notion is of relevance because much externality is due precisely to the "public" qualities of a great many activities. For example, the externality associated with the generation of ideas, knowledge, etc., is due in good part to the public character of these "commodities." Many interconsumer externalities are of this sort: my party is my neighbor's disturbance, your nice garden is any passerby's nice view, my children's education is your children's good company, my Strategic Air Command is your Strategic Air Command, etc. The same consumption item enters, positively or negatively, both our preference func-

8. P. A. Samuelson, *Review of Economics and Statistics*, XXXVII (Nov. 1955), 350.

9. And a third unpublished paper, which was read at the 1955 American Economic Association meetings and to a copy of which I came to have access while this paper was being written. For earlier writings on public goods, by Wicksell, Lindahl, Musgrave, Bowen and others see references in the above cited Samuelson articles.

tions. The consumptions involved are intrinsically and essentially joint.

This kind of externality is distinct from either of the other two pure types. Here technological nonconvexities need in no way be involved. In fact the $MRT = \Sigma MRS$ condition is certain to hold true precisely where production takes place at constant or nonincreasing returns, and hence where the production possibility set is necessarily convex. Further, there are no decentralized organizational rearrangements, no private bookkeeping devices, which would. if only feasibility were not at issue, eliminate the difficulty. It is the central implication of the Samuelson model that where public good phenomena are present, there does not exist a set of prices associated with the (perfectly definable) bliss point, which would sustain the bliss configuration. The set of prices which would induce profit-seeking competitors to produce the optimal bill of goods, would be necessarily inefficient in allocating that bill of goods. Moreover, even abstracting from production, no single set of relative prices will efficiently ration any fixed bill of goods so as to place the system on its contract locus, except in the singular case where at that output and income-distribution MRS's of every individual are identically the same (or zero for all but one). There is failure "by existence."

IV. Comments

Type (1). In a sense, Type (1) is not symmetrical with the other two categories. One can think of some nontrivial instances where the institutional element does appear to be "binding": skill-training of people, for example. But even there, it could be argued that the crucial elements are durability, uncertainty, and the fact that slavery as a mode of organization is itself in the nature of a public good which enters people's preference functions, or the implicit social welfare function, inseparably from the narrowly "economic" variables. In those instances, in turn, where bookkeeping feasibility appears to be the cause of the trouble, the question arises why bookkeeping is less feasible than where it is in fact being done. In the end, it may be that much of what appears to partake of Type (1) is really a compound of Types (2) and (3), with dynamical durability and uncertainty elements thrown in. At any rate, a deeper analysis of this category may cause it substantially to shrink.

Nonproduced scarcities. One particular instance where what appears like Type (1) is really Type (2) warrants special mention. Public ownership of nonproduced resources, e.g., the lakes and mountains of national parks, may make it appear that externality is due

to statutory barriers to private ownership and commercial rental. But this is missing the point. Take, for instance, a community which has available a single source of fresh water of fixed capacity. Assume that the bliss solution gives out a positive ration-price per gallon such as would make sale of the water commercially profitable. Yet a laissez-faire system would fail, "by structure," to sustain bliss. A private owner of the single indivisible well, if given his head, would take advantage of the tilt in the demand curve. The real cause of externality is not the arbitrary rapaciousness of public authority but the indivisibility of the source of supply. This case, by the way, is akin to where indivisibility or increasing returns to scale within a range allow profitable scope for one or a few efficient producers, but for no more. At the bliss price all will do the right thing, but if prices are not administered, oligopoly or monopoly will result. A capitalist Lange-Lerner system with private ownership but administered prices would work fine, but laissez-faire markets would fail.

Meade's "atmosphere." The relation of my tri-cornered ordering to Meade's polar categories is of interest.[1] His first category, "unpaid factors," is identical to my Type (1). But his second, labeled "atmosphere," is a rather curious composite. Meade's qualitative characterization of "atmosphere": e.g., of afforestation-induced rainfall, comes very close to the public good notion.[2] He links this, however, as necessarily bound up with increasing returns to scale in production to society at large, hence a J. B. Clark-like overexhaustion, adding-up problem.[3]

If, following Meade, one abstracts from shared water-table phenomena (let rain-caused water input be rigidly proportional to area) then Farmer Jones' rain is Farmer Smith's rain and we have my Type (3). But nothing in this situation requires that either farmer's full production function (with an axis for rain) need show increasing returns to scale. It may be that returns to additional bundles of non-rain inputs, with given constant rainfall, diminish sharply, and that it takes proportional increases of land, labor *and rain* to get a proportional effect on output. If so, Meade's overexhaustion problem

1. *Op. cit.* (This and the next section can be omitted without loss of continuity.)

2. See esp. bottom of p. 61 and top of p. 62, *op. cit.*

3. Since his argument is restricted to competitive situations, hence necessarily excludes increasing-returns-to-paid-factors such as would require production at a loss, Meade specifies constant returns to proportional variation of labor and land in wheat farming, though the full production function for wheat, including the atmosphere input (rain), exhibits increasing returns to scale. But the individual farmer does not pay for rain, hence his factor payments just match his sales revenue, by the Euler Theorem.

will not arise. But all would not be well: the public good quality of rainfall would cause an independent difficulty, one that Meade, if I understand him correctly, does not take into account, i.e., that rain ought to be "produced" by timber growers until its MC is equal to the sum of all the affected farmers MRS's for rain as an input, whatever may be the curvature of the latter's production functions.[4]

On the other hand, Meade's formal mathematical treatment of "atmosphere," as distinct from his verbal characterization and his example, suggests that it is a nonappropriable, and therefore unpaid, factor which gives rise to increasing returns to scale to society though not to the individual producer. At least this is all he needs for the effect he is looking for: a self-policing though nonoptimal competitive situation, where, because the full production functions (i.e., with an axis for rain) are of greater than first degree, the correction of externality via subsidies to promote the creation of favorable atmosphere requires net additions to society's fiscal burden. If this is the crucial consequence of "atmosphere," then it need have no "public" quality. All this would happen even though Smith and Jones were "competing" for the water from the shared water-table under their subsoil, just like bees competing for nectar.

Scitovsky's "two concepts."[5] Professor Scitovsky, in turn, in his suggestive 1954 article, distinguishes between the statical direct interactions of equilibrium theory and the kinds of pecuniary external economies emphasized in the economic development literature. He classifies the former as consumer-consumer, producer-consumer, and producer-producer interactions, labels the last as external economies and asserts that they are rare and, on the whole, unimportant.

While Scitovsky does not raise the question of what gives rise to such producer-producer interactions, both his examples, and his conclusion that they are of little significance, suggest that he is thinking

4. Formally, Meade denotes "atmosphere" as a situation where the production function, e.g., of farmers takes the form $X_1 = H_1(L_1,C_1)A_1(X_2)$, with L as labor, C as capital and A the atmosphere effect on X_1 of X_2. The full function exhibits increasing returns to scale but the H function alone, with A constant, is homogeneous of first degree. But why can't this be put in terms of Meade's unpaid factor type function where $X_1 = H_1(L_1,C_1,X_2)$? Example: $X_1 = L_1^a\, C_1^{1-a}\, X_2$. All this has nothing to do with whether $A = A_1 + A_2$ or rather $A = A_1 = A_2$. Unfortunately, the example itself tends to mislead. The fact that exclusion of rain-users (farmers) by producers (timber-growers) is hardly feasible, i.e., that rain is like Type (1), distracts attention from the important point that *if* rain is, as Meade tells us, a public good, then rationing it by price would be inefficient even if it were feasible. (It should be said that Meade concludes his article: "But, in fact, of course, external economies or diseconomies may not fall into either of these precise divisions and may contain features of both of them.")

5. *Op. cit.*

primarily of Type (1): nonappropriability. But this is to ignore public goods — surely a more important cause of interaction. Moreover, by taking full account of these, Scitovsky's "fifth and important case, which, however, does not quite fit into . . . (his) . . . classification . . . , where society provides social services through communal action and makes these available free of charge to all persons and firms," can be made nicely to fall into place.[6]

Samuelson on Types (2) and (3). While the public good model helps to sort out the phenomena Meade lumped under "atmosphere," Samuelson himself emphasizes the analytical bond between indivisibility and public good situations. In both an explicit "summing in" is required of "all direct and indirect utilities and costs in all social decisions."[7] In Type (2) situations it is the intramarginal consumer's and producer's surpluses associated with various all or nothing decisions "in-the-lump" that have to be properly (interpersonally) weighted and summed, while in Type (3) it is only utilities and costs at the margin that require adding. But, and this is the crucial shared quality of the two categories, both make it necessary to sum utilities over many people.[8]

Exclusion. One more comment may be warranted on the significance, in a public good type situation, of nonappropriability. "Exclusion" is almost never impossible. A recluse can build a wall around his garden, Jones can keep his educated children away from those of

6. *Ibid.*, fn. 3, p. 144. Scitovsky, following Meade, restricts his "first concept" of external economies to phenomena consistent with competitive equilibrium. He treats indivisibilities and increasing returns to scale as belonging to his "second concept" which has to do with disequilibrium, investment decisions, and growth. It is, of course, entirely legitimate to restrict analysis to competitive situations. But the Scitovsky treatment must not be taken to imply that lumpiness is irrelevant to statical analysis of stationary solution points. If one is interested in the statical efficiency of decentralized price calculations, they are crucial. But this is carping. Scitovsky's important contribution lies in emphasizing and clarifying the point first hinted at by P. N. Rosenstein-Rodan that in a world of disequilibrium dynamics pecuniary external economies may play an independent role — one distinct, that is, from simply being an unreliable signal of monopoly troubles (*Economic Journal*, LIII, 1943, 202–11).

7. *Ibid.*, p. 9.

8. There is one qualification to be made: if all public good and increasing returns to scale industries produce only intermediate products, all externalities may cancel out in intra-business-sector transactions. If so, only total revenues and total costs have to be summed. Incidentally, the exposition may misleadingly suggest another symmetry between Types (2) and (3). In a pure Type (3) situation, *if* there are no public producers' goods, then while prices cannot be used to ration the bliss point output-mix, they can be used efficiently to mediate production. In Type (2), on the other hand, *if* all final consumables are divisible, price calculations, while failing in production, will work in exchange. This symmetry breaks down, of course, as soon as one violates, as does the real world, the two "if's."

Smith, etc. But if thereby some people (e.g., the recluse) are made happier and some (e.g., the passers-by) less happy, any decision about whether to "exclude" or not implies an algebraic summing of the somehow-weighted utilities of the people involved. And if the wall requires scarce resources, the final utility sum must be matched against the cost of the wall. When Type (3) blends with indivisibility in production, as it does in the case of the wall, or in the case of a lighthouse, the comparison has to be made between intramarginal totals. Where no lumpiness is involved (e.g., the decibels at which I play my radio) only *MRS* and perhaps *MC* calculations are called for. But the really crucial decision may well be about how much perfectly feasible appropriation and exclusion is desirable.

Arrangements to offset. It is of interest to speculate what, if any, organizational rearrangements could offset the three categories of externality and avoid the need for centrally calculated tax-subsidy schemes.[9] In concept, Type (1) can be offset by rearrangements of ownership and by "proper" bookkeeping, such as need not violate the structural requirements of decentralized competition. Further, no resort to nonmarket tests would be required.[1]

Types (2) and (3) are not so amenable to correction consistent with decentralized institutions. The easiest possible case occurs where increasing returns obtain on the level of single producers'-good plants, much of whose production can be absorbed by a single user firm. Here vertical integration takes care of the problem. Not every process inside a well-run firm is expected to cover its cost in terms of the correct set of internal accounting (shadow) prices. Total profits are the only criterion, and it may pay a firm to build a private bridge between its two installations on opposite sides of a river yet charge a zero accounting price for its use by the various decentralized manufacturing and administrative divisions; the bridge would make accounting losses, yet total company profits will have increased. As long, then, as such integration is consistent with the many-firms requirement for competition, no extra-market tests are required.[2] The private total conditions: *TR* less *TC*, correctly account for social gain.

9. For illustrative derivation of the formulas for corrective taxes and subsidies in Type (1) situations, see Meade (*op. cit.*).

1. The Emancipation Proclamation could constitute, of course, a substantial barrier.

2. If, however, the "break even" scale of operation of the integrated firm (i.e., where *MC* cuts *AC* from below) is much greater than if the river had not been there to span, or could be spanned by some means of a lower fixed-cost-to-variable-cost ratio, the monopoly problem may simply be "pushed forward" to consumer markets.

Where a producers'-good firm, required to produce at a stage of falling AC, sells to many customer firms and industries, an adding up of all the associated TR's and TC's at the precalculated "as if" competitive prices associated with the bliss point would again effectively "mop up" all social costs and benefits.[3] But the institutional reorganization required to get correct decentralized calculation involves horizontal and vertical integration, and the monopoly or oligopoly problem looms large indeed. The Type (3) case of a pure *producers'* public good belongs here: only input MRS's along production functions require summing.

In the general case of a mixed producer-consumer good (or of a pure consumer good) which is "public" or is produced under conditions of increasing returns to scale, it is impossible to avoid comparison of multiperson utility totals. Explicit administrative consideration must be given, if you like, to consumer's and producer's surpluses for which no market-institution tests exist short of that provided by a perfectly discriminating monopolist. But to invoke perfect discrimination is to beg the question. It implies knowledge of all preference functions, while as Samuelson has emphasized,[4] the crucial game-theoretical quality of the situation is that consumers will not correctly reveal their preferences: it will pay them to "cheat."

Blends. Examination is needed of various blends of Types (2) and (3), such as Sidgwick's lighthouse;[5] or, for that matter, and as suggested by Samuelson, of blends of public and private goods even where all production functions are fully convex. There are many puzzling cases. Do bridge crossings differ in kind from radio programs? Both involve indivisibility and, where variable cost is zero for the bridge, zero MC's. The correct price for an extra stroller, as for an extra listener, is clearly zero. Yet bridge crossings have a distinctly private quality: bridges get congested, physical capacity is finite. This is not true of a broadcast. There is no finite limit to the number

3. Assuming that all consumer goods are finely divisible and require no lumpy decisions by consumers.

4. Cf. any of the three "Public Expenditure" articles (*supra*).

5. Sidgwick, by the way, as also Pigou, thought of a lighthouse as of Type (1). It is, of course, "inconvenient" to levy tolls on ships, but it is hardly impossible to "exclude," for instance by means of "scrambling" devices (though poor Sidgwick could hardly have known about such things). The point is, rather, that it would be inefficient to do so: the marginal cost to society of an additional ship taking directional guidance from the beacon atop the Statue of Liberty is zero, *ipso* price should be zero. In the case of a lighthouse this is twice true: because the beacon is in the nature of a public good: more for the Queen Mary means no less for the Liberté; and because a lighthouse is virtually an all-fixed-cost, zero variable-cost facility.

of sets that can costlessly tune in.[6] Radio programs, then, have a public dimension. Yet, in a sense, so do bridges. While your bridge crossing is not my bridge crossing, in fact could limit my crossings, your bridge is my bridge. What is involved here is that most things are multidimensional and more than one dimension may matter.

V. Efficiency, Markets and Choice of Institutions

All the above has to do with the statical efficiency of price-directed allocation in more or less idealized market situations. Relevance to choice of institutions depends, of course, on the prevalence of the phenomena which cause externality and on the importance to be attached to statical efficiency. Space precludes extensive discussion of these important issues, but a few casual comments, in the form of *dicta*, are perhaps warranted.

How important are nonappropriability nonconvexity and public goods? I would be inclined to argue that while nonappropriability is of small import,[7] the same cannot be said of the other two. True enough, it is difficult to think of many examples of pure public goods. Most things — even battleships, and certainly open air concerts and schools (though not knowledge) — have an "if more for you then less for me" quality. But this is of little comfort. As long as activities have even a trace of publicness, price calculations are inefficient.[8] And it is surely hard to gainsay that some degree of public quality pervades much of even narrowly "economic" activity.

Lumpiness, in turn, and nonlinearity of the increasing returns sort, while in most instances a matter of degree, and, within limits, of choice, are also in the nature of things. The universe is full of singularities, thresholds and nonproportionalities: speed of light, gravitational constant, the relation of circumference to area, etc. As economists we can cajole or bully engineers into designing processes and installations that save on congealed inputs and give smaller maximal service yields, especially when designing for low-income communities. But the economically perhaps arbitrary, not completely physics-imposed quality of indivisibilities associated with standard designs

6. Richard Eckaus has suggested to me that it is possible to exhaust the space to which the broadcast is limited and that this makes the situation a little more like that of a bridge. Neither of us is entirely satisfied, however.

7. Except for labor skills — and these would take us beyond the bounds of reversible statics.

8. This is not to say that there exist other feasible modes of social calculation and organization which are more efficient.

and ways of doing things should not blind. Nonlinearity and lumpiness are evident facts of nature.[9]

More important, at this level of discourse[1] — though perhaps it hardly need be said — is that statical market efficiency is neither sufficient nor necessary for market institutions to be the "preferred" mode of social organization. Quite apart from institutional considerations, Pareto efficiency as such may not be necessary for bliss.[2] If, e.g., people are sensitive not only to their own jobs but to other people's as well, or more generally, if such things as relative status, power, and the like, matter, the injunction to maximize output, to hug the production-possibility frontier, can hardly be assumed "neutral," and points on the utility frontier may associate with points inside the production frontier.[3] Furthermore, there is nothing preordained about welfare functions which are sensitive only to individual consumer's preferences. As a matter of fact, few people would take such preferences seriously enough to argue against any and all protection of individuals against their own mistakes (though no external effects be involved).

All this is true even when maximization is subject only to technological and resource limitations. Once we admit other side relations, which link input-output variables with "noneconomic" political and organizational values, matters become much more complicated. If markets be ends as well as means, their nonefficiency is hardly sufficient ground for rejection.[4] On the other hand, efficient markets may

9. Their quantitative significance is, of course, very sensitive to scale, to "size" of markets. This explains the particular emphasis on the role of "social overheads" in low income countries.

1. Where recourse to strategic considerations of feasibility, crucial though they be, is quite out of order.

2. That it is never sufficient is, of course, well known. Of the infinite Pareto-efficient configurations at best only one: that which gives the "right" distribution of income in terms of the W-function that is to count, has normative, prescriptive significance. Moreover, most interesting W-functions are likely to be sensitive to "noneconomic" factors, such as are, if not inconsistent, at least extraneous to Paretian considerations. Where such additional values of a political or social nature are separable from input-output values (i.e., where the two sets can be varied independently of each other) one "can" of course separate the overall W-function into a "political" and an "economic" component and maximize separately over each.

3. This is different from the usual case of consumer sensitivity to the input-output configuration of producers, e.g., factory soot or a functional but ugly plant spoiling the view. Such joint-product "bads" can be treated as inputs and treated in the usual Paretian fashion. It is a different matter that their public quality will violate duality, hence render market calculation inefficient.

4. This is too crude a formulation. It is not necessary that markets as such be an "ultimate" value. Political and social (non-output) values relating to the configuration of power, initiative, opportunity, etc., may be so much better served

not do, even though Pareto-efficiency is necessary for bliss. Even with utopian lump-sum redistribution, efficiency of the "invisible hand" does not preclude preference for other efficient modes of organization, if there be any.[5]

Yet when all is said, and despite the host of crucial feasibility considerations which render choice in the real world inevitably a problem in the strategy of "second best," it is surely interesting and useful to explore the implications of Paretian efficiency. Indeed, much remains to be done. There is need, in particular, for more systematic exploration of the inadequacies of market calculation in a setting of growth.[6]

by some form of nonefficient market institutions than by possible alternative modes of more efficient organization as to warrant choice of the former. The analytical point, in all this, is that the outcome of a maximization process and the significance of "efficiency" are as sensitive to the choice of side-conditions as to the welfare-function and that these need be "given" to the economist in the same sense that a welfare function has to be given.

5. The above is still strictly statical. For related dynamical problems, e.g., possible conflict between one-period and intertemporal efficiency, cf., "On Capital Productivity, Input Allocation and Growth," this *Journal*, LXXI (Feb. 1957).

6. The development literature on market failure, while full of suggestive insight, is in a state of considerable confusion. Much work is needed to exhaust and elucidate the seminal ideas of Young, Rosenstein-Rodan, Nurkse and others. For important beginnings, see Scitovsky (*op. cit.*), M. Fleming, "External Economies and the Doctrine of Balanced Growth," *Economic Journal*, LXV (June 1955), and Fellner (*op. cit.*).

The view that we should not turn social historian or what not, that the logic of economizing has some prescriptive significance, rests on the belief that narrowly "economic" efficiency is important in terms of many politically relevant *W*-functions, and consistent with a wide variety of power and status configurations and modes of social organization. On the other hand, some may feel that the very language of Paretian welfare economics: "welfare function," "utility-frontier," in relation to choice of social institutions, is grotesque. What is at stake, of course, is not the esthetics of language, on which I yield without demur, but abstraction and rigorous theorizing.