CITATION: "WELFARE ECONOMICS," IN JOHN EATWELL, MURRAY MILGATE, AND PETER NEWMAN, EDS, *THE NEW PALGRAVE DICTIONARY OF ECONOMICS*, THE STOCKTON PRESS, NEW YORK, 1987, VOL. 4, PP. 889-895.

WELFARE ECONOMICS

W000031

In 1776, the same year as the American Declaration of Independence, Adam Smith published *The Wealth of Nations*. Smith laid out an argument that is now familiar to all economics students: (1) The principal human motive is self-interest. (2) The invisible hand of competition automatically transforms the self-interest of many into the common good. (3) Therefore, the best government policy for the growth of a nation's wealth is that policy which governs least.

Smith's arguments were at the time directed against the mercantilists, who promoted active government intervention in the economy, particularly in regard to (ill-conceived) trade policies. Since his time, his arguments have been used and reused by proponents of *laissez-faire* throughout the 19th and 20th centuries. Arguments of Smith and his opponents are still very much alive today: The pro-Smithians are those who place their faith in the market, who maintain that the provision of goods and services in society ought to be done, by and large, by private buyers and sellers acting in competition with each other. One can see the spirit of Adam Smith in economic policies involving deregulation of industries, tax reduction, and reduction in government growth in the United States; in policies of denationalization in the United Kingdom, France and elsewhere, and in the deliberate restoration of private markets in China. The anti-Smithians are also still alive and well; mercantilists are now called industrial policy advocates, and there is an abundance of intellectuals and policy makers, aside from neomercantilists, who believe that: (1) economic planning is superior to *laissez-faire*; (2) markets are usually monopolized in the absence of government intervention, crippling the invisible hand of competition; (3) even if markets are competitive, the existence of external effects, public goods, information asymmetries and other market failures ensure that *laissez-faire* results in the common bad rather than the common good; (4) and in any case, *laissez-faire* produces an intolerable degree of inequality.

The branch of economics called welfare economics is an outgrowth of the fundamental debate that can be traced back to Adam Smith, if not before. The theoretical side of welfare economics is organized around three main propositions. The first theorem answers this question: In an economy with competitive buyers and sellers, will the outcome be for the common good? The second theorem addresses the issue of distributional equity, and answers this question: In an economy where distributional decisions are made by an enlightened sovereign, can the common good be achieved by a slightly modified market mechanism, or must the market be abolished altogether? The third theorem focuses on the general issue of defining social welfare, or the common good, whether via the market, via a centralized political process, or via a voting process.

It answers this question: Does there exist a reliable way to derive from the interests of individuals, the true interests of society, regarding, for example, alternative distributions of wealth?

This entry focuses on theoretical welfare economics. There are related topics in practical welfare economics which are only mentioned here. A reader interested in the practical problems of evaluating policy alternatives can refer to entries on CONSUMERS' SURPLUS, COST-BENEFIT ANALYSIS and COMPENSATION PRINCIPLE, to name a few.

I. The First Fundamental Theorem, or Laissez-Faire Leads to the Common Good

> 'The greatest meliorator of the world is selfish, huckstering trade.' (R.W. Emerson, *Work and Days*)

In *The Wealth of Nations*, Book IV, Smith wrote: 'Every individual necessarily labours to render the annual revenue of the society as great as he can. He generally indeed neither intends to promote the public interest, nor knows how much he is promoting it … . He intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote as end which was no part of his intention.' The philosophy of the First Fundamental Theorem of Welfare Economics can be traced back to these words of Smith. Like much of modern economic theory, it is set in the context of a Walrasian general equilibrium model, developed almost a hundred years after *The Wealth of Nations*. Since Smith wrote long before the modern theoretical language was invented, he never rigorously stated, let alone proved, any version of the First Theorem. That honour fell upon Lerner (1934), Lange (1942) and Arrow (1951).

To establish the First Theorem, we need to sketch a general equilibrium model of an economy. Assume all individuals and firms in the economy are price takers: none is big enough, or motivated enough, to act like a monopolist. Assume each individual chooses his consumption bundle to maximize his utility subject to his budget constraint. Assume each firm chooses its production vector, or input–output vector, to maximize its profits subject to some production constraint. Note the presumption of self-interest. An individual cares only about his own utility, which depends on his own consumption. A firm cares only about its own profits.

The invisible hand of competition acts through prices; they contain the information about desire and scarcity that coordinate actions of self-interested agents. In the general equilibrium model, prices adjust to bring about equilibrium in the market for each and every good. That is, prices adjust until supply equals demand. When that has occurred, and all individuals and firms are maximizing utilities and profits, respectively, we have a competitive equilibrium.

The First Theorem establishes that a competitive equilibrium is for the common good. But how is the common good defined? The traditional definition looks to a measure of total value of goods and services produced in the economy. In Smith, the 'annual revenue of the society' is maximized. In Pigou (1920), following Smith, the 'free play of self-interest' leads to the greatest 'national dividend'.

However, the modern interpretation of 'common good' typically involves Pareto optimality, rather than maximized gross national product. When ultimate consumers appear in the model, a situation is said to be *Pareto optimal* if there is no feasible alternative that makes everyone better off. Pareto optimality is thus a dominance concept based on comparisons of vectors of utilities. It rejects the notion that utilities of different individuals can be compared, or that utilities of different individuals can be summed up and two alternative situations compared by looking at summed utilities. When ultimate consumers do not appear in the model, as in the pure production framework to be described below, a situation is said to be *Pareto optimal* if there is no alternative that results in the production of more of some output, or the use of less of some input, all else equal. Obviously saying that a situation is Pareto optimal is not the same as saying it maximizes GNP, or that it is best in some unique sense. There are generally many Pareto optima. However, optimality is a common good concept that can get common assent: No one would argue that society should settle for a situation that is not optimal, because if A is not optimal, there exists a B that all prefer.

In spite of the multiplicity of optima in a general equilibrium model, most states are non-optimal. If the economy were a dart board and consumption and production decisions were made by throwing darts, the chance of hitting an optimum would be zero. Therefore, to say that the market mechanism leads an economy to an optimal outcome is to say a lot. And now we can turn to a modern formulation of the First Theorem:

> *First Fundamental Theorem of Welfare Economics*: Assume that all individuals and firms are selfish price takers. Then a competitive equilibrium is Pareto optimal.

To illustrate the theorem, we focus on one simple version of it, set in a pure production economy. For a general version of the theorem, with both production and exchange, the reader can refer to Malinvaud (1972).

In a general equilibrium production economy model, there are K firms and m goods, but, for simplicity, no consumers. Given a list of market prices, each firm chooses a feasible input–output vector y_k so as to maximize its profits. We adopt the usual sign convention for a firm's input–output vector y_k:$y_{kj} < 0$ means firm k is a net *user* of good j, and $y_{kj} > 0$ means firm k is a net *producer* of good j. What is feasible for firm k is defined by some fixed production possibility set Y_k. Under the sign convention on the input–output vector, if p is a vector of prices, firm k's profits are given by

$$\pi_k = p \cdot y_k.$$

A list of feasible input–output vectors $y = (y_1, y_2, \ldots y_k)$ is called a *production plan* for the economy. A *competitive equilibrium* is a production plan $\hat{y}$ and a price vector p such that, for every $k, \hat{y}_k$ maximizes π_k's subject to y_k's being feasible. (Since the production model abstracts from the ultimate consumers of outputs and providers of inputs, the supply equals demand requirement for an equilibrium is moot).

If $y = (y_1, y_2, \ldots, y_k)$ and $z = (z_1, z_2, \ldots, z_k)$ are alternative production plans for the economy, z is said to *dominate* y if the following vector inequality holds:

$$\sum_k z_k \geq \sum_k y_k.$$

Finally, if there exists no production plan that dominates y, y is *Pareto optimal*. (The notational conventions are very important for this model; note for example that $y_{11} + y_{21} + \ldots + y_{k1}$ represents an aggregate amount of good 1 produced in the economy, if positive, and an aggregate amount of good 1 used, if negative. Note also that some y_{k1}'s might be positive and some negative, and that the direction of the vector inequality is 'right' whether good 1 is an input, in the aggregate, or an output).

We now have the apparatus to state and prove the First Theorem in the context of the pure production model:

> *First Fundamental Theorem of Welfare Economics, Production Version*. Assume that all prices are positive, and that $\hat{y}$, p is a competitive equilibrium. Then $\hat{y}$ is Pareto optimal.

To see why, suppose to the contrary that a competitive equilibrium production plan $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_k$ is not optimal. Then there exists a production plan $z_1, z_2, \ldots, z_k$ that dominates it. Therefore

$$\sum_k z_k \geq \sum_k \hat{y}_k .$$

Taking the dot product of both sides with the positive price vector p gives

$$p \cdot \sum_k z_k > p \cdot \sum_k \hat{y}_k .$$

But this implies that, for at least one firm k,

$$p \cdot z_k > p \cdot \hat{y}_k ,$$

which contradicts the assumption that $\hat{y}_k$ maximizes firm k's profits.

II. First Fundamental Theorem Drawbacks, and the Second Fundamental Theorem

> 'That amid our highest civilization men faint and die with want is not due to niggardliness of nature, but to the injustice of man.' (Henry George, *Progress and Poverty*)

The First Theorem of Welfare Economics is mathematically true but nevertheless objectionable. Here are the commonest objections: (1) The First Theorem is an abstraction that ignores the facts. Preferences of consumers are not given, they are created by advertising. The real economy is never in equilibrium, most markets are characterized by excess supply or excess demand, and are in a constant state of flux. The economy is dynamic, tastes and technology are constantly changing, whereas the model assumes they are fixed. The cast of characters in the real economy is constantly changing, the model assumes it fixed. (2) The First Theorem assumes competitive behaviour, whereas the real world is full of monopolists. (3) The First Theorem assumes there are no externalities. In fact, if in an exchange economy person l's utility depends on person 2's consumption as well as his own, the theorem does not hold. Similarly, if in a production economy firm k's production possibility set depends on the production vector of some other firm, the theorem breaks down.

In a similar vein, the First Theorem assumes there are no public goods, that is, goods like national defence or lighthouses, that are necessarily non-exclusive in use. If such goods are privately provided (as they would be in a completely *laissez-faire* economy), then their level of production will be sub-optimal. (4) The most troubling aspect of the First Theorem is its neglect of distribution. *Laissez-faire* may produce a Pareto optimal outcome, but there are many different Pareto optima, and some are fairer than others. Some people are endowed with resources that make them rich, while others, through no fault of their own, are without. The First Theorem ignores basic distributional questions: How should unfair distributions of goods be made fair? And on the production side, how should production plans that give heavy weight to luxury items for the rich, and little or no weight to food, housing and medical care for the poor, be put right?

The first and second objections to the First Theorem are beyond the scope of this entry. The third, regarding externalities and public goods, is one that economists have always acknowledged. The standard remedies for these market failures involve minor modifications of the market mechanism, including Pigovian taxes (Pigou, 1920) on harmful externalities, or appropriate Coasian (Coase, 1960) legal entitlements to, for example, clean air.

The important contribution of Pigou is set in a partial equilibrium framework, in which the costs and benefits of a negative externality can be measured in money terms. Suppose that a factory produces gadgets to sell at some market-determined price, and suppose that, as part of its production process, the factory emits smoke which damages another factory located downwind. In order to maximize its profits, the upwind factory will expand its output until its marginal cost equals price. But each additional gadget it produces causes harm to the downwind factory – the marginal external cost of its activity. If the factory manager ignores that marginal external cost, he will create a situation that is non-optimal in the sense that the aggregate net value of both firms' production decisions will not be as great as it could be. That is, what Pigou calls 'social net product' will not be maximized, although 'trade net product' for the polluting firm will be. Pigou's remedy was for the state to eliminate the divergence between trade and social net product by imposing appropriate taxes (or, in the case of beneficial externalities, bounties). The Pigovian tax would be set equal to marginal external cost, and with it in place the gap between the polluting firm's view of cost and society's view would be closed. Optimality would be re-established.

Coase's contribution was to emphasize the reciprocal nature of externalities and to suggest remedies based on common law doctrines. In his view the polluter damages the pollutee only because of their proximity, e.g., the smoking factory harms the other only if it happens to locate close downwind. Coase rejects the notion that the state must step in and tax the polluter. The common law of nuisance can be used instead. If the law provides a clear right for the upwind factory to emit smoke, the downwind factory can contract with the upwind factory to reduce its output, and if there are no impediments to bargaining, the two firms acting together will negotiate an optimal outcome. Alternatively, if the law establishes a clear right for the downwind factory to recover for smoke damages, it will collect external costs from the polluter, and thereby motivate the polluter to reduce its output to the optimal level. In short, a legal system that grants clear rights to the air to either the polluter or pollutee will set the stage for an optimal outcome, provided that transactions are costless.

With respect to public goods, since Samuelson (1954) derived formal optimality conditions for their provision, the issue has received much attention from economists; one especially notable theoretical question has to do with discovering the strengths of people's preferences for a public good. If the government supplies a public judicial system, for instance, how much should it spend on it (and tax for it)? At least since Samuelson, it has been known that financing schemes like those proposed by Lindahl (1919), where an individual's tax is set equal to his marginal benefit, provide perverse incentives for people to misrepresent their preferences. Schemes that are immune to such misrepresentations (in certain circumstances) have been developed in recent years (Clarke, 1971; Groves and Loeb, 1975).

But it is the fourth objection to the First Theorem that is most fundamental. What about distribution?

There are two polar approaches to rectifying the distributional inequities of *laissez-faire*. The first is the command economy approach: a centralized bureaucracy makes detailed decisions about the consumption decisions of all individuals and production decisions of all producers. The main theoretical problems with the command approach are that it requires the bureaucracy to obtain and act upon superhuman quantities of information, and that it fails to create appropriate incentives for individuals and firms. On the empirical side, the experience of Eastern European and Chinese command economies suggest that highly centralized economic decision making leaves much to be desired, to put it mildly.

The second polar approach to solving distribution problems is to transfer income or purchasing power among individuals, and then to let the market work. The only kind of purchasing power transfer that does not cause incentive-related losses is the lump-sum transfer. Enter at this point the standard remedy for distribution problems, as put forward by the market-oriented economist, and our second major theorem.

The Second Fundamental Theorem of Welfare Economics establishes that the market mechanism, modified by the addition of lump-sum transfers, can achieve virtually *any* desired optimal distribution. Under more stringent conditions than are necessary for the First Theorem, including assumptions regarding quasi-concavity of utility functions and convexity of production possibility sets, the Second Theorem asserts the following:

> *Second Fundamental Theorem of Welfare Economics*. Assume that all individuals and producers are selfish price takers. Then almost any Pareto optimal equilibrium can be achieved via the competitive mechanism, provided appropriate lump-sum taxes and transfers are imposed on individuals and firms.

One version of the Second Theorem, restricted to a pure production economy, is particularly relevant to an old debate about the feasibility of socialism, see particularly Lange and Taylor (1939) and Lerner (1944). Anti-socialists including Von Mises (1937) argued that informational problems would make it impossible to coordinate production in a socialist economy; while pro-socialists, particularly Lange, argued that those problems could be overcome by a Central Planning Board, which limited its role to merely announcing a price vector. This is called ‘decentralized socialism’. Given the prices, managers of production units would act like their capitalist counterparts; in essence, they would maximize profits. By choosing the price vectors appropriately, the Central Planning Board could achieve any optimal production plan it wished.

In terms of the production model given above, the production version of the Second Theorem is as follows:

> *Second Fundamental Theorem of Welfare Economics, Production Version*. Let $\hat{y}$ be any optimal production plan for the economy. Then there exists a price vector p such that $\hat{y}$, p is a competitive equilibrium. That is, for every k, $\hat{y}_k$ maximizes $\pi_k = p \cdot y_k$ subject to y_k being feasible.

The proof of the Second Theorem requires use of Minkowski’s separating hyperplane theorem, and will not be given here.

III. Tinkering with the Economy and Voting on Distributions

The logic of the Second Theorem suggests that it is all right, perhaps even morally imperative, to tinker with the economy. And after all, is not tinkering what is done by policy makers and their economic advisers? How often do we choose between a *laissez-faire* economy and a command economy? Our choices are usually more modest. When choosing among alternative tax policies, or trade and tariff policies, or antimonopoly policies, or labour policies, or transfer policies, what shall guide the choice? The applied welfare economist's advice is usually based on some notion of increasing total output in the economy. The practical political decision, in a Western democracy, is normally based on voting.

Applied Welfare Economics

The applied welfare economist usually focuses on ways to increase total output, 'the size of the pie', or at least to measure changes in the size of the pie. Unfortunately, theory suggests that the pie cannot be measured. This is so for a number of reasons. To start, any measure of total output is a scalar, that is, a single number. If the number is found by adding up utility levels for different individuals, illegitimate interpersonal utility comparisons are being made. If the number is found by adding up the values of aggregate net outputs of all goods, there is an index number problem. The value of a production plan will depend on the price vector at which it is evaluated. But in a general equilibrium context, the price vector will depend on the aggregate net output vector, which will in turn depend on the distribution of ownership or wealth among individuals. Economists have always agreed that if q^1 and q^2 are alternative aggregate net output vectors, and if p^1 and p^2 are the corresponding price vectors, then $p^1 \cdot q^1 < p^2 \cdot q^2$ has no welfare implications. Unfortunately they now also agree that if there are two or more individuals in the economy, even $p^2 \cdot q^1 < p^2 \cdot q^2$ may not signify q^2 is an improvement in welfare over q^1.

An early and crucial contribution to the analysis of whether or not the economic pie has increased in size was made by Kaldor (1939), who argued that the repeal of the Corn Laws in England can be justified on the grounds that the winners could in theory compensate the losers: 'it is quite sufficient [for the economist] to show that even if all those who suffer as a result are fully compensated for their loss, the rest of the community will still be better off than before'. Unfortunately, Scitovsky (1941) quickly pointed out that Kaldor's compensation criterion (as well as one proposed by Hicks) was in theory inconsistent: it is possible to judge situation B Kaldor superior to A and simultaneously judge A Kaldor superior to B. The Scitovsky paradox can be avoided via a two-edged compensation test, according to which situation B is judged better than A if (1) the potential gainers in the move from A to B could compensate the potential losers, and still remain better off, and (2) the potential losers could not bribe the gainers to forego the move.

Scitovsky's two-edged criterion has some logical appeal, but it, like the single-edged Kaldor criterion, still has a major drawback: it ignores distribution. Therefore, it can make no judgement about alternative distributions of the same size pie. And worse, as was pointed out by Little (1950), either criterion would approve a change that would make the wealthiest man in England richer by £1,000,000,000, while making each of the 1,000,000 humblest men poorer by £900. In Little's view, the applied welfare economist should adopt Scitovsky's two-edged criterion and *also* requires that the change from A to B not result in a worse distribution of welfare. Unfortunately, what constitutes a worse distribution is, as Little concedes, purely a value judgement – a matter of personal opinion.

Another important tool for measuring changes in the economic pie is the concept of consumer's surplus, which Marshall (1920) defined as the difference between what an individual would be willing to pay for an object, at most, and what he actually does pay. With a little faith, the economic analyst can measure aggregate consumers' surplus (note the new position of the apostrophe), by calculating an area under a demand curve, and this is in fact commonly done in order to evaluate changes in economic policy. The applied welfare economist attempts to judge whether the pie would grow in a move from A to B by examining the change in consumers' surplus (plus profits, if they enter the analysis). Faith is required because consumers' surplus, like the Kaldor criterion, has been shown to be theoretically inconsistent; see for example Boadway (1974).

In short, although the tools of applied welfare economics are crucially important in practice, theory says they must be viewed with suspicion.

Voting

> 'A minority may be right, a majority is always wrong.' (Henrik Ibsen, *An Enemy of the People*)

In most cases, interesting decisions about economic policies are made either by bureaucracies that are controlled by legislative bodies, or by legislative bodies themselves, or by elected executives. In short, either directly or indirectly, by voting. The Second Theorem itself raises questions about distribution that many would view as essentially political: How should society choose the Pareto-optimal allocation of goods that is to be reached via the modified competitive mechanism? How should the distribution of income be chosen? How can the best distribution of income be chosen from among many Pareto optimal ones? Majority voting is the most commonly used method of political choice in a democracy.

The practical objections to voting, the fraud, the deception, the accidents of weather, are well known. To quote Boss Tweed, the infamous chief of New York's Tammany Hall: 'As long as I count the votes, what are you going to do about it?' But let us turn to the theoretical problems.

The central theoretical fact about majority voting has been known since the time of Condorcet's *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, published in 1785: Voting may be inconsistent. The now standard Condorcet voting paradox assumes three individuals 1, 2 and 3, and three alternatives *x, y* and *z*, where the three voters have the following preferences:

$$\begin{array}{llll} 1: & x & y & z \\ 2: & y & z & x \\ 3: & z & x & y \end{array}$$

(Following an individual's number the alternatives are listed in his order of preference, from left to right.) Majority voting between pairs of alternatives will reveal that *x* beats *y*, *y* beats *z*, and, paradoxically, *z* beats *x*.

Recently it has become clear that such voting cycles are not peculiar; they are generic, particularly when the alternatives have a spatial aspect with two or more dimensions (Plott, 1967; Kramer, 1973.) This can be illustrated by taking the alternatives to be different distributions of one economic pie. Suppose, in other words, that the distributional issues raised by the First and Second Theorems are to be 'solved' by majority voting, and assume for simplicity that what is to be divided is a fixed total of wealth, say 100 units worth.

Now let *x* be 50 units for person 1, 30 units for person 2 and 20 units for person 3. That is, let *x* = (50, 20, 30). Similarly, let *y* = (30 50, 20) and *z* = (20, 30, 50). The result is that our three individuals have precisely the voting paradox preferences. Nor is this result contrived, it turns out that *all* the distributions of 100 units of wealth are connected by endless voting cycles (see McKelvey, 1976). The reader can easily confirm that for any distributions *u* and *v*, that he may choose, there exists a voting sequence from *u* to *v*, and another back from *v* to *u*!

The reality of voting cycles should give pause to the economist who studies or recommends tax bills. And it is most disturbing for the economist looking for a political basis for judging among alternative distributions.

IV. Social Welfare and the Third Fundamental Theorem

How then might the distribution problem be solved? One potential answer is to assert the existence of a Bergson (1938) Economic Welfare Function $E(\cdot)$, that depends on the amounts of non-labour factors of production employed by each producing unit, the amounts of labour supplied by each individual, and the amounts of produced goods consumed by each individual. Then solve the problem by maximizing $E(\cdot)$. If necessary conditions for Pareto optimality are derived that must hold for any $E(\cdot)$, this exercise is harmless enough; but if a *particular* $E(\cdot)$ is assumed and distributional implications are derived from it, then an objection can be raised: Why that $E(\cdot)$ and not another one?

De V. Graaff (1957) focuses Bergson's approach by analysing welfare functions of the 'individualistic' type: these can be written $W(u^1, u^2,\ldots, u^n)$ where u^i represents person *i*'s utility level. Graaff makes clear that maximizing a too broadly defined $W(\cdot)$ simply rediscovers the conditions for Pareto optimality, whereas maximizing a too narrowly defined $W(\cdot)$ simply rediscovers the preferences of the economist who invents $W(\cdot)$! Thus a good $W(\cdot)$ is neither too broadly nor too narrowly defined; rather it captures some widely shared judgements about which distributions are desirable and which are not. Maximizing such a welfare function implies both Pareto optimality and an appropriate distribution of wealth. But can a good $W(\cdot)$ function be discovered? Graaff is optimistic that the members of society can agree on the degre of equality to be incorporated in $W(\cdot)$. However, $W(\cdot)$ must also incorporate assumptions about an appropriate horizon (do we include unborn children?), as well as attitudes towards uncertainty, time discounting, and so on. And on these issues, he believes it extremely unlikely that enough agreement can be found to build a $W(\cdot)$. So, at the end of an illuminating book on normative economics, Graaff recommends that we all try positive economics. Which still leaves us with the Bergson social welfare function dilemma: Where do they come from?

In his classic monograph *Social Choice and Individual Values* (1963), Arrow brings together both the economic and political streams of thought sketched above. Arrow's theorem can be viewed in several ways: it is a statement about the distributional questions raised by the First and Second Theorems; it is a remarkable logical extension of the Condorcet voting paradox; and it is a statement about the logic of choice of Bergson welfare functions, and about the logic of compensation tests, consumers' surplus tests, and indeed all the tools of the applied welfare economist. Because of its importance, Arrow's theorem can be justifiably called the Third Fundamental Theorem of Welfare Economics.

Arrow's analysis is at a high level of abstraction, and requires some additional model building. We now assume a given set of alternatives, which might be allocations in an exchange economy, distributions of wealth, tax bills in a legislature, or even candidates in an election. The alternatives are written *x*, *y*, *z*, etc. We assume there is a fixed society of individuals, numbered 1, 2,…, *n*. Let R_i represent the preference relation of individual *i*, so xR_iy means person *i* likes *x* as well as or better than *y*. A preference profile for society is a specification of preferences for each and every individual, or symbolically, R_1, R_2, …, R_n. We shall write *R* for *society*'*s* preference relation, arrived at in a way yet to be specified. *R* is, of course, a much modernized version of Bergson's $E(\cdot)$, appearing here as a binary relation rather than as a function.

Arrow was concerned with the logic of how individual preferences are transformed into social preferences. That is, how is *R* found? Symbolically we can represent the transformation this way:

$$R_1, R_2, \ldots, R_n \rightarrow R.$$

Now if society is to make decisions regarding distributions, it must 'know' when one alternative is as good as or better than another, even if both are Pareto optimal. To ensure it can make such decisions, Arrow assumes that *R* is *complete*. That is, for any alternatives *x* and *y*, either *xRy* or *yRx* (or both, if society is indifferent between the two). If society is to avoid the illogic of cyclical voting, its preference ought to be *transitive*. That is, for any alternatives *x, y* and *z*, if *xRy* and *yRz*, then *xRz*. Following Sen (1970), we call a transformation of individual preference relations into a complete and transitive social preference relation an Arrow Social Welfare Function, or more briefly, an Arrow function.

Anyone can make up an Arrow function, just as anyone can make up a Bergson function, or for that matter a judgement about when one distribution of wealth is better than another. But arbitrary judgements are unsatisfactory and so are arbitrary Arrow functions. Therefore, Arrow imposed some reasonable conditions on his function. Following Sen's (1970) version of Arrow's theorem, there are four conditions: (1) *Universality*. The function should always work, no matter what individual preferences might be. It would not be satisfactory, for example, to require unanimous agreement among all the individuals before determining social preferences. (2) *Pareto consistency*. If everyone prefers *x* to *y*, then the social preference ought to be *x* over *y*. (3) *Independence*. Suppose there are two alternative preference profiles for individuals in society, but suppose individual preferences regarding *x* and *y* are exactly the same under the two alternatives. Then the social preference regarding *x* and *y* must be exactly the same under the two alternatives. In particular, if individuals change their minds about a third 'irrelevant' alternative, this should not affect the social preference regarding *x* and *y*. (4) *Non-dictatorship*. There should not be a dictator. In Arrow's abstract model, person *i* is a *dictator* if society always prefers exactly what he prefers, that is, if the Arrow function transforms R_i into *R*.

An economist or policy maker who wants an ultimate answer to questions involving distribution, or questions involving choices among alternatives that are not comparable under the Pareto criterion, could use an Arrow Social Welfare Function for guidance. Unfortunately, Arrow showed that imposing conditions 1 to 4 guarantees that Arrow functions *do not exist*:

> *Third Fundamental Theorem of Welfare Economics*. There is no Arrow Social Welfare Function that satisfies the conditions of universality, Pareto consistency, independence, non-dictatorship.

In order to illustrate the logic of the theorem, we will use a somewhat stronger assumption than independence. This assumption is called N–I–M, or *neutrality–independence–monotonicity*: Let *V* be a group of individuals. Suppose for some preference profile and some particular pair of alternatives *x* and *y*, all members of *V* prefer *x* to *y*, all individuals *not* in *V* prefer *y* to *x*, and the social preference is *x* over *y*. Then for *any* preference profile and *any* pair of alternatives *x* and *y*, if all people in *V* prefer *x* to *y*, the social preference must be *x* over *y*. In short, if *V* gets its way in one instance, when everyone opposes it, then it must have the power to do it again, under other possibly less difficult circumstances.

A group of individuals *V* is said to be *decisive* if for all alternatives *x* and *y*, whenever all the people in *V* prefer *x* to *y*, society prefers *x* to *y*. Assumption N–I–M asserts that if *V* prevails when it is opposed by everyone else, it must be a decisive group. If the social choice procedure is majority rule, for example, any group of $(n+1)/2$ members, for *n* odd, or $(n/2)=1$ members, for *n* even, is decisive. Moreover, it is clear that majority rule satisfies the N–I–M assumption, since if *V* prevails for a particular *x* and *y* when everyone outside of *V* prefers *y* to *x*, then *V* must be a majority, and must always prevail. (Majority rule is just one example of a procedure that satisfies N–I–M: there are countless other procedures that also do so.)

Now we are ready to turn to a short version of the Third Theorem

> *Third Fundamental Theorem of Welfare Economics, Short Version*. There is no Arrow Social Welfare Function that satisfies the conditions of universality, Pareto consistency, neutrality–independence–monotonicity, and non-dictatorship.

The logic of the proof is as follows: First, there must exist decisive groups of individuals, since by the Pareto consistency requirement the set of all individuals is one. Now let *V* be a decisive group of minimal size. If there is just one person in *V*, he is a dictator. Suppose then that *V* includes more than one person. We show this leads to a contradiction.

If there are two or more people in *V*, we can divide it into non-empty subsets V_1 and V_2. Let V_3 represent all the people who are in neither V_1 nor V_2. (V_3 may be empty). By universality, the Arrow function must be applicable to any profile of individual preferences. Take three alternatives *x*, *y* and *z* and consider the following preferences regarding them:

$$\text{For individuals in } V_1: \quad x \quad y \quad z$$
$$\text{For individuals in } V_2: \quad y \quad z \quad x$$
$$\text{For individuals in } V_3: \quad z \quad x \quad y$$

(At this point the close relationship between Arrow and Condorcet is clear, for these are the voting paradox preferences!)

Since V is by assumption decisive, y must be socially preferred to z, which we write yPz. By the assumption of completeness for the social preference relation, either xRy or yPx must hold. If xRy holds, since xRy and yPz, then xPz must hold by transitivity. But now V_1 is decisive by the N–I–M assumption, contradicting V's minimality. Alternatively, if yPx holds, V_2 is decisive by the N–I–M assumption, again contradicting V's minimality. In either case, the assumption that V has two or more people leads to a contradiction. Therefore V must contain just one person, who is, of course, a dictator!

Since the Third Theorem was discovered, a whole literature of modifications and variations has been spawned. But the depressing conclusion has remained more or less inescapable: there is no logically infallible way to aggregate the preferences of diverse individuals. By extension, there is no logically infallible way to solve the problem of distribution.

Where does welfare economics stand today? The First and Second Theorems are encouraging results that suggest the market mechanism has great virtue: competitive equilibrium and Pareto optimality are firmly bound. But measuring the size of the economic pie, or judging among divisions of it, leads to the paradoxes and impossibilities summarized by the Third Theorem. And this is a tragedy. We feel we know, like Adam Smith Knew, which policies would increase the wealth of nations. But because of all our theoretic goblins, we can no longer prove it.

Allan M. Feldman

See also ***compensation principle; pigou, arthur cecil; public finance; social choice.***

Bibliography

Arrow, K.J. 1951. An extension of the basic theorems of classical welfare economics. *Second Berkeley Symposium on Mathematical Statistics and Probability,* ed. J. Neyman, Berkeley:University of
California Press, 507–32.

Arrow, K.J. 1963. *Social Choice and Individual Values*. 2nd edn, New York: John Wiley and Sons.

Bergson, A. 1938. A reformulation of certain aspects of welfare economics. *Quarterly Journal of Economics* 52, 310–34.

Boadway, R. 1974. The welfare foundations of cost-benefit analysis. *Economic Journal* 84, 926–39.

Clarke, E.H. 1971. Multipart pricing of public goods. *Public Choice* 11, 17–33.

Coase, R.H. 1960. The problem of social cost. *Journal of Law and Economics* 3, 1–44.

Graaff, J. de V. 1957. *Theoretical Welfare Economics*. Cambridge University Press.

Groves, T. and Loeb, M. 1975. Incentives and public inputs. *Journal of Public Economics* 4, 211–26.

Kaldor, N. 1939. Welfare propositions of economics and interpersonal comparisons of utility. *Economic Journal* 49, 549–52.

Kramer, G.H. 1973. On a class of equilibrium conditions for majority rule. *Econometrica* 41, 285–97.

Lange, O. 1942. The foundations of welfare economics. *Econometrica* 10, 215–28.

Lange, O. and Taylor, F.M. 1939. *On the Economic Theory of Socialism*. Minneapolis: University of Minnesota Press.

Lerner, A.P. 1934. The concept of monopoly and the measurement of monopoly power. *Review of Economic Studies* 1, 157–75.
Lerner, A.P. 1944. *The Economics of Control*. New York: The Macmillan Company.
Lindahl, E. 1919. Just taxation – a positive solution. Translated and reprinted in *Classics in the Theory of Public Finance*, ed. R.A. Musgrave and A.T. Peacock, New York: Macmillan, 1958.
Little, I.M.D. 1950. *A Critique of Welfare Economics*. Oxford: Oxford University Press.
Malinvaud, E. 1972. *Lectures on Microeconomic Theory*. Amsterdam: North-Holland.
Marshall, A. 1920. *Principles of Economics*. 8th edn, London: Macmillan, ch. VI.
McKelvey, R. 1976. Intransitivities in multidimensional voting models and some implications for agenda control. *Journal of Economic Theory* 12, 472–82.
Mises, L. von. 1922. *Socialism: An Economic and Social Analysis*. 3rd edn, trans., Indianapolis: Liberty Classics, 1981.
Pigou, A.C. 1920. *The Economics of Welfare*. London: Macmillan, Part II.
Plott, C.R. 1967. A notion of equilibrium and its possibility under majority rule. *American Economic Review* 57, 787–806.
Samuelson, P.A. 1954. The pure theory of public expenditure. *Review of Economics and Statistics* 36, 387–9.
Scitovsky, T. 1941. A note on welfare propositions in economics. *Review of Economic Studies* 9, 77–88.
Sen, A.K. 1970. *Collective Choice and Social Welfare*, San Francisco: Holden-Day.