WELFARE ECONOMICS

(DRAFT, SEPTEMBER 22, 2006)

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, tax reduction, denationalizing industries, and reduction in government growth in western countries; and in the deliberate restoration of private markets in China, the former Soviet Union, and other eastern European countries. The anti-Smithians are also still alive and well; mercantilists are now called industrial policy advocates, and there are intellectuals and policy makers who believe that: (1) economic planning is superior to *laissez-faire*; (2) markets are often 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* will not bring about the common good; (4) and in any case, *laissez-faire* may produce 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. It is the economic theory of measuring and promoting social welfare.

This entry is largely organized around three propositions. The first answers this question: In an economy with competitive buyers and sellers, will the outcome be for the common good? The second 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 abandoned? The third 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 the true interests of society, regarding, for example, alternative distributions of income or wealth, from the preferences of individuals?

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 an end which was no part of his intention.' The First Fundamental Theorem of Welfare Economics can be traced back to these words of Smith. Like much of modern economic theory, the First Theorem 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 mathematical language of economics was invented, he never rigorously stated, let alone proved, any version of the First Theorem. That was first done by 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 that we assume *self-interest*, or the absence of *externalities:* An individual cares only about his own utility, which depends only on his own consumption. A firm cares only about its own profits, which depend only on its own production vector.

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 self-interested 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 versions of the theorem, with both production and exchange, the reader can refer to Mas-Colell, Whinston & Green (1995).

In a general equilibrium production economy model, there are K firms and m goods, but, for simplicity, no consumers. We write $k = 1, 2, ..., K$ for the firms, and $j = 1, 2, ..., m$ for the goods. 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. When we add the amounts of good j over all the firms, $y_{1j} + y_{2j} + \ldots + y_{Kj}$, we get the aggregate net amount of good j produced in the economy, if positive, and an aggregate net amount of good j used, if negative. 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 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:

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

The production plan *y* is said to be *Pareto optimal* if there is other production plan that dominates it. (Note that for two vectors *a* and *b*, $a \geq b$ means $a_j \geq b_j$ for every good *j*, with the strict inequality holding for at least one good.)

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 $\hat{y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_K)$ is not optimal. Then there exists a production plan $z = (z_1, z_2, \ldots, z_K)$ 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. Q.E.D.

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

The First Theorem of Welfare Economics is mathematically true but nevertheless open to objections. Here are the commonest: (1) The 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 theorem assumes competitive behaviour, whereas the real world is full of monopoly and market power. (3) The 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 theorem assumes there are no public goods, that is, goods like national defense, judicial systems, 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 theorem ignores 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 extremely rich, while others, through no fault of their own, are extremely poor.

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 various modifications of the market mechanism, including Pigovian taxes (Pigou, 1920) on harmful externalities, or appropriate Coasian legal entitlements to, for example, clean air (Coase, 1960).

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 bargaining is costless. If bargaining is costly, then the law should be designed with an eye towards minimizing social costs created by the externality.

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 (Clarke, 1971; Groves and Loeb, 1975).

But it is the fourth objection to the First Theorem that may be 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 central bureaucracy makes detailed decisions about the consumption decisions of all individuals and production decisions of all producers. The main theoretical problem with the command approach is that it fails to create appropriate incentives for individuals and firms. On the empirical side, the experience of the late Soviet and Maoist command economies establish 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 money transfer. Enter at this point the standard remedy for distribution problems, as put forward by market-oriented economists, 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 gives the following:

> *Second Fundamental Theorem of Welfare Economics*. Assume that all individuals and producers are self-interested 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 was 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 will not be presented here.

III. Adjusting the Economy and Voting

We rarely choose between a *laissez-faire* economy and a command economy. Our choices are almost always more modest. When choosing among alternative tax policies, or trade and tariff policies, or development 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 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 easily 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.

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 could be justified on the grounds that the winners might 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 around the same time by Hicks) was inconsistent: Consider a move from situation A to situation B. It is possible to judge B Kaldor superior to A (the move is an improvement) and simultaneously judge A Kaldor superior to B (the move back would also be an improvement). This Scitovsky paradox can be avoided via a two-edged compensation test, according to which 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.

However, while Scitovsky's two-edged criterion has some logical appeal, it still has a major drawback: it ignores distribution. Therefore, it can make no judgement about alternative distributions of the same size pie. Even worse, both the Kaldor and the Scitovsky criteria would approve of a change that makes the wealthiest man in society richer by $1 billion, while making each of the million poorest people worse off by $999. This is an judgment that many people would reject as wrong or immoral.

Another important tool for measuring changes in economic welfare 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). Some faith is required because consumers' surplus, like the Kaldor criterion, is theoretically inconsistent; see for example Boadway (1974).

Under certain circumstances, however, consumers' surplus inconsistencies can be ruled out. In particular, if individual utility functions are all quaisilinear, of the form

$u_i(x_i) = v_i(x_{ij,j\neq m}) + x_{im}$, then consumers' surplus paradoxes disappear. (Here $u_i(x_i)$ is person *i*'s utility, as a function of his consumption bundle $x_i = (x_{i1}, x_{i2}, ..., x_{im})$, and the utility function can be separated into two parts, the first one of which is a function $v_i(\cdot)$ which depends on quantities of all the goods *except* the mth, and the second of which is simply the quantity of the mth good. The mth good can be interpreted as the "money" good; all the individuals like it, and value it the same way, with the same marginal utility, of one.) The assumption of quasilinear preferences is a very strong one if we think about "real" commodities like wine and bread, but it has a certain intuitive appeal if we are inclined to believe in utility from "money."

To sum up this section, although the tools of applied welfare economics are widely used and very important in practice, in theory they should be viewed with some skepticism.

Voting

In many cases, interesting decisions about economic policies are made either by government 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 rule is a commonly used method of choice in a democracy, both for political choices and economic ones, and we now turn our attention to it.

The practical objections to voting, the fraud, the deception, the accidents of weather, are well known. To quote Boss Tweed, the infamous 19th century chief of New York's Tammany Hall: 'As long as I count the votes, what are you going to do about it?' We will examine the theoretical problems.

The central theoretical problem with 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 logically 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:

1: *x* *y* *z*

2: *y* *z* *x*

3: *z* *x* *y*

(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*.

It is now 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.

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 recommends legislation about economic choices, especially choices among alternative distributions of income or wealth.

IV. Social Welfare and the Third Fundamental Theorem

How then can economic choices be made; how, for example, 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 Bergson function $E(\cdot)$, and not a different one?

At first, in his modestly titled "A Difficulty in the Concept of Social Welfare" (1950), and later, in his classic monograph *Social Choice and Individual Values* (1963), Kenneth Arrow brought 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 an extension of the Condorcet voting paradox; it is a statement about the logic of voting; and it is a statement about the logic of Bergson welfare functions, compensation tests, consumers' surplus tests, and indeed all the tools of the applied welfare economist. Because of its central 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. From this point onward we assume a given set of at least 3 distinct alternatives, which might be allocations in an exchange economy, distributions of wealth, tax bills in a legislature, or candidates in an election. The alternatives are written x, y, z, etc. We assume 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. (Strict preference is shown with a P_i, and indifference with a I_i.) A *preference profile* for society is a specification of preferences for each and every individual, or symbolically, $R = (R_1, R_2, ..., R_n)$. We write R_S for *society's* preference relation, arrived at in a way yet to be specified.

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

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

Now if society is to make decisions regarding things like distributions of wealth, 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 requires that R_S be *complete*. That is, for any alternatives *x* and *y*, either xR_Sy or yR_Sx (or both, if society is indifferent between the two). If society is to avoid the illogic of voting cycles, its preferences ought to be *transitive*. That is, for any alternatives *x*, *y* and *z*, if xR_Sy and yR_Sz, then xR_Sz. Following Sen (1970), we call a transformation of preference profiles into complete and transitive social preference relations an *Arrow social welfare function*, or more briefly, an Arrow SWF.

Anyone can make up an Arrow SWF, just as anyone can make up a Bergson function, or for that matter a simple moral judgment about when one distribution of wealth is better than another. But arbitrary judgments 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 directly into R_S.

An economist or policy maker who wants an ultimate answer to questions involving distributions, or questions involving choices among alternatives that are not comparable under the Pareto criterion, could use an Arrow SWF 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 NIM, or *neutrality–independence–monotonicity*, defined as follows: 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 *w* and *z*, if all people in *V*

prefer *w* to *z*, the social preference must be *w* over *z*. In short, if *V* gets its way in one instance, when everyone opposes it, then it must have the power to do it again, when the opposition may be weaker.

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 NIM asserts that if *V* prevails when it is opposed by everyone else, it must be decisive. 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 NIM 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 NIM; there are many other procedures that also do so.)

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

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

The proof goes 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:

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

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

Since *V* is by assumption decisive, *y* must be socially preferred to *z*, which we write yP_Sz. By the assumption of completeness for the social preference relation, either xR_Sy or yP_Sx must hold. If xR_Sy holds, since xR_Sy and yP_Sz, then xP_Sz must hold by transitivity. But now V_1 is decisive by the NIM assumption, contradicting *V*'s minimality. Alternatively, if yP_Sx holds, V_2 is decisive by the NIM 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! Q.E.D.

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 the same: there is no logically infallible way to aggregate the preferences of diverse individuals into a social preference relation. Therefore, there are no logically infallible ways to vote, or to solve the problems of distribution of income and wealth in society.

V. Social Welfare After Arrow

Social Choice Functions and Strategy

The Third Fundamental Theorem of Welfare Economics tells us that we cannot find an Arrow social welfare function satisfying certain reasonable requirements. An Arrow function maps preference profiles (that is, preference relations for each and every member of society) into social preference relations. But in order to make judgments about what alternative is *best* for society, it is not really necessary to have a social preference relation. Suppose we just had a rule that tells us, if the set of alternatives is *x*, *y*, *z*, etc., and the preference profile is $R = (R_1, R_2, ..., R_n)$, then the best alternative is such-and-such? Such a rule would be a mapping from preference profiles into alternatives, written symbolically as follows:

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

Such a rule is called a *social choice function*, or SCF for short. An Arrow function produces a social ranking of all the alternatives; an SCF in contrast, just produces a winner. As an example, think of plurality voting, with some kind of rule to break ties.

The essential difficulty with SCF's is that they may create obvious incentives for people to misrepresent their preferences, so as to obtain better (for them) social choices. As an example, consider again the Condorcet voting paradox preferences:

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

Suppose the three people use plurality voting (each person casts one vote for his favorite), and, in case of a tie, the social choice is the outcome closest to the beginning of the alphabet. Under this rule, if 1 votes for his favorite, *x*, and persons 2 and 3 do likewise, there is a 3-way tie, which is resolved with the (alphabetical) choice of *x*. Now put yourself in the shoes of person 2. You will immediately see that, if persons 1 and 3 continue to vote for their favorites, and if you switch from your favorite *y* to your second favorite *z*, then social choice changes, from *x* to *z*, making you better off! You are in effect being asked "what is your preference relation?" Instead of answering honestly (*y z x*), you offer, in effect, a false preference relation (*z y x*).

Reporting a false preference relation in order to bring about an SCF outcome that you prefer to the one you get if you are honest, is called *strategic behavior,* or *strategizing*. It is obviously a bad thing if an SCF produces lots of opportunities for strategic behavior: if individuals are commonly strategizing, there is no reason to believe that the outcome, based as it is on false reports, is truly best for society. If an SCF has the property that it is never advantageous for anyone to report a false preference relation it is called *strategy-proof*. For instance, suppose an SCF always chooses the alternative that is first in the alphabetical list of alternatives. This SCF would be frustrating and idiotic, but it would be strategy-proof.

The Gibbard-Satterthwaite Theorem

This leads to a natural question: Are there SCF's that are immune to strategic behavior, and that satisfy a few other reasonable conditions? Note that the question is

very similar in style to the question that Arrow asked about Arrow SWF's. What would the reasonable conditions be? First (similar to Arrow), the SCF ought to be *universal*; that is, it should work no matter what the profile of individual preferences might be. Second (also similar to Arrow), there should be no dictator. In the SCF context person *i* is a *dictator* if the social choice is always a top-ranked alternative for person *i*. Third (and different from Arrow), the SCF should be *non-degenerate*. This means that for any alternative *x*, there must be some preference profile which would give rise to *x*'s being the social choice. (This requirement excludes the SCF that always chooses the first alternative in the alphabetical list.) Now we can ask the question: Do there exist SCF's which are universal, non-degenerate, strategy-proof , and non-dictatorial?

This question was asked and answered, independently, by Gibbard (1973) and Satterthwaite (1975). The Gibbard-Satterthwaite result turns out to be logically very close to Third Fundamental Theorem of Welfare Economics; in fact Gibbard uses Arrow's theorem to prove his theorem, and Satterthwaite shows that his theorem can be used to prove Arrow's. Following is the Gibbard-Satterthwaite theorem. The proof is omitted; a simplified and restricted version of the theorem, and a simple proof, can be found in Feldman and Serrano (2006):

> *Gibbard-Satterthwaite Theorem*. There is no social choice function that satisfies the conditions of universality, non-degeneracy, strategy-proofness, and non-dictatorship.

Like the Third Fundamental Theorem, the Gibbard-Satterthwaite theorem is starkly negative; it says that if you want a decision-making process, an SCF to be precise, that has desirable characteristics, including being immune to strategic manipulation, you are bound to be disappointed. To put it differently, for any reasonable SCF, there will be circumstances under which some person will want to falsely report his preferences, resulting in a perversion of the process, and an outcome that may not be desirable for society.

If a decision-making process works in a way that offers each individual *no* incentive to misrepresent his preferences, no matter what preferences the other *n*-1 individuals might be reporting, we say that honestly reporting one's preferences (or telling the truth) is a *dominant strategy*. The Gibbard-Satterthwaite result then says that if a social choice function satisfies the conditions of universality, non-degeneracy and non-dictatorship, truth-telling will not be a dominant strategy. That is, there will be some reported preference relations of all individuals except *i*, which will provide an incentive for individual *i* to lie. If everyone else is saying such-and-such (which might be true or false), person *i* will give a false report. This is what strategy-proofness excludes.

But what if we narrowed this broad notion of strategy-proofness; what if we required that *i* not have an incentive to lie when the others are reporting the truth, rather than requiring that *i* never have an incentive to lie, no matter what the others are reporting?

Implementation and the Maskin Theorem

If telling the truth is a best strategy when others are telling the truth, rather than always, then truth telling is a *Nash strategy*, rather than a *dominant strategy*. The theory of implementation, or mechanism design, provides a way out the negativity of the

Gibbard-Satterthwaite theorem; it provides a way to *implement* an SCF, or support its choices, by incorporating truth telling about preferences into Nash equilibrium strategies in games.

The major theorem on implementation is due to Maskin (1999), whose paper first circulated in 1977. In Maskin's model, there is a social planner (or central authority) who wants to bring about choices according to a given SCF, which we now call *F*. The planner knows *F*, as do all the members of society. Given any preference profile *R*, the SCF produces an outcome $F(R) = x$. But the planner does not know the true preferences of the individuals. He must rely on the individuals to report their preferences, and they may lie. We assume for simplicity that every individual knows the true preference relation for himself *and* every other individual; that is, each *i* knows the true preference profile, but the social planner doesn't. (This obviously a strong assumption.) From this point on, when one preference profile may be true and another may be false, we will use the unadorned *R* to represent the *true* profile. The social planner receives reports on preferences, or preference profiles, from the individuals, but they may be lies. We let $\hat{R}_i$ represent a reported preference *relation for* person *i*, which may be false; similarly $\hat{R} = (\hat{R}_1, \hat{R}_2, ..., \hat{R}_n)$ represents a reported preference *profile*, which may be false; and $\hat{R}^i = (\hat{R}_1^i, \hat{R}_2^i, ..., \hat{R}_n^i)$ represents a preference *profile, reported by person i*, which may false. The social planner wants to devise a method, a mechanism, to induce individuals to honestly report preferences. That way he will get hold of the true preference profile *R*, and produce the desired outcome $F(R) = x$.

How might this be done? The intuition is to ask each and every individual to report a preference profile. (Note that since we assume all the individuals know each other's true preferences, it is no more challenging for an individual to report a preference profile, comprising preference relations for everyone in society, than it is to report his own preference relation.) If all the reported preferences profiles agree, there's a good chance they are all true, and the planner might accept the generally agreed-upon profile. If they all agree except for one, the one that's out of line probably comes from a liar, and he should be given a motive to avoid lying. (If the social planner were a despot, the out of line person would be shot. Note also that there must be 3 or more individuals in society to discover whose report is out of line.) Finally, when the reported preference profiles generally disagree, the social planner needs a way to avoid having the process stop at an inappropriate Nash equilibrium.

Let us be more precise. Maskin's algorithm for implementing an SCF works as follows: Each person *i* reports a message m_i, which is composed of an alternative *x*, a preference profile $\hat{R}^i$, and a non-negative integer. (i) If every message is the same, of the form $(x = F(\hat{R}), \hat{R}, 0)$, then the social planner chooses *x*. (ii) If every message but one is the same, of the form $(x = F(\hat{R}), \hat{R}, 0)$ for every person but *j*, while *j* reports a message of the form ($y \neq x$, $\hat{R}^j$, anything), then the social planner chooses *y*, unless person *j* would like *x less than y* according to $\hat{R}_j$, the person-*j* preference relation that all the other people are reporting. If this is the case the planner chooses *x*. (Person *j* is not shot. He simply does not gain, and may lose, from his deviation.) (iii) In all other cases, the social

planner finds the person who proposes the highest integer (with some method for resolving ties), and chooses the alternative named by that person.

Now the questions can be framed: First, given this mechanism, would $m_1 = m_2 = ... = m_n = (F(R), R, 0)$, with *R* the true preference profile, constitute a Nash equilibrium? Second, if $(m_1, ..., m_n)$ is any Nash equilibrium list of messages in this mechanism, can we be sure the resulting chosen alternative will be $F(R)$?

The answers are Yes and Yes, under certain general assumptions. The assumptions of Maskin's theorem are as follows: First, there must be 3 or more individuals (so that a deviant message can be spotted). Second, a mild diversity condition must be satisfied. Maskin uses a condition called *no veto*. Loosely speaking, this means that if $n-1$ people prefer *x* to all the other alternatives, then the SCF must choose *x*. Alternatively, one can assume the existence of a private economic good, that everyone values. This guarantees that individuals will disagree about what alternatives are best. In this entry we will simply assume *diversity*, meaning the following: For any given alternative *x*, there exist at least two people, each of whom prefers something else to *x*.

Third, the social choice function *F* must satisfy an intuitive condition called *Maskin monotonicity*. (The condition is actually a distant relative of the NIM assumption used in the simple version of Arrow's theorem presented above.) Maskin monotonicity means the following: Let $\hat{R}$ and R be any two preference profiles. (These may be true or false; it does not matter in this context.) Suppose $F(\hat{R}) = x$, and suppose that, for all individuals *i* and all alternatives *y*, $x\hat{R}_i y$ implies $xR_i y$. Then $F(R) = x$. (In other words, in a hypothetical transition from $\hat{R}_i$ to R_i, for every person *i* the set of alternatives that *i* likes less than *x* or the same as *x* has expanded, or at least hasn't shrunk. Since *x* was the social choice under $\hat{R}$, *x* must continue to be the social choice under R.) With these three conditions, Maskin proved:

> *Maskin Theorem*. Assume $n \geq 3$. Assume diversity and Maskin monotonicity. Then the mechanism described above implements the SCF *F*, in the sense that truthful messages leading to $F(R)$ comprise a Nash equilibrium, and in the sense that any Nash equilibrium list of messages results in the social planner choosing $F(R)$.

We will not provide all of the proof, but the logic is as follows: First, establish that $m_1 = m_2 = ... = m_n = (F(R), R, 0)$ is a Nash equilibrium, where R is the true preference profile. This is rather obvious, given rules (i) and (ii) of the Maskin algorithm. Second, establish that under rules (ii) and (iii), there are no Nash equilibria. This follows rather easily from the diversity assumption. Third, establish that if $m_1 = m_2 = ... = m_n = (F(\hat{R}), \hat{R}, 0)$ is any Nash equilibrium, then $F(\hat{R}) = F(R)$. That is, given a Nash equilibrium based on a universally reported, but possibly false preference profile, the outcome implemented is the same as if the true preference profile had been reported. This follows from the assumption of Maskin monotonicity.

Maskin also provided a near converse this theorem, which says that Maskin monotonicity is a necessary condition for any SCF *F* to be implementable. Relatively simple proofs of both Maskin theorems are available in Feldman and Serrano (2006).

Last Words

Where does welfare economics now stand? The First and Second Theorems are encouraging results that suggest the market mechanism has great virtue: competitive equilibrium and Pareto optimality are firmly bound. The Third Theorem exposes impossibilities and paradoxes in economic choices, voting choices, and, in general, almost any choices made collectively by society. The Gibbard Satterthwaite theorem, like the Third Theorem, is a starkly negative result: any plausible social choice function will, under some circumstances, produce incentives for someone to lie. But the Maskin theorem is a ray of hope; it suggests a way for a social planner to design a game, whose Nash equilibria will implement a desired social choice function.

Allan M. Feldman

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

.

Bibliography

Arrow, K.J. 1950. A difficulty in the concept of social welfare. *Journal of Political Economy* 58, 328-346.

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.

Feldman, A.M. and Serrano, R. 2006. *Welfare Economics and Social Choice Theory, 2nd edn*, New York, Springer.

Gibbard, A. 1973. Manipulation of voting schemes: a general result. Econometrica 41, 587-601.

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.
Marshall, A. 1920. *Principles of Economics*. 8th edn, London: Macmillan, ch. VI.
Maskin, E. 1999. Nash equilibrium and welfare optimality. *Review of Economic Studies* 66, 23-38.
Mas-Colell, A., Whinston, M.D., and Green, Jerry R. 1995. *Microeconomic Theory*. New York: Oxford University Press.
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.
Satterthwaite, M.A. 1975. Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory* 10, 187-217.
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.