

The Second Fundamental Theorem of Welfare Economics: A Pedagogical Note

by Sergio Parrinello

Received February 1998, accepted March 1998

1. Introduction

Modern textbooks of microeconomics tend to offer a uniform interpretation of the Second Fundamental Theorem of Welfare Economics. This standard interpretation, that many lecturers regularly convey to students, appears not only in elementary, but also in intermediate and advanced textbooks of microeconomics. For example take the following quotation from the volume *Microeconomic Theory* by Mas-Colell -Whinston - Green (1995):

If household preferences and firm production sets are convex, there is a complete set of markets with publicly known prices, and every agent acts as a price taker, then *any Pareto optimal outcome can be achieved as a competitive equilibrium if appropriate lump-sum transfers of wealth are arranged.*

The first welfare theorem provides a set of conditions under which we can be assured that a market economy will achieve a Pareto optimal result; it is, in a sense, the formal expression of Adam Smith's claim about the "invisible hand" of the market. The second welfare theorem goes even further. It states that under the same set of assumptions as the first welfare theorem plus convexity conditions, *all* Pareto optimal outcomes can in principle be implemented through the market mechanism. That is, a public authority who wishes to implement a particular Pareto optimal outcome (reflecting, say, some political consensus on proper distributional goals) may always do so by appropriately redistributing wealth and then "letting the market work" (Mas-Colell-Whinston-Green, 1995, p. 308).

William Bryant (1994) reminds us of the many textbooks of microeconomics and parts of journal literature where this interpretation of the Second Fundamental Theorem of Welfare Economics (called the "Theorem" from now on) can be found and much of the following argument is inspired by his critical article. For the sake of argument let us confine the discussion of the Theorem to a pure exchange economy composed of n consumers and k commodities¹. The following notation will be used:

$\mathbf{x}_i = (x_{1i}, \dots, x_{ki})$ is consumer i 's consumption vector and
 $\boldsymbol{\omega}_i = (\omega_{1i}, \dots, \omega_{ki})$ is his endowment vector, for $i = 1, \dots, n$;
 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a $k.n$ -matrix of consumption allocation;
 $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_n)$ is a $k.n$ -matrix of endowment allocation;
 $\boldsymbol{\omega}^\circ = (\omega^\circ_1, \dots, \omega^\circ_k)$ is the economy's initial endowment vector.
 Allocations \mathbf{X} and $\boldsymbol{\Omega}$ are feasible if $\sum_i x_{hi} = \omega^\circ_h$, $\sum_i \omega_{hi} = \omega^\circ_h$ for every commodity h , respectively; $h = 1, \dots, k$.

Three formulations of the Theorem are presented below by statement (A) followed by one of three statements (B₁), (B₂), (B₃) written in order of increasing generality.

(A) Assume the set of conditions under which a competitive equilibrium exists and assume that preferences are convex. Then for every feasible Pareto optimal allocation matrix \mathbf{X}^* there is a price vector \mathbf{p}^* such that $(\mathbf{X}^*, \mathbf{p}^*)$ is a competitive equilibrium.

(B₁) from the endowment matrix $\boldsymbol{\Omega} = \mathbf{X}^*$ in which:

¹ The 2x2 dimensional case is represented by the Edgeworth box.

$$\omega_i = \mathbf{x}_i^*, \text{ for } i = 1, \dots, n;$$

(B₂) from any feasible endowment matrix Ω such that:

$$\mathbf{p}^* \mathbf{x}_i^* = \mathbf{p}^* \omega_i, \text{ for } i = 1, \dots, n;$$

(B₃) from any feasible endowment matrix Ω and wealth transfers T_1, \dots, T_n such that:

$$\mathbf{p}^* \mathbf{x}_i^* = \mathbf{p}^* \omega_i + T_i, \text{ for } i = 1, \dots, n, \text{ with } \sum_i T_i = 0.$$

The Theorem in version (A), (B₁) is not helpful from a normative point of view². The usual interpretation of the Theorem refers in fact to (A), (B₂). The same interpretation, extended to the more general formulation (A), (B₃), states that "*every Pareto optimal outcome \mathbf{X}^* can be achieved as a competitive equilibrium if any appropriate mix of redistribution of the initial endowments Ω^* and lump-sum transfers of wealth is arranged*"³. This standard interpretation conveys four important ideas to students.

First, the student is led to believe that efficiency and distributive justice are separable and that a sort of division of labour can exist between the public authority, which is responsible for a socially acceptable distribution, and the market mechanism, that guarantees efficiency. In a different metaphorical sense, distributive justice and Pareto efficiency are two targets of the policy maker, who possesses two independent instruments: a redistribution policy with lump-sum transfers and the market mechanism.

Second, the student is inclined to compare two institutional arrangements to achieve a certain Pareto optimal allocation \mathbf{X}^* . One is a market economy in which a public authority intervenes only as a *ministry of finance*; the other is a command economy ruled by a public authority which acts as a *central planner*. The main idea resulting from this comparison is that the former arrangement is in principle superior to the latter because the market mechanism allows the economy to achieve the same optimal outcome by decentralized information.

Third, the student absorbs the idea that the obstacles to implementing a particular Pareto optimal outcome are the circumstances in which the Theorem does not hold, especially the existence of increasing returns to scale and the absence of markets for some externality, which bring about market failures. Indeed the most careful writers of microeconomics textbooks warn students that there exist other obstacles to implementing a Pareto optimum on the basis of the Theorem: the impossible redistribution of some endowments and, most importantly, the state of incomplete information on the part of the public authority. However, these warnings

² Varian (1992, p. 326) adopts formulation (A)+(B₁).

³ "Walrasian equilibrium" and "competitive equilibrium", "Pareto optimal" and "Pareto efficient" are used as synonyms, respectively. "Wealth" means "value of goods measured in units of numeraire". The mix is appropriate if it makes (1) fulfilled.

are usually spread over different chapters or relegated to some footnote of the textbooks and their dispersion does not give them sufficient weight to qualify the strong message transmitted to the student by the dominant interpretation. In particular, the third idea is reinforced by the analytical efforts of those economists who show that non convexities at the micro-level are not so important because aggregation can "convexify" them under mild assumptions.

Fourth, the student tends to believe that at least one point in the Pareto set must be a social optimum in so far as it encompasses both an appropriate distribution of wealth and efficiency. The standard interpretation of the Theorem is silent on this issue, but its optimistic flavour encourages this belief.

Below we shall examine a whole set of difficulties that weaken the usual interpretation of the Theorem itself. We want to argue that the usual interpretation, which most economic literature endorses by acquiescence, is misleading.

2. Equilibrium snags

Bryant (1994) convincingly criticized the usual interpretation of the Theorem because it is based on an arbitrary substitution of words. In fact the sentence "for every feasible Pareto optimal allocation matrix \mathbf{X}^* there is a price vector \mathbf{p}^* such that $(\mathbf{X}^*, \mathbf{p}^*)$ is a competitive equilibrium...", which is found in the Theorem, becomes "every feasible Pareto optimal allocation matrix \mathbf{X}^* *can be achieved* as a competitive equilibrium..." in the standard interpretation. "*Can be achieved*" in such a context presupposes that the uniqueness of and the convergence to a competitive equilibrium are granted. Instead non uniqueness of equilibrium and non convergence to a pre-determined equilibrium are not exceptional features. Non convergence can mean either (global or local) instability of a given equilibrium or undeterminateness of the final equilibrium point, if transactions occur at non equilibrium prices. In the following we shall call the set of these unpleasant properties *equilibrium snags*. Suppose that the teacher warns the students of the *equilibrium snags*. Should any further logical difficulty be emphasized in assessing the implications of the same Theorem? We believe that this is indeed the case.

3. Information snags

In a pure exchange economy two kinds of information should be kept distinct: information about endowments and information about preferences. We shall investigate: 1) the case of complete information; 2) the problem of labour endowment; 3) a case of incomplete information and moral judgements.

3.1 *The case of complete information*

Suppose that the initial endowments and consumers' preferences are known to the public authority. Then it can indifferently choose to act as a central planner or as a ministry of finance to achieve any Pareto optimal consumption allocation \mathbf{X}^* .

As a central planner, it can confiscate the initial endowments $\mathbf{\Omega}^\circ = (\omega^\circ_1, \dots, \omega^\circ_n)$ and reallocate them to achieve \mathbf{X}^* . As a ministry of finance, the Theorem allows it to follow one of the following courses of action to achieve \mathbf{X}^* :

i) impose only transfers of physical endowments $\Delta\Omega = \omega_i - \omega_i^\circ$, such that:

$$\mathbf{p}^*\mathbf{x}_i^* = \mathbf{p}^*\omega_i, \text{ for } i = 1, \dots, n;$$

ii) impose only lump-sum transfers of wealth calculated from equations:

$$T_i = \mathbf{p}^*\omega_i^\circ - \mathbf{p}^*\mathbf{x}_i^*, \text{ for } i = 1, \dots, n; \text{ under the constraint } \sum_i T_i = 0;$$

iii) impose any appropriate mix of i) and ii) consistent with (A),(B3).

If *all* endowments can be freely redistributed, the public authority can freely choose any of the above alternatives and then "let the market work" to achieve \mathbf{X}^* . However, it would not be wise to choose indifferently. In fact, among the many appropriate redistributions, it can pick up the physical transfers $\Delta\Omega = \Omega^\circ - \mathbf{X}^*$. In this case the risk of finding an equilibrium snag becomes immaterial, because the work to be performed by the market would be trivial. The market would only have to validate a consumption allocation equal to the endowment allocation *without exchange*. In this case no gain accrues from the market mechanism and no informational advantage would exist for the ministry of finance, compared to the central planner.

3.2 *The problem of labour endowment*

A well-known difficulty for the implementation of option i) arises, if some endowments cannot be physically redistributed. In particular the individual skills cannot be easily transferred among agents. In this case option i) is restricted and only option ii) remains fully feasible to the ministry of finance. In both cases he must know the amounts of *potential labour endowments*. This is not easy information to get because the potential labour of an individual is not observable; only effective labour performed by him is observable. In the pure exchange model, although the production of commodities is missing, this problem already appears. Suppose that one of the commodities is labour services directly useful for consumption. Suppose that the public authority knows the consumers' preferences and the initial endowment allocation Ω° included the endowment of potential labour services (and then the potential leisure) that each consumer can perform. How can it act to achieve a Pareto optimal allocation \mathbf{X}^* that is different from Ω° , at least in the distribution of the consumption of the labour services ?

As a central planner, it should impose forced labour on some consumer (that is a forced decrease in her leisure) to provide more labour services for somebody else. As a ministry of finance, it can arrange an appropriate lump-sum transfer of wealth without imposing forced labour. The outcome would be the same. The same consumer will provide the same amount of labour services for the same consumers. However this kind of indifference dismisses the risk of *equilibrium snags* and also the circumstance in which the individual prefers to perform *the same action* under an economic incentive instead of command. I shall expand the latter point in the final section.

3.3 *A case of incomplete information and moral judgements*

Suppose that the public authority knows the initial endowment allocation but has no information about the consumers' preferences and equilibrium prices. In this case it cannot find the Pareto optimum locus and no Bergson-Samuelson welfare function can be specified. In the two dimension-case, the Edgeworth box appears to the authority as a black box in which only one point, the initial endowment allocation, is visible.

This is a case in which an interesting interpretation of the Theorem is available under certain assumptions. Suppose that a moral judgement establishes what is a fair allocation and that the public authority wants to achieve an allocation that is fair and at the same time Pareto efficient. In his elementary textbook of microeconomics Varian (1987) assumes that a fair allocation is envy free and that an *egalitarian* allocation, in which each consumer possesses an equal share of total endowments, is fair in this sense. Of course, an egalitarian allocation, in which no consumer strictly prefers the endowment of any other consumer, in general will not be efficient because the consumers' preferences may differ. However, Varian shows in a simple way that this allocation possesses the following feature, which is not common to all fair allocations: any competitive equilibrium from an egalitarian endowment allocation preserves the absence of envy. This result offers a new interpretation of the Theorem⁴. In fact the ministry can implement a fair and Pareto optimal consumption allocation by redistributing the endowments according to the egalitarian principle and then "letting the market work", even if he ignores the consumers' preferences and cannot determine the commodity composition of the ensuing equilibrium allocation.

This seems to be an important advantage of the ministry of finance compared to the central planner, on the grounds of the amount of knowledge required. However, this interpretation of the Theorem does not avoid the main snags. The ministry may miss his target either because he cannot carry out the redistribution of some endowment or because of equilibrium snags. He might implement a fair outcome by appropriate lump-sum transfers of wealth instead of an egalitarian redistribution of endowments, if he knew the appropriate equilibrium prices, but how can he get this information without knowing the consumers' preferences? Furthermore, if the domain of the envious consumer's preferences embeds the consumption of other consumers, then an externality emerges and we are compelled to stretch the assumptions of complete markets and convex preferences. We must assume that a market (not too *thin*) for such an externality exists and that convexity is preserved.

The case in which the ministry knows the statistical distribution of preferences does not help in this respect. Foley (1994) has proved, applying the approach of statistical mechanics to a market economy, that a statistical market equilibrium differs from Walrasian equilibrium in important respects. The former cannot guarantee Pareto-efficiency because it allows agents' trading at different disequilibrium prices.

4. Pareto optimum and social optimum

The student is also tempted to believe that at least one point of the Pareto set must be a social optimum. Hausman (1992) pointed out the premise implicit in this belief:

⁴ Varian (1987) does not mention this interpretation in his textbook..

Other things being equal, competitive equilibria are morally good and market failures are morally bad, and all other moral concerns can be satisfied by adjusting the initial distribution [...]. The argument is not valid, because it contains the implicit premise that there is a Pareto efficient state of affairs that satisfies all other moral constraints (Hausman, 1992, p.68).

This remark, that is shared by distinguished welfare economists, should invite the student to keep the *ceteris paribus* clause under scrutiny. In particular, a social optimum can depend not only on the allocation of commodities but also on the institutions by means of which this allocation is achieved. For sake of argument, it is not necessary to assume that the allocative institutions are public goods over which the agents extend their preferences. It is sufficient to observe that a fundamental externality cannot be eliminated by expanding the commodity space and the corresponding markets. A competitive market can be defined as a state of affairs in which all agents are price-takers, but these markets cannot emerge and persist without the active *competitive behaviour* of the agents. Instead of the passive price-taker conduct, such a behaviour must describe some feature of social interaction implying a Pareto relevant externality ("greedy" is one of the emotional words used to describe such behaviour in common parlance). Then *no competitive market* can logically exist for the negotiation of the rights *to act competitively* instead of acting in a different way. We can speculate whether the basic assumption of complete markets is violated and whether a special market failure emerges in this case: if it is so, the first theorem of welfare economics would be in danger as well. Whatever position is held on this issue, the student should be made aware that no redistribution of endowments and wealth transfers can make a consumption allocation, achieved as a competitive equilibrium, a social optimum too, if an alternative institutional setting, that *ceteris paribus* implies a moral improvement, can be implemented.

5. Conclusion

I have argued that the interpretation of the Second Fundamental Theorem of Welfare Economics may or may not be useful from a normative viewpoint, depending on what is not included in the theorem itself. Only the investigation of the whole set of assumptions required to achieve a socially desired outcome, through the market mechanism supplemented by a redistribution policy, can ascertain the importance or the futility of the Theorem. This must be said to students and should be taken into account by those economists who endeavour to generalize the scope of the Theorem by relaxing the importance of non convexities. The standard interpretation puts exclusive emphasis on the assumptions under which a competitive equilibrium exists and on convexity conditions. These assumptions are of course sufficient conditions *within* the Theorem. By contrast, the same set of assumptions are necessary but not sufficient conditions regarding the interpretation in which we read that *every Pareto optimal outcome X^* can be achieved as a competitive equilibrium if any appropriate mix of redistribution of the initial endowments Ω^\bullet and lump-sum transfers of wealth is arranged*. This conclusion holds *a fortiori* if a certain competitive equilibrium has to be a social optimum as well as a Pareto optimum.

References

- Bryant W. (1994), Misinterpretations of the Second Fundamental Theorem of Welfare Economics: Barriers to Better Economic Education, *Journal of Economic Education*; vol. 25, n.1, Winter, pp. 75-80.
- Foley D. (1994), A Statistical Equilibrium Theory of Markets, *Journal of Economic Theory*, vol. 62, n.2, April, pp. 321-345.
- Hausman D.H. (1992), *The Inexact and Separate Science of Economics*, Cambridge, Cambridge University Press.
- Mas-Colell A. - Whinston M.D. - Green J.R. (1995), *Microeconomic Theory*, Oxford, Oxford University Press.
- Varian H. (1987), *Intermediate Microeconomics. A Modern Approach*, New York, Norton; trad. it. *Microeconomia*, Editrice Cafoscarina, Venezia, 1990.
- Varian H. (1992), *Microeconomic Analysis*, New York, Norton.

Summary: The Second Fundamental Theorem of Welfare Economics: A Pedagogical Note (J.E.L.: A2, D6).

The author extends the criticism that W. Bryant (1994) levelled against the usual treatment given to the Second Fundamental Theorem of Welfare Economics in many microeconomics textbooks and economic journal literature. He argues that the omission of basic caveats makes the usual interpretation misleading and an obstacle to better economic education.