

Another important and original voice on this issue comes from Italy. It is that of the public economist De Viti de Marco (1890), who applies the notion of natural monopoly to the telephone industry¹⁹. He states that such industry always tends to become a monopoly, and he gives different reasons for this: some are similar to those found in the writings of the economists already examined, while others, very inventive, will be discussed later because they are related to technology. We stop here, with the recognition of the network effects by De Viti de Marco²⁰, because all the concrete situations in which natural monopolies occur have already been pointed out, and nothing original was added in this respect by later economists.

3. *The inquiry into economies of scale*²¹

Natural monopoly is due to technological reasons; it is some specific technologically-determined production process which generates it. In the traditional view of natural monopoly, the fundamental characteristic of technology responsible for its emergence is economies of scale²². It is well known that economies of scale are a more general category than increasing returns²³: increasing returns to scale occurs with the same proportional change in all the inputs, while for economies of scale inputs increase by some amount; for example productions with high fixed costs and low marginal costs give rise to economies of scale, without exhibiting increasing returns to scale. It is also worthwhile recalling that until the 1920s, the expression “increasing returns” was used, as it implied changes in input proportions²⁴. In this section we see that economists of the past followed three different paths to identify decreasing costs. The first concerns those who focused on increasing returns, considering at the same time (more or less explicitly or approximately) their symmetry with the reductions in costs. The second is related to those who identified scale economies by the distinction between fixed and variable costs. The third includes

¹⁹ De Viti de Marco's article on the telephone industry is examined in Mosca (2007).

²⁰ It seems to be the earliest recognition of network effects in economic literature: “The consumers enjoy a utility which is greater, the greater the number of subscribers with whom they can communicate when necessary” (De Viti de Marco[1890] 2001: 521).

²¹ While for the aspects examined in the previous sections there is only the secondary literature focused on the specific issue of natural monopoly, the topics dealt with from now on have been widely studied from many historical points of view; the literature cited here only shows a part of this abundance of references.

²² We have seen in the introduction that on this point the perspective changed after the 1970s.

²³ This is true if the price of inputs doesn't change.

²⁴ For the meaning of “increasing returns” in Marshall, see Loasby (1989: 62), while on the terminology concerning the laws of returns used in the cost controversy see Aslanbeigui (1996: 278-280).

those who clearly understood the functional relation between costs and output. Again, we would like to point out that economists' ideas on economies of scale do not necessarily have to do with their ideas about the other elements composing the notion of natural monopoly: for example, not all of them used the expression, or even considered economies of scale related to monopoly. We will see in the next section if the decrease in costs was considered to lead necessarily to monopoly.

We limit our inquiry here to economics scholars, and we do not mention all those who applied the decreasing average cost concept to their businesses, like, for example, Gottfried Härtel, a German music publisher (see Scherer 2001). So we start again with Adam Smith. Despite the role that the division of labor plays in his analysis²⁵, Smith does not take into the slightest consideration the possibility that increasing returns may be at the origin of monopolies²⁶. Schumpeter ([1954] 1986: 585) recalls that from Smith onwards every classical economist, and especially West and Senior, assumes the existence of a law of increasing returns in manufacturing, in contrast to that of decreasing returns in agriculture²⁷. The advantages of the division of labor are well expressed by Babbage (1832), who was to be the fundamental reference point for later economists on this issue (Loasby 1996: 307-308).

Senior had a very clear understanding of economies of scale when fixed costs incur; referring to the spinning of cotton in a mill he writes: "As the quantity produced is increased, the relative cost of production is diminished" (1836: 4.53). A few years later, Cournot, as we know, used the total cost function and its derivative; concerning the latter, *i.e.* the marginal cost, he believed that this function was usually diminishing for "*manufactured articles*", because of "a better organization of the work, ... and ... [the] reduction [of] *general expenses*" ([1838] 1960: 59, author's italics). As for the shape of this function, he adds that: "it may happen however ... that when the exploitation is carried beyond certain limits ... [the function] ... again begins to increase" (60), thus indicating the

²⁵ Of course Smith was not the originator of the idea that the division of labor gives rise to increasing returns.

²⁶ "Adam Smith did not appear himself to be in the least troubled by the thought that competition and increasing returns might not be able to coexist" (Richardson 1975: 354); see also Stigler (1951) and Groenewegen (1999).

²⁷ Schumpeter writes: "Senior - or West and Senior - must be held responsible for the tradition, which took such time in dying, that agriculture was the domain of the latter [decreasing returns] and 'industry' the domain of the former. This quite misleading arrangement was not set right until the next period" ([1954] 1986: 585).

possibility that the marginal cost curve might be U-shaped. It is to Babbage that J. S. Mill (1848, I.9) refers in the discussion on increasing returns, but he goes further, explicitly stating that a large scale of production gives rise to savings in costs: “as a general rule – he writes – the expenses of a business do not increase by any means proportionally to the quantity of business” (I.9.3). On this issue, Walras cites J.S. Mill and especially Dupuit (1949 and 1952-53). In examining the distribution of water and gas, he recognizes that average costs decrease, because: “the expenses of the initial set-up, and up to a certain point in its utilization, can be spread over a varying number of products” (Walras [1875] 1936: 210).

H.C. Adams classifies industries into three types, according to their returns to scale: “All industries, as it appears to me, fall into three classes, according to the relation that exists between the increment of product which results from a given increment of capital or labor” ([1887] 1969: 105). A similar classification, but seen from the costs side, is made by the Italian marginalist Pantaleoni when he writes: “we may ... divide all products into various classes: [1] commodities of which a greater quantity than that available.... may be obtained *by a simply proportionate increase of cost*; [2] products which can be increased at a *less than proportionate cost*; [3 other products require] *a more than proportionately increased cost*” ([1889] 1957: 187). As we can see, he clearly states here the existence of a functional relation between costs and output. But Pantaleoni goes further, claiming that the three cases he has listed are not alternative and do not rule each other out, but can coexist in the same business: “an increase of the quantity produced is, within certain limits, accompanied by a diminution of cost – writes Pantaleoni – but ... beyond those limits, it involves increased cost” (193). We have to add that he never considered the possibility that decreasing costs could generate a monopoly (146)²⁸. Another of the Italian marginalists, De Viti de Marco, makes an important contribution to a much more detailed identification of the cost features related to natural monopoly: he identifies industries with high fixed costs (part of which are described as sunk) and low marginal costs (transport networks, telegraph and telephone industries), or zero marginal costs (the production of non-rival goods, like theatres) and on this basis he explains the reason why costs decline

²⁸ And he will never consider that possibility, even in his later work on combinations and associations, where he doesn't see any actual danger for the competitive process, because he believes that the threat of entry by new firms is always at work (Pantaleoni 1903 [2001]: 164-165). Pantaleoni's view of market power is examined in Mosca (2007).

as the scale of production increases. The importance attributed by Marshall to the “law of increasing returns” is well known (Loasby 1996). The aspect of his theory that interests us in this paper is his internal economies which, according to him, depend on “the resources of the individual houses of business engaged in [the industry], on their organization and the efficiency of their management” (Marshall 1890a: IV.IX.25). Internal economies are very important in Marshall’s *Economics of Industry*, and also in his *Principles*²⁹. The “reconciliation problem” to which they could give rise will be discussed in the next section.

What Pareto describes vaguely in the following terms: “for each type of production, there is a certain size of enterprise which corresponds to the minimum cost of production” ([1906] 1971: 243), is expressed with perfect clarity by Barone. As a matter of fact, perhaps for the first time in the economic literature, he gives a precise, complete description of the shape of the average cost curve: “if [one drew] a diagram with the *unit* costs of production on the y-axis, [the curve] would slope downward until a certain point, and then upward” ([1908] 1936: 23, author’s italics). Apparently, it is the first time in the history of economics that a diagram with the average cost curve is verbally described³⁰. Historians of economic thought like Stigler (1941: 112) and Schumpeter (who quote Stigler) agree that “Edgeworth’s analysis of the law of returns is one of his most important contributions to economic theory”. In fact he examines increasing returns in relation to marginal and average costs, separately considered (Edgeworth 1911a, 1911b). Moreover, his 1913 article has probably the first diagram of the U-shaped average and marginal cost curve, and the demonstration that the marginal costs curve intersects a U-shaped average costs curve at its minimum (1913: 214, fig. 3). We recall briefly that it was first Harrod (1930) and then Viner (1931) who found and diagrammed the relation between the U-shaped average and marginal costs curves for the short and long run, both for the single firm and for the whole industry³¹. After having examined economists’ analysis of the variation in costs as output increases, we pass now to their consideration of *generalized* economies of scale, and to the problem of monopoly that they could imply.

²⁹ Marshall devotes the entire chapter IX of book IV of his *Principles* to internal economies. See Stigler (1941: 76- 83) and Marchionatti (1992: 559-561) among others.

³⁰ Scherer states that only in 1912 did Fisher give “a correct verbal discussion of the short-run relationships between average cost and volume when fixed costs are incurred (2001: 900).

³¹ See Harrod (1931), Robbins (1934) and the literature cited in Aslanbeigui & Naples (1997).