can see here that there has been a gradual shift from the first meaning to the second, with a period of overlapping of the two; we can also see that the new sense crowded out the first in a period in which it was generally thought that the new natural monopolies were much more important than the old ones[17].

2. The singling out of the concrete situations to which it is applied

Natural monopolies typically occur in two kinds of production: the first is characterized by the need of a large infrastructure to start the operation, as in transport networks and some public utilities; the second is due to the presence of network effects (Liebowitz and Margolis 1996). Over the years economists have identified some industries in which monopoly is spontaneously generated for reasons linked to the production process itself. In this section we analyze the writings of the economists who identified new situations in which this phenomenon occurs. We will show here that the singling out of this kind of industry by economists has not necessarily to do with the development of the theory of natural monopoly. In actual fact, the justifications they gave to explain these cases are not always based on technological reasons, such as economies of scale. It should be also remembered that the expression "natural monopoly" was not necessarily used to describe these situations.

Adam Smith, discussing the subject of joint stock companies, explains that businesses cannot expand without running into problems of mismanagement; however he believes there are domains where large size firms can work better than small ones; they are "those of which all the operations are capable of being reduced to what is called a routine, or to such a uniformity of method as admits of little or no variation. Of this kind is, first, the banking trade; secondly, the trade of insurance from fire, and from sea risk and capture in time of war; thirdly, the trade of making and maintaining a navigable cut or canal; and, fourthly, the similar trade of bringing water for the supply of a great city" (1776, V.1.121). Notice that Smith speaks only of "large size firms", not of monopolies[18].

---

[16] The title of the article is: Natural monopolies and the workingman.

[17] See for instance Hadley: "This monopoly, due to the advantages of large organizations of capital, is characteristic of the present day. … Natural monopolies, like that of land ownership, are still important; but they are not the matter of supreme importance in productive industry any more than in transportation" (1886: 40).

[18] Elsewhere, talking about wages, Smith claims that where there are few agents, competition cannot work: "The masters, being fewer in number, can combine much more easily" (1776: I.8.12); the same reasoning

After him, it is J.S. Mill who explains in which sectors production on a large scale is preferable to production on a small scale, giving the example of the postal service, and the supply of water and gas (1848: I.9.1). And he goes further than Smith, when he argues that the possible disadvantages of a change from a small to a large scale "are not applicable to the change from a large to a still larger" ([1848] 1849: 175); this reasoning leads him to conclude that firms belonging to these sectors are in general destined to become monopolies: "where competitors are so few, they always end by agreeing not to compete" ([1848] 1849: 176). As we will see later, this is not the only, nor the most interesting reason J.S. Mill gives to explain why monopolies are spontaneously generated in these sectors.

A French contemporary of J.S. Mill, the engineer Jules Dupuit, identifies another situation of natural monopoly, *i.e.* transport networks, which he calls a "de facto monopoly". The reasons he found for this phenomenon are quite different from those given by the economists already examined. In his opinion, it is impossible for a new firm to enter the market of transport networks because: 1. the huge size of capital requirement cannot be available to more than a very limited number of entrepreneurs; 2. the new firm takes customers away from the monopolist and the profit will not be enough to cover the fixed costs of both; 3. the first business uses the best conditions, leaving the less favorable ones to the new one; in short: "instead of one good business, there will be two bad ones" (1852-53: 340).We have seen in the previous section that also for Walras monopoly is inevitable in transport networks; but the reasons he gives are different from Dupuit's. According to Walras, competition cannot work because the expropriation of the land needed to build the communication networks can be decided only by the Government. The same occurs, he says, to public utilities, due to the impossibility of laying pipes under public roads without authorization; as such permission can be granted only to very few firms, a monopoly is necessarily created because: "Competition between a limited number of entrepreneurs is rationally nothing but a passing phase after which there is the definitive creation of a sole monopoly based on the ruin of the others, or a monopoly of all of them or of some of them in coalition" ([1875] 1936: 202). We will see later that Walras also provides other much more interesting technological reasons for natural monopolies.

---

could be applied also to this case, and we could push his argument to the statement that when firms are large, they are few, and when they are few, competition cannot work; but this would be forcing Smith's meaning.

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 industry[19]. 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 Marco[20], 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 scale[22]. It is well known that economies of scale are a more general category than increasing returns[23]: 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 proportions[24]. 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).