

Europe

The Indivisibility of Telecommunications Networks: A Possible Explanation for Past and Present Trends in Telco Mergers

By Dr. Fernando Herrera González | Telefonica S. A.



Edited by Anna Tzanaki & Juan Delgado

The Indivisibility of Telecommunications Networks: A Possible Explanation for Past and Present Trends in Telco Mergers

By Dr. Fernando Herrera González*

Recently, Orange and MásMovil have announced that they are negotiating an agreement to form a single company in Spain. Experience shows that these transactions are viewed with great suspicion by competition authorities, as they reduce the choice for consumers and thus are considered to impede effective competition.¹ Operators often argue that the larger size acquired with these concentrations will allow them to reduce costs and be in a stronger position for the deployment of new technologies. The invariable response of the European Commission ("EC"), possibly guided more by political than economic reasons, is that they should achieve such dimension through acquisitions in other countries, without reducing the alternatives for consumers in any country. The following lines describe certain elements of economic theory about the productive structure of industries, which easily explain the rationality of the operators' proposals against the more political position of the EC.

Basic Notions About the Economic Theory of the Production Structure

Any good we consume requires a production process, no matter how simple it may be. Even an act of consumption as simple as eating a berry from a bush will require at least a little time and some work, which will have to be combined with the berry in its bush to give rise to a berry ready for consumption.

That is why economists differentiate between two types of goods: goods of first-order or consumer goods, and goods of higher-order or productive factors which, when combined with each other through labor, will result in a good

ready for consumption. It is obvious that all productive factors only enter production processes to the extent that they will end up in a product ready for consumption, otherwise they would not be useful for individuals and would not be produced.

Economists call the various productive factors that are in the process of transformation into final goods "*production structure*".² The production structure can be defined within several scopes— there is the production structure of a company, an industry, or society as a whole. It is important to note that the production structure is a consequence of the fact that all production processes happen over time, not instantaneously. If production processes were instantaneous, productive structure would not emerge since productive factors would immediately be transformed into consumer goods, and thus disappear.

The Complementarity of Productive Factors

There are two fundamental characteristics of the productive factors that shape the productive structure: complementarity and indivisibility³.

Investment decisions are highly conditioned by the existing production structure. Entrepreneurs make investments that, in some way, are complementary to the assets already present in the market. This is evident in the case of investments made by companies in their own structure, but it is also valid in general. For example, investments made by the so-called Big Tech were made at the time based on the existence of a type of asset, telecommunications networks, of which their investments would be complementary. In turn,

* Dr Fernando Herrera-González is a Regulatory Economics Manager at Telefónica S.A.

¹ See the European Union, "*Guidelines on the assessment of horizontal mergers under the Council Regulation on the control of concentrations between undertakings*" (2004/C 31/03)

² It is also called capital. The term "productive structure" is chosen here so as not to lose the connotation that each factor of production has a differentiated place in the process, something that is lost with the denomination "capital", which seems to homogenize all productive factors.

³ Lachmann, L.M. (1956), *Capital and Its Structure*. London: London School of Economics and Political Science.

the applications and services provided by Big Tech have made certain complementary investments desirable in telecommunications networks desirable, such as FTTH or 4G.

The Indivisibility of Productive Factors

In this article, however, we are more interested in the other fundamental characteristic of productive factors, which is the varying degree of divisibility. This is because productive factors cannot be acquired in an infinitely granular way but are usually acquired by indivisible "pieces".

With the simple example of a taxi driver, this will be clearer: A taxi driver needs a car regardless of the number of kilometers he travels per day. It is obvious that the car is an indivisible factor of production, for which the price is the same irrespective of whether 10 km or 10,000 km. Also, fuel will be needed, which is consumed proportionally to the number of kilometers. However, for practical reasons, the driver will not gradually pay for the use of such fuel (that is, kilometer by kilometer), but will normally fill the tank when it runs out. Again, the price of that refill will be the same, 1 km or 800 km. But it is also clear that the degree of divisibility of the production factor "fuel" is much higher than that of the vehicle. Between these two productive factors, there are, for example, the wheels, which have a degree of divisibility midway between the two other ones.⁴ As can be seen, productive factors that are actually variable in the sense that traditional economists give to the term are exceptional.

The Reduction of Costs by Increasing Production Is Due to the Indivisibility of Factors

The indivisibility of the productive factors has an immediate consequence that can be deduced from the definition: the larger the number of units produced, the lower the average cost of production of each of them. It is important to note that this occurs with absolutely all production structures, with its intensity

increasing with the degree of indivisibility of the productive factors.

Since the productive factors are added by discrete pieces, the entrepreneur requires a minimum number of products sold for his investment to be profitable. That is, it will not normally be convenient "to install an indivisible capital good unless there are enough complementary capital goods to justify it." In any case, part of the capacity of the productive factor will likely remain unused in the short term, which implies that its marginal value is zero. This means that it can be sold at a price below average cost or even given away for free, without making the investment unsustainable.

Following the example of the taxi driver, let us assume that the expected life of the vehicle he plans to acquire is 100,000 km, but that he plans to retire after doing 50,000 km. In principle, he would like to adjust the capacity, but that is not possible. In short, his investment, if it remains viable, will add to the production structure 100,000 km of which he only plans to consume 50,000. He may choose not to use the other 50,000 km (that is, send the taxi to the scrapyards once they are done), but he might also leave them to a friend so that he goes on vacation, or someone may come up with other ideas for those kilometers.

Indivisibility of Factors of Production and Economies of Scale

Before proceeding, it is useful to refer briefly to the concept of economies of scale, which is often confused with the phenomenon just described of reducing average costs due to the generalized indivisibility of the factors of production. It is interesting to understand well the concept because this is one of the typical reasons that industrial economy provides for horizontal concentrations (that is, of companies that manufacture the same product), whether in the same geographical market.

Economies of scale are obtained when "output can be doubled for less than a doubling of cost.

⁴ Variable costs are those directly proportional to the number of units produced.

"⁵ Some typical ways to obtain it are the following:⁶

1. Workers can specialize in the activities where they are the most productive.
2. Greater flexibility, so that managers can organize the production process more effectively.
3. The firm may be able to acquire some production inputs at a lower cost because it is buying them in large quantities and can, therefore, negotiate better prices.

The concept can be illustrated by continuing with the example of the taxi. Let us imagine that there are vehicles with two possible technologies. With technology A, the vehicle has a service life of 100,000 km, and costs 10,000 Euros; vehicles with technology B can travel 200,000 km and require an investment of 15,000 euros.

It is clear that technology B allows economies of scale compared to technology A, since traveling twice as many kilometers with Technology B would cost less than twice as much as it would with technology A. Note also that, if you travel 100,000 km or less, technology B would be more expensive than technology A. In fact, economies of scale will only be such if more than 100,000 km are covered, because in that case the taxi driver would be forced to buy

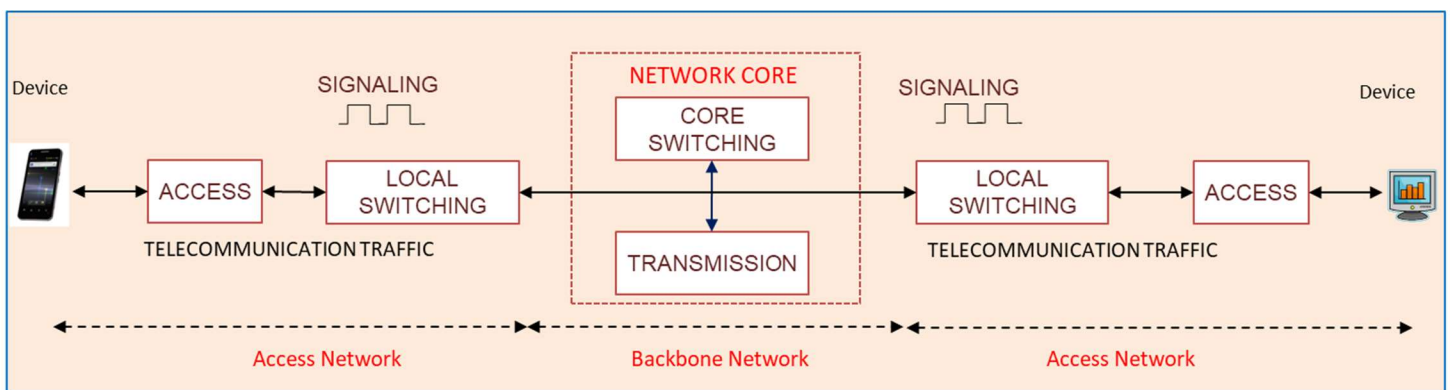
another vehicle of technology A to complete the production.

It is up to the entrepreneur to assume the risk of investing in one or another asset, according to the expectations he/she has of the market. In any case, once the asset with economies of scale has been added to the production structure, the situation will be like that described in the initial section. There will be an indivisible productive factor with a higher capacity which, if used, will result in lower average costs than the asset would have been without economies of scale.

The Production Structure of Telecommunications: The Telecommunication Network⁷

The production structure of telecommunications necessarily has a lot to do with the telecommunication network of the operators. The telecommunication network is the "plant" by which the operator produces the products that its customers demand.

Entire volumes can be devoted to describing the structure of telecommunication networks, but for our purposes, it is enough to focus on the following scheme. Broadly speaking, telecommunication networks currently have two quite different parts: the access network and the core network.



⁵ See Pyndick. R.S. & Rubinfeld D.L. (2018). Microeconomics 9th Ed. See p. 264.

⁶ See Pyndick. R.S. & Rubinfeld D.L. (2018). Microeconomics 9th Ed. See p. 263.

⁷ In the EU and due to the existing access regulation, there are numerous operators that provide services using the networks deployed by network operators on a wholesale basis, such as resellers. These "virtual" operators have a completely different production structure from the one described, which is only relevant for network operators.

The access network consists of the elements that allow potential customers to connect to the telecommunication network and access its services. The resources of the access network are dedicated exclusively or almost exclusively to each client. A very clear example is the segment of optical fiber that reaches the router in each home. This fiber can only be used by that home; absent it, that home would not be able to connect to the Internet or watch Netflix.

The core network, on the other hand, is used by all customers who access the telecommunication network, to a greater or lesser extent depending on the traffic (the gigabytes) each of them generates.

The incorporation of mobile access networks into this production scheme is slightly more complex. In the case of mobile, the productive factors used for the access network (the antennas and the radio spectrum) are shared between several customers, so there is not that degree of exclusivity pointed out for the fixed network. At the same time, since the resource has to be shared between several clients, the capacity available to each will be less than in the case of a fixed network. This is the reason why fixed networks allow higher speeds than mobile networks.

In any case, the telecommunication network is shared by more and more users as we physically move away from the user, until we reach the core of the network, which is shared by all users. Going in the opposite direction, the productive factors are increasingly exclusive for each client, until we reach the end of the cable we have at home, which is reserved for each of us.

The Indivisibility of Telecommunication Networks: Geographical Coverage

Now that we have an idea of the production structure of the sector, we are already better positioned to analyze its degree of indivisibility. To do this, we need to know how the deployment of telecommunication networks is carried out, and what the deployment unit is that makes technical-economic sense for a telco entrepreneur. What we know (and some will likely have suffered when the operator tells us

that we do not have coverage wherever one lives or is on vacation) is that telecommunication networks are deployed by geographical "chunks" that are indivisible units for economic purposes.

In fixed access networks, geographic areas require a relatively small size in comparison to mobile access networks. In fact, each location reached by a fixed network can already provide a functional service to the customer located there, regardless of the extent of the network. On the contrary, mobility being intrinsic to mobile access networks, customers will only contract these services if they have a minimum geographical coverage where they can use the service.

Therefore, at first glance, the indivisible "chunks" are considerably larger in mobile access networks than in fixed access ones. This implies that, comparatively, greater investment in productive factors will be required in the first case than in the second.

As for the core network, in practice, it follows the same divisibility criteria that apply to access networks, thus the extension of the core network is associated with the geographical coverage of the access network.

In short, telco operators add geographic "chunks" of the network to the production structure. These "chunks" of the network have a contained geographical reach in the case of fixed access networks (for example, a small village or a neighborhood) and are necessarily more widespread for mobile access networks (because a minimum population and territorial area are required for the user to see some value in the mobility of the service).

Indivisibility with Respect to What: Traffic vs. Access

As we have seen, network operators add production structure based on geographical "chunks". Only once a certain geographical area is covered are they in a position to provide services to potential users who live (fixed networks) or move (mobile networks) in it.

We already know that the greater the number of units produced in these "chunks", the lower the

cost of production of each unit. We also know that a minimum number of these units will have to be sold for the investment to be sustainable. But what do telecommunication networks produce?

The immediate answer seems obvious: telecommunication networks produce traffic and fabricate the transmission of bytes between distant points. This is the traditional view of the product of telecommunication networks and is basically valid till the beginning of the twenty-first century, although at that moment traffic was measured and paid in minutes and not in bytes.

However, nowadays operators no longer produce traffic. Or, rather, they continue to produce it, but it is not what they measure to charge their customers. A look at the metrics used by regulators, analysts, academics, and investment banks to characterize the sector will reveal that virtually no one talks about traffic. What everyone is concerned about is not traffic, but the number of customers.

This is reasonable because, for a few years now, the income of the operators is not so much related to the traffic they move as to the number of customers they have. The cause is the generalization of flat rates for data services (Internet access), which constitutes about 100% of the demand today. Operators do not bill their customers for traffic, but for access.

Thus, operators do not see themselves as "factories" of bytes or minutes. At the moment, they are "factories" of access, and, therefore, the relevant metric of indivisibility is the number of accesses.

In short: although technically telecommunication networks are indivisible both with respect to traffic and access, at present the "economic" indivisibility is with respect to the number of accesses, which is what operators mainly sell to their customers.

Consequences of the Geographical Indivisibility of Telecommunication Networks with Respect to Access: The Future

Although the following analysis refers to fixed networks (because it is easier to visualize), it is extensible without variation to the case of mobile access networks.

As has been said, the operator carries out the deployment of the network by geographical "chunks". Once a "chunk" is deployed, the operator can provide services in the covered area. A certain number of the households reached by the network (*passed homes*, in telco operator lingo) may contract the offered service (homes in service or *take-up*), but it is highly unlikely that all covered households will do so. However, the investment will only be sustainable if a minimum number of households pay for this service, in the same way that the taxi driver will only recover his investment in the car if he makes a certain minimum number of kilometers.

This minimum number (the *breakeven* point) will depend on two factors, as in any industry: the cost of production (cost per passed home, investment already made in the form of an indivisible asset) and the price the client is willing to pay. In turn, this price depends on the utility of the user and the availability of alternatives.⁸ Both parameters, together, establish the minimum market share that an operator must obtain in the covered geographical area for the investment to be viable. Otherwise, it will not be able to recover the investment, nor maintain or upgrade the network.

From the above description, it is immediately noticeable that the lower the price clients are ready to pay, the greater the market share required for the investment to be viable. But note as well that, for a constant utility for the client, the price decreases with the number of alternatives available. Both features drive the inevitable trend towards the concentration of network operators: the need for a minimum market share to be viable, and the increase of that minimum market share if the price is

⁸ According to the marginal theory of the value, as enunciated by Menger, Jevons and Walras, independently, in 1879.

reduced, which happens precisely with the entry of new competitors, making it more difficult to attain that minimum market share.

Does this mean that the telecommunications market is a natural monopoly? Far from it. The dynamic process in any market is based on a game of prices and profits. If market concentration results in prices that provide extraordinary profits (in our analysis this would translate into a larger market share than is needed to be viable), there will be new entry, which would reduce prices and shares, with the good news for the end customer that such new entry will be carried out with better technology, whether in capacity, costs, or both. Of course, it may also happen that entrants are excessively optimistic,⁹ their assessments reveal as wrong, and we end up in a scenario that requires concentration to restore prices and market shares to viability levels. From where the virtuous circle described above would begin again.

In any case, the geographical indivisibility of the telecommunications network is the cause of the tendency of the market to concentrate. The need for both a minimum market share and a minimum price to recover the investment in the "chunk" of the network limits the number of networks that can be economically viable simultaneously in the same geographical area. Such a number cannot be known a priori and will depend a lot on the value that customers give to each network at any given time. There is therefore not even an optimal number per area, and this number varies over time. Neither is there a single way to establish geographical areas of deployment—each operator can have as optimal different geographical "pieces".

A Brief Look at the Past

The almost inevitable need of telecom operators to concentrate on the geographical areas in which they are present has been explained. Since 2006, the big mergers of telcos have always been nationwide, what we could call "*in-market*". These are the cases of Vodafone-ONO or Orange-Jazztel in 2014 in Spain, O2 Germany and E-Plus in that country, or the

attempted acquisition of O2 UK by Hutchinson, blocked by the EC in 2016. And, of course, the operation with which these lines were opened was that of MasMovil and Orange. All are consistent with the above reasoning.

However, until 2006, numerous mergers and acquisitions between telecommunications operators in different geographical areas were proposed, apparently against the logic just explained. A quick historical review shows that telcos were betting at the beginning of the century on expansion to other geographic markets, not only within the EU but also outside it. Such expansion was intended, in some cases, through organic growth obtaining spectrum licenses in the corresponding contest or auction; in others, by acquiring an operator already in operation.

Thus, Telefónica bought O2 and Cesky Telecom in 2005, in addition to different operations in Latin America (such as Bellsouth in 2004); Vodafone acquired Mannesmann in 2000, in one of the highest volume operations so far, as well as expanding in Asia; Orange (France Telecom at the time), did the same in Eastern European countries, and also in Africa.

What changed in the market to alter the interest of operators? This has already been alluded to. Until the middle of the first decade of this century, operators were traffic manufacturers and most of their revenue came from charging by the minute, as many people will remember. Thus, the analysis of the indivisibility of the network would have to be done with such unity in mind, that of the coursed minute.

Although the geographical "chunk" is still the unit of addition to the production structure, in this case the market share in the access network is much less important because the relevant thing is that a lot of traffic is transmitted. In this context, horizontal cross-border mergers may make sense since they increase the traffic in the operator's network. To the current traffic in its network, the traffic generated by the acquired network will be added, and the new traffic occurring between both networks in a different geographical area. This happens

⁹ I consciously avoid referring to the incentives that wholesale regulation in the market might be causing in this regard.

regardless of the access market share it holds in each of the networks.

In short, the geographic indivisibility of networks seems to support small shares of access clients when the unit of production is traffic, and thus the aggregation of networks of different geographical coverages.

Back to Economies of Scale

Fortunately for users and operators, the telecommunications market, like many others, allows for economies of scale. Let us see what form they can take in light of the analysis of indivisibility that has just been done.

Recall that economies of scale are obtained when "*output can be doubled for less than a doubling of cost,*" and we illustrated it with the example of two technological alternatives for the vehicle that the taxi driver was considering buying¹⁰. If we understand production in a broad sense, not only by the number of produced units, but also their quality, then the economies of scale made possible by concentrations in the same geographical area can be easily understood.

Imagine that operator 1 alone expects to produce, say, 20% of the accesses in the area; operator 2, by itself, the same. That makes each of them, separately, opt for an efficient technology for the production they expect, of that 20% of accesses. However, if the operator's forecast reached 40%, it would surely consider another technology more appropriate to that amount of production and more efficient, in costs or quality, for users.

The example is on the table: 5G technology for mobile access networks has clear efficiencies in quality and costs compared to current technology, but it is only viable with minimum expectations for market share. In the same way that the taxi driver in our example will never consider buying the technology A car to make less than 100,000 km, telecommunication operators cannot consider that investment unless they have minimum production expectations, in this case measured by the

number of customers, always due to the indivisibility of the asset.

Regarding mergers of networks with different geographical coverage, which, as has been seen, had logic in the case of traffic production, economies of scale as those referred to in Section "Indivisibility of factors of production and economies of scale" appear. Thus, the greater volume in acquisition of productive factors improves the power of negotiation with suppliers. On the other hand, synergies due to the centralization of common services (for example, commercial) allow for a greater specialization of workers and consequently greater productivity. It is very likely that the expectation for these economies of scale may have contributed to the wave of cross-frontier mergers at the beginning of the century, on top of the cost reduction due to a greater volume produced with indivisible factors.

Conclusions

In summary:

1. The production structure of companies and industries is composed of productive factors. Productive factors present varying degrees of divisibility. Because of this, the higher the production, the lower the average cost of each of the products.
2. Telecommunication networks are installed covering specific geographical areas. Each of these geographical units is practically indivisible for production purposes, whether for access or for traffic production.
3. Today, contrary to what happened in the past where firms produced and charged for traffic, telecommunications networks produce and charge for access. Given the indivisibility of the network in its geographical scope, this implies that the investment is only sustainable if a minimum market share is reached in the covered area, a minimum that depends on the price that can be obtained from the service in the same.
4. In these conditions, and, as the entry of new networks in the same geographical area

¹⁰ *Supra* footnote 4.

decreases both the expectations of market share and the price obtained, there is an almost irresistible tendency for the concentration of operators to be sustainable.

5. Additionally, if high production is not expected in terms of access, it is difficult to

undertake the deployment of new technologies, given their greater indivisibility (in terms of quantity and quality), which will only make them viable with a minimum dimension of production higher than the current one.