BUSINESS & INDUSTRIAL ECONOMICS

Professor Luca Grilli

Professor Mattia Jr. Pedota

GeMeBoRi

Academic Year: 2022

INDEX

6.4 Bain, Sylos Labini and Modigliani model

6.5 Bain, Sylos-Labini & Modigliani with Economies of Scales

6.6 Dixit Model

7. EXTERNALITIES

7.1 Examples and types of Externalities

7.2 Externalities and Efficiency

7.3 Externalities and Property Rights, Coase's Theorem

7.4 The market signal for Merging

8. PUBLIC GOODS

8.1 When should a public good be provided?

8.2 The free-riding problem

9. COMMON GOODS

9.1 The problems

9.2 Possible Remedies

10. ASYMMETRIC INFORMATION (3RD MARKET IMPERFECTION)

10.1 Adverse Selection

10.2 Moral Hazard

10.3 Remedies to market asymmetric information

10.4 Asymmetric Information: Some remarks

11. TRANSACTION COSTS

11.1 The Transaction Cost Economics (TCE)

11.2 The hold-up problem

12. INNOVATION AND TECHNOLOGY: THE SCHUMPETER LEGACY

12.1 Brief Taxonomy

12.2 Schumpeterian Marks

13. INNOVATION AND TECHNOLOGICAL CHANGE: MARKET STRUCTURE AND INNOVATION

13.1 Which market structure favors innovation

13.2 To what extent innovation should affect market structure?

14. ALTERNATIVE THEORIES OF THE FIRM: CONTRACTUAL AND HOLISTIC APPROACHES

14.1 The contractual approach

15.2 Holistic approach

15. FIRM GROWTH: DIVERSIFICATION AND VERTICAL INTEGRATION

15.1 Diversification

15.2 Vertical Integration

15.3 Firm Growth: Alternatives to Integration and Internationalization

15.4 The Economic Theories of MNCs

16. NETWORK ECONOMICS

16.1 An introductory example: *The story of the garbage bin.*

16.2 Network markets' characteristics

16.3 Innovation in a network market: the start-up problem.

16.4 Further remarks on the possible *non-emergence* of "winner takes all" market equilibria

16.5 Strategic choice between compatibility and incompatibility

16.6 Cases on standard wars

16.7 Third concept: Switching costs & technology replacement

16.8 Final thoughts

17. INDUSTRIAL AND COMPETITION POLICY

17.1 Industrial Policy

17.2 Competition policy

18. CONCENTRATION INDEXES

18.1 Link to the market structure

18.2 A multitude of CIs

1. EFFICIENCY, COORDINATION, ECONOMICS ORGANIZATION

Why is trade also beneficial for the society as a whole and not solely for the parties involved within both inter- and intra-transactions?

"*Economic organizations* are entities created by individuals (not natural thus) through which people interact to reach their goals" (Migrom and Roberts, 1992).

In other words, they are instruments through which people reach *goals*. This implies that people know which are their goals (i.e., preferences). Such goals are not only economic ones, however: they could be related to the capacity of satisfying the economic needs of participants such as the consumption of goods. The individuals must know their *preferences*, which means needs and priorities. We can describe their satisfaction through a *utility function* and the goal is to maximize it.

We live in a context of *scarcity of resources*: if we choose to consume more of one good, we will have to reduce the consumption of another good. Moreover, the satisfaction of the utility function of an individual is at the expense of the others. We assume that economic agents have specific preferences.

Economy as a whole is the highest-level organization. Furthermore, it is comprised of the lower levels such as *markets* and *firms*. However, there are other formal economic organizations such as labor unions, government agencies, universities and so on. They are independent legal entities: they can enter binding contracts and provide enforcement of those contracts in their own name.

The places where individuals and firms interact are:

- *Product market*: households express a demand of goods and producers supply these goods.
- *Market for inputs*: producers demand labour and components. These markets are clearly interrelated: a decision in the product market has an impact on the market for inputs.

Every market is connected to many others. For instance, If I take a decision of consuming a specific good, this will have effects on other markets, which will cause, in turn, further effects on others.

1.1 Pareto efficiency

How can we allocate resources in the best possible way within a society?

An allocation of resources A is inefficient if there is some other available allocation B that everyone concerned likes at least as A and that one person strictly prefers. In such a case A is Pareto dominated by B (B is Pareto superior to A) and it is clearly wasteful from a society point of view. Otherwise, A is Pareto efficient (or Pareto optimal).

This is a way to evaluate different choices leading to different allocation of resources and products. The Pareto efficient solution depends on the group of people and set of available resources considered. Specifically:

1. If there are more people and resources, the Pareto efficient solution changes. Giving all resources to a single insatiable and completely selfish individual could be Pareto efficient, despite of *no ethical considerations*.
2. Furthermore, given a set of resources, there are *many efficient allocations*. Therefore, the Pareto efficiency has a *normative power*: it allows to compare different situations, although its predictive powers are not that big (not possible to predict which pareto efficient condition will emerge from a given situation). Despite this, it has still the power to predict, because of the efficiency principle. But what does such principle state?

Efficiency principle: If people can bargain together effectively and can effectively implement and enforce their decision, then the outcomes of economic activity will tend to be efficient (at least for the parties of the bargain).

If people recognize an inefficient allocation, in the long run, they will migrate to a Pareto efficient position through coordination, without hurting anybody (i.e., those that are already satisfied). Indeed, since efficient

choices and allocations are less vulnerable, we should expect inefficient arrangements being supplanted over time, while efficient ones survive.

Despite this, there are some conditions in which if let working alone, the system will not reach a desirable outcome. Under this scenario, there will be the necessity of an exogenous intervention (the government).

1.2 The Edgeworth box

To highlight the benefits of trading and understand how a Pareto efficient solution can be achieved, the Edgeworth box will be used.

If we consider two consumers, A and B, and their endowment of goods 1 and 2, which are:

For example: $\omega^A = (6,4)$ and $\omega^B = (2,2)$ $\omega = (\omega_1, \omega_2)$ and $\omega^B = (\omega_1^B; \omega_2^B)$.

Then, let's assume that the total available quantities will be: $\omega_1^A + \omega_1^B = 6 + 2 = 8$

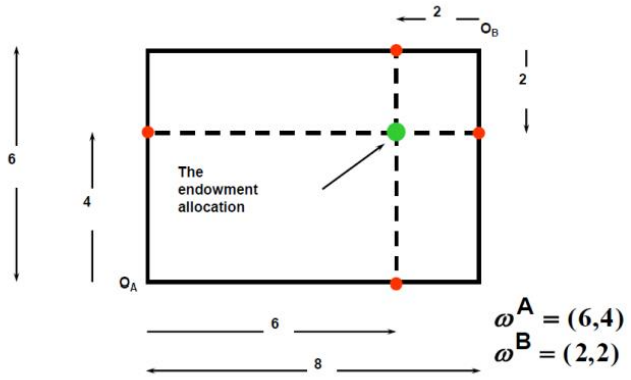
$\omega_2^A + \omega_2^B = 4 + 2 = 6$

$\omega_2^A + \omega_2^B$

The dimensions of the box are the quantities available of goods.

$\omega_1^A + \omega_1^B$

Notice that the box includes all the feasible allocations of goods between the two consumers:



The allocations to consumers A and B are $(X_1^A; X_2^A)$ and $(X_1^B; X_2^B)$. An allocation is feasible only if:

$$X_1^A + X_1^B \leq \omega_1^A + \omega_1^B$$

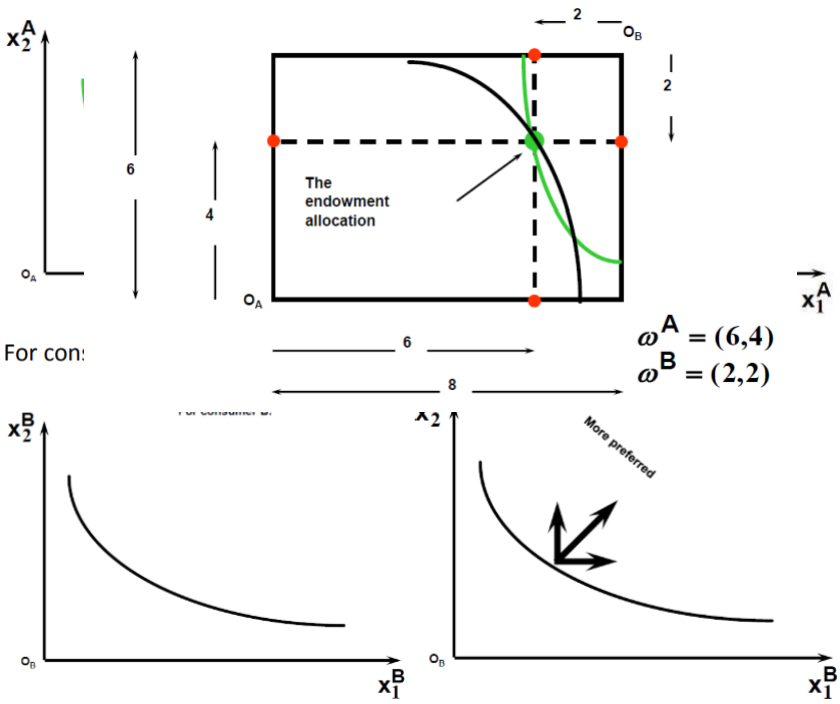
$$X_2^A + X_2^B \leq \omega_2^A + \omega_2^B$$

Which allocations will be blocked by one or both consumers? Which allocations make both consumers better off?

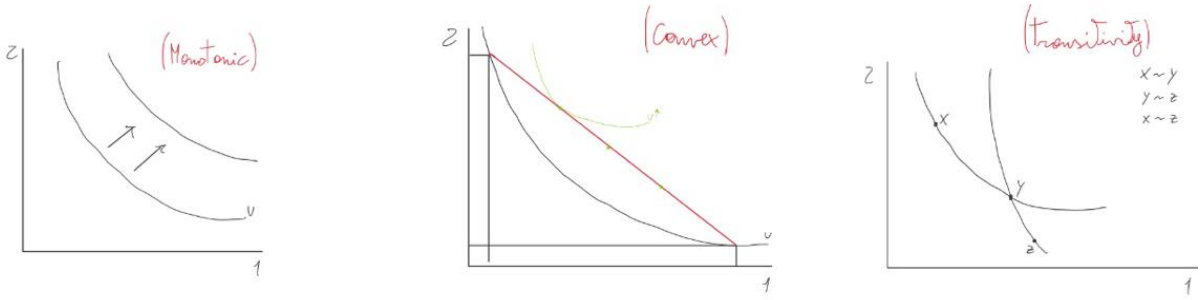
There are several feasible allocations. One or both *consumers will block an allocation*. The consumers prefer to move to an allocation rather than another on the base of their *preferences*.

We can add the preferences to the Edgeworth box, drawing the *indifference curve*, which represents the allocations of goods that give the same level of satisfaction to the consumer. We always assume *well-behaved* indifference curves with (1) *monotonic preferences*. This means that more consumption will always be preferred to less consumption (non-satiety assumption). Furthermore, we assume (2) *convex preferences*, which means that consumers prefer having a balanced mix of goods rather than extreme quantities.

According to the previous example, for consumer A the indifference curve will be the following:

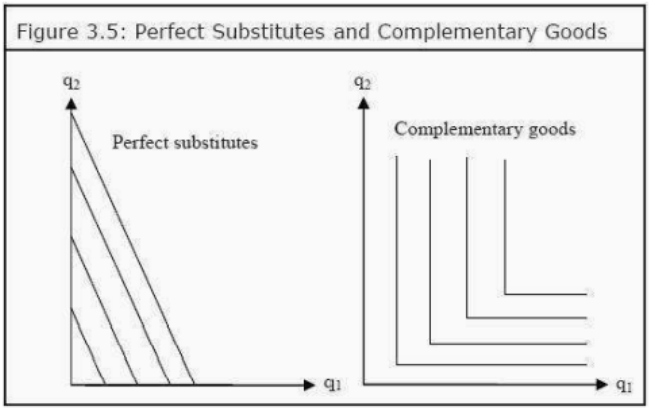


The indifference curves have a *negative slope* because if we want to maintain the same level of utility increasing the consumption of one good, we must reduce the consumption of the other. The indifference curves of a consumer never cross each other – (3) *preferences' transitivity*. Moreover, a balanced mix is always preferred instead of a rather unbalanced one.



Notes: we can observe the indifference curves of the perfect substitutes and complementary goods. In the case of perfect substitutes, the extreme quantities of the goods (on the axes) give the same level of utility to the consumer. In the case of complementary goods, if the consumer buys one of the two goods, he must buy also the other (for example shoes).

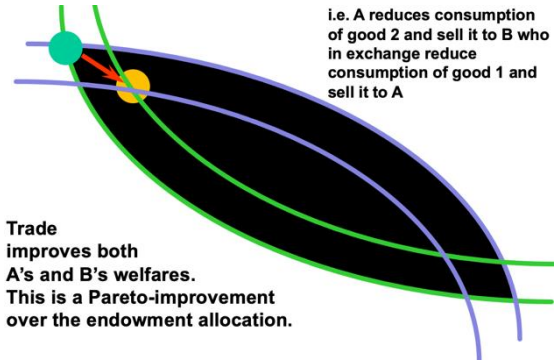
By mixing the indifference curves of the two consumers and adding them to the Edgeworth box, we obtain the following graph:



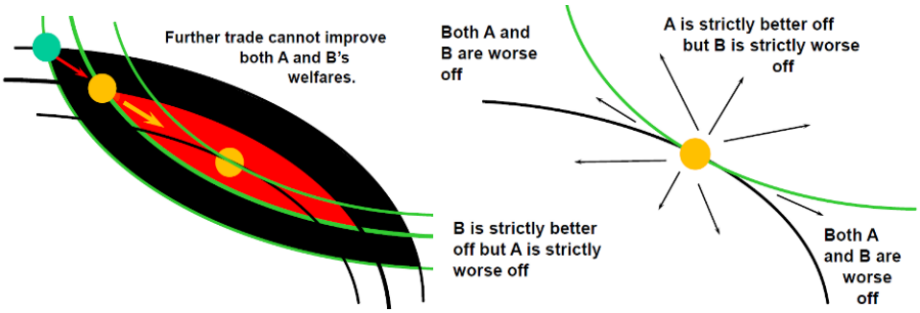
Is there a way to improve the utility of both economic agents?

Indeed. Within the area between the two curves, both consumers may find allocations that give them a better level of utilities. Therefore, an allocation that improves the welfare of a consumer without reducing the welfare of the other is a *Pareto-improving allocation*.

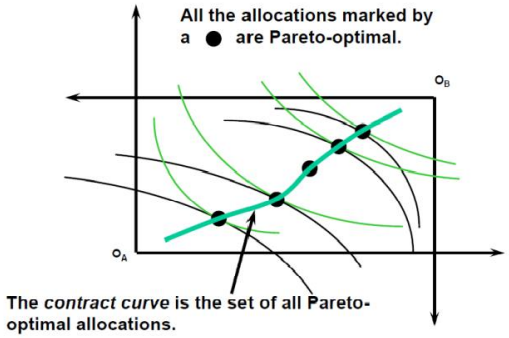
Trades between the two customers allow them to move to a better allocation.



These trades end when another trade cannot improve the utility of one (or both) consumer (s) without hurting an involved party. This happens when the curves of the two consumers are tangent.



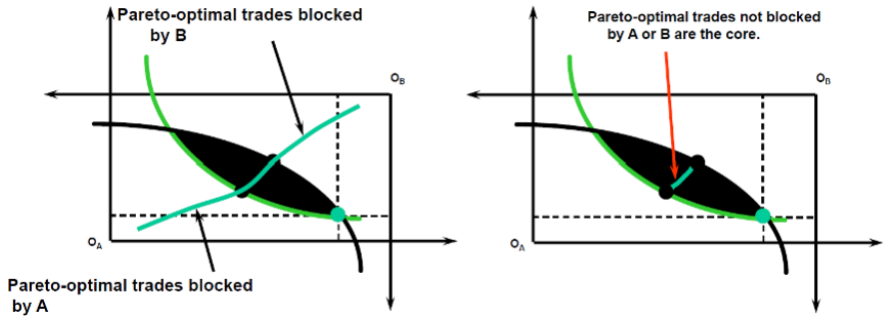
This allocation is *Pareto Optimal* because the only way one consumer's welfare can increase is to decrease the welfare of the other.



There are many points in the box where the curve of one consumer is tangent to the curve of the other: all these points form *the contract curve.* If there is an initial allocation of resources, there is *limited set of*

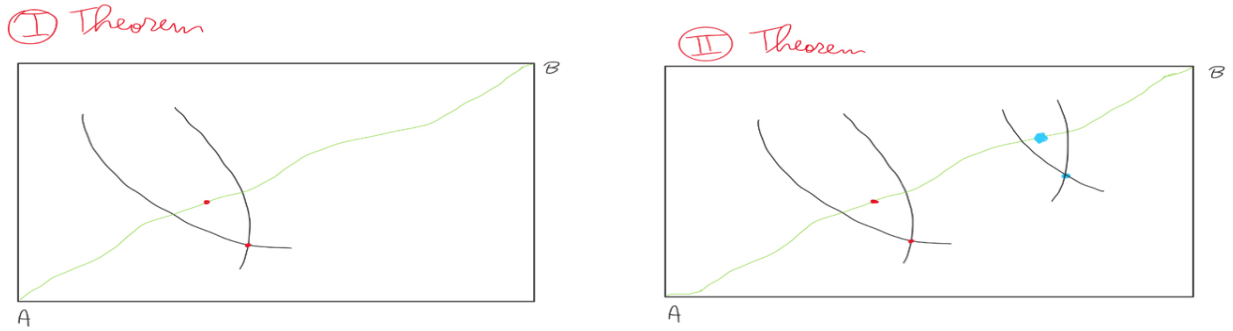
allocations reachable in the contract curve. Mixing the contract curve and the indifference curve within the box, we can find the *Pareto efficient allocations not blocked by a consumer*. These Pareto optimal allocations, called *the Core,* are welfare improving for both agents relatively to their own endowments.

Rational trades should achieve a core allocation. The ultimate trade should be into the core. Therefore, trades can increase the *welfare of a community*.



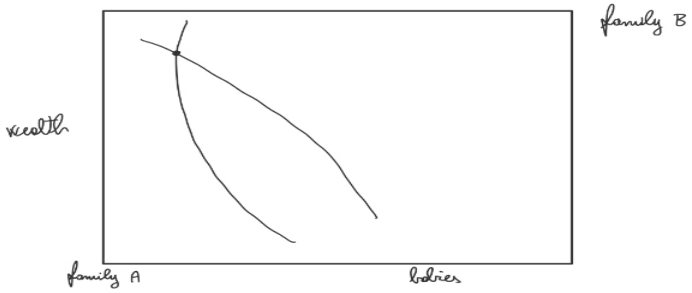
Thanks to the Edgeworth Box we can scrutinize the following considerations:

1st Consideration:



- *1st* Fundamental *theorem* in Welfare Economics: any competitive equilibrium (CE) is Pareto efficient.
- *2nd* Fundamental *theorem* in Welfare Economics: under certain conditions, every Pareto efficient (PE) allocation can be achieved as a competitive equilibrium. This implies that the problems of distributions and efficiency can be separated. The market mechanism is distributional-neutral: whatever your criteria for a good or a just distribution of welfare, you can use competitive markets to achieve it.

2nd Consideration:



Market is not ethical. Take for example the situation in which a very rich family cannot have any children, whereas another family has a lot of children but is extremely poor. The market's forces should take place and let the trade begin, but is it ethical at all? Can you trade children as they would be goods?

Trade is also important because it leads to:

1. Specialization – theory of comparative advantage
2. Specialization increases productivity

1) *Theory of comparative advantage*

Let's consider 2 individuals (Bob and Ann) and 2 goods (fish and bananas). Bob in 24 hours produces 10 fishes or 10 bananas whereas Ann in the same amount of time produces 30 fishes or 10 bananas. Here, we are assuming that both economic agents have monotonic and convex preferences. Let's take into consideration two scenarios:

1. *No trade* → Bob produces 5 fishes and 5 bananas and Ann 15 fishes and 5 bananas. Total production 20 fishes and 10 bananas.
2. *Trade* → Bob specializes in producing bananas (10) and Ann in fishes (30). Total production will be 30 fishes and 10 bananas. Here, Bob has an *opportunity cost* – the forgone benefit that would have been derived from an option not chosen – of 1 fish by producing 1 banana, whereas Ann has an *opportunity cost* of 3 fishes for 1 banana. Thus, Bob will be keen to trade 1 banana if he can receive back more than 1 fish, whereas Ann will be keen to receive 1 banana back for less than 3 fishes. Thus, they will trade 5 bananas for 10 fishes, ending up with: Bob, 5b and 10f, whereas Ann 5b and 20f. *Trade maximises welfare*!

The aforementioned scenarios embed implicitly the definitions of both comparative and absolute advantage.

a) *Absolute advantage* is the ability of an individual, company, region, or country to produce a greater quantity of a good or service with the same quantity of inputs per unit of time, or to produce the same quantity of a good or service per unit of time using a lesser quantity of inputs, than its competitors. Thus, in our example, we can say that Ann has an absolute advantage with respect to Bob in fish production (in the same time laps, it is able to produce 3 times Bob's quantity).

b) *Comparative advantage*, instead, stems whenever the opportunity cost (that is, the potential benefit that has been forfeited) for one company is lower than that of another. The company with the lower opportunity cost, and thus the smallest potential benefit, which was lost, holds this type of advantage. Hence, we can say that in our example Bob has a comparative advantage with respect to Ann in producing bananas (1 vs. 1/3).

2) *Specialization increase productivity*

Through specialization and coordination, people will be able to produce and transact more to acquire goods or services they desire. However, we shall always remember that **specialization requires coordination**: every task requires a complementary task on which someone else is specialized. People can produce more if they specialize **and** cooperate.

Time and efforts of specialists are wasted **unless** (1) they can be sure about the fact that the other specialists are doing their part; and (2) they will be able to buy on the market what is necessary for their needs.

Coordination is achieved whenever information is shared. Thus, producers must gather relevant information for instance about the complementary tasks. There are two ways to collect information:

a) *Centralized planning* (liked by socialist economists): it can be applied to the highest level of economic organization with a centralized decider (communism).

b) *Autonomous decentralized* decisions: everyone takes choices to maximize its utility. If coordinated, these choices lead to the optimal allocation (capitalism). Markets are costless mechanisms to achieve efficient allocations because prices act as "information vehicles" by signalling scarcity. But this is 100% valid only if markets are perfectly competitive: only in in this case prices signal true benefits and costs for the use of resources by the economic system.

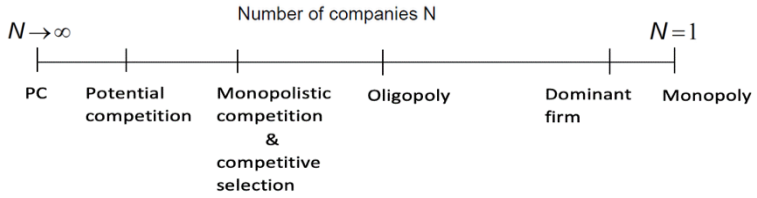
From a social welfare point of view, hence, the best general economic equilibrium is reached if all the markets are perfectly competitive (both for inputs and outputs). Unfortunately, in many circumstances this is not the case due to market imperfections such as: (1) market powers' presence, (2) externalities, (3) asymmetric information, and (4) transaction costs, which may lead to *market's failure.*

2. INSIGHTS ON COMPETITIVE STRUCTURES

Markets can be classified according to their competitive structures and, more specifically, in terms of *market power*, that is, the capability of a firm to set a price above its cost.

Markets can be characterized by monopolies or constituted by a dominant firm when no competitors challenge the throne/position of such firm in terms of market power or perfectly competitive, for instance.

Specifically, we can classify market structures considering the # of firms in such markets.



- *Perfect competition* (PC) consists of infinite companies producing homogeneous products
- *Oligopoly* consists of limited # of companies producing homogeneous products or differentiated products. To study oligopolies, it's common to make use of game theory.
- *Monopoly* consists of one single company producing a product without substitutes.

Specifically, the focus of this lecture will be on the two extremes: PC and Monopoly.

2.1 Perfect competition

In PC, each firm has (1) equal market power. A perfectly competitive market is a market where firms are (2) *price-taker*, i.e., they do not determine the price to which sell their products, price is settled by the market, i.e., by the interaction of demand and supply. For this typology of market to exist, 5 central *assumptions* must hold:

1. *Atomicity.* An infinite number of firms compete in such markets. Moreover, since they have the same size, each firm has a limited production capacity to satisfy the demand of customers and consequently, not sufficient power to affect the market price.
2. *Product homogeneity*. There's no differentiation among sellers. Furthermore, such homogeneity should not be perceived at a physical level only but by the customer as well.
3. *Perfect information*. Each agent, firm, and consumer know the price charged by every firm.
4. Firms have access to all production technologies required to stay in the business. However, for simplicity, one can assume the extreme form of *technology symmetry,* but it's not necessary. (Though, it might ease the analysis).
5. *No entry and exit barriers* (free entry and exit). No costs shall be incurred by companies to enter or exit the market.

These assumptions lead to two main results or, better, market's *dynamics*:

1. Companies are *price-takers*.
    - Products are homogenous, so if a company raises the price, consumers will buy the products from competitors and the demand for that company is null (perfect information). Therefore, no (rational) firm will raise the price.
    - Firms cannot collude, given their high number (atomicity).
    - If the company reduces the price, consumers will try to buy all the products from that company. But the company is not able to serve the entire market due to its limited production capacity (atomicity).
2. *Firms cannot realize any economic profits* (economic costs take into consideration opportunity costs as well, hence we are assuming that firms are already choosing their best options, i.e., minimizing opportunity costs. They could increase their profits in implementing their inputs in other activities):
    - If a firm makes extra-profit, this extra-profit will attract other firms to acquire the technology required (equal access to technology) and enter the market (no barriers).
    - This induced competition will erode in the long run any possibility of extra-profits for the firms with the adoption of the best technologies possible by firms. Extra profits will work as a signal, attracting new firms and eroding any extra-economic profits.

      Remark: no extra profits does not mean that a firm does not gain anything by competing in such a market!

Under such market configuration, in the long run the equilibrium will be achieved through:

$$p = MC = \min AC$$

To understand why, we need to first recall that companies do seek profit (π) maximization. Then:

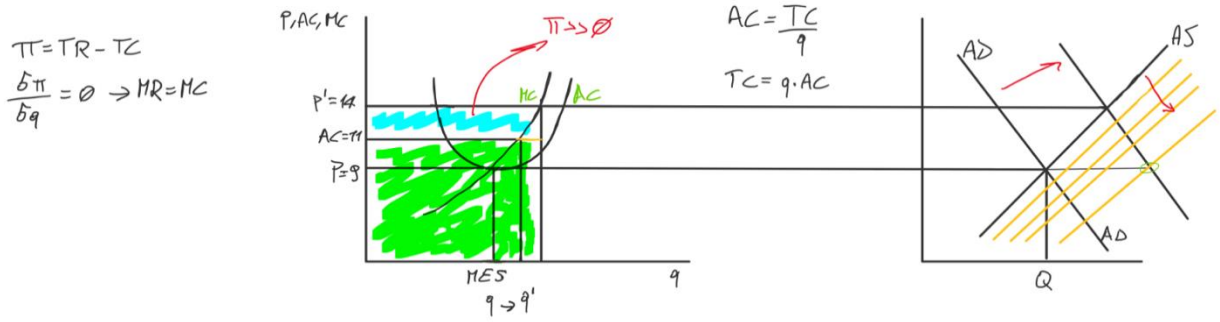
$$\max \pi = p \cdot q - TC(q)$$

$$\pi = TR - TC$$

$$\frac{\partial \pi}{\partial q} = \frac{\partial TR}{\partial q} - \frac{\partial TC}{\partial q} = 0$$

$$MR = MC$$

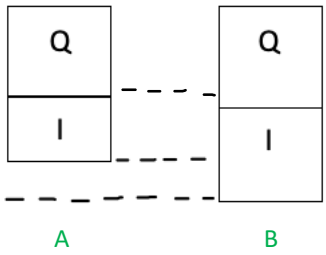
$$p = MC$$



More specifically, In the long run, PC will maximise the three components of *social welfare*, namely:

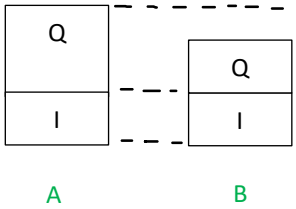
1. *Productive efficiency* → being able to produce a given number of outputs with the lowest number of inputs (costs) possible (static concept)

   Let's consider 2 firms: A and B. let Q be the total quantity of goods produced and sold and I the number of inputs available.



Since Q is the same in both cases but the amount of I used by B are greater than the amount of I used by A, A is more efficient than B from a productive POV.

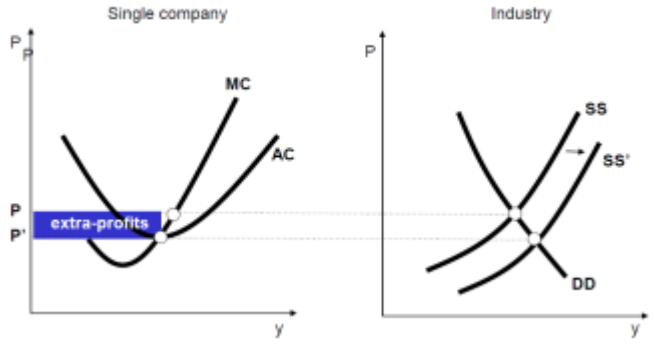
2. *Allocative efficiency* → being able to produce the largest quantity of goods (sold at lowest price) with a given (fixed) number of inputs (static concept)



Since I is the same in both cases but the Q produced and sold used by A is greater than the Q produced and sold by B, A is more efficient than B from an allocative POV.

3. *Dynamic efficiency* → capability of an economic system to introduce new goods or service into a market system (i.e., innovation)

To better understand the dynamics of a perfectly competitive market, we look at the Demand (DD) and Supply (SS) curves.



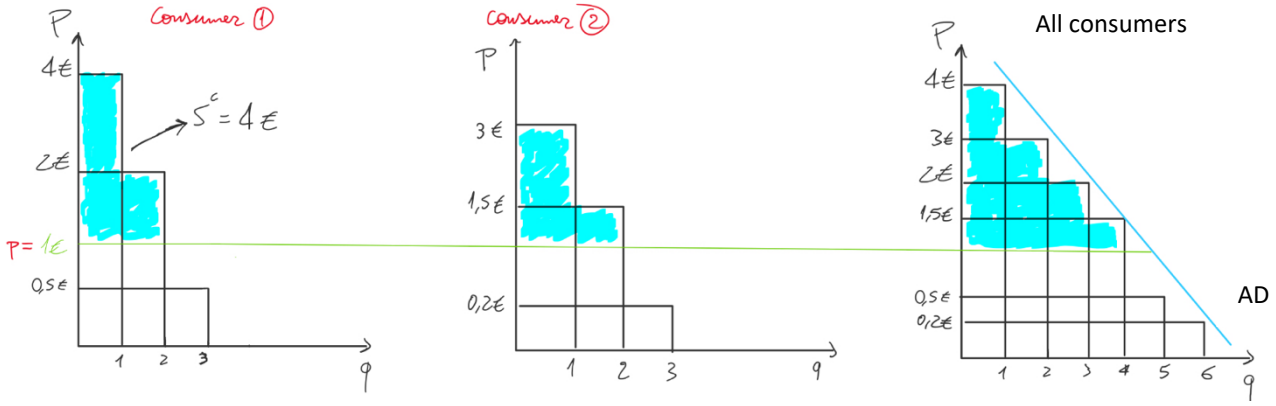
At price P, the profits realized attract new firms in the industry over the long run. Entrants will enter, start operations with easy access to technology and compete. The SS shifts to the right, quantity goes up, and the equilibrium price decreases. Over the long run p = min (AC) is the equilibrium. In the opposite situation, if the firms incur in a loss, they want to move inputs to an alternative use. Some firms will exit the market and the SS will shift to the left.

Shutdown condition: AVC(y) = $\frac{c_v(y)}{y}$ > p

If average variable costs are still greater than p, the firm would be better off producing zero units of output. This makes good sense, since it says that the revenues from selling the output y don't even cover the variable costs of production, cv(y). In this case the firm might as well go out of business. If it produces nothing it will lose its fixed costs, but it would lose even more if it continued to produce

2.1.1 Surpluses concepts

Consumer's surplus: it is the difference between the price an individual is willing to pay to have a certain good or service and the market price for the same good or service. It measures the consumers' welfare.



Remark: aggregate demand (AD) is given by the horizontal sum of the amount of goods each consumer is willing to buy at a given price level P. The aggregate supply (AS), similarly, is given by the sum of individual supply curves (willingness to sell) and is equal to their marginal cost curves.

Producer's surplus: it is the difference between the price of a certain good or service paid to the producer and the price the producer is willing to accept for selling the same good or service. It measures producers' welfare.

⇒ Social Welfare achievable in each market: $W = S^c + S^p = \Sigma v\pi$

2.2 Potential competition – contestable markets

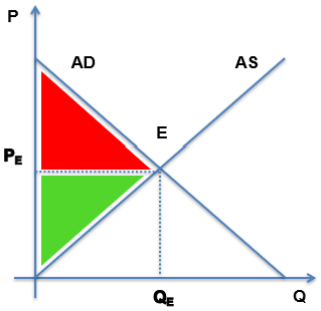
A contestable market is a market where firms from other markets/sectors can perform a hit and run competition with no costs of entry and exit. Pre-requisites required:

1. No atomicity (on the contrary of perfect competitive markets)
2. No entry and exit barriers (no sunk costs and non-redeployeable investments)
3. Perfect information for consumers (they can react immediately to price differentials between companies)
4. Time requested for the incumbent to retaliate to the entry of the new firms (by lowering price) is superior to the time needed for the entrant to make all the investment necessary to operate in the

focal market (i.e., making market contestable) -> fundamental assumption for the creation of a "eat and run" market.

Results:

- Incumbent Firm(s) are forced to settle a price near to the average cost in order not to "turn on" the signal of extra-profits.
- As a matter of fact, every extra-profit will be captured and exploited by new entrants with a hit and run competition.
- If the market is contestable, the n° of firm is a poor predictor of the market power, and even a market with only 1 firm may behave more similarly to perfect competition rather than monopoly.



Remark: The hypotheses underlying perfect competition and contestable markets are difficult to observe in the real world. These models become only useful as benchmark models for comparing results (in terms of efficiency and social welfare) of more realistic market structures, where firms do have market power (oligopoly, monopoly).

2.3 Monopoly

Under this typology, there is only on firms serving the market, thus that is entitled to the entire supply. In this case, the firm is not a price-taker anymore since no competition exist. Hence, it will have the market power to set the price. How does it choose? First, we shall note that choosing price is equivalent to choosing quantity to produce.

Max π = TR(q) – TC (q) → p(q)*q – TC (q) → MR = MC → F.O.C. $p + q \bullet \frac{\partial p}{\partial q} = MC$ or also: $p\left[1 + \frac{q}{p}\frac{\partial p}{\partial q}\right] = MC$

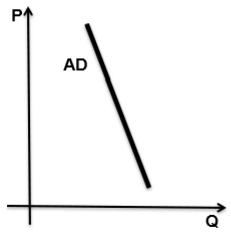
where the term in brackets equal to the inverse of the elasticity of demand, which is the variation of demand consequently to price changes

$$\varepsilon_p = \frac{\frac{\Delta q}{q}}{\frac{\Delta p}{p}} = \frac{\Delta q}{\Delta q} \cdot \frac{p}{q}$$

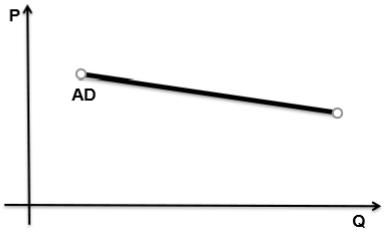
Optimum price, where the markup (maximized) is given by the inverse of the demand elasticity. The monopolist will settle a higher price if it faces an inelastic demand (ε smaller than one), while the more the demand curve is elastic the more it will fix a price close to MC, like the one that would emerge under perfect competition.

$$p\left[1 - \frac{1}{|\varepsilon|}\right] = MC$$

A) P = 50 -0,8Q B) P = 25 – 0,2Q

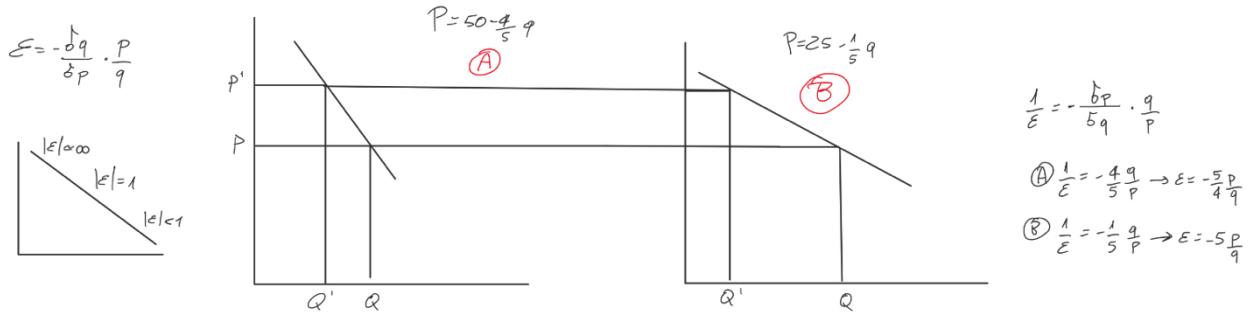


Rigid AD



Elastic AD

We expect that the monopolist will price its goods higher in case of an inelastic demand with respect to an elastic demand. Indeed, the elasticity of demand defines how much the customers do need the specific good. If very inelastic it means they will be "almost" indifferent to pay, because requiring a given amount.



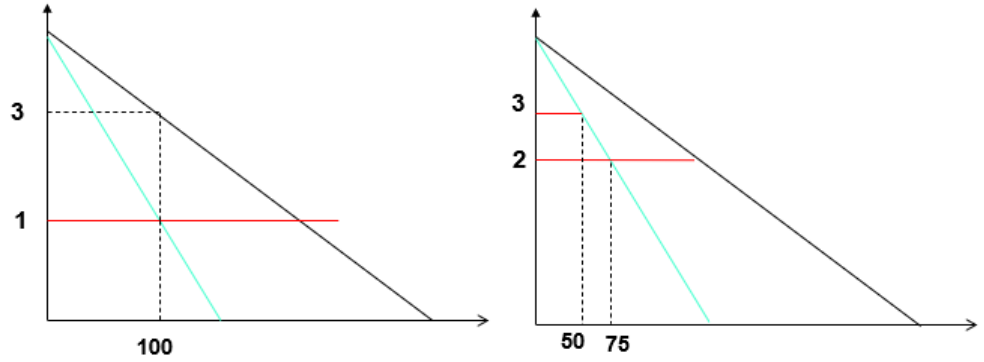
Remark: in perfect competition the (perceived) aggregate demand is horizontal and thus elasticity tends to infinity (P= MC).

Additional notes:

1) The analysis on pure monopoly we conducted so far can also be applied to the market context in which there's only one very large firm and a set of very small firms with a limited production capacity. If K is the total production capacity of the set of small firms, these latter typically fix a price only marginally inferior to the one fixed by the large firm and produce a quantity such as their capacity is saturated

2) 1 monopoly is better than 2. This is known as the "double marginalization" problem (DMP). Under this circumstance, both firms' and consumers welfare are affected negatively. To understand it better, let's see an example: We have a monopoly in a retail market with p = 5 – (1/50) *q, no fixed costs and use of only 1 input for each unit of output (1 engine for 1 car). Now suppose MC for using this input is constant and under perfect competition in the wholesale market the retailer buys the good for 1 Euro. Thus,

MR = 5 – (2/50) *q = MC = 1 → q = 100; p = 3



Here we need to observe that the Marginal Revenue curve for the retailer is the demand curve for the wholesaler.

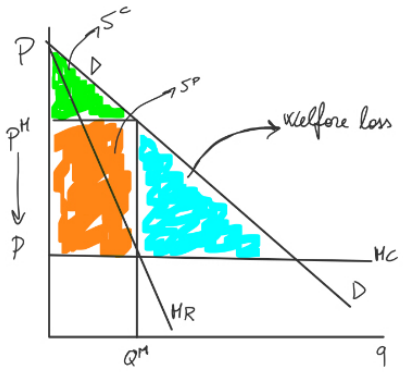
Now suppose that wholesale market is not anymore in perfect competition but instead is dominated by a monopolist. What is the price (and quantity) that the monopolist would settle? The demand faced by the monopolist in the wholesale market is p = 5 – (2/50) *q (The marginal revenue, i.e., the marginal cost, of the retailer)

MR = 5 – (4/50) *q = MC = 1 → q = 50; p = 3

What the monopolist in the retail market will do? Now its marginal cost is 3 and not 1 as before.

MR = 5 – (2/50) *q = MC = 3 → q = 50; p = 4

So, in the 1° scenario (wholesale market perfectly competitive): $\pi = 200$; $S^C = 100$, whereas in the 2° scenario (wholesale market in monopoly): $\pi_{wholesaler} = 100$; $\pi_{retailer} = 50$; $S^C = 25$. This example clarifies the problems of having two monopolists that are both trying to internalize a margin.



2.4 Natural Monopoly

First, we must bear in mind that sometimes having a natural monopoly might the best scenario possible.

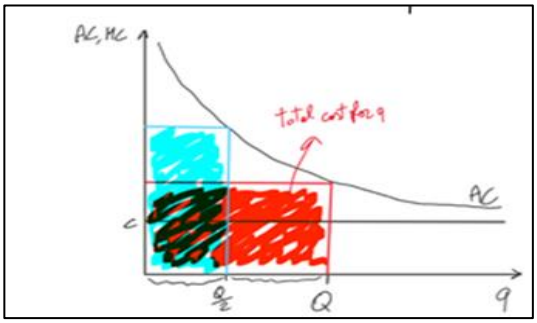
This is due to the "subadditivity" of the cost function, i.e., assigning the overall costs of the production to a single firm is more efficient (less costly) than partitioning the costs among several firms (if every firm faces same cost structures, i.e., is adopting the same technology).

$$\sum_{i=1}^{n} q_i = Q, \forall n \geq 2 \qquad TC(Q) < \sum_{i=1}^{n} TC(q_i)$$

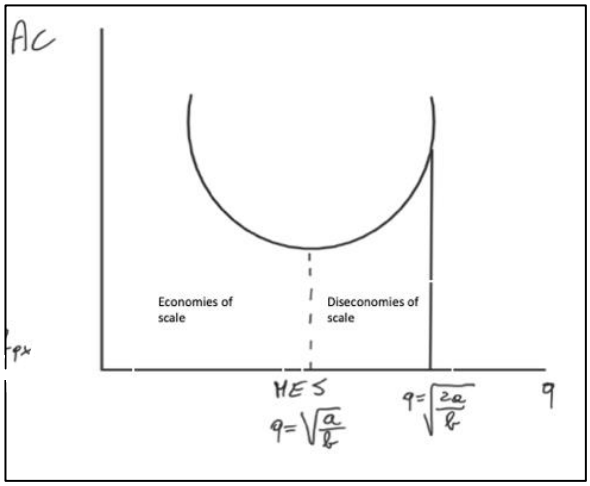
i.e., the total costs to produce Q (as sum of qi) is larger if n is at least 2

This, because of FCs will be "spread" among each additional product that is produced.

Assumption: affine cost curve, i.e., a cost curve where the average costs decrease asymptotically. Hence, it is efficient to allocate all the costs to one single firm, i.e., a monopoly. *Remark*: scale economies (i.e., affine cost curve) are sufficient but not a necessary condition to prove subadditivity. Sub-addictive characteristics, indeed, can still be present even in absence of economies of scale.



Affine $TC(q) = F + Cq \rightarrow AC = F/q + c$ and $MC = C$. Thus, $\frac{dAC}{dq} = -F/q^2 \ll 0$, i.e., is always decreasing. If this was the case, then an affine curve is present. A Natural Monopoly would be the most efficient outcome in terms of productive efficiency



Which is the optimal production level that minimizes industry's total costs if more players are involved?

If $TC = a + bq^2$, then $AC = a/q + bq$ and $MC = 2bq$.

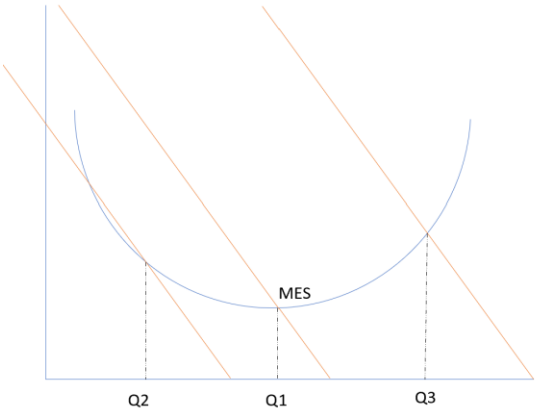
$TC_{industry} = TC(x) + TC(q-x) = a + bx^2 + a + b(q-x)^2$

$\frac{dTC}{dx} = 2bx + 2bx + 2bq = 0 \rightarrow x = q/2$

$$\frac{dAC}{dq} = 0 \rightarrow -a/q^2 + b = 0 \rightarrow \quad q = \sqrt{a/b}$$

Hence, $a + bq^2 < 2\left[a + b\left(\frac{q}{2}\right)^2\right]$, which leads to $q < \sqrt{2a/b}$.

This shows that to minimize the Total Costs of the entire industry, q shall be divided equally among the players. Thus, we will find ourselves in a case of monopoly if $a + b*q^2 < (a + b * (q/2)^2)$ which leads to $q < \sqrt{\frac{2a}{b}}$



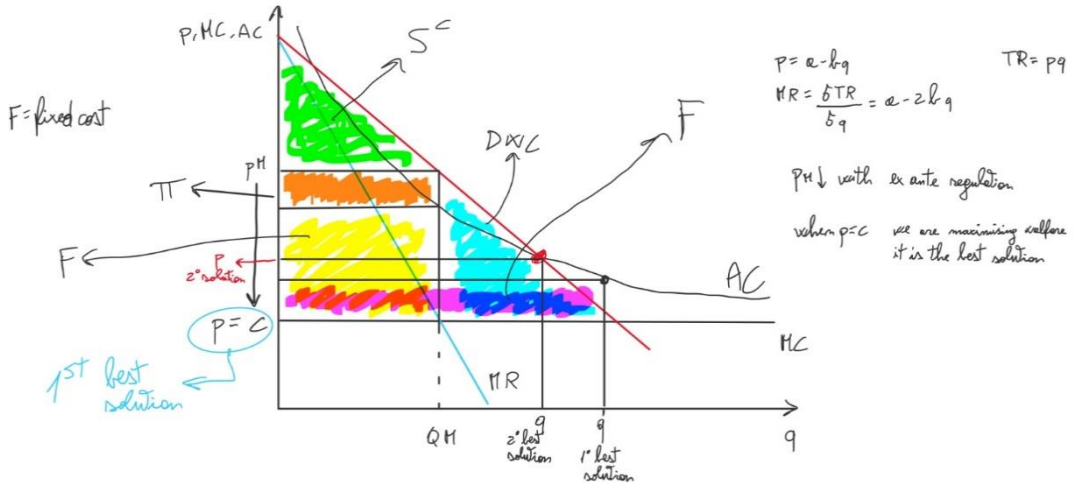
If we look at the picture above, we can notice that the quantity that shall be delivered in the market will be represented by the intersection between the demand curve and the average cost curve. Hence, it will be the market demand one of the main drivers deciding whether it is convenient to have a monopoly or not. Consumers will be the deciders.
Three cases can be considered under this scenario:
Case Q1: Natural monopoly, because letting one firm producing the given level of quantity is less costly than letting more firms producing Q1.

Case Q2: Let's suppose that we have two firms, where one (A) produces at MES. To that, we need to add what is not yet produced by the first firm but is required by the market.

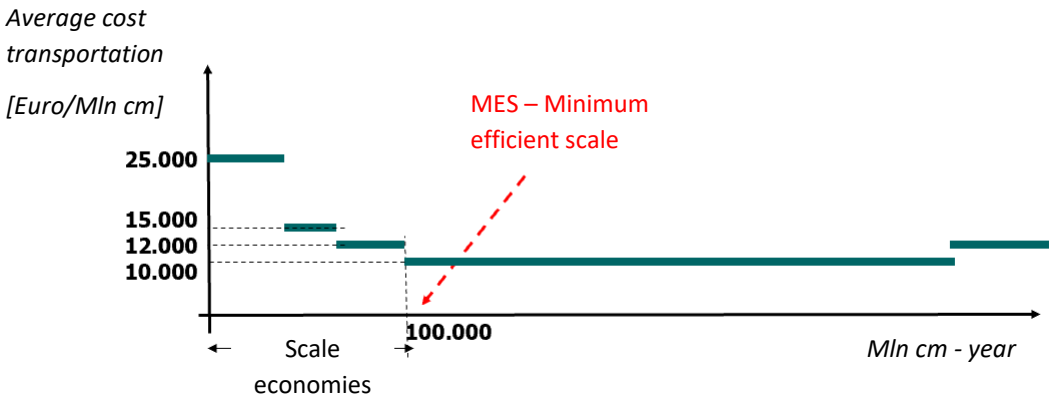
Case Q3: Let's suppose that Q3 is 2 times MES. The simple graphical representation shows that the further you move from the minimum efficient scale quantity, the less likely a monopoly will take place.

For the sake of simplicity but keeping the insights intact we will still consider an affine cost curve, i.e., we are sure that a monopoly is present. What is the problem with a natural monopoly?

If no competition arises, because of high entry barriers, for instance, the firm will behave as a proper monopoly, hence maximising its own surplus (revenues). To ensure allocative efficiency (minimizing price charged to the consumer), thus, ex-ante, **regulations** are required to reduce the DWL that arises from monopolist's behaviour. Such regulations will force *pm* to decrease and get much closer to *c*, indeed, as it can be noticed from the graph below.



2.4.1 An example of Natural Monopoly: Natural Gas



Note that there is not one single point in which MES is achieved. Furthermore, in this specific industry strong *scale economies* might be stemming from:

- Great amount of indivisible resources (resources that should be present even for small scale operations. E.g., central delivery points, rapid interventions team, regulatory office, etc.).
- Specialization of workforce
- "Volumetric returns to scale of the pipelines
- Administrative costs: Natural monopoly was and still is the crucial economic concept behind public service utilities and many network industries. Natural Gas, Electricity, Water, Telecommunications, Railroads: all businesses that delivers an essential good or service through a wide network infrastructure.

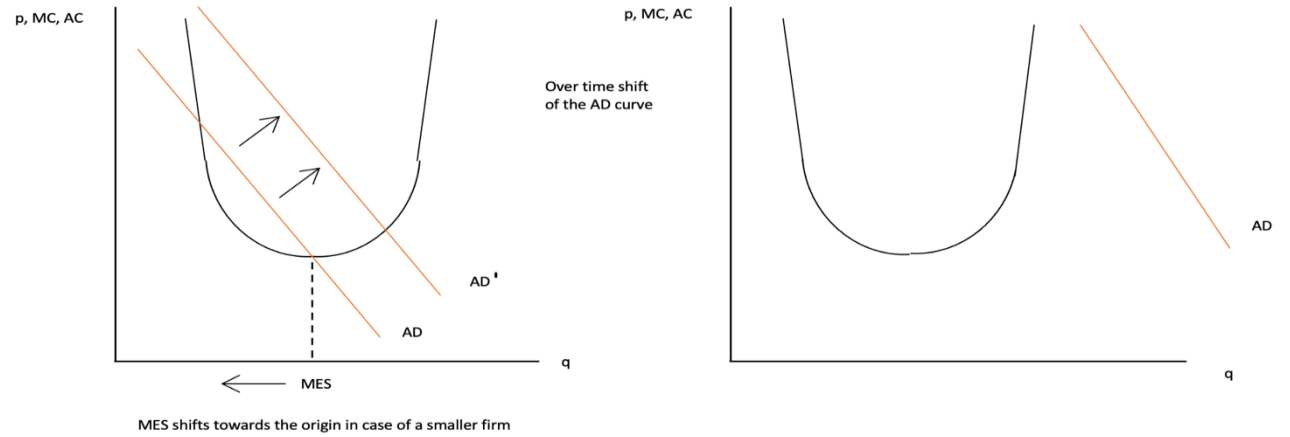
For what concerns Gas, Electricity and Telecommunications in specific, different phases to deliver the related services are involved:

1. Production/Generation
2. Transportation
3. Distribution
4. Sales

Before 2000, *all* these stages were delivered by *one* single firm, which represented a natural monopoly (Eni, Enel, and Telecom). From an efficient production point of view, it has always been more convenient to allocate all the stages to one single firm (cost minimization). We know that whenever facing a monopoly, a trade-off between productive and allocative efficiency occurs. Thus, how could the loss in allocative efficiency be faced? By making such firms government-owned (or state-owned) enterprises. This, indeed, is precisely because the goal of the government is not profit maximisation but rather social welfare maximisation (prices were set to reach the latter objective). If p M=MC, the firm can recover only variable costs losing the fixed costs. From a social point of view, this is the best solution. Often the regulation imposes p M=AC: therefore, a firm does not incur in a loss without having an extra-profitability. From a social point of view, this is the second-best solution.

Starting from 2000, instead, *production* (Upwards stage) and *sales* (downward stage) were not assigned to a single firm anymore (no natural monopolies anymore), whereas the *infrastructural stages* (transportation and distribution) were (local monopolies rather than national ones). The loss of monopoly conditions allowed to any firm to enter in these specific stages of the production (*liberalization*). Simultaneously to the liberalization process, *privatization* of Government-owned companies begun. This latter movement enhanced the need of setting up regulatory agencies (e.g., Arera, Agicom, etc.), i.e., public agencies that were setting the rules of the game, more specifically the prices charged to the final consumers. But why the upwards and downwards stages compared to the intermediate ones have been opened to competition? We

said that this happened because the natural monopoly conditions did not hold anymore…but why? Two reasons: One related to the demand-side and one to the supply-side.



Demand Side: The demand curve shifted to the right because of an increase in demand for the services supplied.

Supply Side: The MES shifted to the left due to technological progresses which made production and services' management more efficient.

Remark: these two effects are mutually reinforcing.

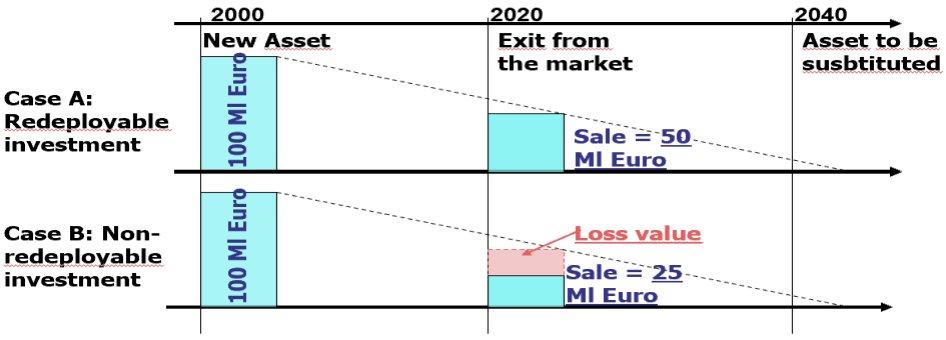
Economic Regulation can take the form of:

1) *Ex-post* regulation: ANTITRUST (later in the course): collection of laws which regulates the conduct and organization of business corporations to promote fair *competition* for the benefit of *consumers.* Two areas of great importance: anticompetitive practices (e.g., cartels) and abuse of dominant position.

2) *Ex-ante* regulation: specific REGULATORY COMMISSIONs that right from the beginning set invasive and pervasive "rules of the game" (e.g., including prices of final or intermediate services/products).

There are 2 important reasons reinforcing the need to regulate network industries rather than only natural monopolies.

First, the presence of (1) inelastic demand (demand side), and second, the presence of (2) non-redeployable investments (sunk costs – supply side). For what concerns the first one, services provided by natural monopolies are usually necessity goods: we need them badly. There are not substitutes and if there are some, they might be far from perfect ones (imperfect substitutes). This characterizes the rather inelastic demand.



For what concerns "non-redeployable" investments (very specific to the context they have been made), it is quite difficult to assess how much residual value one can get from the investment in case of exit from the market. Their value outside the context is very much penalized if not null. In case A shown aside, we made an investment that will last 40 years (depreciation is assumed to be linear). In case of a redeployable asset we can realize the entire residual value (50 mil in 2020) by selling it to a buying party. In case B, non-redeployability is the cause for a 50% value loss. Secondhand markets are highly imperfect because of the specificity of the investment (geographical, product related but also institutional) reducing the number of credible acquires (high bargaining power of potential acquires); and because it is difficult from the other side

(potential acquirer) to infer the "true" value of the asset (information asymmetries). These are the reasons because of which network infrastructure are scarcely redeployable. Thus, the implications related to redeployable assets are under-investment and no competition (high entry barrier for new entrants and high exit barriers fir incumbents: if you want to entry into a market making a huge investment into such an asset, thus being stuck afterwards, you will consider it twice. The same reasoning can be applied to the incumbents whenever thinking to exit the market. High incumbents' resilience will be a natural consequence of non-redeployablility thus.

3. FIRM HETEROGENEITY, MARKET POWER AND PRICE DISCRIMINATION

3.1 Monopolistic competition and competitive selection

Monopolistic competition and competitive selection are two market forms originating from an attempt to conciliate the results that we achieve in perfect competitive markets with what we see in the real world. Indeed, in reality, there are many markets which seem to be perfectly competitive although they do not actually reach the long run the result expected in perfectly competitive markets, i.e., MC = P and where every firm produces at the MES.

3.2 Competitive selection model

Model that stems from three *stylized facts* that we should observe in perfectly competitive markets, but which are not realistic, i.e., do not appear in reality:

1) Although theoretically every firm involved in the competition will realize 0 economic profits in the long run, in reality some firms will realize larger profits than competitors.
2) The second result that characterizes perfect competition, but which does not cope well with reality is that in theory we observe simultaneous entry *or* exit of the firms, whereas in reality the markets might contemporaneously enter *and* exit the market (turbulence)
3) All firms are of similar size (in the long run) and thus productivity, while conversely in the real-world firms do differ substantially regarding the aforementioned two dimensions.

Jovanovich tried considering a model in which perfect competition is assumed but where the 3 real-world's aforementioned aspects are taken into consideration as well. This is called: *Competitive selection model*, where each firm is characterized by θ (estimate of the value of own productivity, capability): $\Pi = pq - \frac{q^2}{\theta}$, where θ underlines the capability of the firm of organizing resources. This ability is assumed to be known AFTER the firm has entered the market of interest. Time is made by several periods. At the beginning of each period, each firm decides whether to remain active or not. Next, active firms will decide how much to produce, which they do by choosing the quantity that maximize profits.

- 1° order condition: $p - \frac{2q}{\theta} = 0$; so: $q^* = \frac{1}{2} p\,\theta$
- Maximum profit levels: $\Pi = \frac{1}{2} p^2\,\theta - (\frac{1}{2} p\theta)^2 / \theta$; so: $\Pi^* = \frac{1}{4} p^2\,\theta$

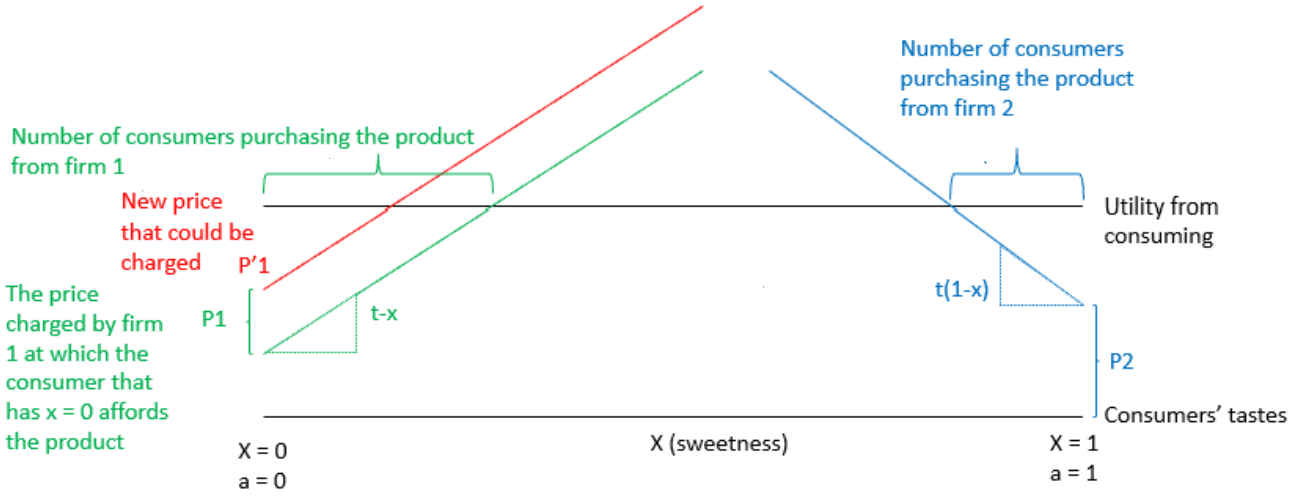
Note that the more productive firms (higher θ) are those that will produce more, and each firm is a price-taker (as perfect competition)

3.3 Monopolistic competition

"It is evident that virtually all products are differentiated, at least slightly, and that over a wide range of economic activity differentiation is of considerable importance".

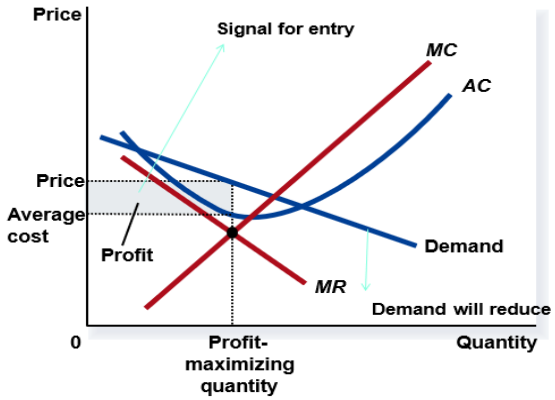
In this model, Chamberlain takes the 5 assumptions behind perfect competition and gives a twist to one of these, namely *product homogeneity*. Indeed, the perceived demand by each firm in case of product homogeneity should be flat (firm is a price taker, i.e., has no influence on the market), whereas if product is *different* (introduced by Chamberlain), the perceived demand is not as elastic as before (i.e., not flat) thus granting the firm some market power

Condition for which the consumer buys the product: U > p + t*|x-a|, where a = *"location choice"*, i.e., the farer from the consumer's preferences, the less likely she will purchase that good

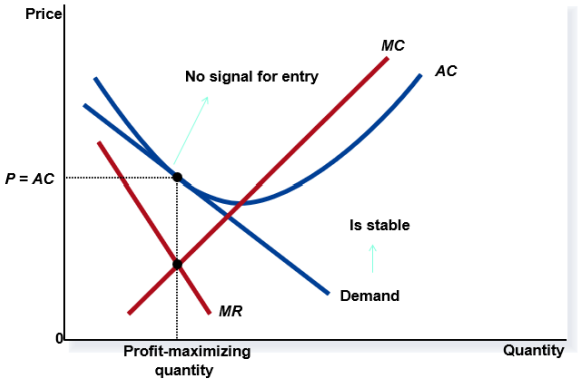


Remark: this graphical representation shows that firms do not actually perceive a flat demand curve, but rather a negatively sloped one. Thus, each firm, according to the product it offers, and thus the preferences that it targets, has some market power. In the case of firm 1, if it decides to increase the price to P' demand would not be null as it would be in case of perfect market competition where the firm is a price taker, but it will shrink to a number still larger than 0.

Short run



In the short run, since not facing a perfectly elastic demand curve, the firm will try to maximize its income streams setting the MR = MC (first condition), achieving some economic extra profits. This possibility will perform as a signal attracting other firms that will start to compete, shrinking the extra profits down to 0 (no more signals to entry). But, since the conditions of no entry and no exit barrier still apply these levels of extra profits will function as a signal for entry, hence new firms will enter the market and will compete with the existing firms and the demand curve that each company faces will shrink until it reaches the final outcome, i.e. a situation in which the market price will be set equal to the average costs: no signals for entry will be present anymore. Still, we are facing some sort of inefficiencies, namely in terms of both allocative and productive point of view. From the latter point of view firms will <u>not</u> produce at their MES, but rather at less than that, thus achieving a "proliferation" of firms, i.e., too many competitors producing small quantities which are not enough to maximize their production efficiency. They are not exploiting their productive capacities. From the allocative point of view, price is not the minimum anymore, as we had in perfect competition, but it is above the minimum



Long run

average costs. There could be further transactions that could be exploited to achieve larger social welfare. Results, tough, are not so far from perfect competitive markets.
NOTE: The main difference between perfect competition and monopolistic competition lies in the short run thus, where the firms are able to achieve some extra-profit. Furthermore, whereas in perfect competition firms produce where MC=P, here P>MC, since not producing at the MES anymore.

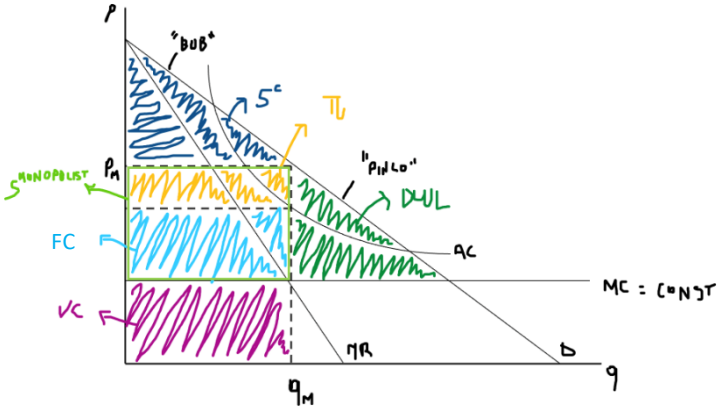
3.4 Final taxonomy

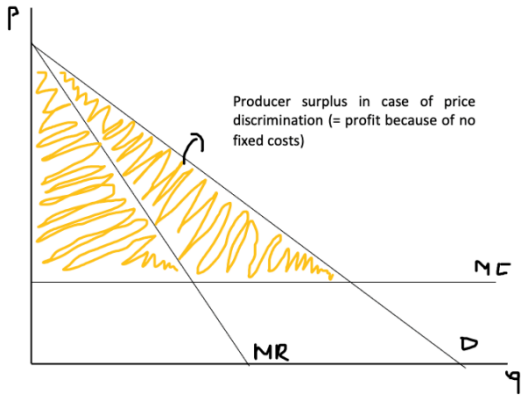
1. Perfect competition (PC): p min; Q max; consumer surplus max; null extra-profits; SW is max.
2. Potential competition: close (or not too distant) results to (from) PC.
3. Monopolistic competition (with moderate differentiation) and competitive selection: close (or not too distant) results from PC.
4. Oligopoly: p intermediate; Q intermediate; consumer surplus intermediate; positive extra-profits; intermediate SW.
5. Dominant position: close (or not too distant) results to (from) monopoly).
6. Monopoly: p max; Q min; consumer surplus min; extra-profits max; SW min.

3.5 Price discrimination

Why do firms want to "discriminate"?

"Each customer is charged a different price – exactly matching his/her willingness to pay for each unit". Indeed, we shall ask ourselves whether it is convenient or not from a social welfare point of view if the monopolist opts for producing that quantity allowing it to maximize profits. Recalling that social welfare is composed by the monopolist's and the consumers' surpluses, whenever a monopolist opts for a quantity (hence a price P_m) that maximizes its own profits it generates a dead weight loss, i.e., a loss in social welfare, along a small (but still present) consumers' profits (those willing to pay a larger amount than what is actually asked). From the monopolist point of view, it would be more convenient to internalize these two sections (DWL and Consumer Surplus). How? Through *1st degree price discrimination* (also referred to as perfect PD) By doing so, both Bob, who is capitalizing a surplus at P_m, and Pinco, who is not included in any transaction because not offered a price he is willing to disburse (P_m is too high), *will be charged the price they are actually willing to pay* generating larger welfare for the monopolist and, at the very end, for the society since incrementing the number of transactions.





More specifically, as shown by the second picture, this will maximize producer surplus, maximize social welfare, BUT no consumer's surplus will be realized (ethical aspect). Moreover, note that the monopolist will sell until the intersection between the demand curve and the marginal costs, since below it the utility of the further consumers is smaller than the actual cost faced by the producer to provide that specific good.

Implications of 1st degree price discrimination:

1. The monopolist should know exactly each consumers' willingness to pay

2. Difficulties to avoid arbitrage (re-sale): people that have a lower willingness to pay could be charged their relative price but then sell the good to those consumers willing to pay a greater price, making a profit, and breaking down the positive outcomes of price discrimination.

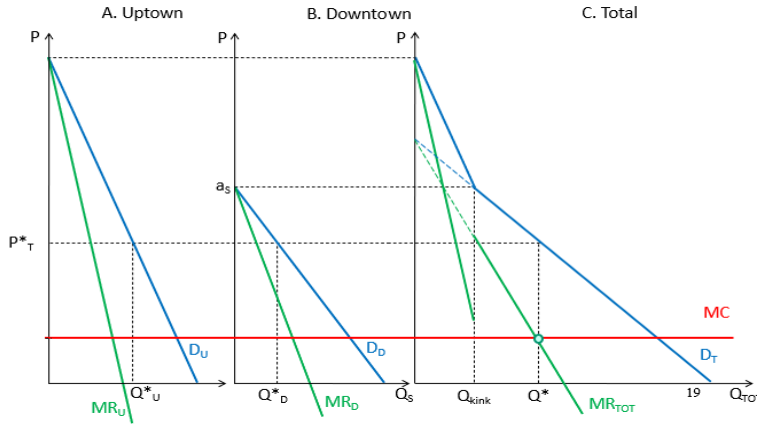
Nowadays, thanks to the latest technologies, price discrimination is possible (think of the flight industries, e.g., yield management). They have the possibilities to track our behaviours, analyse our data and predict our (persona's) actual willingness to pay. Of course, a perfect 1st degree price discrimination is not achievable.

Here, we are talking about a *3rd degree price discrimination*, which relates to "group pricing", i.e., different prices for different groups of consumers but same price within the same group. Objective? Create different sub-markets with different elasticities of the D curve. So, each group will bear different prices depending on its elasticity if inelastic, higher price. If elastic, lower price). Groups might be selected according to some exogenous indicators such as age, occupation, geography, or device (Apple, Samsung etc.). To find the suitable price for the given targeted group, the "elasticity rule" shall be applied.

In each sub-market *i*: max π_i = TR (q_i) – TC (q_i) o $p_i(q_i) \cdot q_i$ – TC (q_i)

First order condition: $p_i + q_i \bullet \frac{\partial p_i}{\partial q_i} = MC \rightarrow p_i\left[1 + \frac{q_i}{p_i}\frac{\partial p_i}{\partial q_i}\right] = MC \rightarrow p_i\left[1 - \frac{1}{|\varepsilon_i|}\right] = MC \rightarrow \frac{p_i - MC}{p_i} = \frac{1}{|\varepsilon_i|}$ (Lerner formula)

Of course, the absence of resale possibilities is required in this case as well. Why is this scenario convenient for the monopolist with respect to a homogenous pricing strategy, i.e., no discrimination at all?



Market is divided into "Uptown" and "Downtown". First, we notice that the elasticities are different. Indeed, in the uptown market, demand is more inelastic with respect to the downtown market (relatively poorer). In the case if the firm cannot perform any price discrimination, Q* will be the quantity produced, which as shown in the graphs is too large in the uptown market and too low in the downtown one. Indeed, this can be seen by the different intersections of the marginal revenues' curves in both markets. This can be seen in a more formal way as well. Let's assume that the inverse demands are:

- Uptown market: $P_u = 8 - q_u$
- Downtown market: $P_d = 5 - q_d$
- Total market: P = 8 – q (in the first part) and P = 6.5 – 0.5q

Scenario A: price discrimination

- f.o.c in U: $8 - 2q_u = 2$; $q_u = 3$; $p_u = 5$; $\pi_u = 9$
- f.o.c in D: $5 - 2q_d = 2$; $q_d = 1.5$; $p_d = 3.5$; $\pi_d = 2.25$
- Total π_{ud} (in this case = S^p) = 11.25
- Total S^c: S^{cu} (= 4.5) + S^{cd} (= 1.125) = 5.625
- W = 16.875

Scenario B: no price discrimination

- f.o.c in T: $6.5 - q_T = 2$; $q_T = 4.5$; $p_T = 4.25$; π_T (again = S^p) = 10.125
- $S^c = 4.5 + [(4.5+3) * 0.75] / 2 = 7.3125$

- W = 17.4375

Note: To obtain the Total market demand, you simply sum the two demand functions paying attention to the fact that $P_d + P_u = P_{TOT}$ and not 2P

This simple example shows us that for a firm, 3rd price discrimination increases its profits, but creates a larger DWL (Total social welfare is smaller), since less trade possibilities are seized.

Remark: If price discrimination can increase the quantity produced to the market, we cannot take for granted that the social welfare decreases with respect to no-discrimination. In case of equal quantity, tough, price 3rd degree price discrimination will deliver a less efficient result in terms of social welfare with respect to homogenous pricing.

On the other hand, *2nd degree price discrimination* revolves around the concept of self-selection by the consumers. Sellers cannot directly identify consumer type, but they can still induce consumers to distinguish themselves. This selection may be based on the willingness of consumers to consume:

- Different quantities (thus, the price paid by consumers depends on the quantity of the good consumed: non-linear pricing, bundling. The more you buy, the less you spend per unit.
- Different versions of the same product – *Versioning*: (often) case of *vertical differentiation*, i.e. where the premium version is without any doubts better than the basic one (rather than *horizontal*, where tastes, preferences, are taken into consideration). Timing might be another variable for versioning as well (e.g., think about watching a new movie: if you want to watch it immediately you go to the cinema and pay more than waiting a little bit to purchase the DVD)

3.6 Versioning

Versioning: the aim is to sell more than one version of the same product (e.g., premium, and basic) at different prices targeting different segments of consumers (experts and beginners). It is a pure strategic-driven decision. Indeed, this applies also when the low-quality version has the same cost of production of the high-quality one (or even higher):

- Software: Basic version obtained by degrading the premium through the disablement (bearing some costs) of some functions.
- Information services about share prices: the "delayed" version is produced with some additional costs with respect to the "immediate" version.

Leave those consumers choose the version they prefer (self-selection). Versioning can produce a better result with respect to homogenous pricing:

	Impatient customers	Patient customers
Immediate Version	100	50
Delayed Version	40	30
Number of customers	40	60

Here we have two different product versions and two different types of customers. Note that the patient customers, despite not interested in the product itself, still do value the immediate version more than the delayed version. In case of uniform pricing, the supplier will sell the immediate version at 50, realizing a profit of 100*50 = 5000.

- In case of 1st degree discrimination, the supplier knows what type of customer it has in front of it, hence it will maximise its profits 100*40 + 50*60 = 7000.
- In case of 2nd level discrimination, instead, we will have an intermediate solution, by selling first the immediate version to the impatient customers 100*40 and then the delayed version to the patient customer 60*30…WRONG. The fallacy here lies in thinking that the impatient customers will buy the

immediate version instead of the delayed one, but, if rational enough (tough assumption) it will prefer to wait for the delayed version which will be sold at 30 and realize a surplus of 10. Thus, the final revenues will total to 3000.

But, still, there is a way to reach larger profits through 2nd degree price discrimination with respect to uniform pricing. By setting the price of the immediate version a price infinitesimally smaller than 90, impatient customers will be willing, compared to the first case, to buy this version since their surplus will be slightly larger than 10. For what concerns the patient, instead, the price charged will be exactly 30. The profits achieved will equal to 90*40 + 30*60 = 5400.

The last solution (right one) concerns a pricing strategy that respects the so-called *Incentive compatibility constraint*, which requires that a version designed for a specific customer is actually bought by the targeted consumer.

3.7 Price Discrimination

RECAP: 2nd Degree Price discrimination consists of the consumers self-selecting themselves. The seller cannot directly identify consumer type but can still induce consumers to distinguish themselves. This selection may be based on the willingness of consumers to consume:

- Different quantities (so price paid by consumers depends on the quantity of the good consumed: nonlinear pricing, bundling).
- Different versions of the same product (Horizontal or Vertical).

The art of making a profitable Versioning. Key constraint (incentive): you can't make the inexpensive version too attractive to those willing to pay more.

On the other hand, two other important constraints shall be taken into consideration by the seller to perform a proper versioning strategy: you need to (1) lower price of the premium version and (2) lower the quality of the basic version. Additionally, the cheap version (of decent quality) must be relatively inexpensive that those with relatively low (but still sufficient) willingness to pay wish to purchase (to “be in the market”) [PARTICIPATION] and it should be impossible for consumers to transform the basic into the premium version, really different versions [GOOD DESIGN].

Example: Windows IT (Shapiro & Varian, 1999): Microsoft offers two versions of its Windows NT software: the Windows NT Workstation, which sells for about $260, and the Window NT Server, which sells for $730-$1,080, depending on the configuration. Workstation NT can run a Web server but accepts only ten simultaneous sessions; the server version will accept any number of simultaneous sessions. According to an analysis by O'Reilly Software, the two operating systems are essentially the same. In fact, the kernel (the core component of the operating system) is identical in the two products and relatively minor tuning can turn Workstation NT into Server NT. In response to O'Reilly's analysis, Microsoft claimed that the two operating systems differ on more than 700 counts. According to one reporter: "*While the Big 'M' folks in Redmond maintain the products are vastly different, critics allege Workstation can be switched into the Server version with a few easy tweaks. An official Microsoft marketer suggests that's like arguing the only difference between men and women is a Y chromosome. We think it's more akin to discovering your date is indeed a drag queen. “*

3.8 Unprofitable Versioning: an Unprofitable Versioning scenario:

	Impatient customers	Patient customers
Immediate version	100	50
Delayed version	60	30
Number of customers	40	60

Versioning: Profits = 4,600 (30*60 + 70*40)

No versioning: Profits = 5,000 (50*100) → Uniform pricing leads to larger profits!

And what about consumer welfare? Is versioning good or bad for consumers? "It might seem like product sabotage should be banned. But it can make customers better off. *The key test is whether the practice means more goods are sold*. Suppose the French had regulated trains so that all carriages had roofs. All those in second class might have switched to third class, potentially rendering both uneconomical to provide. Altering quality, even if that means damaging goods, can make total supply rise. So, the fact that apps are crammed with annoying adverts is a sign of good economics. If that is little comfort, perhaps it is time for an upgrade."

Difficult to say a priori. The answer crucially depends on whether versioning can *enlarge the customer base or not*:

a) If the number of consumers increases, Sc may not be inferior with versioning, and social welfare could be higher.
b) If the number of consumers does not increase, then versioning enables the producer to gain more at the expense of consumers. In this case versioning leads Sc to be lower, and social welfare is unlikely to increase.

General rule of thumb: if price discrimination increases output, it can be beneficial. If output does not increase, welfare is reduced

Example:

	Impatient customers	Patient customers
Immediate version	20 (Willingness to pay)	6
Delayed version	5	5
Number of customers	100	1000

a) 1 version. Best choice: p = 6 (Profits= 6600) (Note that the output will equal to impatient + patient customers and costs are not considered)

b) 2 versions. Best choices: p =20 and p = 5 (Profits = 7000). Technically the immediate version should be sold at 20-2ξ (where ξ is an infinitesimal amount) and the delayed version at 5- ξ. So, impatient customers will achieve higher consumer surplus buying the immediate version thus opting for that one rather than the delayed version.

Remark: In both cases sold quantity is always 1100, but in b) firm is better off (Producer surplus: +400) and consumers worse off (Consumer surplus: -1400).

On the other hand, if

	Impatient customers	Patient customers
Immediate version	20	6
Delayed version	5	5
Number of customers	1000	100

a): 1 version. Best choice: p = 20 (Profits= 20.000). Only 1000 units are sold.

b): 2 versions. Best choices: p =20 e p = 5 (Profits = 20.500). 1100 are sold here.

Note that now in b) also patient customers enter the market, no reduction in consumer surplus, profits increase, more social welfare.

Similarly in the case of 3rd price discrimination

RECAP: Uptown (Inelastic) and Downtown (Elastic) markets.

Look: for P > 5, only the uptown market will be served. Moving from scenario B to scenario A we notice a decrease in social welfare and consumers' surplus, whereas monopolist's profits increase. Note that in these scenarios the overall quantity has not changed (=4.5 in both cases).

Scenario A: price discrimination

Slide from previous lecture:
$p_u = 8 - q_u$
$p_d = 5 - q_d$
$p_T = 8 - q_T$ for $p > 5$
$p_T = 6.5 - 0.5q_T$ for $p \leq 5$
No fixed costs, MC = 2

f.o.c in U: $8 - 2q_u = 2$; $q_u = 3$; $p_u = 5$; $\pi_u = 9$
f.o.c in D: $5 - 2q_d = 2$; $q_d = 1.5$; $p_d = 3.5$; $\pi_d = 2.25$

Total π_{ud} (in this case = S^p) = 11.25
Total S^c: S^{cu} (= 4.5) + S^{cd} (= 1.125) = 5.625
W = 16.875

Scenario B: no price discrimination

f.o.c in T: $6.5 - q_T = 2$; $q_T = 4.5$; $p_T = 4.25$; π_T (again = S^p) = 10.125

$S^c = 4.5 + [(4.5+3)*0.75]/2 = 7.3125$

W = 17.4375

If we consider a scenario in which the downtown market is even poorer ($P_d = 4-q_d$, i.e., at P = 4 rather than 5, no one will purchase the item), we notice that in case of no price discrimination the firm's profits will total to 8 if considering the downtown market but 9 if focusing on the uptown market (q = 3). On the other hand, thanks to price discrimination firm's profits will increase to 10, social welfare rises from 13.5 to 15 (Note that the consumers' surplus increases because now the downtown market is considered as well), and total quantity augmented from 3 to 4.

Remark: because of this example, price discrimination is not forbidden by regulatory authorities!

Both examples, i.e., versioning and 3rd price discrimination, validate the rule of thumb mentioned before, which says that discrimination might be beneficial if and only if it increases output.

Scenario A: price discrimination

But suppose now:
$p_u = 8 - q_u$
$p_d = 4 - q_d$
$p_T = 8 - q_T$ for $p > 4$
$p_T = 6 - 0.5q_T$ for $p \leq 4$
No fixed costs, MC = 2

f.o.c in U: $8 - 2q_u = 2$; $q_u = 3$; $p_u = 5$; $\pi_u = 9$
f.o.c in D: $4 - 2q_d = 2$; $q_d = 1$; $p_d = 3$; $\pi_d = 1$

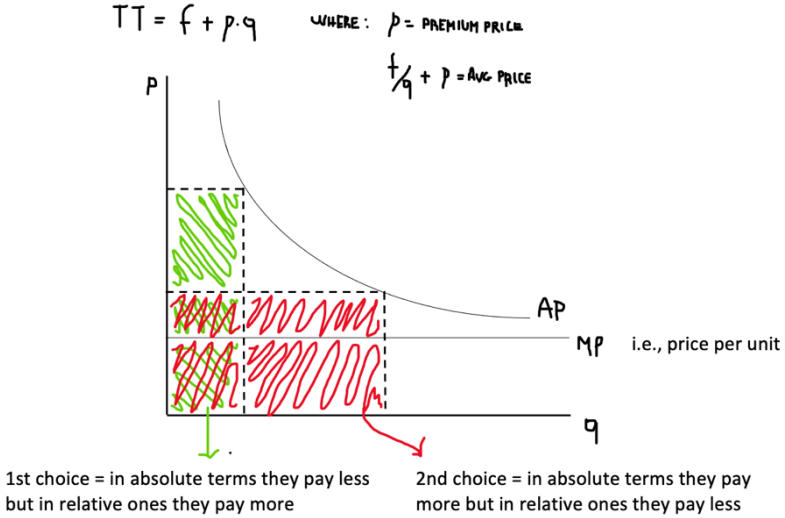
Total π_{ud} (in this case = S^p) = 10
Total S^c: S^{cu} (= 4.5) + S^{cd} (= 0.5) = 5
W = 15

Scenario B: no price discrimination
Operating in the downward part of the total demand ($p_T = 6 - 0.5q_T$) is not convenient for the firm:
f.o.c: $6 - q_T = 2$; $q_T = 4$; $p_T = 4$; π_T (again = S^p) = 8

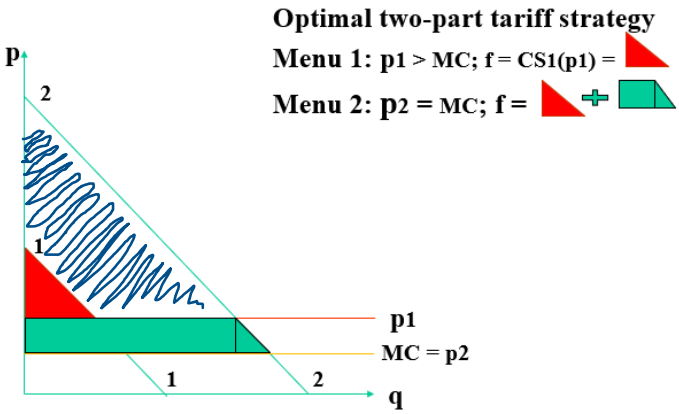
In this case it will just focus on the uptown market:
f.o.c in U: $8 - 2q_u = 2$; $q_u = 3$; $p_u = 5$; π_u (again = S^p) = 9 (> 8)
$S^c = 4.5$
W = 13.5

Versioning on *quantity* is achieved by firms through a *two-part tariff,* which:

- Entails a fixed entry fee (f) and a per-unit price (p)



- Total average unit price depends on quantity
- Customers self-select depending on their preferences
- Simplifying, it is like having 2 main dichotomous options: a) consume few units (and spend less in absolute terms) but paying a high average price; b) consume more units (and spend more in absolute terms) but paying a lower average price. Hence, there is not much difference with versioning price discrimination.



Be aware of the *participation constraint* (to choose the right tariffs to still involve the customers): To better meet the participation constraint the seller may also propose different combinations (f,p), i.e., a (relatively) low f and high p for "low" demanders (those who consume few units), and a (relatively) high f and low p for "high" demanders.

It can be shown that the optimal "menu" scheme is organized in such a way that the price P_1 is higher than the marginal costs and the fee $_{(low)}$ equal to the area of the red triangle. Thus, through the fee, the producer can extract the consumer's surplus. For the second menu, P_2 will be set equal to the marginal costs and the access fee will total to the sum of the green and red areas. Why? This is the only way for sellers to meet the *incentive -compatibility constraint* (consumer *i* prefers menu *i* to menu *j*). This is the maximum the firm can extract from high demanders avoiding them self-selecting themselves for the other version. By consuming Q2, high demander will total a surplus equal to the blue area. But if the access fee would be higher than the green + the red area, they would self-select themselves for menu 1.

How many versions should be launched into the market? There is no general rule. Let's see why.

More versions mean more possibilities to capture value from consumers by both enlarging the consumers pool and extracting value from high willingness to pay- consumers. But more versions mean more personalization costs (which in many cases might be convex, i.e., once you have differentiated one product, it will be tougher to personalize even more) and there exists the risk of "cluttering" effect for potential consumers (detachment from homo economicus theory on which classical economics is based on), i.e., the consumers cannot decide which is the best choice for her. Consequently, the greater the number of versions, the larger the latter effect.

To alleviate cluttering (especially when the number of versions increases) many firms think is important to describe functionalities in a very detailed way for each version with the aim of helping the customer to understand which is the best option for her.

Product features	Stata/IC	Stata/SE	Stata/MP 2-core	Stata/MP 4-core	Stata/MP 6+
Maximum number of variables					
Up to 2,048 variables	✔	✔	✔	✔	✔
Up to 32,767 variables	-	✔	✔	✔	✔
Up to 120,000 variables	-	-	✔	✔	✔
Maximum number of observations					
Up to 2.14 billion	✔	✔	✔	✔	✔
Up to 20 billion	-	-	✔	✔	✔
Runs most estimation commands...					
Fast	✔	✔	✔	✔	✔
Twice as fast as Fast	-	-	✔	✔	✔
Almost four times as fast as Fast	-	-	-	✔	✔
Even faster	-	-	-	-	✔

According to Shapiro and Varian (1999), *"If you can't decide how many versions to have, choose three."*

3 can be better than 2 because of the phenomenon of *"extremeness aversion":* The risk that 2 versions (heavy and light) are felt by potential consumers as "too big" or "too small" with the risk that a high percentage will opt for the light version thus generating less revenue for the company. To add another category (super-gold-premium) and make the previous-premium version as the medium one can produce some advantages.

Suppose McDonald's does not offer the larger version anymore. Hence, the medium will become the new large. Will the latter be the most sold size? Apparently not.

According to the experiment conducted by Simonson e Tversky (1992), two possible sets of choices for two identical groups of people (n = 60) were possible (Microwave ovens). Two homogenous groups have been selected.

<u>1° group:</u>

1) basic (Emerson): 110 $ → 57% of consumers

2) premium (Panasonic): 180$ → 43%

<u>2° group:</u>

1) basic (same): 110 $ → 27%

2) premium (same): 180$ → 60%

3) High-premium (Panasonic): 200$ → 13%

"In media stat Virtus"

(Note, moreover, that by putting a third, more costly option some consumers have been willing to spend even more than what could have had with only two options). If switching from 2 to 3 is not possible for whatever reason, note that a firm might always artificially increase the number of versions and sell the preferred version exploiting the cognitive biases of consumers (*Decoy Effect*).

Subscription from The *Economist* some years ago:

3 versions:

- 1. on-line subscription for one year and access to all issues from 1997 for US $59,00;
- 2. Print subscription for one year for US $125,00;
- 3. Print subscription for one year and on-line access to all issues from 1997 for US $ 125,00.

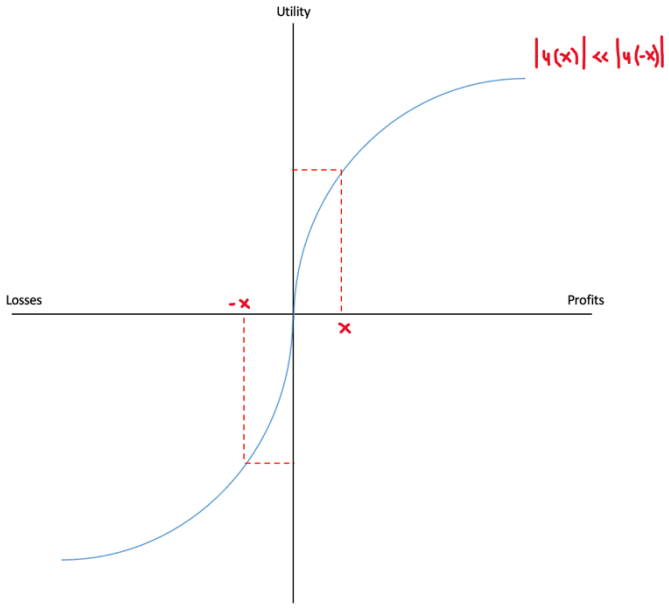
The Economist subscription experiment (on the left):

First Group has been presented with 3 options (As depicted on the right): Out of 100 students, 16 students choose option 1, 0 option 2, and 84 the last option.

Second Group has been presented with the first and the latter (no 2nd option). 68 students opted for the first and 32 for the latter, i.e., the third (which now is the 2nd). Note that options 1 and 3 were dichotomous with respect to their attributes. Whereas the first performed good in terms of price convenience but bad in terms of quality (offer) the other one was the exact opposite. Thus, the consumers could not decide between these two. On the opposite, the second option is without any doubt worse than version 3. This gives the consumer a sort of indication (or a nudge) that the third option should be the right one because better than another option offered.

3.8.1 Prospect Theory

Generally, individuals (and so consumers) are much more sensitive to losses than to gains. The decoy effect exploits this psychological trait: the average consumer does not know if s/he "wins" by choosing "3" but s/he is sure (or at least more confident) that s/he does not lose by choosing that option. Thus, more generally, our choices may depend on how outcomes are presented ("framed") to us. Invariance (and possibly other axioms) of "rational choice" may not hold.



Experiment: Problem 1 (N = 152): Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternatives to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

If program A is adopted, 200 people will be saved (72%)

If program B is adopted, there is one-third probability that 600 people will be saved and two-thirds probability that no people will be saved (28%)

Which of the two programs would we favour?

"The formulation of Problem 1 implicitly adopts as a reference point a state of affairs in which the disease is allowed to take its toll of 600 lives. The outcomes of the programs include the reference point and two possible gains, measured by the number of lives saved. In this case, preferences are risk averse. A clear majority of respondents prefer saving 200 lives for sure over a gamble of that offers a one-third chance of saving 600 lives." Now consider another problem in which the same cover story is followed by a different description of the prospects associated with the two programs:

Problem 2 (N=155): If program C is adopted, 400 people will die (22%)

If Program D is adopted, there is one-third probability that nobody will die and a two-thirds probability that 600 people will die. (78%)

"It is easy to very that options C and D in Problem 2 are indistinguishable in real terms from options A and B in Problem 1, respectively. The second version, however, assumes a reference state in which no one dies for

the disease. The best outcome is the maintenance of this state, and the alternatives are losses measured by the number of people that will die of the disease." *People in this case are risk seekers*.

3.9 Bundling

Another important type of price discrimination by self-selection is Bundling, which concerns offering two more distinct products as a package at a single price. There is:

- *Pure bundling*: sale of the bundle of goods but not of a good singularly. Hence, in a strict sense, it is not price discrimination since consumers do face the choice of either buying the whole package or not buying anything. Example: Netflix, where you pay a subscription fee, and you receive a bundle of movies.
- *Mixed bundling*: sale of the package or of single components (package price < sum of prices of single components). It's a type of a 2° price discrimination/versioning. Example: Microsoft office bundle. You can buy word, excel, and outlook, for instance, singularly but also all of them together.

Since the logic behind the two is intrinsically similar, and thus can be both considered examples of 2nd price discrimination, we will focus primarily on pure bundling to make some general considerations.

Profitable budling can reduce the "willingness to pay" dispersion:

	Word Processor	Spreadsheet
Marco	$120	$100
Giorgio	$100	$120

In the scenario presented above, the best strategy if no price discrimination would be possible to implement would be to price both products for 100 selling both to both consumers. Instead, if bundling is possible, the bundle could be priced at 220 (100 +120), achieving profits (disregarding production costs) of 440. Through bundling, the producer can eliminate or at least reduce the dispersion of willingness sot pay (in this case 20 for both products). To do so, three conditions shall be in place:

1) Heterogeneous willingness to pay
2) Negatively correlated willingness to pay of consumers for single components (someone must prefer more one component, someone else more another one)
3) (Overall) Willingness to pay of consumers for the products should be similar.

When one of these three conditions is not met, profits become senseless.

The sum of the dispersions of the willingness to pay single components must be greater than the dispersion of the willingness to pay the bundle. If the latter condition is verified, profitable bundling is performable.

	Word Processor	Spreadsheet
Marco	$120	$100
Giorgio	$120	$100

Useless Bundling: Case of homogenous preferences (no heterogeneity between the consumers' preferences)

In the first case, bundling would not achieve a higher outcome in terms of profits. By selling both separately and together the firm would always make 440 as profits.

	Word Processor	Spreadsheet
Marco	$120	$100
Giorgio	$80	$50

Useless Bundling: If not negatively correlated

bundling: Π = 260

No bundling: Π = 260

In the second case, both consumers prefer without any doubt Word processors to the spreadsheet, i.e., no negative correlation. Again, by bundling the products the firm could achieve a profit equal to no bundling.

	Word Processor	Spreadsheet
Marco	$120	$100
Giorgio	$40	$30

Overall willingness to pay, i.e. both consumers should be willing to purchase both products at similar prices: Useless or Harmful Bundling

bundling: Π = 140

No bundling: Π = 220

In the third case, the willingness to pay of the consumers differs so

largely that it does not make any sense to bundle. Actually, the best strategy here would be the one of disregarding completely Giorgio by selling both products to Marco at its specific willingness to pay.

	Word Processor	Spreadsheet
Marco	$120	$100
Giorgio	$20	$60

Useless or Harmful Bundling

bundling: Π = 80*2= 160

No bundling: Π = 240

Despite the first two conditions being met, if the third is not met, bundling will not be profitable. Here the Spreadsheet would be priced at 60 and sold to both, whereas the Word Processor should be sold solely to Marco.

Note that, whenever a small group of potential consumers shows relatively a very low willingness to pay (the bundle and single components), trying to target these consumers and sell them the bundle may not be convenient (caveat to the respect of the participation constraint).
Bear in mind 2 compromises:

1) Bundling can be used strategically to foreclose a market. A dominant firm can bundle its dominant product with a nascent markets' product, limiting thus competition in the nascent market. Hence, it can exploit its advantageous position in one market to impose itself in a new market where competition might be fierce.
2) Bundling might be more advantageous compared to versioning. In the latter, you need to estimate the willingness to pay for every single item with respect to every single customer…quite tricky. In bundling the variables that must be estimated are fewer (especially for pure bundling), because you need to estimate the willingness to pay solely for one bundle, thus allowing the company to avoid cumbersome calculations (market analyses). Even if you misprice one good in the bundle, on average the error might even cancel out.

Propositions:

1. It is better to avoid bundling *Superstars* (Killer applications). Killer applications (Software or game that is so appreciated by people that can drive consumption of the hardware as well that comes with the software). In case a game for instance is so fantastic, and despite a consumer is not that into games she might be willing to try it. To do so, she will be "forced" to purchase a console as well (Hardware).

Willingness to pay	Game 1	Game 2	Game 3	Game 4	Game 5
Consumer 1	60	10	10	10	70
Consumer 2	10	10	10	10	70
Consumer 3	10	10	60	10	70
Consumer 4	10	10	10	60	70
Consumer 5	10	10	10	10	70

Game 5, as shown in the table is the "Killer application" because wanted by everyone. Everybody agrees that the fifth game is by any chance the best one. Even Consumer 5, who is not interested in gaming (apparently) would be willing to try game 5.

→ Best Profits with complete bundling (640$: $p_{bundling}$ = 160$ for 5 games (10+10+10+70+60), bundling sold at first 4 consumers since the 5th's willingness to pay is smaller, namely 110).

→ Best profits with partial bundling (710$: $p_{bundling}$ = 90$ for first 4 games, p_{game5} = 70$. In this way, the 4 games can be sold to the first 4 consumers, whereas the 5th game can be sold uniquely to every consumer, including consumer number 5).

2. Mixed bundling might be more complex to implement but it may lead to greater profits than the pure one (note that the maximizing $p_{bundling}$ may be different in pure bundling vs. mixed bundling) from the firm's perspective.

Consumers/Software	Word	Excel
A	40	20
B	20	40
C	45	5
D	5	45

- No bundling: Profits = 160; $p_{word} = p_{excel} = 40\$ * 4$ Consumers (At 40 Word will be purchased by A and C while B and D will buy Excel at 40)
- Pure bundling: Profits = 200; $p_{bundling} = 50\$$ (obtained by the minimum willingness to pay among consumers, i.e., 45+5). This is not the best strategy that can be implemented, tough!
- Mixed bundling: Profits = 210; $p_{bundling} = 60\$$; $p_{word} = p_{excel} = 45\$$. Consumers A and B will buy the bundle (60*2), while C will purchase Word (45) and D Excel (45)

3. Auction: Pricing strategy that does not belong in principle to price discrimination but is a relative, i.e., there are some similarities between auctions and 2nd price discrimination, wince it bases on the consumers revealing their willingness to pay.
There are different types of auctions:

- Ascending (aka English): Most common one
- Descending (Dutch): You start from a very level and the first that stops the price confirms the order.
- First price
- Second price

Suppose a seller must sell an item. There are only 2 potential buyers. Buyers' evaluations for the item are either both 100$ or both 150$ **(**i.e. they have the same willingness to pay)**.** Buyers of course know their evaluations, while the seller doesn't. The seller only knows that each value is equally likely (50%).
Now let's consider 2 Scenarios:
1)An auction cannot be performed: Under a "Fixed Price" scenario, if the seller sets a price of 100$ he is certain to sell the item (gaining 100$). If he sets a price of 150$ he has 50% probability to get 150$, and 50% to end up with 0$. This results in an expected value of 75$, which is lower than 100$. Thus, in this scenario, he will opt for setting p=100$.

2)In case an English auction can be performed, he starts the auction by calling out 100$. If buyers value, the item 100$ they will make a sign of acceptance. As the seller asks for higher bids, buyers will be silent, and the auction ends with a winning bid of 100$ (let's suppose that the quicker buyer wins). This occurs with a probability of 0%. However, if buyers' valuation is 150$, they will continue outbidding each other, until the price reaches 150$. This event will occur with probability of 50%. So, the expected value, in this scenario is 125$ > p=100$ (100*0.5 + 150*0.5) and this is possible solely through an auction.

Theoretical Questions' Sample:

Excursus on the premium model: You got something for free and then you have a premium version (2nd Price Discrimination model, Versioning). The free version of course is done in such a way that users are still willing to use it but not that well to convince professional users who will be willing to purchase the premium version to have more, better working, features.

1. Question from the Economist: First-time subscribers to the *Economist* pay a lower rate than repeat subscribers. Is this price discrimination? If yes, of what type?

3rd Degree price discrimination, where the exogenous indicator is the purchase history of the company, and we can expect that repeat customers will be more than first comer because they show a more inelastic demand than the latter one

2. Please examine the evolution of the movie industry from the recent past to nowadays in terms of pricing strategies put in place by the firms.

Timeline: before.

1. Once you could watch the movie in the cinema if you weren't willing to wait any moment paying relatively high prices.
2. Otherwise, you could wait for the VHS or DVD or Blue Ray
3. If you were still willing to wait you could watch the movie on the TV (bearing the cost of advertisement)

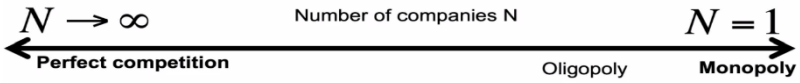
A second price discrimination based on Versioning where Time was the variable of interest.

Timeline: nowadays.

1. Most movies are watched on platforms (e.g., Netflix or Disney Plus) Bundling**.** Despite Netflix, for instance, releasing movies in theatre as well, they release in parallel to the platform one. This is done solely to be eligible for winning the Oscars.

The same story goes for the Music Industry. Netflix à Spotify (Bundling of thousands of songs). Once, an album could have been purchased which was a pure bundling technique (fewer songs).

4. OLIGOPOLY - A BROAD INTRODUCTION



The terms come from the Greek *oligos = fell* and *polein = sell*. Here the number of companies in the market (N) is small, where the market power with which firms are endowed with is comparable, le, i.e., similar. Note: There exist different types of Oligopolies, namely symmetric (perfectly equal) or asymmetric. In any case, though, every competitor within an oligopoly has substantial market power. Here, on the contrary of monopolistic or perfect competition, the behaviour of each firm significantly affects the behaviour of other firms, which leads to *strategic interdependence*. Indeed, under a monopoly, for instance, the behaviour of the firm is independent of any other player apart from the consumer: Here, the company will try to set a strategy that maximizes its own profits. Under an oligopoly, each player' strategy depends on the competitors.

4.1 Game theory

To analyse such market structure, we will borrow the neoclassical profit maximization logic (profit maximization rule) and a tool that has been explicitly developed for oligopoly, namely game theory: it is the formal modeling of optimal decision-making in contexts of strategic interaction. In its simplest version (simultaneous game, two players, pure strategies):

- There are 2 players: the most simplified scenario.
- Each player has a set of possible actions, which may be discrete or continuous: In the case of a rock-scissors-paper game, we have three possible, discrete actions. On the other hand, there might be a scenario in which the decisions cannot be discretely chosen.
- Different combinations of actions **unambiguously** determine different outcomes: e.g., paper beats rock, scissor beats paper and so on.
- Each outcome is unambiguously associated with a pay-off for each player.
- Players are **perfectly rational** and perfectly informed.
- Players decide their actions simultaneously, aiming for pay-off maximization.

Let's define two key concepts:

1. *Reaction function*: it is a function associating each possible strategy of one player with optimal response of the other player. The reaction function of player A is going to give us the best strategy given all the possible actions of player B.

2. *The Nash equilibrium*: configuration of strategies where each player's strategy is the best response to the strategy of the other player. A configuration of strategies is a Nash equilibrium if no player could improve the resulting pay-off by unilaterally deviating from the selected strategy.

A discrete game example: A prosecutor is trying to extrapolate information from a couple. To do so, it offers the following plan.

Note that each prisoner is posed in front of the dilemma independently and simultaneously to the other, hence it cannot consult anyone but himself.

THE PRISONER'S DILEMMA

	B stays silent (cooperates)	B betrays A (defects)
A stays silent (cooperates)	Both serve 1 year	A serves 3 years, B goes free
A betrays B (defects)	A goes free, B serves 3 years	Both serve 2 years

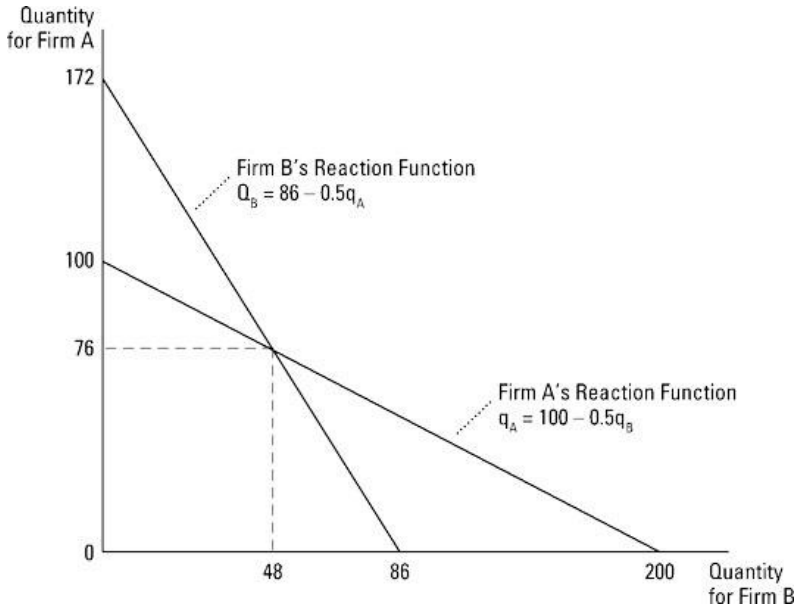
There are multiple scenarios:

- A stays silent, B betrays
- B stays silent, A betrays
- A betrays, B betrays
- B betrays, A betrays

A betrays B betrays is the only Nash equilibrium

In this case, the reaction function of B always leads to B betraying A. If A stays silent, B has the opportunity to be freed, whereas if A talks, B better talks as well, otherwise will serve for 3 years instead of 2. Intuitively, the reaction function of A is perfectly symmetric to B's. Note, that a Nash equilibrium is NOT Pareto efficient. Indeed, they could achieve a better outcome if both would have stayed still.

A continuous game *example:*



As mentioned before, a reaction function is necessary (fundamental) in case of a continuous set of implementable actions, which tells the optimal quantity given the production of competitor. Indeed, to solve Cournot oligopoly problem, the latter function will be strictly necessary, otherwise, no solution can be found. A simple example might be to assume the case in which A decides to produce 10: here B will opt for producing 81 goods. The problem here is that since now B produces 81, A will change its production levels to 59.5 (100-81*0.5). Both firms will change their quantities supplied until they reach a specific point: the Nash equilibrium, which will be obtained by equating the two reaction functions to each other (q_A=76 and q_B=48).

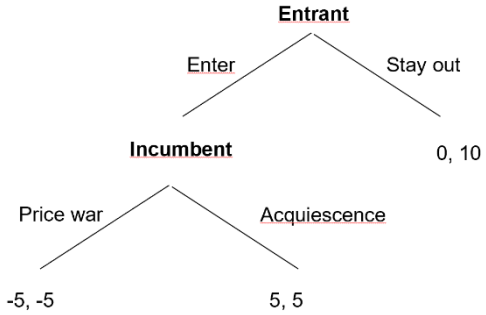
The intersection will be the optimal result for both firms, which, as mentioned in the discrete case, is not Pareto efficient. Indeed, firms could be better off by cooperating rather than competing.

A variant to the simultaneous case: **A sequential game**

- There are 2 players
- Each player has a set of possible actions, which may be discrete or continuous.
- Different combinations of actions unambiguously determine different outcomes.
- Each outcome is unambiguously associated with a pay-off for each player.
- Players are perfectly rational and perfectly informed.
- Players decide their actions sequentially, aiming for pay-off maximization. The player acting first is known as the leader, while the other is known as the follower.

To solve such case,

- *Discrete sequential* games are typically represented as **decision trees.**
- Sequential games can be easily solved by ***backward induction****:* you start by determining the pay-off maximizing strategy of the follower. Then, you base the strategy of the leader on this information. So, you start from the end, not from the beginning.
- The sequence of optimal actions obtained through backward induction is known as ***subgame perfect Nash equilibrium.***



- Here we have *sequential* strategic interdependence, because first we have the entrant that must decide whether to enter or stay out and in case it decides to do so, we will have the market leader (monopoly) trying to decide whether to deter such entrance through a price war or to accept it and divide the market (here assumed to be equal). So, at the very beginning, the potential entrant will demand itself which would be the actions that the market leader will commit. Here, both will lose -5. Since we have assumed that both market players are perfectly rational and informed, no rational agent would prefer -5 to 5, hence the incumbent should in theory accept the new entrant. On the other hand, bearing this information in mind, the new entrant will face the decision of either entering the market and obtaining 0 profits, thus leaving the monopolist with the whole loot, or to enter and gain 5. Thus, the outcome will be an equal division of the market between the two agents.

What about continuous sequential games? Are we more interested in leaders' or follower's reaction function? If I am the leader, I MUST know the follower's reaction function to know which quantity I have to produce to reach the desired goal. For each possible action that the monopoly implements, the latter will know the reaction of the entrant. This information will allow the market leader to formalize all the possible scenarios (in terms of combination of quantities produced by both firms and choose the one that allows the latter to maximize its profits.

5.OLIGOPOLISTIC MARKETS: CLASSICAL DUOPOLISTIC MODELS

Beware those models can be:

1. *Collusive* (e.g., cartel, where company organizes themselves to maximise their combined profits, which is better than starting to compete. Note that within collusive agreements, the parties are incentivized to break the agreement to steal further market shares. Of course, in case this would happen, the other player would retaliate)
2. *Competitive*, i.e., there is no mean to achieve a collusive agreement (Bertrand, Cournot, Stackelberg). There exist two categories:

- Simultaneous models: decisions are concurrent
- Sequential models: company decides in sequence, i.e., there would be a leader and a follower

3. Key (focal) variables, i.e., the object of the decision, can be either

1. Prices adopted by each company or
2. Quantities offered by each company (Cournot, Stackelberg)

5.1 Oligopolistic models: Bertrand model (decision on price choice)

Such model is based on very restrictive assumptions which anyhow allow us to reach powerful analytical results.

1. There are solely 2 players
2. No potential entrants (closed markets)
3. Homogenous good Implication: if a player reduces the price by a small delta, he will bear the gains of the entire market demand
4. Perfect rationality
5. Perfect information (both players know all the assumptions of the games involved, they know their costs structures, the opponents' cost structures, etc.)
6. Symmetrical technology, hence, they base their production on the same cost function à MC = AC = c (since no fixed costs are present, there are no economies of scale achievable à every additional unit has the same cost)
7. Only 1 strategic variable, namely price
8. Price is decided simultaneously

What are the options for firm *I?* If I set a price:

- Lower than j, it captures the entire market demand
- Equal to j, it shares the market demand with j (equally, look below at the formula)
- Greater than j, it has a null market demand (consumers demand the good from j)

Analytically:

$$D_i(p_i,p_j)=\begin{cases} D(p_i) & if \quad p_i < p_j \\ \frac{1}{2}D(p_i) & if \quad p_i = p_j \\ 0 & if \quad p_i > p_j \end{cases}$$

The assumed cost function is TC = c*q

In this way:

Fixed costs are zero; Average Cost (AC) and Marginal Cost (MC) coincide:

$$MC=\frac{\partial TC}{\partial q}=c \qquad AC=\frac{TC}{q}=c$$

Thus, the profit will be: $\pi=(p-AC)\cdot q \Rightarrow \boxed{\pi=(p-c)\cdot q}$

i and *j* choose their price to maximize profits $\max_{p_i} \pi_i(p_i,p_j)$ $\max_{p_j} \pi_j(p_i,p_j)$

Under our cost assumptions, profit functions are:

$$\pi_i(p_i,p_j)=(p_i-c)\cdot D_i(p_i,p_j)$$
$$\pi_j(p_i,p_j)=(p_j-c)\cdot D_j(p_i,p_j)$$

The game is *simultaneous* and *competitive* (each company tries to maximize its own profit)

Nash Equilibrium:

- Couple of strategies where none of the players find it convenient to change strategy given the other's strategy. It is the best response to the response of the competitor
- No one can unilaterally change its position and improve its situation

Each company's price maximizes profits given the other's choice

Nash Equilibrium: (p_i^*, p_j^*) if two conditions are satisfied: $\begin{cases} \pi_i(p_i^*, p_j^*) > \pi_i(p_i, p_j^*) \quad \forall p_i \neq p_i^* \\ \pi_j(p_i^*, p_j^*) > \pi_j(p_i^*, p_j) \quad \forall p_j \neq p_j^* \end{cases}$

In few words, the best response prices are those prices for which there is no better solution (i.e., higher profits) given the best solution decided by the competitor. For instance, firm's one profits are larger than any other profit achievable by changing p_i given p_j^* fixed. The same holds for firm 2 as well.

It is possible to demonstrate that in the Nash equilibrium each firm chooses a price equal to c:

$p_i^* = p_j^* = c$ Which, recall, it's equal to the ACs and since no FCs, to the MCs as well

Why? Because in this case, none of the two companies has an incentive to change its choice, given the other's choice.

- Price higher than c: loss of the entire demand
- Price lower than c: the firms make losses instead of profits

Given the market conditions, firms are identical, and the game is symmetric; the reasoning developed for one player is perfectly applicable to the other.

Are there any other equilibria? Let's check why $p_i = p_j = c$ needs to hold

There are three possible alternatives:

- Case I $p_i > p_j > c$
- Case II $p_i = p_j > c$
- Case III $p_i > p_j = c$

(p_i, p_j can't be < c because c=MC=AC)

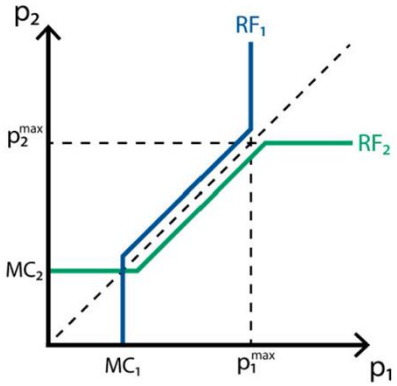
If at least one of the two firms sets a price higher than c, there is no equilibrium. That's because the firm setting p > c is either getting none or half of the demand. In both cases it is incentivized to lower the price to the point it is infinitesimally lower than the price charged by the other firm.

Each firm has always the incentive to revise its price decision, unless the price for both firms is equal to c = MC = AC

Thus, the Nash equilibrium is represented by the following couple of strategies:

$$(p_i^*, p_j^*) = (c, c)$$
$$\begin{cases} \pi_i(p_i^*, p_j^*) > \pi_i(p_i, p_j^*) \quad \forall p_i \neq p_i^* \\ \pi_j(p_i^*, p_j^*) > \pi_j(p_i^*, p_j) \quad \forall p_j \neq p_j^* \end{cases}$$

Graphically, even if the reaction functions are not mandatory to reach an equilibrium, still this game can be represented in terms of these. Recalling that a reaction function is a function that gives the best response



(locus of pints) to the strategy of the competitor, i.e., the price set, we can see that the reaction function of firm 1 tells which is the optimal response in terms of P_1 given P_2. We notice that it has a ceiling and a floor, i.e., a maximum and a minimum: those points beyond which price won't change anymore. For instance, given P_2^{MAX}, firm 1 will set its own price at the ceiling as well (which is exactly equal to P_2^{MAX}). So, why do we reach an equilibrium at MC_1=MC_2? Because, as we have seen previously if firm 1 sets a price higher than firm 2, the latter will absorb the entire demand, whereas if the price is set below the marginal costs, then the profits would be negative! The subtilty here lies in why does a P^{MAX} even exists after which both firms won't raise their prices anymore. This is because if we think at monopolies, thus disregarding completely the competitor, there is an optimal price after which profits will start to sink again, because demand will decrease more than proportionally with respect to the revenues increments due to an increase in price. What about all the points in the middle? The bisector of the quadrant is that locus of points where P1 = P2. So, we can see that the reaction function of firm 1 is slightly left to the bisector and that is a simple representation of the reality: both firms (the reaction function of firm 2 is slightly on the right) will set a price slightly smaller than the other firm (P1 < P2 and vice versa).

Some critiques: Note that there are several assumptions that are hardly coping with reality.

5.2 Oligopolistic models: Cournot Model (Quantity choice: continuous model)

First, we must bear in mind that this model shares many assumptions with the Bertrand model:
Cournot duopoly assumptions:

- Only 2 firms, duopoly
- No potential entrants (closed markets)
- Homogenous good
- Perfect rationality
- Perfect information
- Only 1 strategic variable: quantity (q), hence both firms will decide the quantity (which will set price consequently depending on the demand function) to maximize profit which; instead of directly price
- Production levels are **simultaneously decided**
- The price is determined by the market at a level where the demand equals the joint production of the two firms

Since there is a homogeneity of goods, there will be a unique market price, which will be determined by the quantity, which is an aggregated value, i.e., the sum of both firms' production levels.

Thus, the strategic variable here is quantity:
Firms choose how much they want to produce, and the price is given by the aggregated market demand (under the hypothesis of standard goods, the DD has a negative slope):

$$p_1 = p_2 = P(Q) = P(q_1 + q_2)$$

The two firms strategically interact by influencing the (unique) market price through the quantity they set.

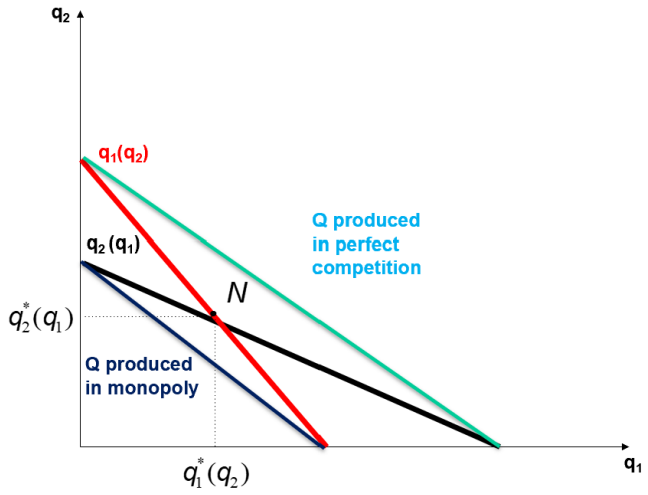
Equilibrium: given the competitor's choice, firms choose the best strategy to maximize their profits (the main objective of rational agents)

Assume that:

- Firms can choose the quantity they prefer in the interval: $[0,+\infty)$
- Both the profit functions can be differentiated in quantity

Goal: derive the equilibrium (2 steps):

1. Determine the set of optimal choices (i.e., those that maximize the profits) of each firm given the rival's behavior determine *reaction functions.*
2. Intersect the two reaction functions to find the *combination* of *mutually compatible* decisions (i.e., the Nash-Cournot equilibrium of the game)



This graph underlines two aspects:

1. The graphical derivation of the equilibrium (as mentioned previously given the solution to a continuous game).
2. *Red line*: reaction function of firm 1, i.e., the optimal response in terms of Q1 for any possible Q2. *Black line*: optimal Q2 as function of Q1. *Green line*: quantity produced in perfect competition (level is constant). *Blue line*: quantity produced in monopoly.

1st Notice that the red and the green line intersect at Q2 = 0. Why? If firm 1 produces the quantity of perfect competition where price is equal to marginal costs (and in case no fixed costs are existing, to average costs), the response of firm will be equal to 0 → It is better to stay out of the market, otherwise it would incur in losses (no one would buy it because the market is saturated. The only way would be to sell it at less than the marginal costs, which would make no sense). The same idea applies to the case in which Q1 = 0.

2nd: Firm will produce the monopolistic quantity if and only if the other firm will produce 0 output, since being the only one in the market producing anything.

3rd: A Cournot duopoly solution the two firms will produce a higher quantity than what happens in the monopoly, to the benefit of the consumers. This because both firms must consider the opponent reactions (they are strategically interdependent), thus limiting each other further possible benefits (thus market power).

Given the following inverse demand function (we consider the simplest case): $P(Q) = a - bQ$

Where Q is the total industry output, equal to: $Q = q_1 + q_2$

And the following (assumed) simple cost functions: $TC_1(q_1) = c_1 \cdot q_1$

$TC_2(q_2) = c_2 \cdot q_2$

Firm 1's profit is:

$$\max_{q_1} \pi_1 = [P(q_1+q_2)\cdot q_1 - TC(q_1)] =$$

$$= \{[a-b\cdot(q_1+q_2)]\cdot q_1 - c_1\cdot q_1\}$$

The first order condition for a maximum is:

$$\frac{\partial \pi_1}{\partial q_1} = 0 \Rightarrow a - 2bq_1 - bq_2 - c_1 = 0$$

Or simply:

$$q_1^* = \frac{a-c_1}{2b} - \frac{1}{2}q_2 \quad \longleftarrow \text{Firm 1's reaction function}$$

Although in reality fixed costs might be present, the reasoning would not change.

Note that here firm 1 maximises profits according to its own production since it is the one that can control. Thus, firm 1 wants to know which is the best quantity of goods it can produce given any quantity of Q2 (which will be a variable of the profit function).

The result will be the reaction function of firm 1, i.e., the optimal quantity produced by firm 1 given the quantity of firm 2.

By symmetry, firm 2's reaction function can be found immediately: $q_2^*(q_1) = \frac{a-c_2}{2b} - \frac{1}{2}q_1$

- The equilibrium is given by the couple of values → (q_1^N, q_2^N)
- To identify the equilibrium, firm 1 decides its output based on the conjectures regarding firm 2's behavior; for example, if firm 1 expects that Firm 2 will produce the quantity q_2^e, then firm 1 will have to produce $q(q_2^e)$
- (q_1^N, q_2^N) → Equilibrium values: to be a Nash equilibrium, we need both firms to produce the optimal quantity given the best response of the opponent.
- (q_1^e, q_2^e) → Expected values
-

Thus, the *Nash equilibrium* is the solution to the following system:

$$\begin{cases} q_1^N = q_1(q_2^e) \\ q_2^N = q_2(q_1^e) \end{cases} \Rightarrow \begin{cases} q_1^N = \frac{a-c_1}{2b} - \frac{1}{2}q_2^e \\ q_2^N = \frac{a-c_2}{2b} - \frac{1}{2}q_1^e \end{cases} \wedge \begin{cases} q_1^e = q_1^N \\ q_2^e = q_2^N \end{cases}$$

$$q_1^N = \frac{a-2c_1+c_2}{3b}$$

$$q_2^N = \frac{a-2c_2+c_1}{3b}$$

We know that q_1^N must lie on the reaction function of q_1 and the same holds for q_2^N. But then the firms ask themselves: Which is the value that I should expect from the other? → The Nash Equilibrium Value! Hence the expected value of quantities produced will exactly be given by the Nash equilibrium quantities.

5.3 Oligopolistic models: Stackelberg model (Decision on quantities: sequential model)

Sequential competition: the competition is not simultaneous anymore but articulates in two steps.

The second step is contingent on the first one.

2-step model solution: backward induction:

- The second step is contingent on the decisions taken in the first step: we know that the follower will behave rationally, thus maximizing profits given the leader's choice.
- The leader optimizes the first step by maximizing profits given the follower's reaction in the second step (which is known a priori).

As usual, it is assumed that:

- Both the leader and the follower know everything (perfect information)
- Both the leader and the follower are perfectly rational

Who has the advantage here? It depends whether the game that is played is imperfect in terms of information or not. If it is, the follower will have gathered the information from the leader's move (think to

the game of *Poker*, where position is power!), thus having a considerable advantage. On the other hand, if the competition is assumed to be perfectly informed the follower will become a simple "Appendix" of the leader, since the latter will already know the reaction function of the follower, thus being able to "constraint" (control/nullify) the actions of the follower.

- The leader considers the (future) follower's choice when it chooses its own output level
- Its profit maximization depends on $q_2 = f_2(q_1)$

$$\Rightarrow \quad \max_{q_1} \pi_1 = p[q_1 + f_2(q_1)] \cdot q_1 - c(q_1)$$

Follower's RF

$$\pi_1 = \left(a - b\left(q_1 + \frac{a - c_2}{2b} - \frac{1}{2} q_1\right)\right) q_1 - c_1 q_1$$

$$\pi_1 = \left(a - b\, q_1 - \frac{1}{2}(a - c_2) + \frac{1}{2} b\, q_1\right) q_1 - c_1 q_1$$

$$\frac{\partial \pi_1}{\partial q_1} = 0 \rightarrow \quad a - 2b\, q_1 - \frac{1}{2}(a - c_2) + b\, q_1 - c_1 = 0$$

$$q_1 = \frac{a - 2c_1 + c_2}{2b} \quad \Rightarrow \quad q_2 = \frac{a - 3c_2 + 2c_1}{4b}$$

Here, the profit maximization function for firm 1 will take as quantities the sum between its production levels and firm two levels, which will depend directly on q_1 ($q_2 = f_2(q_1)$), which has been found in the continuous game (Cournot model). Note that by maximizing q_1, the leader controls the quantity produced by firm 2 as well, thus having larger power with respect to the follower: it will be able to achieve higher profits than the follower. At the very end we will end up with one equation and one unknown, thus being easily able to find the optimal quantity.

In the real world, firms' strategic behaviors depend on several variables:

- *Price* (Bertrand model)
- *Quantity* (Cournot and Stackelberg models)
- R&D investments; Product features; Commercialization modes

All these decisions imply strategic interaction and interdependence.

6. ENTRY BARRIERS, ENTRY DETERRENCE AND LIMIT PRICING

6.1 Entry defined

First, how do we define *entry*? Let's suppose we are Ferrari – thus, we're competing in the automotive industry – and we face a new entrant such as Hyundai. Are we threatened? No, since we're competing in a completely different market. In our case, a threat might be caused by companies such as Lamborghini or Aston Martin. More specifically, a threat might be represented by the entry of a firm producing a good that is a *perfect substitute* to the ones already produced in the industry we're operating (recall: if the good is a perfect substitute, the price charged by both companies must be the same).

Remarks:

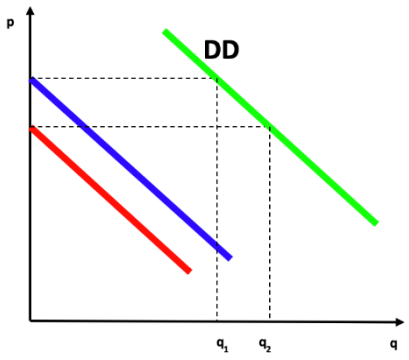
1. The degree of *substitutability* depends on *consumers' preferences*.
2. New entry does not necessarily imply the creation of a new firm (e.g., entry from another industry: entry in industry A of a firm belonging to industry B (diversification).

We are dealing with a scenario in which 2 players must take a decision. On what does the entry decision depend? In neoclassical economics, it will depend on the *expected profits*, which are a function of:

1. Production costs – exogenous variable
2. Revenues which crucially depend on demand conditions post entry. Indeed, the entry has an impact on the market quantity and price. Hence, new entrants must forecast the reaction of incumbents to understand whether it is profitable or not to enter the game.

However, we've to bear in mind that the leader will always try to deter the entrance in the market of the follower since such entrance might lead to a loss of market shares and, hence, profits.

Let's try to understand the actions that the incumbent will perform to prevent, thus, the just mentioned situation.



Green: classical market demand.
Blue: demand for entrant if incumbents are producing q1.
Red: demand for the entrant if incumbents are producing q2.

The blue and red curves concern the residual demand that remains to the new entrant given the production levels of the incumbent.

Why such curves? If the incumbent produces a quantity equal to q1, that portion of the market demand will be already satisfied. Hence, it won't be possible for the new entrant to reap out some benefits from the already targeted consumers. Q=qi+qne , qi=100 >> Q=100+qne >> P=400-(100+qne)=300-qne >> PD (res. D) =300-qne

Example: If MD is $P = 400 - Q$ and q(incumbent) = *100*, then PD is $P = 300 - Q$. The new entrant will face a Potential Demand given by: PD = Market Demand - q(incumbent)

With q(incumbent) equaling the quantity that incumbents choose to produce after the new entry.

The concept of entry barriers: intuitively, *entry barriers* are obstacles preventing new firms from entering a market and competing against the incumbents.
In a market with no entry (and exit) barriers:

- Every firm can enter the market, compete with the incumbents, and make profits. In the long run, P=AC. Hence, any profit opportunity will vanish.

A) If p > AC, firms make extra-profits. Without entry barriers, new firms will be willing to enter the market.

$$\pi = p(q)\cdot q - AC(q)\cdot q = [p(q) - AC(q)]\cdot q > 0$$

- As a result of the increase in supply, price goes down. So, the remaining profits will shrink.
- New firms will continue to enter until P = AC, i.e., as long as there is a unitary positive margin.

B) If $p < AC$, firms make losses. Without exit barriers, the most inefficient firms leave the market.

- As a result of the decrease in supply, prices go up.
- Firms will continue to exit until p=AC.

Entry barriers allow the incumbents to keep the price higher than the average cost.

However, the *threat* of new entry affects the price set by incumbents.

- After new entries, incumbents risk incurring losses.
- Incumbents may tend to set lower prices to prevent new entries.

If an incumbent allows a new entrant to step a foot in the market, two risks would be ahead:

- Profits would shrink.
- The new entrant would have the opportunity to enhance its economies of scales, economies of learning and so on, becoming thus more efficient. This might threaten the position of the leader in the long run.

6.2 Entry barriers taxonomy

1. *Institutional/legal barriers*: why? The government tries to maximize social welfare. Thus, it would raise barriers if such goal can be reached.
    - Administrative authorizations needed to conduct business
    - Patents: an institutionally granted right to exclude others from using the fruits of your inventing activity. They can be exploited to gain a competitive advantage or to gain royalties by selling it. Why do we have? The aim of patents is to incentivize innovation. Indeed, if we

would not have patents, what would be the reason behind spending resources in R&D when competitors could imitate it without incurring any costs?

2. *Structural barriers*: barriers that are present due to the market intrinsic characteristics.
    - *Economies of scale* (real or pecuniary), *scope* (commitment in terms of the range of operations) and *learning* (commitment in terms of time: if in an industry you have to become acquainted with a peculiar technology, you will incur in some losses at the very beginning because you won't be able to exploit that technology efficiently). There are some industries that require you to make a strong investment upfront (very high fixed costs) in the hope of making considerable profits in the long run (example: pharmaceutical industry). Hence, capital requirements are a barrier because not everyone can afford to spend so much to enter a market.
      As for the economies of scale, a distinction must be made between:
        - *Real:* related to factors of production and, in particular, to the indivisibility of factors. I.e., fixed costs spread across the high number of units produced.
        - *Peculiarity*: related to bargaining power and perception. As you grow bigger & bigger, you can gain a set of advantages with suppliers and banks. You can be perceived as more trustworthy and reliable and, consequently, face lower costs.
    - *Customer loyalty* (e.g., structural switching costs, brand loyalty). Brand loyalty is the idea that customers might get attached to a given brand and keep purchasing from that brand independently of what might happen to that specific brand (or possible opponents' entries). switching costs can be related to a cost that a consumer incurs from changing from one supplier to another one or can be structural, i.e., inherent in the market, and, to some extent, strategic, i.e., artificially increased. Examples:
        1) Structural SC: banking industry and its intrinsic bureaucracy needed in case a customer wants to shift from one bank to another.
        2) Strategic SC: see Apple example below
    - *Access to key resources* (e.g., distribution channels): to compete effectively and efficiently in a business, you might need complementary resources as well.

3. *Strategic barriers*: built up by the incumbent to protect the market from the attack of a new entrant. These can be further divided into:
    - *Ex-ante*: capacity investment, artificially induced switching costs, long-term binding contracts, product proliferation, vaporware. Example: let's consider Apple. It has succeeded in creating a very valuable system of interacting technologies, which enhance their functionalities reciprocally – *Lock-In effect.*
    - *Ex-post*: predatory pricing, vaporware. Purposely preventing the new entrant to get access to a given, necessary resource.

        1) Capacity investment: who is investing in further capacity? The incumbent (do not confuse it with entry costs). It concerns fundamentally a threat, since the incumbent increases its capacity to produce a lot of additional units, dragging down prices and thus deteriorating the possible new entrant's profit.

        2) Product proliferation: range of segments that your product can cover. It aims at covering all the possible consumers' preferences, eliminating all the possible niche in which the new entrant might enter, which will not have any opportunity to enter.

        3) Vaporware (ex-ante): it stems from the software industry. What is it about? A product is announced by the incumbent, but it's fake! This means that the product will never be released. By doing so, ther is less space for the new entrant to conduct its business both from a commercial perspective (targeting a certain portion of

consumers) and an ecosystem perspective (setting a given standard; setting suppliers of complementary products to produce for my given product removing them from the opponent)

4) Vaporware (ex-post): to scare the new entrant and induce it to exit

5) Predatory investment: incumbents sharply decrease price down to a level in which the new entrant will have no possibilities to compete because they can't cover their costs. This revolves around the efficiencies that the incumbent can exploit.

Definition of entry barriers:

According to Bain, *anything that allows incumbents to rise prices above competitive levels without inducing entry.*

Any problems with this definition? Note that, 1st, a definition must be precise and 2nd, it must be informative. This definition is too comprehensive, it does not convey that much information.

According to Stigler, *entry barriers are given by the cost that a firm entering new markets must bear, but that the incumbents do not have to bear.*

Any problems with this definition? Again, it is not that precise. Indeed, there are costs that are born by the incumbents as well (think of predatory pricing).

Correct definition: obstacles preventing new firms from entering a market and competing against the incumbents. Entry barriers allow the incumbents to keep the price higher than the average cost.

6.3 Sylos Labini postulate

New potential entrants behave as if they were able to forecast that the incumbents will keep their production at the same level as before the new entry. This means that the incumbent is not keeping any space to the new entrant.

Thus, the new entrant will assess whether entering or not considering:

- The residual demand diagram
- Its own cost function

How do incumbents react? Which are new entrants' conjectures? How much will incumbents produce after the new entry?

- Potential entrants hypothesize that incumbents will not vary their production after the new entry.
- Incumbents do not vary their production after the new entry.

→ The potential demand for the new entrant is given by the difference between the market demand and the quantity already offered by incumbents

6.4 Bain, Sylos Labini and Modigliani model

Assumptions under the Sylos Labini and Modigliani model (B-SL-M):

A) Perfect information
B) Sylos Labini's postulate (no production changes after the entry)
C) 2-steps competition. In t=1 the incumbent is the monopolist, and it decides both price and quantity. In t=2 a new potential entrant decides whether to enter the market.
D) Constant MC and AC
E) Incumbent's absolute cost advantage: the incumbent's (constant) AC is strictly lower than the new entrant's AC → Crucial assumption: The incumbent is more efficient. The costs curve, which is a flat line equal to the marginal costs (no fixed costs assumed) will lie below the new entrant's one.

The reason because of which the latter assumption does make sense:
Absolute cost advantages: incumbents' costs are always lower than new potential entrants' costs.

Many factors may underlie new entrants' cost disadvantages:

- Product differentiation (higher differentiation may be needed to compensate for switching costs)
- Institutional barriers (e.g., payment of the royalties related with a certain patent)
- Less advantageous contracts due to lack of prior relationships with suppliers

The entry decision is taken considering the potential demand and the cost function (including implicit costs)
The potential entrant will enter the market if it can obtain positive profits:

$$p - AC_i > AC_e - AC_i$$

unitary profit for the incumbent — incumbent's cost advantage

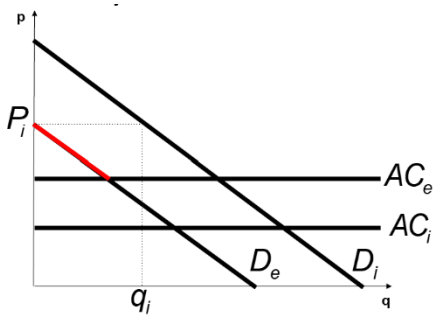
P > ACe >>> P-ACi > ACe-ACi

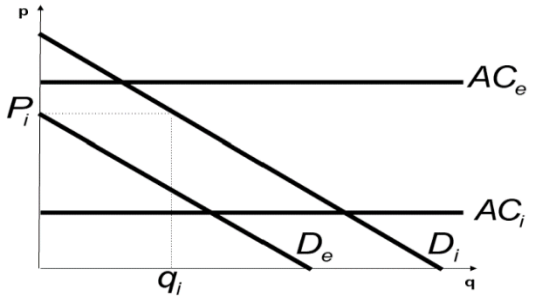
- p is the price before the new entry
- AC_i is the average cost for the incumbent
  AC_e is the average cost for the potential new entrant.

Thus, the new entrant will enter whenever the price is higher than its average costs. By entering and producing even a single unit, the price will indeed shrink. Still, the new entrant might be able to get a profit from that additional unit sold (but if $P > AC_e$).

If the *entrant's residual demand diagram* has a part that is *above* the *average cost* curve, positive profits can be made, and the new firm may enter the market. As long as there exist a positive unitary margin (red line) the entrant will try to enter the market. If P > AC, there is a part of De allowing to the new entrant positive profits.

→ The new entrant enters the market and the price decreases.

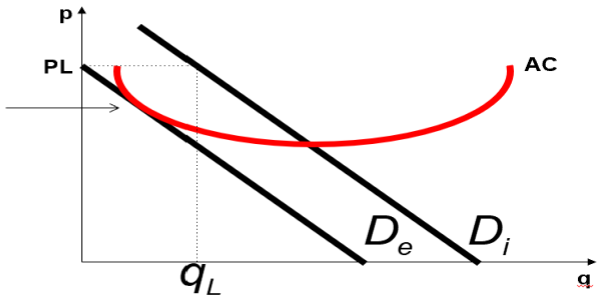




In this case the new entrant average cost curve lies above any existing point of the demand curve. Thus, by producing a single unit, it would incur in a loss since it would be forced to sell the unit at where AC > P to reach consumers.

The higher the difference between AC_e and AC_i, the greater the possible difference between P and AC_i without incurring any new entry.
With a high difference between AC_e and AC_i, the incumbent can charge higher prices without attracting new potential entrants.
P - ACi is a proxy for the height of entry barriers.

Where AC and the Residual Demand are tangent, the profit for the new entrant is null



Which is the optimal price an incumbent must set to maximise profits, subject to the constraints of limiting the new entrant? The limit price will be where:

$$p - AC_i = AC_e - AC_i$$

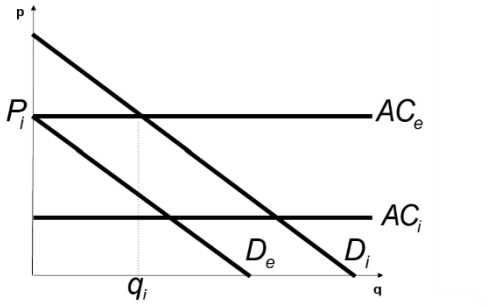
unitary profit margin — incumbent's cost advantage

Price limit : P=ACe

Highest price that can be charged without incurring new entries:

- For *higher prices* the entrant may enter the market.
- For *lower prices* the entrant cannot enter the market, but the incumbent is making suboptimal profits.

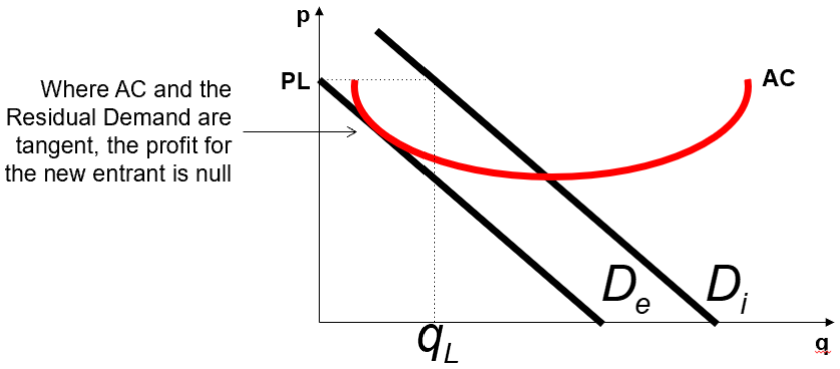
Graphical representation:



6.5 Bain, Sylos-Labini & Modigliani with Economies of Scales

Economies of scale → decreasing average cost before the Minimum Efficient Scale.
→ in the B-SL-M model with ES, the higher the *economies of scale*, the higher is the *price limit* (i.e., the optimal price enabling new entry deterrence). Note that here the cost advantage will revolve around the capability of the incumbent of reaching a higher (more efficient) economies of scale. Thus, the assumption of costs advantages still holds.
→ Extra assumption with respect to the previous case: The two firms face the same u-shaped AC curve: no absolute cost advantages.
Essentially, the line of reasoning is the same:

- SL postulate: the new entrant knows that the incumbent will keep production unchanged.
- The potential new entrant's decision depends on residual demand (i.e., the difference between the market demand and the incumbent's production).
- Considering the average cost curve of the new entrant, the incumbent sets the quantity so that residual demand will not allow any profits for the new entrant.
- As a result, the potential new entrant refrains from entering.

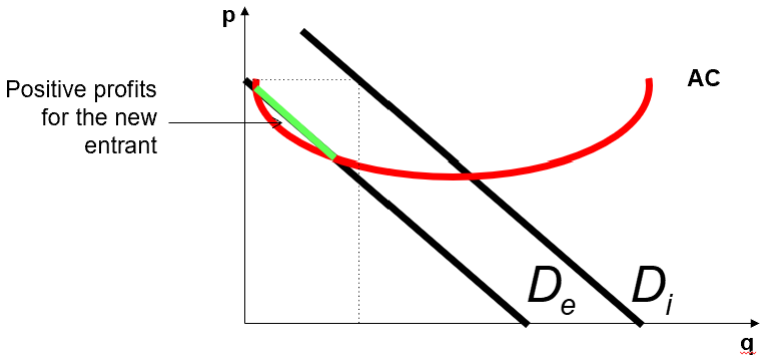


AC is higher than the residual demand → the price is lower than the AC → the entrant would incur losses.
q_L is the minimum quantity allowing new entry deterrence.
PL is the Price Limit allowing new entry deterrence.

Note that the advantage that the incumbent has on the new entrant is that at the very beginning the average cost curve is very high. The cost curves are the same, but the new entrant will serve the residual demand. The quantity that allows the incumbent to maximize its profits while keeping the new entrant far from entering the market: Where the residual market demand is tangent to the cost curve.

As in the model with an absolute cost advantage, the *price limit* is the highest price the incumbent can set to prevent new firms' entry into the market.

- Charging a price *higher* than PL → *entry becomes feasible*



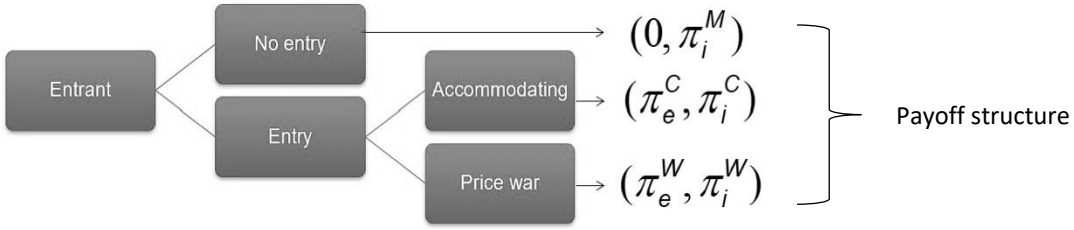
Charging a price *lower* than PL → *extra profits reduction* for the *incumbent.*

If the price is *higher* than the price limit, there is a part of the Residual Demand that is above the Average Cost → Positive profits for AC, meaning that the new entrant. will incur positive profits.

Limitation and critique: the *Sylos-Labini* postulate implies *irrational* conjectures of the potential entrant regarding the leader's behavior:

- After the entry, the incumbent might not find it convenient anymore to produce the quantity q_L at the price limit PL. → It would be a suboptimal quantity, thus not rational, thus not credible.
- *Ex-post*, an *accommodating* strategy by the incumbent may be the most rational outcome, as a *price war* would harm both players. A rational decision, hence, would be to decrease slightly production and recalculate the optimal level given the new entrant presence.

6.6 Dixit Model



6.6.1 Dixit model (1982): removal of the SL postulate

1) First step: the new entrant decides whether to enter the new market or not.
2) Second step: If the entrant decides to enter, the incumbent decides whether to engage in a price war or adopt an accommodating strategy

[The profits in a price war (π_e^w, π_i^w) would be smaller than the profits in the accommodation case (π_e^c, π_i^c)]

Backward induction: (Note that the solution is a short-term "snapshot")

- If the potential entrant does not enter, the incumbent will make monopolistic profits.
- If the potential entrant decides to enter:
  - *Price war* with low profits for both players equal to
  - *Acquiescence* → Cournot duopoly profits equal to
- As long as π^c is greater than π^w, the incumbent *will choose an accommodating strategy* in the event of new entry, and *the new entrant knows it*.
- Since the new entrant is aware of point 1 and it acts first, it just selects the best alternative between *no entry* and *entry* → *accommodating strategy*.
- *No entry* is preferable to *entry* → *accommodating strategy* if and only if the payoffs associated with the former are. Thus, in this version of the model, *if 0 is greater than* π^c (i.e., Cournot profits are negative).

Is there a way through which the incumbent can do to induce the new entrant that the price war might be a credible response? → There is only one way: We must force to push the profits achievable in the accommodation case to a lower level than the one achievable under a price war. How? Through capacity

investments. In doing so, we are anticipating some of the costs that would be incurred anyhow in case of a price war. By doing so π^w remains the same, while π^c decreases. The company commits to the price war before the new entrant enters. In doing so, a price war becomes preferable! This would be a sub-game perfect Nash equilibrium. Now the new profits would be equal to π^m less the costs incurred in implementing the strategy, i.e., the investments to augment the production capacity. Intuitively, the strategy will be implemented if and only if the costs are smaller than the difference between the profits obtainable as monopoly and those in accommodating the new entrant, i.e., c < $\pi^m - \pi^c$. Note that if the costs are larger, i.e., c > $\pi^m - \pi^c$, than the firm would prefer to accept the new entrant and share the demand.

Brief Exercise:

A market demand is characterized by a demand function $Q = 1 - P$ and by a single firm whose fixed and *marginal cost is 0*. The monopolist is facing potential entry from a new firm having a *marginal cost of 0* and a *fixed cost of 0,1.* If the incumbent accepts the entry passively, then *Cournot competition* is played. However, the monopolist can also adopt an aggressive behavior and threaten to set a quantity such that *price = 0.* If the new entrant does not enter, the incumbent will behave as a monopolist.

1) Compute the payoffs for both firms (i.e., profits) in the cases of Monopoly, Cournot duopoly, and aggressive behavior.

In case of *no entry,* the incumbent will set the monopolistic level of output:

$$\pi_i^M = TR - TC \quad \textbf{and} \quad \pi_i^M = P \cdot Q - 0 \quad \textbf{and} \quad P = 1 - Q$$
$$\Rightarrow \pi_i^M = Q \cdot (1 - Q) \Rightarrow \frac{\delta \pi_i^M}{\delta Q} = 1 - 2Q = 0$$
$$\Rightarrow Q = 1/2$$
$$\Rightarrow \pi_i^M = \frac{1}{2} \cdot \left(1 - \frac{1}{2}\right) = \frac{1}{4} \qquad \textbf{and} \qquad \pi_e^M = 0$$

If the firm *enters* the market and the incumbent plays passively, *Cournot competition* arises:

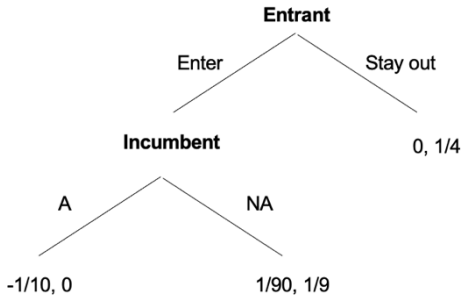
$$\pi_i^C = \pi_e^C = TR - TC = P \cdot Q - FC \quad \textbf{and} \quad P = 1 - q_i - q_e$$
$$\pi_i^C = (1 - q_i - q_e) \cdot q_i \qquad \pi_e^C = (1 - q_i - q_e) \cdot q_e - 0.1$$
$$\pi_i^C = q_i - q_i^2 - q_i q_e \qquad \pi_e^C = q_e - q_i q_e - q_e^2 - 0.1$$

$$\begin{cases} \frac{\delta \pi_i^C}{\delta q_i} = 1 - 2q_i - q_e = 0 \\ \frac{\delta \pi_e^C}{\delta q_e} = 1 - q_i - 2q_e = 0 \end{cases} \Rightarrow \begin{array}{l} q_i = q_e = 1/3 \\ \pi_i^C = 1/9 \\ \pi_e^C = \frac{1}{9} - \frac{1}{10} = \frac{1}{90} \end{array}$$

If the new firm enters the market and the incumbent behaves aggressively (in this case, by producing the quantity such that p=0), then the new entrant pays the fixed costs, but it is not able to make any profit

$$\pi_i^A = 0$$
$$\pi_e^A = -1/10$$

2)Using extensive form (game tree) representation, describe this entry game as a two-stage game: in the first stage the new entrant decides whether to enter or not, while in the second stage the incumbent decides whether to behave aggressively in the event of entry.

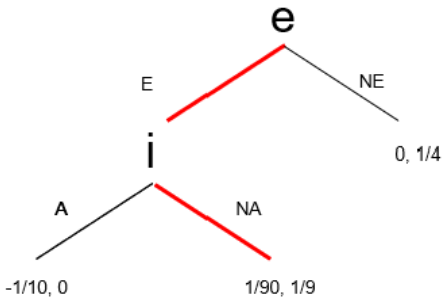


3)Is the threat of aggressive behavior by the monopolist credible?

Assuming the first firm enters the market, the new entrant knows that the incumbent prefers an accommodating response (payoff of 1/9) to an aggressive one (payoff of 0).

Therefore, *the threat of aggressive behavior is not credible: given perfect information and perfect rationality*, the new entrant knows that, by choosing to enter, it is going to get a payoff of 1/90. Conversely, by not entering, it gets a payoff of 0. As 1/90 > 0, it is going to enter.

4) Determine the subgame perfect Nash equilibrium.



Note: e = entry, i = incumbent, E = entering, NE = not entering, A=aggressive, NA = not aggressive

Remark: e = entry, i = incumbent, E = entering, NE = not entering, A=aggressive, NA = not aggressive

7. EXTERNALITIES

7.1 Examples and types of Externalities

An externality (spillover) is a cost, or a benefit imposed upon someone by actions taken by others (with no compensation)

- An externally imposed benefit is a positive externality
- An externally imposed cost is a negative externality

Examples (Negative):

- Air & water pollution.
- Loud parties next door.
- Traffic congestion.
- Second-hand cigarette smoke.

Examples (Positive):

- A well-maintained property next door that raises the market value of your property.
- Network externality
- Vaccines

e.g. you are only man in the 90s who has mobile phone , what's the benefit ?
so others' decision to buy a phone has a positive externality on you

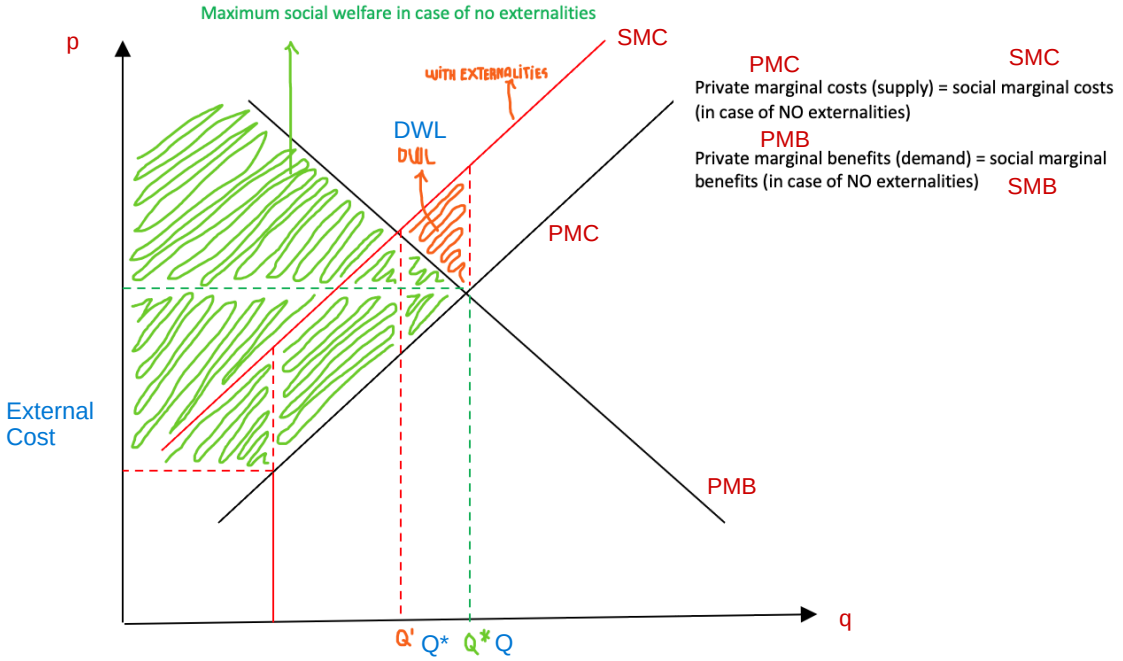
There are different types of externalities:

- Consumption externalities: Consumption of a good by agent *A* has a direct impact on agent *B*'s utility (e.g., smoking or listening to loud music)

- Production externalities: Production actions by agent *A* have a direct impact on agent *B*'s utility/profit (beekeeper and apple orchard)

IMPORTANT TABLE

	Production	Consumption
Positive	Bee-keeping	Vaccines Network externalities
Negative	Steel-mill Fishery	Loud music Second-hand smoke



Private Marginal Benefits: Extra portion of utility got by consuming a further quantity → in case of no externalities, this will be equal to the social marginal benefits (collective marginal benefits).

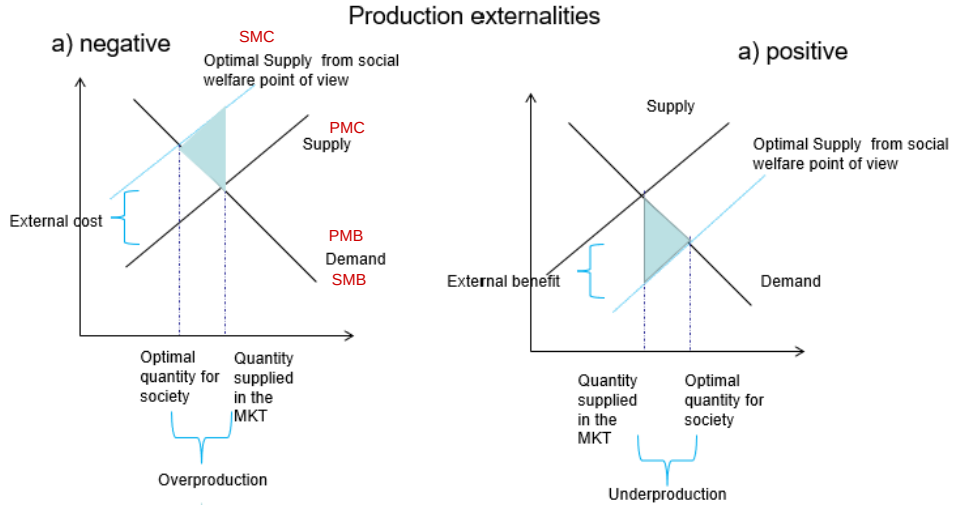
Private Marginal Costs: Extra cost required to produce a further unit → In case of no externalities, this will be equal to the social marginal costs, i.e. the cost that the entire society has to face to produce that quantity. There are no extra costs aside from the one strictly related to production.

Private Marginal Costs: Extra cost required to produce a further unit → in case of no externalities, this will be equal to the social marginal costs, i.e., the cost that the entire society must face to produce that quantity. There are no extra costs aside from the one strictly related to production.

What is then the problem with externalities? That PMB = SMB and PMC = SMC do not hold anymore. Take for instance the production of steel. To the firm, the only cost that will be faced is the one related to the production of steel, but this will also have some negative impacts (production externalities) on the society. Thus, SMC will not be equal to PMC anymore (Red Drawings). In that case, there is a too great production that damages the society (Dead Weight Loss). The social optimal quantity to be produced would be Q' and Q* anymore since the costs to the society consider the externalities as well.
So, externalities cause welfare losses:

- too many resources are allocated to an activity which causes a negative externality (overproduction)
- too little resources are allocated to an activity which causes a positive externality (underproduction).

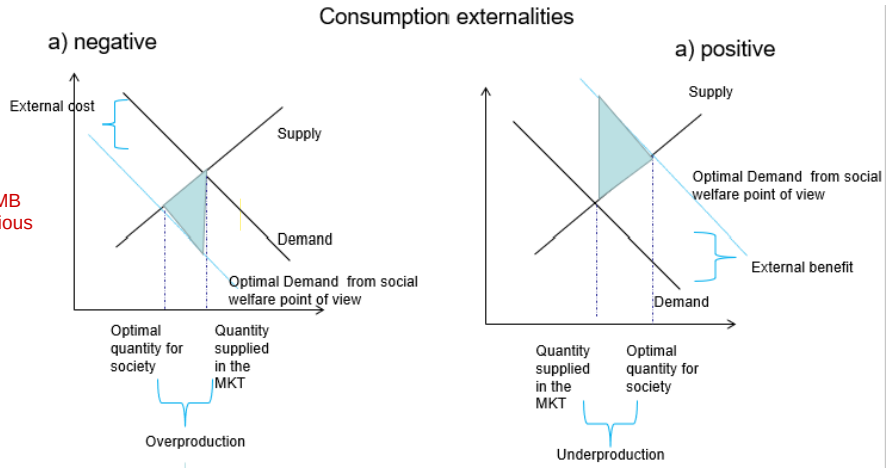
IMPORTANT GRAPHS



a) Welfare loss (blue triangle): Marginal private (= social) benefit of consumers (DD) exceeds marginal private costs suffered from firms (SS) but its inferior to the social marginal costs suffered from the society (which also includes external costs): producing those units has a cost which is larger than the benefit from a social welfare perspective

b) Welfare loss (blue triangle): Marginal private (= social) benefit of consumers (DD) is below marginal private costs suffered from firms (SS) but it is superior to the social marginal costs suffered from the society (which also includes external benefits): producing those units brings a benefit which is larger than the cost from a social welfare perspective

External cost decreases Benefit ,hence, the new SMB curve goes below the previous one

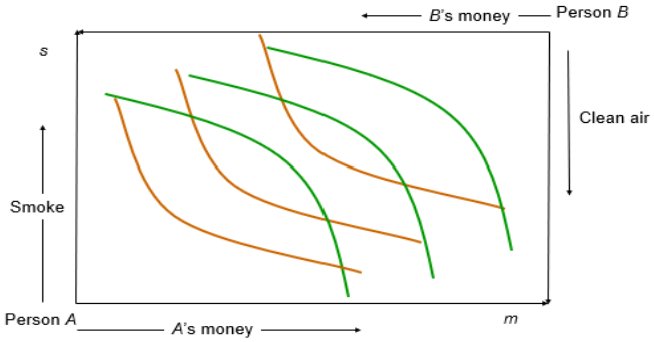


a) Welfare loss (blue triangle): Marginal private (= social) cost of firms (SS) is below marginal private benefits of consumers (DD) but it is superior to the social marginal benefits of the society (which also includes external costs): producing those units has a cost which is larger than the benefit from a social welfare perspective.

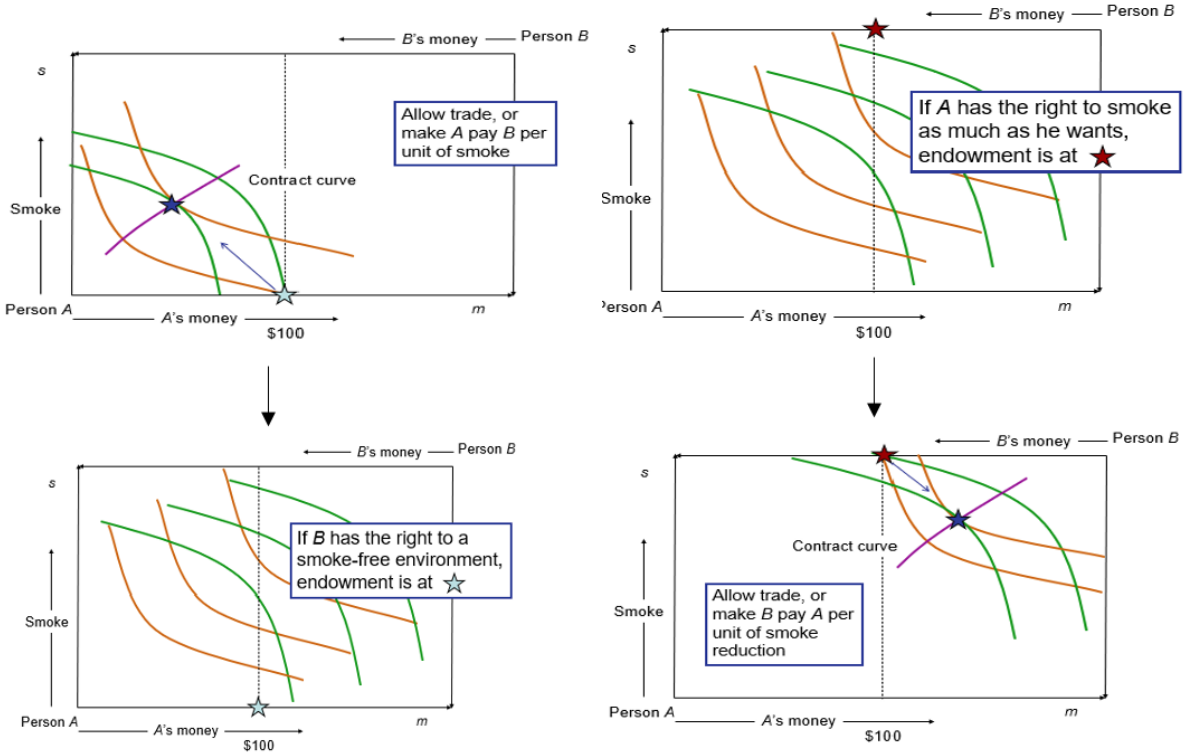
b) Welfare loss (blue triangle): Marginal private (= social) cost of firms (SS) is above marginal private benefits of consumers (DD) but it is inferior to the social marginal benefits of the society (which also includes external benefits): producing those units brings a benefit which is larger than the cost from a social welfare perspective.

7.2 Externalities and Efficiency

How can we solve such problem? → the problem is one of "missing" markets (Coase). By creating a market, the system might help to tackle the problems stemming from externalities. Do understand how, let's consider an example: We have to roommates, A and B and two goods. One is generic and can be approximated to money and its endowment with respect to the roommates is equal. m_A and m_B : Endowments = \$100; The other good is smoke, which concentration varies within the range of [0,1]. Here, A is a smoker, i.e., $u_A(m_A,s)$ → He has a utility both from money and smoking. B is not a smoker; thus its utility function will depend on money and on the inverse of smoke t (where t = 1-s) → $u_B(m_B,t)$, thus the less A smokes, the better he will be. (Note: s + t = 1). The situation can be represented with the Edgeworth Box:



Firstly, we know that both agents are endowed with m = 100.
According to Coase, what is missing is an entitlement of rights → neither A nor B knows who is entitled to do what. Is A entitled to smoke? Is B entitled to have a clean room? This lack of clarity causes the two agents to fight. The intuition of Coase consists in finding an external authority that basically gives to right to each agent to behave to some extent.

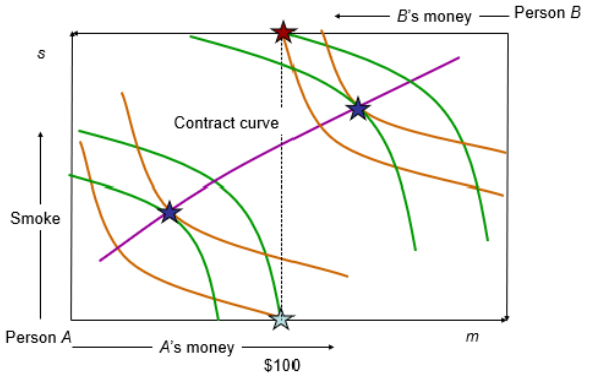


7.3 Externalities and Property Rights, Coase's Theorem

After having assigned a right, then we allow the two agents to trade up until they find an agreement (here given by the points in which the two utility curves are tangent on the contract curve. Note that the outcome will depend on who is entitled with what. On the left quadrant we entitled B with a smoke free environment, thus, if A would like to smoke at least a cigarette (or cigar, or whatever she smokes) it will have to purchase that quantity. A similar reasoning applies in the opposite case in which A is endowed with the right of smoking freely. In both cases we know that anyhow a Pareto Optimal solution will be obtained through allowing them to trade.
Ronald Coase's intuition was that most externality problems are due to an inadequate specification of property rights and, consequently, an absence of markets in which trade can be used to internalize external costs or benefits.
Note: Since Coase did not come up with a proper, rigorous theorem, this has been developed later by others. Therefore, there are actually two versions of it:

Weak version: efficiency proposition) *When parties can bargain without cost (no transaction costs), the resulting outcome will be efficient, regardless of how the property rights are specified* → This holds

Strong version: efficiency + invariance propositions): *When parties can bargain without cost, the resulting outcome will be efficient and the level of the externality generated the same, regardless of how the property rights are specified.* à It can be verified only in case of quasi-linear preferences (quite a peculiar case). Thus, it is very difficult that it holds.



Strong version generally not verified: *A* and *B* care about who gets the property rights and the level of smoke generated changes (from a theoretical point need of quasi-linear preferences).

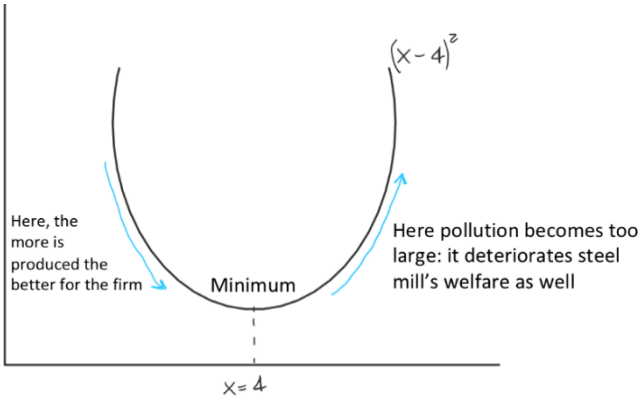
Key message of Coase theorem is that Pareto improvements are possible to the extent that there are ways that allow the economic system to internalize the externality through market mechanisms. Either by means of Bargaining of consumers (Edgeworth box example) or merging of firms (steel mill vs. fishery example).

7.4 The market signal for Merging

Let's suppose that we have a steel mill and a Fishery. Now, let's consider that we have two different scenarios:

A) Steel mill does not care whether the wastes are dumped in the river: Its objective function will be thus: max $\pi (S, x) = p_s S - C_s (S, x)$, where profits are a function of steel and wastes (stemming from core business) produced.

Suppose $C_s (S, x) = S^2 + (x-4)^2$ and $p_s = 12$. à F.O.C_1 $(\partial\pi/\partial s) = 0$; $S^* = 6$ à F.O.C_2 $(\partial\pi/\partial x) = 0$; $x^* = 4$ à $\pi_S = 36$



The cost curve is represented by an inversed parabola where the optimal production is given by $S^* = 6$ (steel to be produced) and $x = 4$ (pollution).

Note: If the steel mill produces more than that quantity the impacts of pollution will also impact the steel mill as well more than the benefits it obtains.

Similarly, Fishery: max $\pi (f, x) = p_F f - C_F (f, x)$ [Note that x is exogenously given, i.e. the Fishery does not control the pollution quantity! On the other hand, it is damaged by it!]

Suppose $C_F (f, x) = f^2 + xf$ and $p_F = 10$. à F.O.C $(\partial\pi/\partial f) = 0$; $10 - 2f^* - x = 0$; $f^* = 5 - (1/2)x$

Given the choice of $x^* = 4$ by Steel mill, $f^* = 3$. à $\pi_F = 9$ à Total profits will be given by $\pi_S + \pi_F = 36 + 9 = 45$.

B) Merger. Now the new objective function will be given by max $[\pi (S, f, x) = 12S + 10f - S^2 - (x-4)^2 - f^2 - xf]$

F.O.C_1 $(\partial\pi/\partial s) = 0$; $S^M = 6$

F.O.C_2 $(\partial\pi/\partial f) = 0$; $10 - 2f^* - x = 0$; $x^M = 10 - 2f^M$

F.O.C_3 $(\partial\pi/\partial x) = 0$; $-2(x^M - 4) - f^M = 0$

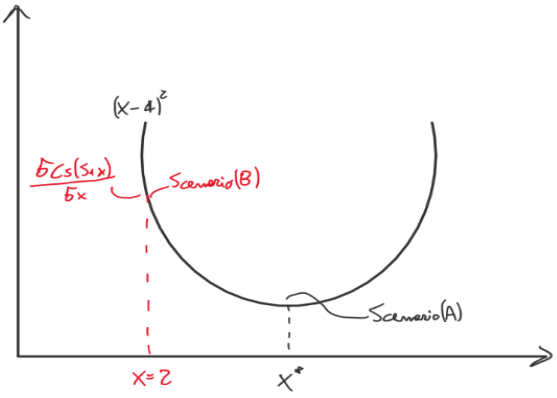
Substituting F.O.C_2 in F.O.C_3: $x^M = 2$ and $f^M = 4$

Before polluting units were equal to 4, in the second scenario are down to 2.

Note: $\pi_M = 48$ while in scenario A it was $\pi_S + \pi_F = 36 + 9 = 45$. The intuition lies in the fact that the market through the profit signal that provides the incentive to the steel mill to merge with the fishery. The merged firm by taking also care of the social cost of steel production makes everybody better off. The second scenario clearly pareto dominates the firs case. Indeed, now the steel mill now cares about the damage provoked on the fishery.

In scenario A, Steel mill F.O.C_2 $(\partial\pi/\partial x) = 0$ à $(\partial C_s(S, x)/\partial x) = 0$

In scenario B, Steel mill F.O.C_3 $(\partial\pi/\partial x) = 0$ à $(\partial C_s(S, x)/\partial x) + (\partial C_F(f, x)/\partial x) = 0$



Now the quantity of pollution produced by steel mill is achieved where the derivative in that particular point is negative, thus, here, 2.

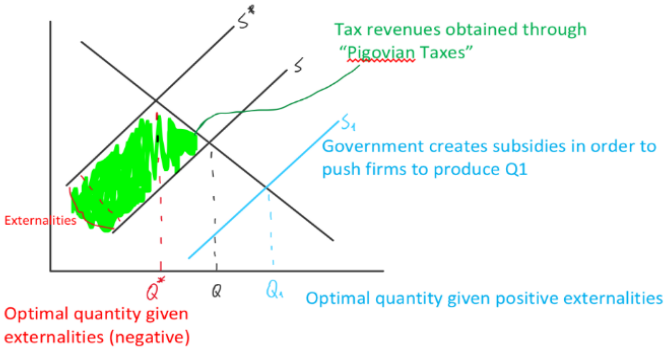
But allocation of property rights and trade (Coase theorem) or redistribution of property rights (Merging) as a way to solve externality problems hinges upon small numbers involved, absence of bargaining (transaction) costs and/or the ability of the firms to recognize profit market signals and act accordingly.

In the real world: Numbers involved are often very large (externalities affect whole society), thus bargaining costs are extremely high and the merge between a steel mill and a fishery is far from being obvious.

How can we cope with these problems in reality? Public policy has a more invasive role than just assigning property rights or leave the market exploit profit signals

What to do if Coase theorem signals mechanism breaks down? More solutions might be implemented:

- Command & Control: The public policy sets a limit in terms of pollution for instance that all the private players cannot exceed. If so, they are entitled to pay a severe fine → negative incentive (similar to speed limit on the highways).
- Pigouvian taxes (or subsidy): Attempt of the public policy to make what the optimal supply curve from social welfare point of view visible
- Tradeable permits (artificial markets in the spirit of Coase



Despite the theory might be simple, the government has no precise idea whether the target in terms of pollution will be met or not. The other aspect regarding the *Command-and-Control* mechanism and Pigouvian taxes is that they are imposed on the entire industry without any distinctions. The last way regards the implementation of tradeable permits ("cap and trade"), which basically consists in giving a permit to each company regarding a maximum level of pollution that can be reached which then will be traded in the market.

A better way to understand how it works is through an example:

Suppose that there are two companies, here Toxic and sludge which equally pollute (1000 gallons). Suppose the state imposes a limit to this level at 500 gallons. In a command and control both firms have to stay within this imposed limit. But these firms have different costs to meet the state's request: cleanup costs. In the case of a command-and-control solution, both firms will have to reduce by 500 their pollution. Toxic will bear clean up costs up to 5000 dollars and Sludge equal to 1000. Total cost = 6000. In case of a cap and trade, the state does not impose a limit but gives the firms 500 permits to pollute (1 permit = 1 gallon of pollution). If we think about it, the aforementioned result can be replicated without any problems, where each firm uses the permits for itself. But, since the state allows to trade permits, firms might have an incentive to do so according to the willingness to pay for an extra (501st) permit. The price Toxic is willing to pay will be at maximum equal to the price it would pay to clean up. The same reasoning applies with respect the price Toxic will be wiling to sell the permit. If larger than 10 it will be actually more convenient to sell them and incur the costs of cleaning. See that whereas Toxic is willing to purchase at max P = 10, Sludge is willing to sell a permit at a price slightly larger than 2 → There exists room for trading. In this simplified example, they will indeed trade 100% of the permits. The result depends solely on the different specializations of the firms: Since

sludge has a "special talent" to clean up, a better result with respect to a command-and-control solution. Note: A boundary condition must hold: The effect of the externality is not localized, i.e. geographically bounded.

Companies	Status quo pollution (gallons)	Clean-up cost (gallon*$)	New Pollution allowed
Toxic	1,000 units	10$	500 units
Sludge	1,000 units	2$	500 units

	C&C
	Clean-up cost
Toxic	(1,000-500)*10$ = 5,000$
Sludge	(1,000-500)*2$ = 1,000$
Total cost for society with C&C	**6,000$**

Market (Cap& Trade)

For 1 gallon of pollution firms need 1 permit
For 1 gallon of pollution reduction a firm obtains 1 permit

	Permits issued	WTP (max price a firm would pay for a permit)	WTS (min price a firm is willing to sell a permit)
Toxic	500	10$	10$
Sludge	500	2S	2$

	Transaction	N° of new permits	N° of clean-up gallons needed	Cost of clean-up
Toxic	buy 500 permits	1000	0	0$
Sludge	sell 500 permits	0	1000	2,000$

Cost for companies	
Toxic	0$ (cost of clean-up) + costs of permits
Sludge	2,000$ (cost of clean-up) - revenue from permits
Total cost for society with Tradeable Permits	**2,000$** (Since cost of permits for Toxic = revenue from permits for Sludge)

8. PUBLIC GOODS

What is a public good? A situation in which an action performed by an agent X to maximise her objective (utility or profit) function has the same impact on others' objective functions (and it is positive) → it can be seen as a positive externality stretched at its maximum.

Non-excludable *from* consumption and non-rivalry *in* consumption, i.e. you cannot prevent someone to consume those goods and if you are consuming these goods you cannot actually prevent someone to use that good simultaneously. Some examples might be Street Lighting, Pedestrian Walk, Defense or Clean air. An economic (rather than a political or technical) concept:

Its public nature depends on its intrinsic characteristics:

a) Do we have the technical capability to exclude non-payers from non-rival goods consumption?
b) If technically possible, is it economically feasible? (Pedestrian walks with gates?)
(Pedestrian walks with gates?)

IMPORTANT IMAGE

Once provided, quantity of the public good is the same for everyone by definition

Basic Taxonomy (image beside)

8.1 When should a public good be provided?

Let's suppose we have two roommates, 1 and 2, who are deciding whether or not to purchase a TV. Given the size of the apartment, the TV will be placed in the living room where both roommates will be able to watch it. Here, the TV is a public good, i.e., neither rival nor excludable. Since it is a public good, we shall ask ourselves how the two of them should consider when to buy the TV or not. Let's use w1 and w2 to denote each person's initial wealth, g1 and g2 to denote each person's contribution to the TV, and x1 and x2 to denote each person's money left over to spend on private consumption.

The budget constraints are given by:

$X_1+g_1 = w_1$

$X_2+g_2 = w_2$

We suppose that TV costs c, so to purchase the TV, the sum of the contributions must be at least c:

$g_1 + g_2 \geq c$ [Perchè "at least" e non 'equal']

Reservation price "r" (i.e. willingness to pay) of Person 1 for the TV is the price according to which the utility he gets from paying TV and watch it equal to utility he gets from not paying and not watching –> Price that makes Person 1 indifferent between the two scenarios:

$U_1(W_1 – r_1,1) = u_1(W_1, 0)$, where 1 = TV has been bought

Analogously for Person 2:

$U_2(W_2 – r_2,1) = u_2(W_2, 0)$

In this TV problem, there are 2 allocations of interest:

1) Each person spends his wealth only on his private consumption: $(w_1,w_2,0)$

2) They decide to buy the TV: $(x_1,x_2,1)$

In which case will the TV be provided? Surely, when allocation 2) Pareto dominates allocation 1), which means that:

$u_1(W_1, 0) < u_1(x_1, 1)$

$u_2(W_2, 0) < u_2(x_2, 1)$

Using reservation prices and budget constraint we can write:

$U_1(W_1 – r_1,1) = u_1(W_1, 0) < u_1(x_1, 1) = U_1(W_1 – g_1,1)$

$U_2(W_2 – r_2,1) = u_2(W_2, 0) < u_2(x_2, 1) = U_2(W_2 – g_2,1)$

Therefore, it has to be that:

$W_1 – r_1 < W_1 – g_1$

$W_2 – r_2 < W_2 – g_2$

Which in turn implies that:

$r_1 > g_1$

$r_2 > g_2$

and $r_1 + r_2 = g_1 + g_2 = c$

The result is pretty intuitive: The price for which both agents would be indifferent between purchasing the TV or not, has to be larger than the actual price they have to spend in order to purchase the TV.
Allocation (2) will be Pareto dominant if each reservation price is greater than each payment and the total amount that the roommates are willing to pay is at least as large as the cost of the purchase.
This is so straightforward we may ask why we have been through all these mathematical formalisms for this intuitive result.
Mainly for one reason: In the real world, the reservation price (willingness to pay) of each person depends on his/her personal wealth. Whether or not to provide a public good, will depend on reservation prices and in turn on the distribution of wealth among members of a community. It is perfectly possible that for some wealth distributions
$r_1 + r_2 > c$ and for others $r_1 + r_2 < c$.

Example: Imagine a situation where one roommate really loves the TV and the other roommate is nearly indifferent about acquiring it. Then if the TV-loving roommate had all of the wealth, he would be willing to pay more than the cost of the TV all by himself. Thus it would be a Pareto improvement to provide the TV. But if the indifferent roommate had all of the wealth, then the TV lover wouldn't have much money to contribute toward the TV (i.e., low reservation price), and it would be Pareto efficient *not to provide* the TV. What is the problem with the provision? Suppose reservation prices exceed personal payments and the sum of personal payments is enough to cover the cost of the good. Thus allocation 2) is Pareto efficient. Are we sure that this equilibrium will emerge? Of course not…

8. The free-riding problem

Suppose that each person has a wealth of 500 €, that each person values the TV at 100 € , that the cost of the TV is 150 € and that there is no way for one of the roommates to exclude the other one from watching (as before).

But now suppose that each roommate has to decide independently from the other whether or not to buy the TV and they do not know each other. They need to understand whether they want to buy the TV on their own.

Free riding game matrix.

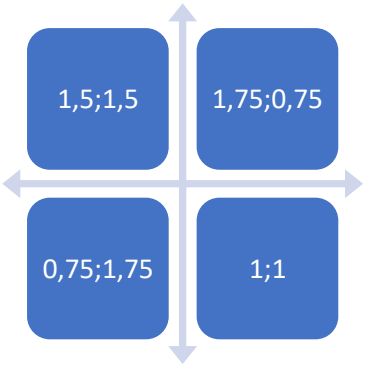
		Player B	
		Buy	Don't buy
Player A	Buy	–50, –50	–50, 100
	Don't buy	100, –50	0, 0

A Free rider: Person who receives the benefit of a good but avoids paying for it

Here the game is trivial but it makes the point: each player has an incentive to free-ride on the expenditure of the other (since by doing this, it enjoys the TV without paying anything). But if both players think in this way the result is that the goodwill not be provided, even if the scenario where TV is bought and the payment is equally shared would be a Pareto improvement. Indeed, if they agree dependently to purchase the TV, they would both pay solely 75, thus obtaining a 25 gain. Note: -50 is achieved since both value the TV 100 but it costs 150 and we are considering the case in which both players are purchasing the TV independently.

Similar example: A Public Good Game:

Suppose we have two players endowed with 1 Euro each that could be put into a public fund which will multiply what put in it by 1.5 and the outcome will be split equally between the two players. According to Game Theory the Nash Equilibrium will be the one in which both players will not put any money in the fund, depicted on the right with [1;1]



In reality, we can expect things not to unfold in this way (2 roommates are friends, contribute equally, and the TV will be bought). But in other circumstances the possibility to free-ride may become more likely and "no provision" a serious possibility. Who can solve the Problem then? The Government can remedy: Providing the public good and paying it with tax revenues. But still problems with free riders in paying taxes.

How much of a public good should be provided? How should this riddle be tackled? It a complex issue: By definition the public good has to be provided in the same quantity for all individuals, even though all individuals have different preferences regarding their "perfect" quantity. But "one size should fit all".

In principle, one could collect the willingness to pay/contribute of each individual for different quantities (e.g., with exit polls) or present individuals with different options to choose from, for then making them vote on the preferred size.

But still, some problems do still arise:

Person A	Person B	Person C
x	y	z
y	z	x
z	x	y

- How to collect trustable and credible "willingness to pay" by citizens?
- Voting is not perfect (e.g., social preferences may not be transitive)
- At the aggregated social level:
  x > y, y > z, z > x
- This is a version of the "Condorcet Paradox" from which Arrow started his reasoning that led him to the so-called "Impossibility theorem" which is as follows:

1. Given any set of complete, reflexive, and transitive individual preferences, the social decision mechanism should result in social preferences that satisfy the same properties.
2. If everybody prefers alternative x to alternative y, then the social preferences should rank x ahead of y.
3. The preferences between x and y should depend only on how people rank x versus y, and not on how they rank other alternatives.

All three of these requirements seem eminently plausible. Yet it can be quite difficult to find a mechanism that satisfies all of them. In fact, Kenneth Arrow has proved the following remarkable result:1 Arrow's Impossibility Theorem. If a social decision mechanism satisfies properties 1, 2, and 3, then it must be a dictatorship: all social rankings are the rankings of one individual. Arrow's Impossibility Theorem is quite surprising. It shows that three very plausible and desirable features of a social decision mechanism are inconsistent with democracy: there is no "perfect" way to make social decisions.

9. COMMON GOODS

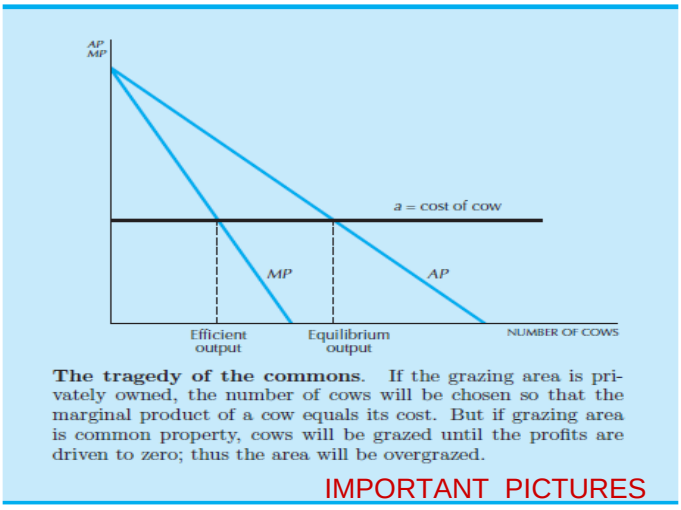
9.1 The problems

What is the problem with 'commons? The definition of Commons: grazing land when the villagers graze their cows on a common field. Individuals are free to graze their cows in the fields with no restrictions.

Hardin (1968): This situation typically determines the so-called tragedy of the commons, i.e., over-exploitation of the field:

- Any villager is tempted to graze more than her (hypothetical) share
- If all succumb to the same temptation, the grass ceases to grow and the value of the pasture to everybody disappears.

More formally: Each villager must decide how many cows bring to graze. The variables to be taken into consideration are *c* is the number of cows, where each cow costs *a; y (milk) = f(c) with f'(c) > 0 and f'' (c) < 0,* i.e., a strictly positive concave function (might be logarithmic, for instance). Furthermore, each villager is a price taker: p of a liter of milk = 1 €. As many problems that we have already seen, villagers will bring their cows to graze until MR = MP. Now, we will distinguish two different scenarios:



The tragedy of the commons. If the grazing area is privately owned, the number of cows will be chosen so that the marginal product of a cow equals its cost. But if grazing area is common property, cows will be grazed until the profits are driven to zero; thus the area will be overgrazed.

IMPORTANT PICTURES

- Private Property: Here, one villager owns the field and decide how many cows should enter:

  $\underbrace{max}_{c}$ π = f(c) – ac → f.o.c. : f'(c*) = a, i.e., marginal costs = marginal value produced by the cows

Thus, *if f'(c) > a, a new cow will enter; if f' (c) < a, a cow will exit, eq.:* f'(c*) = a

- Commons (none owns the field; everybody can freely enter with cows). Suppose that there are *c* cows currently being grazed so that the current output per cow is equal to f(c)/c. When a villager contemplates if adding a new cow, (s)he will compare f(c+1)/(c+1) > a. If this inequality is verified, the new cow will be added; otherwise not. It follows that the total number of cows will be a specific c^ that verifies f(c^)/c^ = a. This c^ is certainly greater than c*

In other words f(c)=ac or f(c+1)=a(c+1) or f(c+2)=a(c+2) , ... which means until the profit will be zero , i.e. TR = AC

9.2 Possible Remedies

Are there remedies to this tragedy of commons?

Transforming the common into a private good (enforce excludability) with precise definition of property rights. Private property provides such a mechanism. Indeed, if the field is owned by someone who can control its use and can exclude others from overusing it, then there are by definition no externalities. The market solution leads to a Pareto efficient outcome (and surely dominant with respect to a situation where the good has become totally useless to everyone because it has been over exploited).

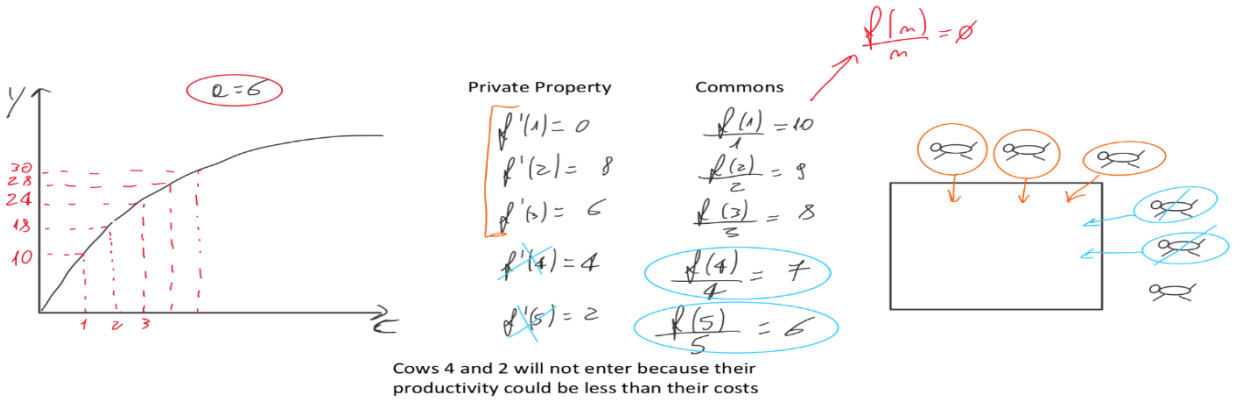
Of course, private property is not the only social institution that can encourage efficient use of resources. As Hardin (1968, p. 1245) notes "we might keep them as public property but allocate the right to enter". Nobel prize Elinor Omstrom defined 8 managing principles for governing commons in her famous book Governing the commons, 1990, freely available in Internet). For example, rules could be formulated about how many cows can be grazed on the village common. If there is a legal (or strong social) system to enforce those rules, this may be an alternative solution for providing an efficient use of the common resource.

Example: Overfishing: However, in situations where excludability mechanisms are not implementable and the law is ambiguous/difficult to enforce, the tragedy of the commons can easily arise. Overfishing in international waters is an example: each fisherman has a negligible impact on the total stock of fish, but the accumulated efforts of thousands of fishermen result in serious depletion.

When "Commons" are left "Commons", and formal institutions (rules, laws, norms) are unable to enforce restrictions, Commons survival will crucially depend on whether informal institutions characterizing a society (culture, sense of belonging to a community) are strong enough-

An appeal to conscience […] (can produce) two communications, and that they are contradictory: (i) (intended communication) "If you don't do as we ask, we will openly condemn you for not acting like a responsible citizen"; (ii) (the unintended communication) "If you *do* behave as we ask, we will secretly condemn you for a simpleton who can be shamed into standing aside while the rest of us exploit the commons." (Hadin, 1968) → in these situations, social (i.e. solidaristic) values have to prevail in order to make things work.

Intuitive graphical explanation of the previous problem's result:



10. ASYMMETRIC INFORMATION (3RD MARKET IMPERFECTION)

In purely competitive markets all agents are fully informed about traded commodities and other aspects of the market. Imperfectly informed markets with one side better informed than the other are markets with asymmetric information. Examples:

- A doctor knows more about medical services than does the buyer.
- An insurance buyer knows more about his riskiness than does the seller.

In what ways can asymmetric information affect the functioning of a market? 2 problems arise:

- Adverse selection
- Moral hazard

And two remedies can be implemented in order to mitigate the aforementioned problems:

- Signalling
- Incentives

10.1 Adverse Selection

Let's start with the Adverse Selection issue (Akerlof, 1970). Consider a used car market: 100 people want to sell their used car; 100 people want to buy. There are two types of cars: "lemons" (bad cars) and "plums" (good, reliable cars). Everyone knows that 50 of the cars are plums, 50 are lemons. Each lemon seller will accept $1,000; a buyer will pay at most $1,200. Each plum seller will accept $2,000; a buyer will pay at most $2,400.

Now, suppose that the quality of the car can be verified ex-ante. If every buyer can tell a plum from a lemon, then lemons sell for between $1,000 and $1,200, and plums sell for between $2,000 and $2,400. Hence, gains-to-trade are generated when buyers are well informed.

Now, let's suppose that buyers are not that accurate and cannot neither verify nor recognize the quality of the car before buying it. Every buyer will have to guess which kind of car it is purchasing. If we suppose that buyers are risk neutral, in this case the buyers have to guess about how much each car is worth. We'll make a simple assumption about the form that this guess takes: we assume that if a car is equally likely to be a plum as a lemon, then a typical buyer would be willing to pay the expected value of the car. Using the numbers described above this means that the buyer would be willing to pay: (1/2)* 1200 + (1/2)* 2400 = $1800.

But who would be willing to sell their car at that price?

- The owners of the lemons certainly would
- The owners of the plums certainly not…they need at least 2000 to part with their cars.

At 1800 only lemons would be offered for sale.

Conclusions: No plums are sold, Consumers soon realize that only lemons are sold, thus they'll revise their expectations (lemons prob. = 1; plums prob. = 0)→ Lemons are the only cars sold: the equilibrium price will be somewhere in-between $1000 and $1200.

Even though the price at which buyers are willing to buy plums exceeds the price at which sellers are willing to sell them, no such transactions will take place.

Note: In the used car market, producers do not choose endogenously whether to have "plums" or "lemons". But note that adverse selection may also occur when producers do have the option to produce low-quality or high-quality goods and face different costs for producing the two.

10.2 Moral Hazard

If adverse selection is a problem of hidden information, whenever we talk about moral hazard, we are underlining a hidden action problem. Moral hazard in a transaction occurs when the party with more information about its actions or intentions has a tendency or incentive to behave inappropriately from the

perspective of the party with less information. Let's think of an example on bike' s insurance. If one has full bike insurance, what are the probabilities that the buyer will leave the buy unlocked (i.e., not caring)?

- No insurance → Consumers have the maximum incentive to take the maximum possible amount of care (large expensive locks, avoiding go in high-risk areas, etc.)
- With insurance → Consumers have no incentive to take any care. In case of theft, they can get money from the insurance company and buy a brand-new bike. This lack of incentive to take care is what is called MORAL HAZARD.

Note: Moral hazard is a *hidden action* problem: if the amount of care is observable, there would be no problem. The insurance company faces a trade-off: full insurance (higher immediate returns) means too little care will be undertaken by ensured (higher risk of bearing great costs later). How can this problem be solved? By incentivizing the insured to take some minimum level of actions in order to take care of the object, for instance by letting her face part of the risk (i.e., through bearing costs → "deductible").

Where does the market inefficiency lie here? The possibility that moral hazard might occur can lead to less trade than the optimum → Consumers may want to buy more insurance, and the insurance companies would be willing to provide more insurance if the consumers continued to take the same amount of care . . . *but this trade won't occur because if the consumers* were able to purchase more insurance, they would rationally choose to take less care!

10.3 Remedies to market asymmetric information

Signalling: Adverse selection is an outcome of an informational deficiency. What if information can be improved by high-quality sellers signaling credibly that their goods are high-quality? E.g., warranties, professional credentials, references from previous clients, etc.

	Lemons	Plums
Willingness to pay	1200	2400
Willingness to sell	1000	2000

Let's suppose sellers would offer buyers a warranty of 1500. They know that their plums will not break down so easily, thus they will actually will not disburse anything in addition. On the other hand, lemons' sellers will bear huge amount of costs because of their goods breaking to fast. In this case, warranty plays a fundamental signalling effect and will allow plums' sellers to distinguish themselves from the others.

Example: Signalling (Spence, 1974, 'Job Market Signalling')

A labor market has two types of workers: high-ability and low-ability. A high-ability worker's marginal product is a_H. A low-ability worker's marginal product is a_L. Note: $\boldsymbol{a_L < a_H}$. A fraction h of all workers is high-ability.1 - h is the fraction of low-ability workers. Each worker is paid his expected marginal product. If firms knew each worker's type, they would:

- pay each high-ability worker wH = aH (wH = high ability worker's wage)
- pay each low-ability worker wL = aL.

If firms cannot tell workers' types then every worker is paid the (pooling) wage rate; i.e. the expected marginal product: $w_P = (1 - h)*a_L + h*a_H$.

As long as the good and the bad workers both agree to work at this wage (and high ability workers maintain their high level of productivity) there is no problem with adverse selection in the eyes of the firm. But suppose the firm fears that the pooling wage scheme may depress high-ability workers and be detrimental to its profit and therefore it is very much willing to look for a *separating equilibrium*, i.e., → $w_H = a_H$ and $w_L = a_L$

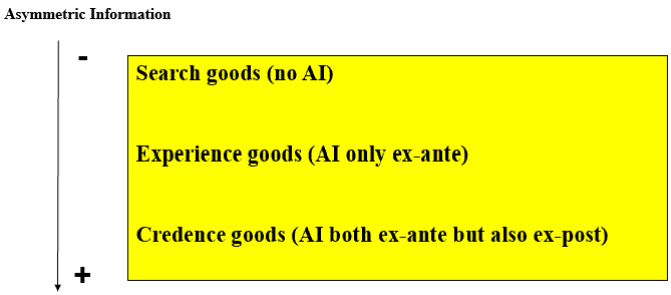
How can this separating equilibrium be achievable? If w_H and w_L is the wage scheme offered, how is it possible to select the "right" types of workers for the "right" wage? How is it possible for high-ability workers being selected for w_H and not w_L? How can they credibly demonstrate their nature? → Through an action (signal) which consists in a costly method available to signal you are a good type such that the cost of signaling is too high for bad types that they won't do it. Hence, it must be a mechanism that has to be more costly for bad

workers than for good workers. Workers can acquire "education" as signal. Education costs a high-ability worker c_H per unit and costs a low-ability worker c_L per unit. → $c_L > c_H$. Suppose that education has no effect on workers' productivities (extreme assumption, just to make the point and make it simple, obviously this should not be necessarily true in real life): i.e., the cost of education is a deadweight loss.

High-ability workers will acquire e_H education units if
(i) $w_H - w_L = a_H - a_L > c_H e_H$ → acquiring e_H units of education benefits high-ability workers, since the cost to purchase e units of education is less costly than the gain they obtain (wH – wL).
(ii) $w_H - w_L = a_H - a_L < c_L e_H$. → acquiring e_H education units hurts low-ability workers, thus not having any incentive to purchase it.
Both equations must hold.

$a_H - a_L > c_H e_H$ and $a_H - a_L < c_L e_H$, which together will require $\frac{a_H - a_L}{c_L} < e_H < \frac{a_H - a_L}{c_H}$.

Acquiring such an education level credibly signals high-ability, allowing high-ability workers to separate themselves from low-ability workers.
Food for thoughts: How much education should low-ability workers acquire (in this very simplified framework)? Zero. Given that acquiring e_H units is not convenient for them, they will be paid $w_L = a_L$ in any case, even if they acquire a lower level of education (say e_L).
Further Note: Signaling can improve information in the market. But it involves a cost, and for this reason, the equilibrium is sub-optimal with respect to a full information scenario. If the pooling equilibrium is accepted by high-ability workers (and does not affect their high productivity) total output does not change. Education is costly so signaling may hurt market's efficiency: to some extent resources are wasted in order to enforce a separating equilibrium.
What is the role of the Government in signaling?



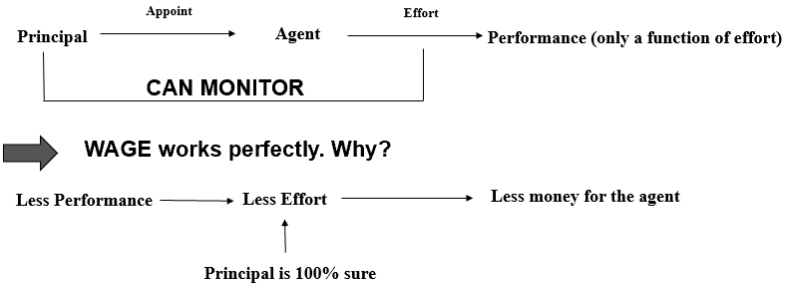
When we dealt with externalities, government had a very invasive role (if not crucial to solve externalities). Here, the signaling actions are put in place by private players, thus the government plays a very limited role, but not absent in case of one exception (credence goods). A further way to distinguish goods next to the non-excludability and non-rivalry in consumption concerns the level of symmetric information they imply 3 typologies can be distinguished on this regard.
Search Goods, in which no information asymmetry is involved.
Experience Goods, where information asymmetry is present only ex ante.
Credence goods, in which information asymmetry is present even after having bought the good, i.e. it is never solved.
Some examples:

a) A pen is a search good. As buyer you are perfectly aware of what you are buying. You do not suffer of any information asymmetry.
b) A car might be an example. Before you are buying it (second-hand) you do not know exactly the qualities, but after having bought, you actually experience whether there were hidden problems or not.
c) A medical mask: You'll never know whether the mask's characteristics will actually prevent the user to get sick or not. As buyer you will not have the competences to evaluate it, thus I will demand the government to check the quality on my behalf.

The use of incentives is particularly suitable to prevent asymmetric information. The principal-agent theory is the area of economics that deals with all situations where there is a principal who wants an agent (hiring) to act in the principal's interest to achieve some goal. When the principal possesses less information than the agent, the two have different objective functions, principal cannot monitor perfectly agent's behavior (i.e., cannot assess whether the agent is behaving 'optimally' to achieve the desired objective) and resulting

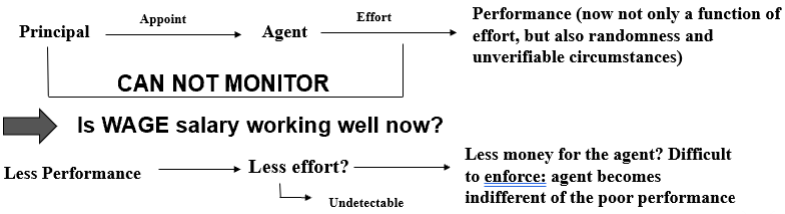
performance of agent's action is noisy (it may clearly depend on the actions put in place by the agents but not only on these. It can be the result of a random component that is exogenous to the agent's behavior, i.e. beyond the responsibility of the agent). In such situations, the agent can indulge in moral hazard, but proper incentives set by the principal, by re-aligning objective functions between the two, can help mitigate the problem. Let's analyze the following two scenarios:

Full info scenario



In case of full information scenario, wage will work properly. The principal knows whenever the agent is not committing 100% to the job → in such case, wage will be adjusted consequently without any problems.

Hidden action scenario



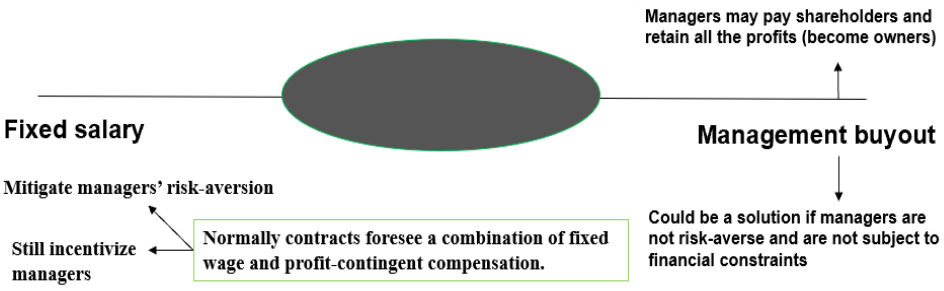
Here, the principal has no clear idea whether the poorer performances are given to the actual smaller effort put by the agent or due to exogenous variables.

Solution: **Flexible remuneration schemes**.

How can the principal mitigate the problem? Choosing another form of remuneration, making the "agent" participate to the sharing of the performance (i.e. part of the agent's salary depends on the final performance).

Sharecropping in agriculture: Stock-options for managers in public companies

Many corporations are owned by many small shareholders. This considerably decreases the incentives for each single shareholder to monitor the behavior of managers: the effort necessary is very high compared to the benefit. Board of directors only partly defend shareholders, and managers are better informed anyway.



10.4 Asymmetric Information: Some remarks

Problems (and partial remedies) in the market for finance of innovation (especially for start-ups), where the imperfections are highly evident:

→ Adverse Selection in the market for finance of Innovation: There are good and bad innovation projects (bad projects = high risk of failure); Those who provide external finance (i.e., banks) cannot perfectly discern good vs. bad projects, and generally proponents are much more informed about odds of success. To shield from the risk, banks pose average (unfavorable) conditions for lending to everybody (i.e., high interest rates or credit-rationing). Only (often low-skilled and high-risk loving) "kamikaze" innovators will ask for money while capable innovators (those with sufficiently good quality but at the same time realistic projects) may

prefer to give up searching for external debt finance (discouraged borrowers) or search for other alternative financing sources. Good innovation projects may risk not to be financed

→ Moral hazard in the market for finance of innovation: For an outsider (external investors) may be difficult to monitor the strategies, choices, decisions taken by the prospective innovative managers/entrepreneurs. Managers/entrepreneurs may put in place actions not in the interest of the investors, or that investors would not have agreed (e.g., divert funds to different aims, exert less effort than required).
Signals & Incentives: Established firms may exhibit their track record, while innovative start-ups may show their goodness to external investors: Of course, these dimensions are not pure signals, but they may still have a signaling function:

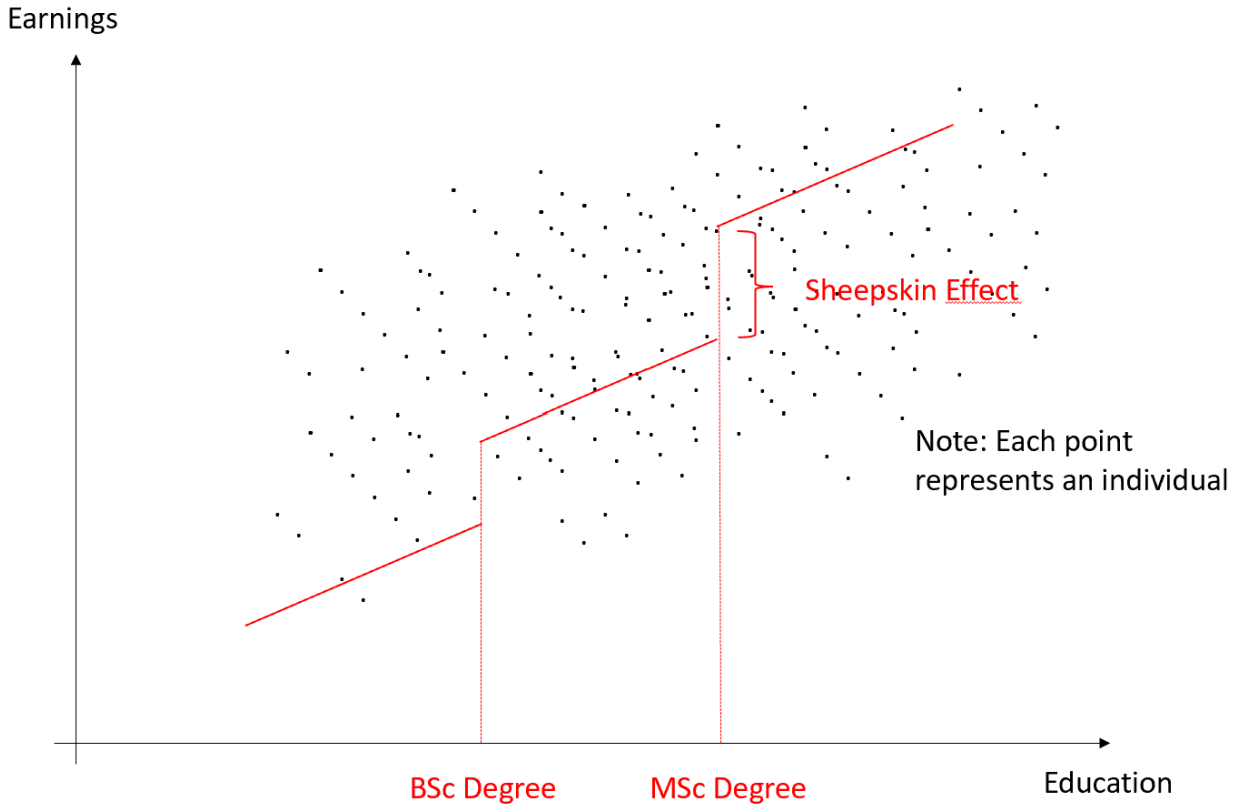
- Patenting (Hsu and Ziedonis, 2013, SMJ) → You have a patent it might mean your project is not that bad
- Endorsement by a reputable alliance partner (Stuart et al., 1999, ASQ; Stuart, 2000, SMJ) with whom the firm might have stipulated an alliance

Banks: Use of collateral to secure debt → Incentive, that leverages on the fact that the bank would seize the collateral in case the debtor would not repay the debt back.

VC: "active investors" (i.e., work side by side with entrepreneurs), co-investment with entrepreneurs as a guarantee of high effort, use of milestones (staged rounds of financing).

1st Remark: This might be all mechanism to correct partially market imperfections. But, in this specific domain, they cannot be considered enough. Think of the collateral case: If an innovator does not possess anything how would it work? That is why usually the intervention of the government which plays the role of fixing market imperfections that target specifically innovators.

2nd Remark: We have seen in the Spence's model that education might be used for signalling. The more you are educated the more you can earn in the labour market in terms of income, i.e., the higher the salary you can pretend. → Graduate overeducation as a sheepskin effect (takes its name from the fact that the parchment is printed on sheepskin):



As it can be observed, the scatter plot shows a positive relationship between education and earnings. But such relationship does not look continuous. Indeed, there are some jumps involved in such relationship. These jumps are recognized as Sheepskin effect. If you think about it, the day before discussing the master thesis you are equally competent as the moment in which you receive the degree, BUT the only fact of obtaining it works as signalling effect to the labour market.

3rd Remark: Signalling and Incentives do help but do not replicate fill information**.** Signaling can improve information in the market. But it involves a cost, and for this reason, we expect the equilibrium to be sub-optimal with respect to a full information scenario. In the Spence's model, education is costly for high-ability workers (they would have preferred to be recognized as such without the need to "signal" instead of having forced to invest in education) so signaling may hurt market's efficiency: to some extent resources are wasted to enforce a separating equilibrium.

A similar logic can be applied to incentives. They can point the objective function of the agent towards the one of the principals, but they are not able to "transform" the agents' objective function exactly into the one of the principals.

This result can be shown analytically as well. Incentives mitigate but do not solve the problem (OPTIONAL)

Principal: output y = f(x) [p = 1 → output = value; x: agent's effort]

Agent: payment s(y)

Principal's $\pi = y - s(y) \rightarrow \pi = f(x) - s(f(x))$

Agent's u: $s(y) - c(x) \rightarrow s(f(x)) - c(x)$

The agent will be willing to «work» for the principal as long as its $u \geq 0$

Principal's problem $\underbrace{max}_{x}\ \pi = f(x) - s(f(x))$

s.t. $s(f(x)) - c(x) = 0$

$\underbrace{max}_{x}\ \pi = f(x) - c(x)$ MP(x*) = MC(x*)

Optimal incentive scheme: Marginal product of Agent's effort = its marginal cost

If principal can observe the amount of effort exerted by the agent, wage works perfectly:

Agent's problem $\underbrace{max}_{x}\ u = wx + K - c(x)$ w = MC(x) principal sets w = MP (x*).

But if principal cannot observe x and y = f (x, ε), w is highly inefficient since received w, the agent has an incentive to shirk (e.g., ↓x since ↓c(x)).

In these cases, «sharecropping» or similar incentivizing methods (Agent gets s = αf(x) and Principal (1- α) f(x)) might be preferrable. But bear in mind that they are not optimal compared to the full information scenario:

Agent's problem $\underbrace{max}_{x}\ u = \alpha f(x) - c(x)$ $\alpha MP(\underline{x}) = MC(\underline{x})$ with $\underline{x} \neq x^*$

Incentives as "signals": If signals are exogenously absent, firms may always endogenously look to put in place some incentive schemes in order to make workers' type reveals ("signals") themselves.

Example: Zappos founded in 1999 in Las Vegas for selling shoes on-line. Management team: "customer obsessed culture", i.e., customer is king (or queen, as we prefer). Customer service: key asset. 365 day-return window, free-shipping, but in team's view the Call centre 24/7 with no limits talking ("protracted talk therapy", as one observer noted) would have been their core advantage. A call-centre job isn't typically very desirable, nor does it pay well (e.g., 11$ per hour) …. How could they obtain that? Paying workers more was not an option. They pursue their objective in two ways:

1. Making the call center a more pleasant place where to work in.
2. The “offer”: Once the new workers have been trained for one month, the company offered the possibility either to stay and continue the job or they offered money (2000 dollars) to quit the job. They created a sort of signal, which underlined if a worker was a good one or not. Not accepting the offer signaled indeed that the worker was keen to work and that her objective function was aligned to the one of the companies.
Asymmetric Information are actually decreasing thanks to the technological developments: Internet, IoT, Big Data and so on. But at which price from a societal point of view? If I own a restaurant, for instance, I am not willing to be ranked. Or maybe, I do not want that everyone knows about my health problems. Thus, applying this technology to reduce market imperfections comes at a cost: Being always under exam, thus relinquishing privacy.

11. TRANSACTION COSTS

Simon’s parallelism (1991, “Organization and Markets”)

Suppose that a mythical visitor from Mars approaches the Earth from space, equipped with a telescope that reveals social structures. The firms reveal themselves, say, as solid green areas with faint interior contours marking out divisions and departments. Market transactions show as red lines connecting firms, forming a network in the spaces between them. As our visitor looked more carefully at the scene beneath, it might see one of the green masses’ divides, as a firm divested itself of one of its divisions. Or it might see one green object gobble up another. No matter whether our visitor approached the United States or the Soviet Union, urban China or the European Community, the greater part of the space below it would be within the green areas, for almost all the inhabitants would be employees, hence inside the firm boundaries. Organizations would be the dominant feature of the landscape. *A message sent back home, describing the scene, would speak of "large green areas interconnected by red lines." It would not likely speak of "a network of red lines connecting green spots.”*

11.1 The Transaction Cost Economics (TCE)

In the neoclassical vision of markets, the exchange of goods in the market is without frictions or costs. If markets are perfectly competitive, the economic system reaches Pareto optimal allocations.

The Transaction Cost Economics (TCE) doubts the existence of these “perfect” markets: transactions are not instantaneous; *the use of markets is costly (no frictionless and costless transactions*). The use of the market may imply transaction costs which can differ: time & money to search for sellers & buyers, negotiate exchange terms, write *contracts*, inspect results, enforce deals. When these costs become too big, market may become sub-optimal.

In many circumstances, market is not the best economic organization to govern a transaction and alternative governance structures can be preferred.

The TCE lies on 3 fundamental pillars: When all these three conditions are in place simultaneously, market becomes costly. So, if one of these pillars do not present itself, market turns out to be efficient. These are:

- Bounded rationality
- Opportunism
- Relationship-Specific Investments

1) Bounded Rationality:

Boundedly rational agents experience limits in formulating and solving complex problems and in processing (receiving, storing, retrieving, transmitting) information“. “Does bounded rationality mean that people (and therefore their actions) are irrational? Not at all. People making choices are intendedly rational. They want to make rational decisions, but they cannot always do so”

In decision making, rationality of individuals is limited by:

- the information they have: the economic agent does not perfectly know the present state of the nature and he is not able to forecast all the possible states of nature on which he will finds himself after a decision (substantial limit caused by environmental uncertainty).
- the cognitive limitations of their minds: even if it is possible to know the present and forecast all the possible future states of nature, there is not an algorithm that enables to find optimal solutions (at least) in a reasonable time. There is a "cognitive" and "time" problem (procedural limit). *Agents rely on heuristics and approximations for their behavior.* Agents do not look for optimal solutions but rather for "satisfying" ones. If a given strategy has proven itself to be successful in each scenario, agents will tend to reply that strategy. In doing so, they will reduce costs.

Concrete Example for Substantial limit: Poker
The player must play but he does not know the other players' cards and the sequence of the cards in the deck. He has a very limited information set. He has to make a move, but he does not know present and future states of nature
Concrete Example for procedural limit: Chess
It's more like chess game, players have good knowledge on all possible states of nature (rules are settled, distribution of figures in the chess is known, other's player move perfectly observable), but it is very difficult to find optimization's algorithms (sometimes even with computers).

In the TCE perspective the bounded rationality implies two important facts.

If 2 or more parties want to regulate a transaction by a contract:

- It is very difficult to include in the contract all possible details of interest for the object of the contract: they cannot include every likely (and less likely) circumstance in the present and in the relevant future that is of interest.
- Even if this could be possible, the parties are not able (or it would be too costly) to negotiate everyone of this single detail. And if this could be possible (but it is not), details should be written down in each language, which whatever it is, offers some degrees of imperfection. Language is ambiguous. It is not mathematics

Thus, Contracts are incomplete: the parties of a contract have and can exert a certain degree of discretion over the fulfilment of contractual clauses.

→ Bounded rationality thus leads to contract incompleteness (Note: there are different levels of incompleteness)

2) Opportunism

(Some) Economic agents: Pursue <u>selfishly</u> their own utility and interest even if this is detrimental to the utility of interest of someone else
Note: Not all economic agents are opportunists, but it is enough that some of them behave opportunistically to incur in additional costs of transactions. Ex-ante it is not possible to distinguish between opportunistic and not opportunistic economic agents. Opportunistic behaviors can appear during the writing of the contract and ex-post contract stipulation (*Post-Contractual Opportunism*<u>)</u>. What does the latter point mean? After having agreed on the contract I might notice that there are some points that omit specific circumstances that might personally beneficial, thus I remain opportunistically silent.

3) Relationship-Specific Investments (Non-redeploy-able investments)

Investments that are specific to the relationship between the parties and to the nature of the counterparts

Examples (among the myriad possible).

- A supplier working for FIAT Chrysler specializes itself in the production of a car-component that can be sold to FIAT Chrysler only (Dedicated/Physical Asset).

- A supplier locates the plant close to one single specific customer firm to facilitate just-in-time processes, but there are no other firms in the surroundings to whom the good can be sold (Sites-specificity).
- A firm acquires a software and trains its employees for its usage, while this type software is sold and its usage taught only by a specific software house.
- I am invited to hold lectures in the small island of Neverland, if I accept, I have to translate all my slides and learn how to speak Neverlandish.

The 3 most important types of relationship-specific investments:

- Dedicated/Physical asset specificity (supplier-buyer relationship along the supply chain)
- Site specificity (high transportation costs)
- Human asset specificity

Dedicated/Physical asset specificity refers to assets whose (physical or engineering) properties are specifically tailored to a particular transaction.

- Fiat example of before (supplier has to invest in machines who are specific for FIAT cars)
- Glass container production requires molds that are custom tailored to particular container shapes and glass-making machines.
- Ports investing in assets to meet the special needs of some customers

Site specificity refers to assets that are located side-by-side to economize on transportation or inventory costs or to take advantage of processing efficiencies.

- Cement factories are usually located near limestone deposits
- Can-producing plants are located near can-filling plants
- Steel manufacturing: side-by-side location of blast furnaces, steelmaking furnaces, casting units, and mills saves fuel costs (no need for re-heating)

Human Asset Specificity:

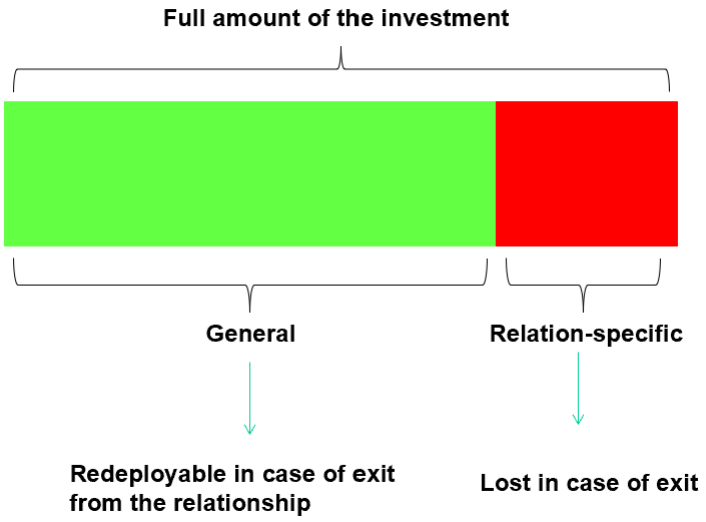
Some of the employees of the firms engaged in the transaction may have to acquire relationship-specific skills, know-how and information:

- A manager who has become a skilful administrator within the context of one organization's routines may be less effective in an organization with completely different routines.
- Workers acquire the skills to use a particular enterprise resource planning software

About these relationship-Specific Investments, we have to make two considerations:

First, there are no all black and white investments. Some might be partially non-re-deployable as shown below.

Second, most of the problems reside in the circumstances in which one party must make larger specific relationship investments than the. When there is a misbalance in such sense, i.e. red and green area are not equally shared, problems might arise. The one with the smaller red area will try to "take by the neck" the other party, which is the one making most of the relation-specific related investment.



The relation-specific part of the investment (red area) is not recognized outside the relationship (non-redeployable investment) → The more this red area is large the more a firm would be reluctant to exit from the relationship (once entered) → The more asymmetric these red areas are between parties, the more the firm facing low relationship-specific investment may "exploit" the firm facing high relationship-specific investment.

11.2 The hold-up problem: Situation in which to run the transaction one party must bear most of the relation-specific costs.

- Suppose your company contemplates building a factory to produce cup holders for Ford automobiles. The factory can make up to 1 million holders per year at an average variable cost of *C dollars per unit.*
- The construction of your factory is financed with a mortgage from a bank that requires an annual payment of *I dollars. The loan payment of I dollars thus represents your (annualized) cost of investment* in this plant. *I* is an unavoidable cost: You have to make your payment even if you do not do business with Ford.
- You will design and build the factory specifically to produce cup holders for Ford. Total cost of making 1 million cup holders is thus:

I + 1,000,000C dollars per year.

Of course your *expectation is that Ford will purchase your holders at a profitable price (you will make non negative extra-profits out of this business).*

- But suppose that you also want to consider outside options (i.e. Market). The "market price" you can expect to get from selling your cup holders is *Pm.*

Suppose that:

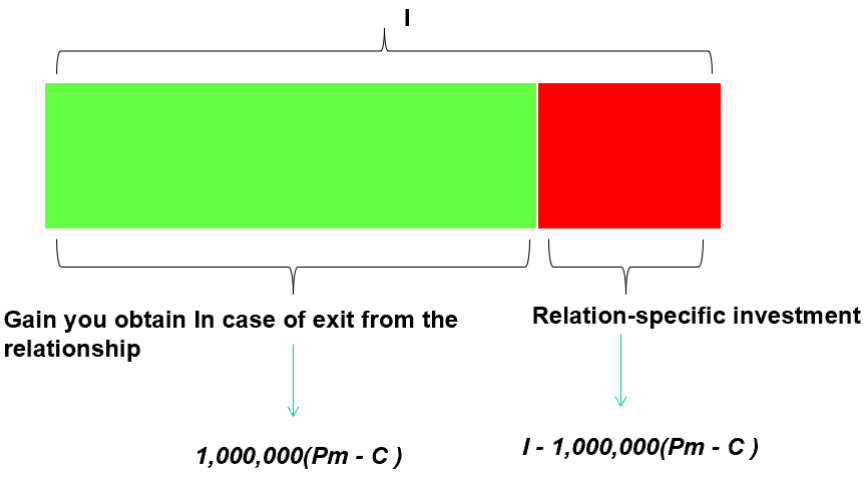
- *Pm > C* (Variable Costs).
- *Variable profits (without considering fixed cost I):* 1,000,000(*Pm - C) > 0*
- But that the annual investment cost *I > 1,000,000(Pm - C).*

Which implications do arise?

1) We know that producing and selling to FORD is the best option since the price Ford will pay us is profitable
2) Since *I* is an unavoidable cost (has to be repaid) once in the market you would still keep producing and selling cup holders even if the relationship with FORD breaks down: <u>selling to the market is better than not producing at all</u>.

The Green area will determine the profits that you will obtain despite existing the relationship, whereas the red area are relationship investment costs.

Suppose that *I = $8,500,000, C = $3, and Pm = $4, then the RSI* (Relation-specific Investment) *is $8,500,000* - 1,000,000(4 - 3) = $7,500,000.



Of your $8,500,000 investment cost, you lose $7,500,000 if you do not do business with Ford and sell to the market. Thus, the red area is way larger than the green area (Note: it is not the picture depicted on the rights).

Rent: Now suppose that before you take out the loan to invest in the cup holder plant, Ford agreed to buy 1 million sets of cup holders per year at a price of *P* per unit.* You will accept these terms as long as extra-profits are non-negative: *1,000,000(P* - C) – I ≥ 0 which implies that P* > Pm*

This is what this stream of literature refers to as "Rent": the rent is simply the profit a company expect to get from a contractual relationship if everything goes as planned.

But is Ford willing to exploit after the contract the existence of that red area? And if yes, to what extent?

To understand if and to what extent (and so the magnitude of the holdup problem) one has to define the concept of quasi-rent, which consists in *the extra profit that one gets if the deal goes ahead as* planned, versus the profit one would get if he had to turn to his next-best alternative (in our case Market)

[1,000,000(*P* - C) - I]{Rents} - [1,000,000(Pm - C) - I] {Next best option} = 1,000,000(P* - Pm)* (The latter term is the Quasi rent)

Suppose your company stipulate the contract with Ford with P = $12 per* unit, and like before *Pm = $4 per unit, C = $3 per unit, and I = $8,500,000.*

At the original expected price of $12 per unit, Rent is: (12 - 3)1,000,000 -8,500,000 = $500,000 per year.

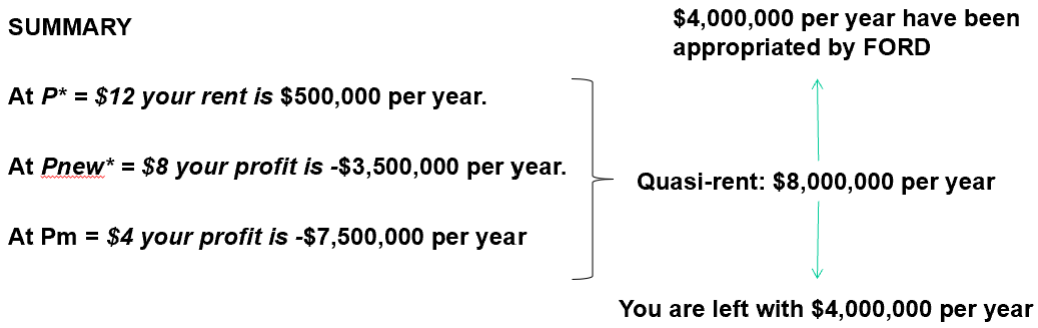
The contract leads you a positive profit, so you think "let's do it". But…… After contract signed, the quasi-rent is (12 - 4)1,000,000 = $8,000,000 per year.

Ford wants that quasi-rent. Ford would like to renegotiate the contract down to let say, *Pnew* = $8.*

→ If you break up the relationship with Ford your profit is: *[1,000,000(Pm - C) - I] = [1,000,000(4 - 3) - $8,500,000]= -$7,500,000.*

→ If you stay in the relationship with Ford (even with deteriorated conditions) profit is: *[1,000,000(Pnew* - C) - I] = [1,000,000(8 - 3) - $8,500,000]= -$3,500,000*

Thus, you decide to stay into the relationship since -3.5 mio. > -7.5 mio. , despite being still at loss



Of course, Ford, if it could, would not stop at Pnew* = 8$ but it will fix the minimum possible price given that you are still willing to serve Ford rather than the market. In other words, Ford, if had the possibility to do that, would fix a new price such that Pnew* = Pm + ε = 4 + ε where ε is sufficiently low. → The fact that Ford is willing to behave in this way, i.e. "opportunistically", has a great incentive to do that, does not mean that it is capable to do it. → "We have signed a contract, with precise obligations from both parties…… we are safe……we are sure that FORD cannot behave in this way". But are we really sure?

I am afraid you cannot be sure…… → CONTRACTS ARE INCOMPLETE (and thus potentially ambiguous).

Ford could assert that, in one way or another, circumstances have changed and that it is justified breaking the contract.

It might, for example, claim that competition for that particular model of cars you are producing cup holders for is facing fierce competition, and if terms with suppliers do not change (enabling FORD to settle lower prices for the final car) it will face bankruptcy risks.

Or it might claim that the quality of your cup holders fails to meet promised (potentially unwritten or ambiguous) specifications and that it must be compensated for this lower quality with lower prices.

You may want to consider the possibility to fight Ford in court for breach of contract.

1) This is itself a potentially expensive move
2) Contractual incompleteness leads almost by definition to ambiguity in contractual terms so the outcome of the trial (even if you are in principle right) can not be taken for granted!!!!! Actually, it can be highly uncertain.

Ford knows both 1) and 2) (that's why it has "held you up")

End of the story: You may be better off accepting Ford's revised offer than not accepting it (this of course also depends on the extent to which Ford wants to appropriate your quasi rents, i.e. to what extent Pnew* is far or close to P*) .

1st Note: This example makes evident how your market transaction with Ford can be problematic and inefficient by the presence of:

- Specific-relationship investment you have to make with Ford
- Bounded rationality that leads to an incomplete contract between you and Ford
- Opportunistic behavior by Ford

But is one of the three pillars do not hold, the contract does not pose any threats.

- Bounded rationality (No?: Then the contracts will be complete)
- Opportunism (No?: No need of complete contracts since Ford would not exploit its stronger position. It will not ask you to rediscuss the terms)
- Relationship-Specific Investments (No?: no need to defend from opportunism, because there is no loss in switching from a transaction to others, i.e. the investment we are incurring is "all green")

2nd Note: But if all pillars are working and you know that you may likely encounter all these problems in the relationship with Ford, would you ever be willing to enter into the market relationship at these conditions with Ford? And if anybody is not willing to do that who would provide cup holders to Ford? One possibility is that Ford produces cupholders itself (hierarchy). That is what the transaction cost economies approach

Given that markets are not always perfect and their use may entail substantial costs the main aim of the TCE perspective is to individuate for each transaction the best governance structure possible.

3 alternatives:

1)Market which may present costs (related to the use of contracts), i.e. market imperfection
2)Hierarchy (transactions are internalized and carried out inside the firm, costs arise from the administrative bureaucracy that internalize and monitor the exchanges)
3)Hybrid forms: This might be some equity alliances where A makes an agreement with B and buys also part of its shares to show its commitment to the agreement. The most efficient solution (the market vs. planned solutions) is identified by a Darwinian selection where the best option finally survives.

→ Note that from this perspective, firm is seen as an efficient response to a market failure.

→ The famous Fisher Body-GM case study is emblematic in this respect:

The dawns of Car Industry a the end of 1800 centuries. In this period the bodies of the cars were made out of wood through a craftsmanship work (artisans). Let's suppose we had GM asking to many, different suppliers among which Fisher-body, to produce 1 by 1 the body for the cars. These types of transactions were governed by short-term contracts. Thus, once a body is provided, the contract ends. Note, that no supplier

had to make any relation-specific investments for a given producer. At the beginning of the 20th century, the cars' market boomed. It moved from the necessity to construct cars' bodies in wood, to cars made out of steel. Now, to produce such bodies, suppliers must invest into new machineries that would be relation-specific to a car manufacturer since dedicated to a specific model (Dedicated/Physical asset to the specific relationship that was existing with specific car producers). Fisher-Body was asked to produce specifically bodies for GM. How to govern such transactions? From short-term to long-term (10 years long) contracts combined with equity alignments (15% of Fisher-Body's equity has been purchased by GM). But this was solely an intermediate situation. The car industry developed further (think of the Model-T produced by Ford). GM had to buy many other cars, it needed to speed up the production process. To do so, it asked Fisher-Body to locate its plants next to GM's ones. So, in addition to dedicated assets, Fisher-Body was asked to add site specificity to their relationship. In the time lapse between 1924 and 1926, GM took the decision to buy Fisher-Body and locate it as it was more convenient to it.

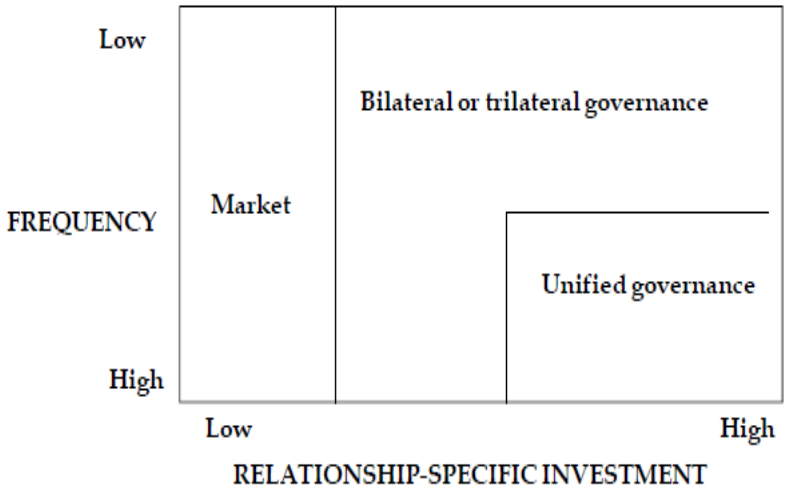
With this respect many scholars tried to classify the preferred ways to govern transaction according to two different characteristics of transactions:

1) Relation-specific investment needed for carrying out the transactions
2) Frequency of transactions

Note: If frequency of transactions is high along with high relationship-specific investments, it will be more likely to see a unified governance (as it happened at the end of the Fisher-Body – GM story)

Further note: Trilateral governance is referred to a joint venture.

An increase in uncertainty of transactions (which may affect the degree of incompleteness of contracts) is likely to lead to less hybrid forms.



Important from a managerial perspective:

1. Be able to spot areas of "contractual incompleteness": This does not mean to be opportunistic but rather to be able to defend ourselves from opportunistic behaviours of others
2. Organize transactions accordingly, to limit the negative backlashes of opportunism

12. INNOVATION AND TECHNOLOGY: THE SCHUMPETER LEGACY

12.1 Brief Taxonomy

Note: "*Talking about Innovation without knowing Schumpeter is like talking about football without knowing who Maradona is".*

So far, we have analysed a static situation, hence where production techniques were given and consequently goods and services to be traded. In this situation there is a market force that pushes towards the best social outcome (maximising social welfare) that should be pursued by policy makers, namely perfect competition → p = MC = AC min, where:

- The maximum quantity of the good produced at the lowest possible cost= Max Productive efficiency (goods and services are produced using the least costly combination of resources and technology)
- The maximum quantity of the good sold at the lowest price= Max Allocative efficiency (resources are dedicated to the combination of goods and services that best satisfy consumer needs)

But in real life things are complicated → dynamicity is involved. The domain of goods that can be traded is not fixed as we have assumed so far, but it changes over time. New goods are created, and old products become obsolete. He same reasoning applies to production processes. These are the outcomes of Dynamic Efficiency: We cannot evaluate an economic system according to the two static concepts seen so far, namely allocative and productive efficiency (which consider solely the Stauts QUo) without considering the lastly introduced concept. Perfect competition, thus, will not be the best outcome if we take into consideration dynamic efficiency. It is called dynamic since the innovation that will occur tomorrow is grounded in the decisions taken today regarding the investments that we are willing to incur.

"INNOVATION has become the industrial religion of the late 20th century's lives. Business sees it as the key to increasing profits and market share. Governments automatically reach for it when trying to fix the economy." (The Economist).

Innovation has become a sort of mantra that we should always followed. Classical and neo-classical economic perspectives do not consider innovation but rather concentrate on the flow of existing goods and services. Innovation is seen as a disturbance factor in their analysis. Something exogenous (random) that occurs to the economy which did not deserve any attention. But, instead of taking innovation as an exogenous factor, it should be controlled: Future Innovation shall be planned today.

The Lisbona Strategy (2000)

- "To become the most dynamic and competitive knowledge-based economy in the world"

Europe 2020 (2010)

- "A European strategy for smart, sustainable and inclusive growth"
- ✓ Research and innovation are placed at the center of the Europe 2020 strategy (Horizon 2020). This includes the headline objective of increasing spending on R&D to 3% of GDP by 2020.

R&D ⟶ **Growth (Sustainability, Inclusion, etc.)**

The Economic background (milestones). Solow, one of the great contributor to innovation economics, discovered that most of the dramatic growth that the US experienced in a specific scrutinized period was not given solely by capital and labour growth but majorly by something unexplained called "The Solow Residual", the rate of technical change.

$$\Delta Y/Y = \theta(\Delta K/K) + (1-\theta)(\Delta L/L) + \Delta A/A$$

- $\Delta Y/Y$: Rate of economic growth
- $\Delta K/K$: Rate of growth of Capital
- $\Delta L/L$: Rate of growth of Labor
- $\Delta A/A$: Technical change

What did Schumpeter came up with? Three concepts that are conceptually related but that should be kept separated:

- **Invention** (telephone):
  - (technological) materialization of an intuition which can be engendered by scientific knowledge. Might be exogenous and largely random in nature and can not be influenced by market signals. Can be a process of inspiration, perspiration, both of them, and also a result of "stumbling" over.
- **Innovation** (telecommunication service):
  - enabled by the invention of the telephone
  - transformation of an invention into a new product/service/process. → it takes time to implement: penetration rate did not jump from 0 to 100% instantaneously.
  - It addresses new needs (often latent needs) on the part of consumers or firms; influenced by economic signals (like pursuit of profit).

- **Diffusion** (telecom penetration)
  - spread process of an innovation into the economy

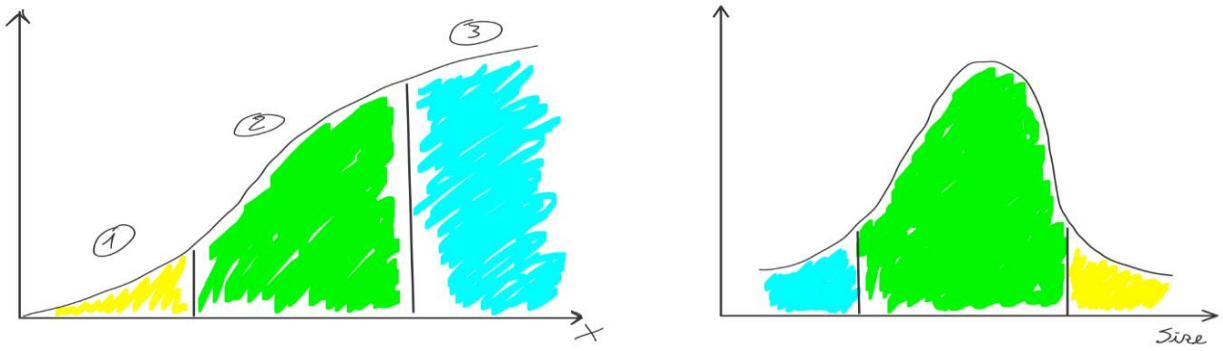
Example: Telephones became a great innovation in the moment in which they could be used to connect people that were geographically far away and not simply separated by a wall (as shown by the video of Mr. Bell calling Watson from the other room).

Note: It is always easy (*a piece of* cake) to understand ex-post where the invention can be implemented in order to become a successful innovation. Ex-ante, on the other hand, it is extremely difficult to identify the right commercial appeal to that invention, i.e., where could the telephone actually be exploited effectively.

Why does the diffusion of inventions not happen immediately? 2 reasons can be found in the roots of Schumpeter's legacy (despite not being found by him directly):

Innovation's penetration rate can be depicted as a logistic curve (pattern of innovation diffusion), in which 3 distinctive stages can be found:

1. The launch phase
2. Rapid growth
3. Maturity



Why these patterns? 2 possible explanatory reasons:

1. "Lack of incentives" story (Probit models of diffusion):

Economic agents are heterogeneous in terms of benefits that they can get by implementing the innovation in the market → Intrinsic characteristics of these economic agents that influence the adoption. They will do if and only if the benefits stemming from the implementation will be greater than the costs of Innovating. Note that it is assumed that the cost of adopting innovation decreases over time, hence $P_t < P_{t-1}$. If we think of the players that are willing to implement are normally distributed (Gaussian distribution) in terms of size and if we assume that the larger the firm, the larger the benefits that can reap from the market by implementing the innovation, at the very beginning the firms that will try to effectively innovate will be only the larger ones. Remember at the beginning the costs are higher than in later stages of the innovation process, thus only for these particularly large firms there will an actual economic sense to join the market: This explains a slow start (yellow part). Subsequently the cost for the implementation will start to decrease up until even medium-sized firms would see the point to join the market. Since assuming a normal distribution, the these will represent the large majority of the firms present in the market, thus the innovation's rate of penetration will increase rapidly (green area). The last step will consist in the last minority of small firms that have not yet joined the market because of too high costs of implementation. After the great majority of firms have adopted the new invention, the cost will be so small that even the remaining realities will benefit from joining the market.

2. "Lack of information" (epidemic models)

Here, the assumptions are that economic agents are not immediately perfectly informed about the existence of the innovations. The knowledge regarding this innovation is possessed solely by few firms. Hence, in the first stage there will be few mouths (those that know about the invention) that will spread the word and some ears that will listen to the spread words. In the second stage, some of the ears that listened carefully and understood (learned) the invention will start to spread the word as well. Thus, more agents will be spreading the innovation's existence → higher chance that more ears will carefully listen to them. The third stage is characterized by the vast majority of economic agents that are spreading the word an few that still have not but that are willing to listen (ears). These will be the last one adopting the innovation.

A further contribution of Schumpeter regards the taxonomy of innovations:

Product vs. Process Innovations

- Product Innovations refer to the creation of new goods and new services, e.g., DVD's and cell phones
- Process Innovations refer to the development of new technologies for producing goods or new ways of delivering services, e.g., robotics and CAD/CAM technology.

Radical (disruptive) vs. incremental Innovations (often PRODUCT)

- Technological breakthrough that might create new markets (radical)
- Small adjustments over existing products (incremental)

Drastic vs. Non-Drastic Innovations (PROCESS)

- Drastic innovations have such great cost savings that they permit the innovator to price as an unconstrained monopolist
- Non-drastic innovations give the innovator a cost advantage but not unconstrained monopoly power

Example: Suppose that demand is given by: $P = 120 - Q$ and all firms have constant marginal cost of $c = \$80$

- Let one firm have innovation that lowers cost to $c_M = \$20$
- This is a *Drastic* innovation. Why?
  - Marginal Revenue curve for monopolist is: $MR = 120 - 2Q$
  - If $c_M = \$20$, optimal monopoly output is: $Q_M = 50$ and $P_M = \$70$

Innovator can charge optimal monopoly price ($70) and still undercut rivals whose unit cost is $80: Thanks to this innovation the firm can kick out any other competitor and become a real monopoly

- If cost fell only to $60, innovation is *Non-drastic*
  - Marginal Revenue curve again is: $MR = 120 - 2Q$
  - Optimal Monopoly output and price: $Q_M = 30$; $P_M = \$90$
  - However, innovator cannot charge $90 because rivals have unit cost of $80 and could underprice it

Innovator cannot act as an unconstrained monopolist, i.e. it cannot pursue its best strategy (maximizing profits) but still can kick out competitors:

  - Best innovator can do is to set price just under $80 and supply all 40 units demanded.

Note: Schumpeter considered other innovations sources that we are not considering anymore: New input availability, Mergers and entry of a new firm.

12.2 Schumpeterian Marks

The contribution for which Schumpeter is mostly known is because of two opposite views that characterize the capitalist economy and have been labelled as:

Schumpeter Mark 1

Schumpeter identifies in innovation the engine of the economic growth in capitalist markets. The fundamental figure is represented by the entrepreneur, who is the synonymous of "Innovator". Entrepreneurship is not a DNA characteristic but rather an attitude, a state of mind. The paleo-

Schumpeterian entrepreneur: Individual capable of transform an invention into an innovation through an entrepreneurial act. What moves the Entrepreneur to innovate? 3 main rationales where economic does play an important part (but is not the only driver):

1. "The dream and will to found a private kingdom" → AKA, you want to gain profits
2. "There is the will to conquer: the impulse to fight, to prove oneself superior to others, to succeed for the sake, not for the fruits of success, but of success itself. From this aspect, economic activity becomes akin to sport" →To be recognized by other peers as a man who "made-it", who had success (measured by profits and economic wealth)
3. "Finally, there is the joy of creating, of getting things done" → Hedonistic motive (engaged in the pursuit of pleasure)

Since these incentives are so rooted in economics, in order to have innovation in an economic system there should be an expectation of supra normal profits that one can obtain from these eventual innovative efforts. In order to ensure these supra normal profits the innovation should basically help the innovator to achieve these profits without feeling the risk to be imitated by others → Absence of these extra-profits as a pay-off of innovative entrepreneurial activity = low innovation rate, low incentive to innovate. So, in order to stimulate innovation there should be a mechanism in place that protects the innovator and enable him to reap the benefits from the innovations → Patents do play a role on this, even if Schumpeter never put a strong emphasis on IPR protection (He emphasizes the "first mover" advantages).

Over time the "entrepreneurial profit" of the innovator is eroded by competition from imitators. The process of "creative destruction", which consists in innovative ideas that destroy the status quo market: competition through innovation results in high level of dynamic efficiency (and high social mobility). It is destructive because it kicks out incumbents. → Innovation stems mainly from new firms (non-existing firms, startups). Entrepreneurship can be seen as a proxy for "social elevator".

Schumpeter Mark 2

It relies on a very "late" Schumpeter (moved from Europe to the US finding a completely different context with respect the European one) → New vision which has been influenced by the context in which he was immersed in. → High innovative performance of the capitalist system, characterized by the presence of large oligopolistic companies which has reduced innovation to "routine" with no need of "leadership". Since innovation became routine (assimilable to any function in a firm), large firm will be thus enjoy a better position. Key role of large scale techno-structures (R&D labs) in the innovative process: R&D benefits from economies of scale. Capital market imperfections are particularly suffered by new firms, thus innovation mainly stems from big companies (in concentrated markets), who have the capabilities to overcome such imperfections.

"Innovation itself is being reduced to routine. Technological progress is increasingly becoming the business of teams of trained specialists who turn out what is required and make it work in predictable ways. The romance of earlier commercial adventure is rapidly wearing away, because so many more things can be strictly calculated that had of old to be visualized in a flash of genius."

"Thus, economic progress tends to become depersonalized and automatized. Bureau and committee work tends to replace individual action […]. Rationalized and specialized office work will eventually blot out personality, the calculable result, the "vision. The leading man no longer has the opportunity to fling himself into the fray. He is becoming just another office worker—and one who is not always difficult to replace."

13. INNOVATION AND TECHNOLOGICAL CHANGE: MARKET STRUCTURE AND INNOVATION

Three questions pop out as a consequence of the aforementioned discussions:

1. Which market structure is more conducive to innovation (to have a good innovation rate)?
2. How much do we want innovation to affect the market structure?

3. Which role should public policy play?

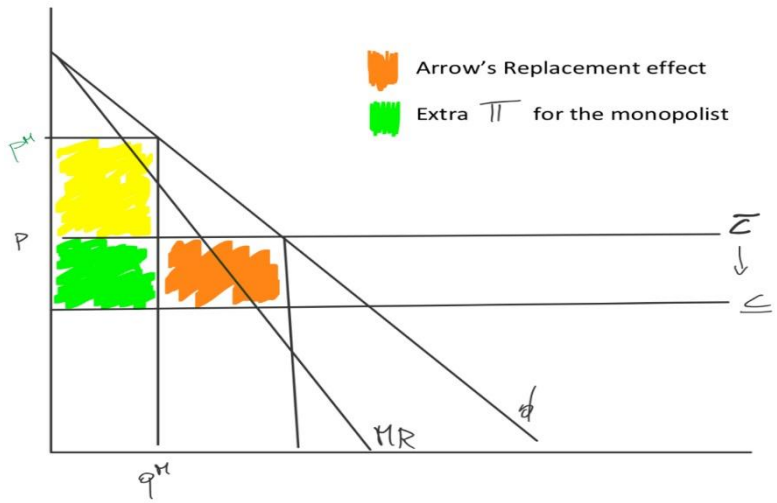
13.1 Which market structure favours innovation

Two different views: Schumpeterian hypothesis against Arrow's view.
The first one (SM1) can be synthetized with the following words: "Monopoly deadweight loss is the price that must be paid for high level of innovative activity going on the economic system" → Monopolies are fundamental (necessary evil) in order to have dynamic efficiency. If we want to dig in a little bit more, within the Schumpeterian view we have to distinguish two different monopolies' powers:

1. **Anticipated monopoly power** refers to an innovator's ability to enjoy the full benefits of its research by preventing imitation (SM1). To achieve high levels of innovation we need to ensure the innovator the possibility to reap out the benefits from its actions (not debatable). What can be discussed on is actually the way through which the situation can be ensured, i.e. Intellectual Property Rights
2. **Actual monopoly** refers to the current monopoly position of a firm. This position can have (positive) direct/indirect effects on innovative activity (SM2), where these advantages that the monopoly can possess can distinguished in:
    - **Indirect** consists in: "A current period monopolist will generally be well placed to erect barriers to future entry, and that the entry barriers which sustain a current period monopoly position may be durable enough to protect the monopolist into the future. Further, whenever the results of future innovations complement innovations already made by the monopolist, then it will gain more than rivals will from their introduction
    - **Direct**: "Monopolists may be possessors of various types of material advantages. In particular, high current period profits generated by the exercise of actual monopoly power may enable a monopolist to hire more highly qualified personnel, and may provide internal finance which facilitates a rapid response to events and weakens the firm's reliance on costly external finance" → Directly related to SM2 and it consist with the achievable economies of scales: They can reach better researchers thus achieving higher R&D results. Hence, monopolist is because of the latter reasons the best forms to achieve the highest dynamic efficiency.

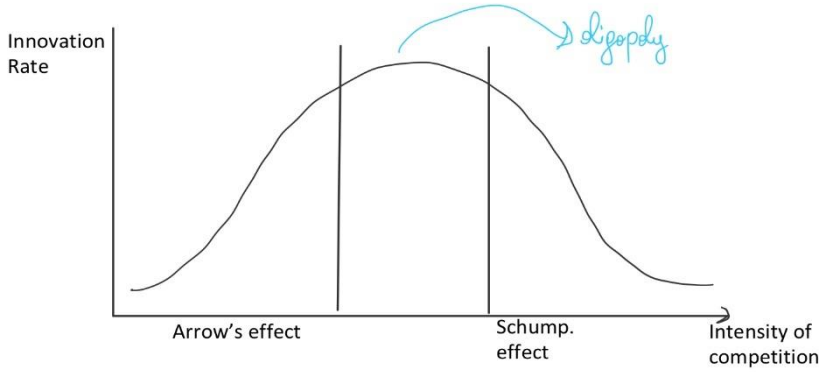
According to Arrow, on the other hand, monopoly provides less incentive to innovate than competitive industry because of the "Replacement Effect". The innovation that would step in, would simply replace the profits that the company was already making. So high degree competition markets are a preferable firm because in that scenario firm do not obtain any economic profits and hence they will be pushed to innovate.

Arrow's theorem can be represented graphically as well.



In the picture depicted above, we know that a monopolist facing marginal costs (equal to c in this case) would produce such a quantity to maximise its profits (yellow area). Suppose now that a new technology steps in the market and helps decrease marginal costs down to $\subseteq$. The extra profits will be thus represented (approximately) by the green area. Now, we assume that we are not anymore in a monopoly situation but rather in a situation where more firms are involved. These will set the price (not marginal revenues) equal to the marginal costs. The same innovation process considered before would actually allow to the innovator to gain larger profits than before since to the green area we should add the

orange one. The innovator will indeed to price an epsilon less than the current price of the market and kick out all other competitors which have not been able to innovate. Monopoly is thus less incentivized to innovate since the profits will cannibalize at least in part the current profits. Note: In the long run It might be the case in which the firm becomes a monopolist, thus increasing even more its profits. This can be perceived as a further inventive for a small firm competing in perfect competition to come up with an innovative solution.



Who is right? Empirics show that both are partly right or partly wrong. If we assume that the innovation rate can be represented by U-Shaped curve (concave) which depends the intensity of competition, as long as competition is kept low, innovation will consequently be low as well (Arrow's effect). Whereas on the other extreme, thus if competition is too high, then the Schumpeter view would be the right one (less competition would be more beneficial to the dynamic efficiency). If we look at the two perspectives, Arrow is more concerned with the incentives to innovate whereas Schumpeter more with the actual capabilities to innovate. Combining these two ideas, the most innovation-conducive market structure would be thus represented by oligopoly, where a non-singular amount of big firms compete in the market.

Suppose we have an Incumbent that has gained monopoly profits.

1) I → π_M

Suppose that a new firm can enter the market by acquiring a patent from an R&D company (outsourcing the invention).

2) NF

1. Incremental innovation (π_D, π_D)
2. Radical Innovation (0, π_M)

ρ= prob to I that NF will not enter the market

$\pi_{D=}$ profit in duopoly

Let's start with Incremental Innovation:

New entrant pushes from π_{NF} =0 (not in the market) to π_D (for push we intend the willingness to pay (outsource R&D)= $\pi_D - \varepsilon$)

Incumbent: π_M ---> $\rho\ \pi_M + (1- \rho)\ \pi_D$

Note:

- If ρ is high, pretty sure no one will enter the market (incentive to bid 0)
- If ρ is very low (perceived probability that NF will not enter), high incentives to bid
- If ρ = 0 (extreme case), incentives= $\pi_M - \pi_D$

By definition, $\pi_M - \pi_D > \pi_D \rightarrow \pi_M > 2\pi_D$ (Profit in monopoly is greater than 2 times the profit in duopoly)

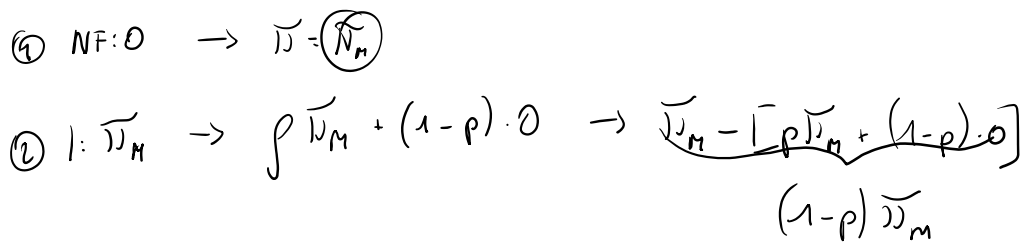
→ competition erodes always some producers' surplus → colluding > competing

Another important stylized fact that we observe in reality is that radical innovation is introduced by start-ups whereas incremental one by incumbents. Even in this case, we can provide a possible theoretical explanation to this pattern.

Suppose that we have an incumbent (I) that is gaining monopoly profits, thus I → π_M . Now, suppose that a new firm (NF) can actually enter the market by purchasing a patent from an R&D company (outsourcing the invention). In this circumstance we can think of two possible scenarios: Either the innovation has incremental characteristics or radical ones. In the first case, the market will become a symmetric duopoly in which both firms will achieve Π_D → (Π_D , Π_D). In the other scenario, the new entrant will be able to kick out the monopoly and become the new one, thus achieving monopolist profits Π_M. Let's suppose that the incumbent perceives the probability that new firm will NOT enter the market is equal to P. How much will these two companies wiling to bid for the patents?

In the case of incremental innovation, by acquiring the patent the new firm would pass from 0 profits to Π_D. How much will it be willing to disburse? → the incentive will be equal to Π_D (or an epsilon smaller than Π_D). What about the incumbent? If the firms enters it will switch from Π_M to P(Π_M) + (1-P)(Π_D). Thus the incentive of the monopolist to bid will be equal to Π_M – [P(Π_M) + (1-P)(Π_D)]. Two considerations are in order: First, when 1-P is very low, the incentive to bid for the patent will be very low. So, if P is high, the Monopolist will actually will not care whether a new firm enters and thus it will not be that willing to bid. The reverse (P is low) case underlines a high willingness to pay. If P = 0, the incentive to bid will equal to (Π_M) - (Π_D). And by definition, the difference between the two will always be larger than Π_D. This due to the fact that competition (even if little as in a duopoly) erodes producers' surplus. Or, we could also say, that colluding (choosing monopoly price and then splitting the gains) is always better than competing.

The second scenario of Radical Innovation is slightly different from the previous



The incentive of the new firm will equal to $\boldsymbol{\Pi_M}$ (now it will be able to reap all the gains), whereas for the incumbents will be equal to Π_M - [P(Π_M) + (1-P)(**0**)] = $\boldsymbol{(1\text{-}P)\Pi_M}$. Note that here the monopolist would not gain a penny if the new firms enters the market. The two firms' incentives will be equal only in the case in which the firm is 100% sure that the new firm will enter the market. But if there is a very small uncertainty about it, the incentive will be less. This why is the incentives for radical innovations is larger for start-ups and this is also why in reality we see new firms pushing for radical innovations and incumbents for radical one.

13.2 To what extent innovation should affect market structure?

This is a tricky issue: Let's consider first what appropriability is. It concerns the ability of the innovator to capture the benefits engendered by its own innovation (to the detriment of other firms).

→ Appropriability instruments that can be used by the firm:

We usually distinguish between two classes of protection mechanisms:

- *Obtained exogenously (formal)* through regulation of property rights: to the innovator is assigned the right (from an authorized super partes) to exclude others to make unauthorized use of the innovation (temporary monopoly). → Patents, Copyrights, Trademarks, etc.
- *Obtained endogenously (also informal)* by an innovative firm through the establishment of strategic barriers to imitation.

Practical problems of patents:

First of all, a patent concerns the right that the innovator has from excluding others from using its innovation (lasts for 20 years on average): There are 3 characteristics:

1. It shall be in an eligible domain (e.g., mathematical theorem cannot be patented)
2. It should be novel in a non-trivial way→ should advance the knowledge in a significant way
3. Industrial applicability → It should have in principle an industrial use and present any commercial appeal

- Diffusion of valuable information (spillovers) relating to the technology, because that patent will publicly release thus anyone will be able to see which innovation has been created
- "Contractual" incompleteness: It consists in a written document → it may suffer from some contractual incompleteness
- Inventing around → for some innovations it is impossible to extract the idea, reverse engineer the innovation, to avoid to clearly copy it
- Difficult enforcement → If you think that someone has infringed your patent, you must demonstrate it in court, where everything you cannot forecast the outcome.

This is why firms have come out with further ***alternative strategic mechanisms*** (there is not any super parte, e.g., a state) to assure appropriability:

- Secrecy (e.g. Coca Cola formula; or see this scene from the movie *Flash of Genius*: https://www.youtube.com/watch?v=IKMELT29qL4). → We will never know which the secret ingredient is to get Cola's flavor.
- Lead time (Intel microprocessor) → To be ahead of imitators. To be able to come up with something new whenever imitators reach you. Intel has been very able to protect innovation through lead time advantage, but in the last years it lost its advantages against competitors
- Complementary investments (in brand, sales & distribution and customer care, e.g. Nespresso) → In order to reap the whole benefits stemming from the innovation complementary assets that are possessed solely by you are necessary. Thus imitators can imitate your innovation but cannot replicate the availability of these complementary assets. Nespresso protected its machines through patents, but specifically through complementary assets such as marketing (George Clooney).

Intellectual Property Rights (IPR): Historically, 2 different views on patent rationale, their lenght, breadth and enforcement:

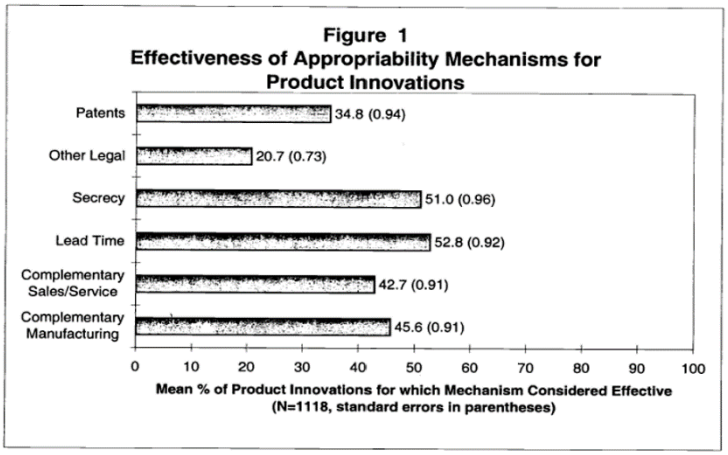
- *Incentives View:* to have that anticipated monopoly power we need patents in place. Otherwise, dynamic efficiency will suffer.
- *Openness View*: IPR regime is not required in the economy since the innovator an get benefits from innovations without protection mechanism, which, on the other hand, might be even detrimental for innovation mechanisms.

This debate has been historically solved much more in favor of the incentives view rather than the openness view. But the openness view has certainly gained some momentum over the past few years.

In the recent years we have seen an incredible upsurge of patents. "Everybody is patentic".

In recent research (2000) it has been discovered that patents are not used that much to protect their innovation with respect to informal mechanisms. Of course, even this latter is far from being perfect.

Things do not change (and are probably reinforced) looking more recently to start-ups in other geographical contexts (Italy), according to a study conducted by the professor.



Firms largely use Intellectual Protection mechanism (IP) but the most of them are informal. Moreover, these two are correlated.
Patents might not be used for defensive reasons but for offensive reasons: I do not want you to innovate.

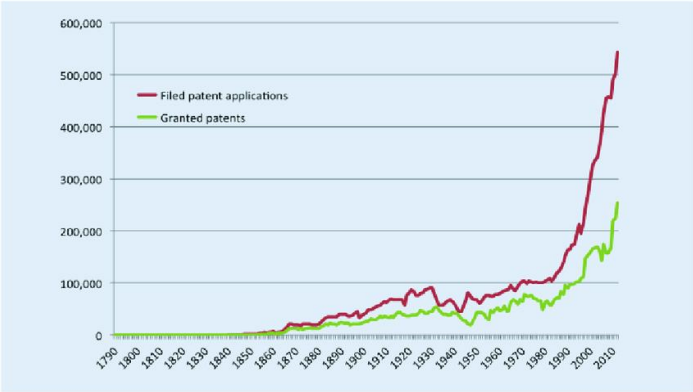
2 Stylized Facts that corroborates the latter findings:

1. Innovation is becoming more cumulative and more complex. In many sectors (like for example ICT) technologies are becoming more and more complex: the number of patents on the same product is increasingly growing. Example: 3G mobile phones (n. of patents to protect the standard is 7796).

Almost impossible that a single firm owns all IPRs on these different technological parts that uses (or wants to use).

Thus, having a patent on a specific and possibly key component of a technology might lead to:

a) Block others' innovation ("blocking patents", Heller and Eisenberg 1998, Science) → Tragedy of the "Non-Commons".
b) Give more bargaining power in a licensing transaction (what Jaffe e Lerner (2004) define as "Rembrandts in the Attic") → Firms patent anyways. If things evolve in such a way that this patent becomes important this will allow them to block or gain from selling them (extremely costly that prevents others to innovate).



2. The probability of weak patents (patent that should not be granted because it covers something stupid) is increasing.

The combination of the latter stylized facts leads to the view that even a weak patent may severely block innovation. Also, weak patents can be used for offensive strategic purposes.

This might be well explained by the following abstract from *"Patently Absurd"* (Forbes): My own introduction to the realities of the patent system came in the 1980s, when my client, Sun Microsystems--then a small company--was accused by IBM of patent infringement. Threatening a massive lawsuit, IBM demanded a meeting to present its claims. Fourteen IBM lawyers and their assistants, all clad in the requisite dark blue suits, crowded into the largest conference room Sun had.

The chief blue suit orchestrated the presentation of the seven patents IBM claimed were infringed, the most prominent of which was IBM's notorious "fat lines" patent: To turn a thin line on a computer screen into a broad line, you go up and down an equal distance from the ends of the thin line and then connect the four points. You probably learned this technique for turning a line into a rectangle in seventh-grade geometry, and, doubtless, you believe it was devised by Euclid or some such 3,000-year-old thinker. Not according to the examiners of the USPTO, who awarded IBM a patent on the process. After IBM's presentation, our turn came. As the Big Blue crew looked on (without a flicker of emotion), my colleagues--all of whom had both engineering and law degrees--took to the whiteboard with markers, methodically illustrating, dissecting, and demolishing IBM's claims. We used phrases like: "You must be kidding," and "You ought to be ashamed." But the IBM team showed no emotion, save outright indifference. Confidently, we proclaimed our conclusion: Only one of the seven IBM patents would be deemed valid by a court, and no rational court would find that Sun's technology infringed even that one.

An awkward silence ensued. The blue suits did not even confer among themselves. They just sat there, stonelike. Finally, the chief suit responded. "OK," he said, "maybe you don't infringe these seven patents. But we have 10,000 U.S. patents. Do you really want us to go back to Armonk [IBM headquarters in New York] and find seven patents you do infringe? Or do you want to make this easy and just pay us $20 million?". After a modest bit of negotiation, Sun cut IBM a check, and the blue suits went to the next company on their hit list.

"If national patent laws did not exist, it would be difficult to make a conclusive case for introducing them; but the fact that they do exist shifts the burden of proof and it is equally difficult to make a really conclusive case for abolishing them" – Roger Penrose.

14. ALTERNATIVE THEORIES OF THE FIRM: CONTRACTUAL AND HOLISTIC APPROACHES

First, what is a neoclassical firm? It is basically a Black-Box. It is characterized by a mathematical production function. It bases the decisions on the demand and costs functions and takes the structure of the market in which operates. But in the end, such firm behaves in a very predictable way (optimizing costs and maximising profits). But what lies behind the revenues and costs functions? How does the firm turn the inputs into outputs? How is it governed? Are there any conflicting interests within the firms? Are shareholders' preferences respected? In reality firms cannot be represented by "a simple" mathematical function. They are very complicated economic agents.

The key issues not considering by the neoclassical firm theory:

1. Firm existence: Why do actually firm exist? Neoclassical theory is very silent in this respect. If we think about it really carefully, all needs could be satisfied by simple market transactions (e.g., If I want to have an iPhone I can buy all its components and hire a person or more to have the phone built for me. Why is Apple needed then?)
2. "The black box": What is the exact functioning of a firm? What happens within this black box? Although the assumptions that economists have come with, such as perfect rationality or information are not respected in most of the real-life scenarios. Instead of having perfect rational behaviours, we might have *profit seeking* (try to seize as many profits as possible), thus agents that do not have the profit maximisation function clear, but simply try to exploit some heuristics (rule of thumbs) to increment profits as much as possible.
3. Firm heterogeneity: According to neoclassical theory, if all the "relevant" equations are the same, then all the firms are equal, thus achieving the same performances in the same market structure. This is NOT empirically true! In any given industry there are more successful and less successful firms. Heterogeneity is present and thus we need to develop some tools to deal with it
4. Profits as the only goal: Is this always true? Not always. Anyways, we can say that profits belong to the principal goals pursues along with some other such as sustainability (easiest issue to deal with among the 4).

To take into consideration these aspects, we need to develop alternative theories. Economic (allocation of scarce resources) and managerial (organization of people in hierarchical structures) theories. Their aim is to explain firm's existence, behaviour and goals.
Firstly, what is a firm? It can be conceptualized as a separate entity that performs certain actions. It makes sense to say "the firm raises prices" although in reality the firm is not an agent. It cannot be considered an anthropomorphic entity. It is not the firm that takes decision, but it is a group of people gathered within the firm that do so.

Firms consist of different kinds of agents:

- Owner(s)
- Providers of inputs: managers, workers

1. The contractual approach

- Bundle of contracts: Harmonize peoples objectives (induced to maximise firm's profits)
- Diverse objectives within: Within the firm every member has different objectives
- No firm-level objectives: We are talking about the objectives of the people, not of a firm

2. The holistic approach (including the neoclassical view): a little more abstract, it looks at reality from a higher level. It goes beyond the interactions of people within the firm and provides a summary of these interactions:

- Beyond linkages
- Anthropomorphic vision
- Firm-level objectives (no diverse objectives within): In this case we are not considering people's objectives but rather at higher level, firm's ones

14.1 The contractual approach

1. Principal-Agent problem
    a. Managerial theory
    b. Team productivity theory
2. Contractual imperfections
    a. Transaction costs theory
    b. Property rights theory

1. Principal-agent problem: In very general terms, a principal is an agent that has some objectives which are delegated to an agent, who will have to act in the best interest of the former receiving a compensation to do so. Thus, the agent should be morally and legally bound to principal's needs. Note that those that possess the firm are not necessarily those who control the firm → Ownership (shareholders) ≠ Control (Managers). Managers should behave in the best interest of shareholders, thus maximizing their value (profits). Why do we then a separation? Wouldn't it be better if shareholders would run their own firm? Well, expertise come into play here. Managers are more suited to drive daily operations of an organization. Giving the firm's steering wheel to well-educated agents is in the interest of shareholders. Moreover, shareholders are usually so many that it would extremely be cumbersome to run the firm efficiently. But, delegating has a huge drawback: Asymmetric information. If we had perfectly informed relationships between agent and principals, the latter would perfectly know how agents behave, thus being able to check and assess whether agents are behaving in their actual best need. With asymmetric information, agents might decide to maximize their objectives which to some degrees are correlated with principals' ones but not equal. Example: Deciding to maximize revenues instead of profits. Moreover, note that ownership fragmentation is negatively correlated with the incentive and ability to overcome these difficulties. If the company' shareholders are a huge number, the latter won't have any incentive to monitor agents' behavior since having a small share at stake. Moreover, a shareholder having a very small share won't even have the power to do so.

So, how do we mitigate principal-agents' problems? 2 ways:

1. Monitoring though a board of directors
2. Partial alignment of incentives (profit-contingent compensation). Making sure that managers have a stake in the profits of the company so that if they manage to increase company's rifts they will get higher bonuses (e.g. through stock options).

a. Managerial theory:

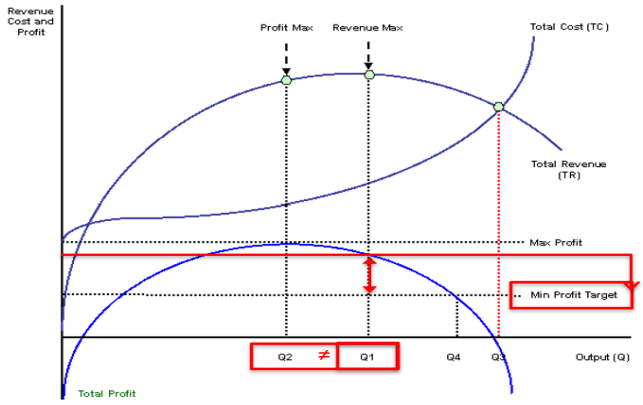
Managers goals ≠ Owners' goals

- Acceptable levels of profit
- Maximization of their own utility
    - Managers' goals: sales maximization
    - Owners' goals: profit maximization

Why do managers maximize revenues instead of profits? Firstly, because normally compensations (bonuses or career promotions) are linked to revenues instead to profits. Secondly, because of personal prestige and visibility. It mikes a bigger impression both on others and on managerial ego to show bigger numbers (revenues). Moreover, managers want to expand the boundaries of the firm. This is rooted into the intrinsic characteristics of the human being. Managers want to enlarge their empire, increasing their revenues (not necessarily profits). Thirdly, every decision that every manager does is connected to a cost. If you think about it, the larger the revenues a firm generates, the larger the costs the firm can face, thus allowing the manager

to a larger amount of decision (even risky ones). Lastly, Profit maximization may entail cost-cutting, which is often a difficult decision since it involves in giving away something (e.g., firing an employee or shutting down a plant)

Baumol's model (1959) of sales maximisation. What is the outcome of maximising sales instead of profits? During the decision-making period, the CEO attempts to maximize sales, where sales maximization is subject to the provision of a minimum required profit to ensure a fair dividend to shareholders. If they would strictly prefer revenues on profits, it would mean that they would be keen to have infinite revenues and 0 profits. Very unrealistic since shareholders are paid from profits.
In this way, the stability of CEO's job is assured.
Note: Conventional cost and revenue functions are assumed and Total costs increasing + Demand curve is downward sloping
Baumol's model graphical representations →



1. Total Revenues are represented by a concave function (inverted U-Shaped function). Starting from 0 as quantity sold, selling an extra unit will increment the total revenues. But, at a certain point, if the firm decides to enlarge output, it will have to lower price to reach further customers (thus increasing the prices for all the consumers that have been reached so far as well).
2. The Total Cost function: At the beginning they are increasing at a decreasing rate because economies of scale are exploitable. But, after a certain the firm will incur diseconomies of scale: The firm will have to incentive the labour force for extra work by increasing wages → Cost increase.
3. Profits curve is constructed by subtracting the first curve to the second one.

The point here is that if we maximise revenues, we are going to reach the peak of the revenue function which DOES NOT coincide with maximum profits. This, because along revenues, costs are increasing as well and at that specific point (maximum), costs are increasing more than what revenues. Thus, firm should consider to stop increasing revenues before reaching its peak.

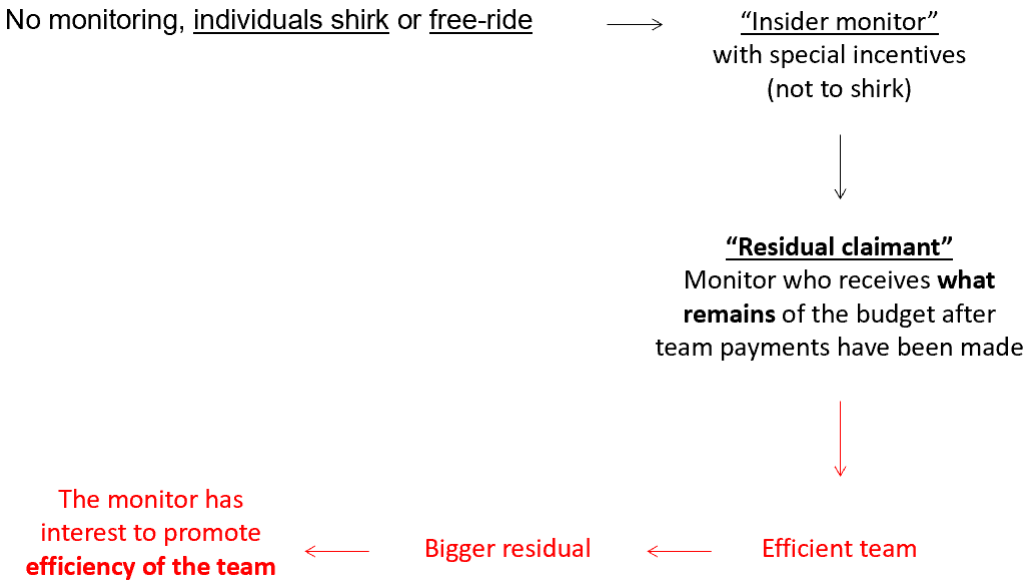
What the implications of this theory?

- Profit rates are higher in owner-controlled firms than managerial firms: If we have suboptimal outcome due to management, if we had a firm solely controlled by owners will stop where profits are maxed
- The higher ownership concentration, the higher profits, ceteris paribus: A higher concentration means that owners have a higher incentive to monitor managers closely and thus minimizing the objectives misalignment

→ Empirical evidence is mixed

b. Team productivity theory

This theory explains the existence of the firms as a *matter monitoring device* that indeed monitors and measures the contribution of people. The focal problem is the following: If we gather people to share productive ends, one problem arises: We cannot exactly isolate the individual contribution and measure it. Team's productivity can be measured through KPIs, but not the singular contribution. This exacerbates the problem of Free riding (Shirking). This can be crippling in terms of productivity. If everyone behaves this way, no production will exist.



How does this theory explain the existence of firms? It conceptualizes the owners of the firm as monitoring devices. They have indeed a huge incentive (depending on the amount of shares at stake). The problem here is that even the managers that have been assigned the duty to monitor employees have no incentives that are aligned with principals' ones → They are not residual claimants. In principal only the owners are the only one to get the profits. This is why they have an incentive to monitor people below them in order to increase the residual. The only way to do so is to organize transactions within the firms.

A critique: The theory's weak spot.

- This idea is effectively applicable to small firms, because here very powerful owners are present and who will have large incentives to monitor performances. In large firms, the incentives get lost in the multitude of layers within the hierarchy.
- It has a very narrow vision of the entrepreneur → Entrepreneur is seen primarily as a "monitor"
    - Similarly to early vision of entrepreneur as "coordinator of factors of production" (e.g. Jean-Baptiste Say, 1828) or "receiver of non-fixed and risky income" (Cantillon, 1755) However, entrepreneurship is much more than that (Ricketts, 1994)
    - Kizner (1979): entrepreneur is alert to hitherto unexploited opportunities that only she's able to recognize and exploit
    - Knight (1921): entrepreneur is willing to tolerate uncertainty
    - Schumpeter (1934): entrepreneur is a revolutionary, an innovator, overturning tried and tested convention and producing novelty

2. Contractual imperfections (Focuses more on the drawbacks of contracts rather than the problems between principals and agents) → The series of imperfections findable within contracts sometimes make it easier to organize transactions within a firm

A. Transaction costs theory

According to this theory, firms arise whenever the sum of production costs and coordination costs is lower than the sum between market price and transaction costs. → It is more sensible to organize transactions within to firm. Why? Because if I organize transactions within a firm I have to pay for inputs and coordination costs (ensuring all the factors of production are adequately coordinated among each other). If instead, I prefer to organize transactions in the market, I'll have to pay the market price for the inputs on and on top of that, transaction costs.

Market transactions are plagued by transaction costs (Arise due to contracts imperfections):

- Search and information costs: Time and effort shall be spent to find a counterparty

- Bargaining costs: Once a party is found, a bargaining shall be stroke → Time costly
- Policing and enforcement cost (e.g. drafting a contract)

Firms arise whenever the cost of transacting through the market is greater than the cost of organizing transactions internally. This explains the existence of the firms. Of course transaction costs increase whenever the relevant determinants increase:

- Bounded rationality: The larger, the larger searching and bargaining costs
- Opportunism: If the other party is more opportunistic, I will have to spend more resources in bargaining and monitoring her.
- Relationship-specific investments: The more RS, the larger transaction costs
- Frequency of transactions: If I have to perform a transaction multiple times, it makes even more sense to organize it within a firm

B. Property rights theory [Similar to the transaction costs theory, but it differs in terms of focus. Both include contracts' incompleteness]

Contracts are necessarily incomplete: it is impossible to specify provisions for any possible state of the world. The PRT focuses on the fact that renegotiations are likely to occur ex-post. Since contracts are incomplete, we might also expect that in the future something unexpected happens in the future, which might invalidate some provisions in the contract. In this case, property rights are crucial. Property rights influence the relative bargaining position of parties during renegotiations, as they allow residual rights of control. Once we enter a contract we have to distinguish between specific and residual rights of control. Specific rights are those provided in the contract. If something unexpected happens (say increase in demand) I have no contractual rights to ask for changes in the production (e.g.). Here, the property rights define the rights of control. Ownership allows residual rights of control. Firms exist because ownership of assets they allows residual rights for control. If, on the other hand I turn to the market, I will have only specific rights. Comments crucial to read

Like, transaction costs theory, this particular theory is well-suited to explain firms existence and vertical integration, which concerns increasing the reach of the firm along the supply chain. Integration solves the problems arising from contract incompleteness.

14.2 Holistic approach

Firm is an autonomous entity able to make decisions on its own.

A. The Resource Based View:

1. Profits are heterogeneous and sticky within markets and industries: There are companies that are more successful than others

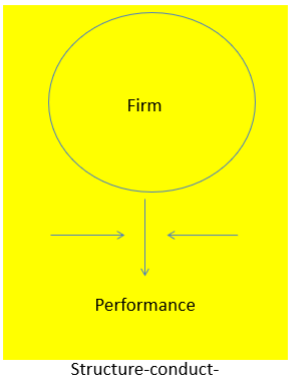
2. Companies are able to achieve SUSTAINED COMPETITIVE ADVANTAGE (SCA)

→ Persistently high profit levels over time.
RBV does not challenge the neoclassical profit max assumption, but it is interested in what are the drivers for SCA and how can these be achieved in a persistent way.

In the neoclassical approach the firm tries to adopt its behaviour to the market structure, which impacts the conduct of the firm, it competes in.

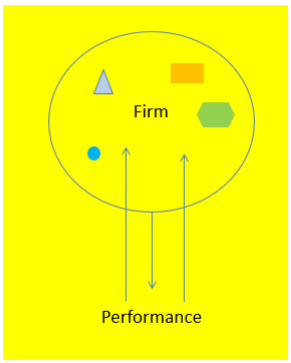
In the RCV the firm is not a simple "circle" (impersonal) but is rather crowded with other geometrical figures which represent every resource and capability that the firm represents. The external environment is not everything thus. Any resource and capability differentiate a firm to the others. There is no firm that is identical to the others. The mix of the resources and capabilities has a crucial impact on performances. If these are extremely valuable, the firm might achieve competitive advantages. But the weight and the adherence to the external environment are important as well. Performances do impact on valuable resources → Through profits (high performances) the firm can foster the pool of resources, i.e. can acquire new competences and resources creating thus a virtuous circle that in the long run will allow it to keep sustainable competitive advantages

Neoclassical



Structure-conduct-performance paradigm

RBV



Firm is a bundle of resources

Where does competitive advantage of stem from?

1. Possession of one resource, such as triangles, that other firms do not possess.
2. Possession of a way to combine resources, such as triangles and points, that other firms cannot replicate.
3. Possession of one (or combination of several) resource(s) that create the best "strategic fit" with the external environment. Although the RCBV puts a lot of emphasis on the internal characteristics of the firm, it still considers the external environment as well.

A resource is everything that a firm possesses: from physical to intangible assets. Capabilities concern the WAY things are done.

Some examples:

- Physical: plant equipment, location, access to raw materials
- Human: training, experience, judgment, decision-making skills, intelligence, relationships, knowledge
- Organizational: Culture, formal reporting structures, control systems, coordinating systems, informal relationships

1° point: external environment matters

It sets what "game" firms are playing (on innovation, cost savings, etc.)

2° point: some resources are key for a wide range of environments, others are not (Money, for instance, are crucial for every industry). Versatility is a very important matter.

3° point: sometimes it is a matter of combining resources rather collecting resources

Think of Ducati: Every particular, from the performances, to the design, to the beauty, to the sound of the engines (and so on) combined is what makes the motorbikes extremely valuable.

What makes a resource (or a mix of resources) a firm possesses contribute to its competitive advantage?

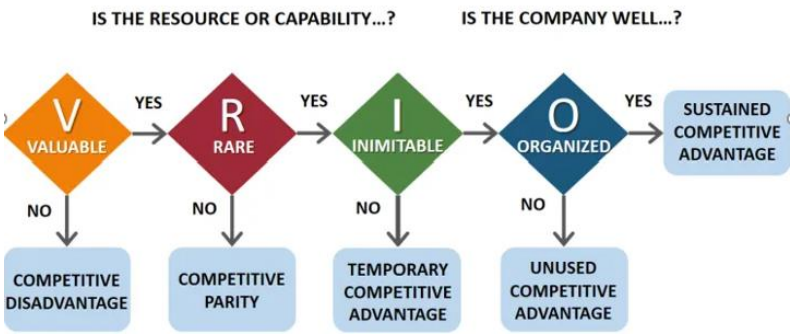
✓

✗

If a resource is not valuable, we are talking about fried air. It is nothing that could help the company. It basically consists in a competitive disadvantages.

If a resource is not rare, then by definition it means that I am not the only one that possesses that resource or competence.

If instead the resource is rare, we should ask ourselves whether the resource is also inimitable or not. Is that something that could last in the long term or not? If it can be imitated rapidly (no patents' regime, for instance), competitors will acquire it.



If there is no organized structure that could help to exploit the valuable, rare and inimitable resource, then this precious resource becomes unused.

Isolating mechanisms:

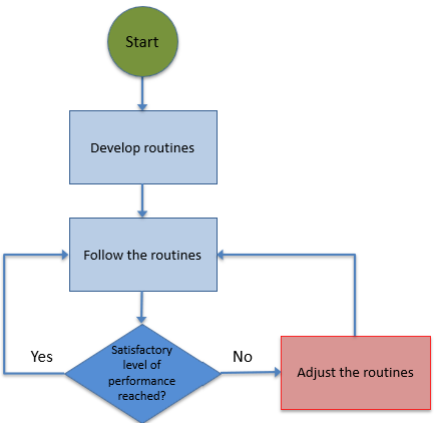
1. Path-dependence of production processes and resources accumulation (e.g., ecosystem, brand reputation). If you manage to have a first mover advantage, even if a competitor can follow with a better product, it will be very difficult to let him kick you out of the market. Path dependence is a first and fundamental mechanism for protection.

2. Intellectual property rights mechanism (patents or secrecy). Note: Secrecy consist basically in a situation in which the authorities allow you to keep a specific resource secret.

3. Causal ambiguity: You cannot really figure out where the advantage is coming from. Of course, if you cannot figure it out, it cannot be imitated.

B. The evolutionary theory

Heterodox (theories that are in contrast with leading one) theory. The key point of contrast with the neoclassical theory, is that according with the evolutionary theory it does not make any sense in talking in profit maximization. Have you ever seen a manager that tries to take the derivative to see where the maximum of the profits curve lies? NO. We are human beings, not computers. Our rationality is bounded. Maximization becomes unfeasible. So, the way real world decisions are taken is through heuristics, rule of thumbs. Secondly, neoclassical theory is very static in nature. It deals a lot with stationary equilibria. On the other hand the evolutionary theory deals with dynamicity. It lies on the legacy of Schumpeter. So, the key will be to develop strategies and practices that lead to a satisfactory level of performance. Over time, strategies and practices become ROUTINES → Firm is a bundle of routines. Such routines are kept until they help to reach satisfactory levels of performances. Once they do not, they have to adjusted. If not able to do so, failure is imminent (e.g. Kodak). So this is the key problem. The new routines that might be necessary that help us adapt to the new strategic environment might be constrained by the current routines. → "Firm's genetic heritage". It is difficult to transition to the new routine. The farer it is from the current, the larger the difficulty.



How wrong is the neoclassical approach then?

It is able to make general predictions. Example: "Assume a change of driving conditions occurs, say, that the roads have become wet and slippery and the fog has reduced visibility." → "The theory is able to predict that traffic will be slower and accidents more frequent, but not able to predict that any particular driver will drive

more slowly or have an accident." In order to understand that we should turn to theories that allow to analyze individual behaviors such as Resource and Competence Based View and Evolutionary Theory.

15. FIRM GROWTH: DIVERSIFICATION AND VERTICAL INTEGRATION

Firms are not always the same. They change. They grow.

Short-run	Long-run
1 fiscal year	More than 1 fiscal year
"Tactic"	"Strategy"
• Changes in **production levels** • Optimization and variation of the **production mix** • Make-or-buy choices for contingent reasons (e.g. demand's peak)	• **Organizational structure** • Adopted **technology** • **Plants size and new machinery** • **New products or processes** • **Entry** to / **exit** from the industry • **Diversification** • **Delocalization** of some activities
• max (Revenue - Avoidable Cost) • Fixed Costs are not considered	• Expected present value of revenue and cost flows

Focus of this lecture and the following

Different types of boundaries:

Horizontal boundaries of the firm: Quantity of products: Keeping the business the same, but reaping economies of scale.
Lateral boundaries of the firm: Variety of products → Crossing the threshold of your business: Diversification. This can be wither related or unrelated. If a pizza restaurant is opening a burger store, it is still in the food industry. If instead opens a library, then we are talking about unrelated diversification. Related diversification focuses on economies of scope: Achieving synergies, where A + B = A + B + C, where C consist in extra benefits stemming from combining activities A and B together. Example: Exploiting the distribution channels. What is the rationale behind unrelated diversification since usually it is better to stick to what the company is able to do best? To create value.
Note: Risk hedging would not make much sense since Shareholders can already do so on their own.

Vertical boundaries of the firm: Number of integrated steps along the supply chain. Vertical boundaries are those boundaries that determine the reach of the company. If the company integrates backwards, for instance, it means that it is becoming its own supplier. Conversely, if expanding forward, it means it is becoming its client. Example: If the pizza restaurant is buying a mozzarella factory is integrating backwards. If establish a delivery company, forwards.
International boundaries of the firm: Geographical span of firm activities

15.1 Diversification

Why do firm diversify? It basically boils down to understanding costs and benefits: In order to understand if a diversification decision is profitable, benefits will have to be larger than the costs.

For what concerns costs, these can be either ex-ante or ex-post. Ex-ante costs can be subdivided into explicit and implicit, where the latter consist in those costs that the firm incurs in letting other opportunities forgone. Example: Growing in the current market is larger than in a new market, hence you take the decision to remain in the current one. This decision of course entails the costs of letting the opportunity to enter in the new market. Within the ex-post costs, it is interesting to see what is meant with reduction in incentives. The larger the firm becomes, the smaller the incentives for the employees to behave optimally. Their objectives might not be aligned with the firms' (shareholders') ones.

Benefits	Costs
Economies of scope (spreading the firm's resources, capabilities and managerial talent over many uses)	**Ex-ante costs** (implicit and explicit costs incurred to acquire a company or to set-up an entrepreneurial venture)
Other synergies (Addition of value to extant businesses, transfer of skills and knowledge, cross-business collaboration)	**Ex-post costs** (coordination costs, reduction in incentives, slower information flows)

How do firm finance it:

1. Internal capital markets: Potentially more limited, especially if the firm is not that large, they allow much freedom. Think of exploiting the profits stemming from a cash cow to invest in a more risky business.

2. External capital markets: Very large, but extremely rigid and risk averse. Investors do indeed prefer lower risks levels.

The choice to diversify ultimately depends on a cost and benefit analysis: are synergies and economies of scope sufficiently high to overcome financial, entry and coordination costs? This is not that simple.

Three conditions should be satisfied:

- Attractiveness test: the target industry should be structurally attractive → You must target industry that is structurally attractive (Porter's 5 forces)
- Cost of entry test: the cost of entry should not exceed (expected) future profits stemming from the target industry → in other words, even though you might get synergies, you should not overpay for an acquisition. Attention: You need to take into account synergies in calculating the benefits stemming from diversification!
- Better-off test: strong synergies should be present between the two businesses and clear and evident → 2+2=6. If 2+2=5, maybe synergies could be too low.

The purpose of these test is to force you to consider all the possible aspects that might undermine the profitability of the business. If all these tests are "passed" you can enter the diversification strategy.

However, the cost-benefit analysis is often biased, and diversification happens much more often than it should…. WHY? Who decides about diversifying? Managers can significantly benefit from diversification:

1. Social prominence, public prestige, and political power (building an empire)
2. Limit managerial risk: diversifying limits managers' risk of achieving a poor profitability overall

How do limit this risk?

1. Board of directors (BoD)

Monitor the management to ensure that actions are taken to increase shareholder value. However, high-level managers may exert considerable control over the selection of new directors
Important to have external / independent members of BoD

2. Incentives to managers

During 1980s' & 1990's: increase in leveraged buyouts (LBO), whereby cash flows of firms were used to undertake unprofitable acquisitions instead of paying dividends

→ Need to realign managerial incentives

- One popular way to partially realign managerial incentives is to link their compensation to the value of shares of the company they direct (e.g. through employee stock options)

3. Capital market discipline → Mechanism that are present in efficient markets that come in help If bad managers are running a company

- Management tends to overpay for diversifying acquisitions
- The stock market ends up internalizing this information, expecting managers to overpay for additional acquisitions in the future → The firm's shares' market price falls immediately

→ The opportunity arises for another entity (raider) to try a hostile takeover and appoint its own managers, with the potential to increase profitability thanks to a better management

→ Capital market discipline deters inappropriate management

4.Labor market discipline

- Managers' reputation circulates in the labor market
- Potential employers know whether managers pursue personal goals or organizational ones

Key takeaways

- Diversification allows the business to grow and benefit from synergies
- Diversification is optimal only when its benefits (i.e. economies of scope and other synergies) outweigh its costs (both ex-ante and ex-post)
- Empirical studies show that the performance of diversified firms is, on average, lower than the one of more focused firms: this is probably since diversification happens more often than it should, due to self-interested managers

15.2 Vertical Integration

In principle, there might be two firms that do these activities. On extracts and refines the oil, the other takes on with the logistics activities.

1.Forward Integration: Firm 1 becomes the final distributor → It gets closer to the final customer.

2.Backwards Integration: Firm 2 starts to extract and refine crude oil, then it becomes its own supplier.

Property Rights Theory (PRT): Integration determines:

A. Who controls resources, who makes decisions and to whom profits are allocated? → Appropriate ownership structure is conducive to efficiency

B. The residual right of control is the right to decide on all the situations that are not included in the contract (typically firm owner)

- Residual rights of control over assets belong to the owner of such assets
- Vertical integration transfers the residual rights of control over the assets of the vertically integrated firm to the vertically integrating firm

→ Transaction costs decrease (higher strategic control, no hold-up problem)

(But: Coordination costs rise)

Example: PepsiCo has two types of bottlers:

1. Independent (no PepsiCo authority on how operations are managed)

2. Company owned (PepsiCo has the ultimate authority over how the bottling assets are deployed)

In case of a marketing campaign (e.g. Pepsi Challenge), production should match (increased) demand

1. But independent bottlers can refuse to participate (no residual rights of control over independent bottlers' assets)
2. Instead, PepsiCo-owned bottlers can be replaced with a more cooperative team → BIG ADVANTAGE

15.2.1 Further vertical integration advantages:

- Lower transaction costs (search, negotiating, enforcement and monitoring costs) → Larger bargaining power
- Possible synergies between different stages of the supply chain: An example might be Amazon which distributes products for many business and deals also with analytics. It can thus analyze costumers' behavior. If it was to integrate backwards, it would have a competitive advantage since knowing perfectly consumers preferences.
- Strategic independence: Is related to the insights to property rights theory. If you own your assets you can be strategically flexible → Better control on multiple business dimensions (e.g. quantity, quality and timing of supply):
- Know-how protection (e.g. protection against potential intellectual property losses and knowledge spillovers). If suppliers were able to reverse engineer your product, they could "steal" your product and eliminate your competitive advantages.

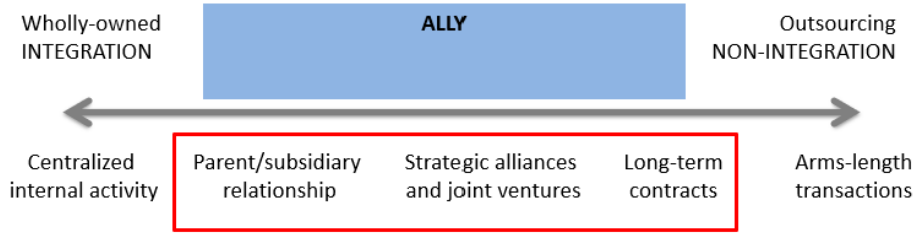
15.2.2 Vertical integration disadvantages

- Often lower technical efficiency (especially at the beginning): external providers usually benefit from strong economies of scale and learning, due to specialization and demand aggregation of several customers
- Higher coordination costs
- Higher commitment: it is typically more difficult to divest a business than to terminate a contract

15.3 Firm Growth: Alternatives to Integration and Internationalization

The Make-or-Ally-or-buy Choice:

→ Reality is more complex than the traditional **make-or-buy** dilemma would suggest:



On one hand of the spectrum we notice a series of completely internalized activities, which do not rely on any transactions with the market. On the other hand, arms-length transactions, which consist basically in short term relationships with the market to get additional components necessary for production processes. In the middle there exist different facets/shades of integration. A long-term contract is technically a market transaction that happens outside the boundaries of the firm, but since talking about something long-term it consists in a transaction in which something more is invested (in terms of relationship) with respect to short ones. Moving leftwards, parent/subsidiary relationships consist in a situation in which transactions between the two players (subsidiary and headquarter) happen within the boundaries of the firm, since the former are part of the latter. Still, we are talking about two physically (maybe also culturally) separated entities.

Intermediate forms between integration and market to have a broad understanding that reality is never black or white:

1. *Tapered integration* (making some and buying the rest): The company buys part of the pieces and makes the rest
2. *Licensing* (e.g., franchising, the right to use a firm's business model and brand for a prescribed period of time)
3. *Joint ventures* (cooperation on a new joint firm) and *strategic alliances* (cooperation on a joint project)

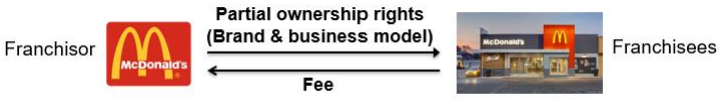
A. Tapered Integration: *Mixture* of vertical integration and market exchange (i.e., co-sourcing, making some and buying the rest)

Note:

- Flexibility is kept since smaller capital commitments are done
- If I can produce efficiently and the market cannot (maybe too costly), there is still the possibility to rely on internal production channels. This might also have impacts on bargaining powers (suppliers cannot fool the company) and protection from supplies' blocks.
- Loss of economies (the most serious problem) of scale are lost on both sides, supplier and the company itself
- Products must be identical
- Biases (less serious problem): If I have an external supplier that partially covers my product's need and in case I am inefficient (no optimal production), then I might treat my supplier as inefficient (like myself) despite being optimally producing.

Advantages	Disadvantages
Additional input/output channels without massive capital investments	**Loss of economies of scale**
Credible threat of substituting internal channels with the market in case of poor performance (and vice versa)	**Coordination** between the two production units (product specs, delivery times…)
Information about costs and profitability from internal operations → **better negotiations with market firms**	**Possible bias** in judgement (e.g., the firm may mistakenly establish the performance of an inefficient internal supplier as the standard)
Internal supply capabilities → **protection against potential hold-up**	

B. Franchising: The big firm sells the rights to use the brand and the business model in exchange for royalties.



Split the tasks:

→ Franchisor performs tasks involving substantial scale economies (e.g. purchasing and branding)

→ Franchisees build and operate the business (does not have to come up with any business model nor brand identity)

Problem: Underperformance is limited through tight quality controls and frequent surprise inspections to avoid brand images

C. Strategic alliances and joint ventures

Organizing complex business transactions collectively without sacrificing autonomy:

- *Strategic alliance:* cooperation, coordination, information, and resource sharing for a joint project by the participating firms (e.g., Google & Luxottica)
- *Joint venture:* strategic alliance where a new independent organization is created and jointly owned by the promoting firms.

Alliances and joint ventures can be:

- *Horizontal* (firms in the same industry)
- *Lateral* (firms in different industries) → to acquire resources and capabilities from other industries to come up with something different without committing too much to the new industry
- *Vertical* (firms at different stages of the supply chain)

Alliances rely more on *trust, reciprocity, cooperation, and information sharing* than arm's length contracts do due to their long-term objectives. Corporate cultures that blend perfectly are required. *Disputes* are rarely litigated and tend to be resolved through *negotiation.*

Factors leading to strategic alliances and joint ventures that can be divided into resource (information) related and governance related:

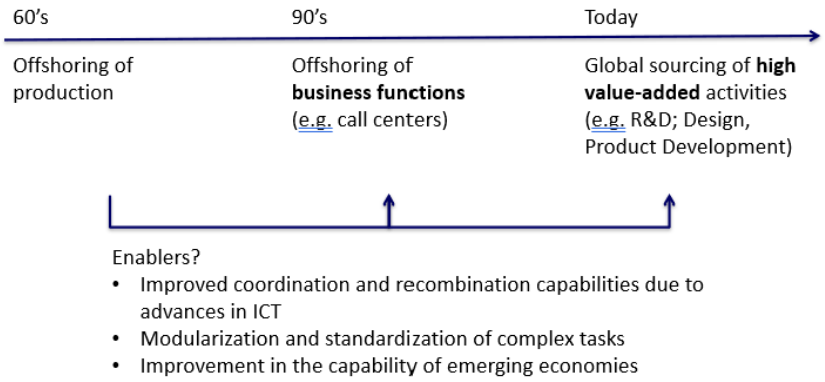
1. *Impediments* to comprehensive contracting → The more complex contracts are, the more I could require a strategic alliance. If it is almost impossible to list everything in a contract, stipulating a strategic alliance would put both parties on the same side, thus, to some extent, aligning their objectives without the necessity to formalized everything by means of a contract.
2. *Relationship-specific assets* by both parties → Strategic alliance → Objective alliance → Why should my ally hold me up?
3. Complementarities between the *resources and capabilities* of the parties involved: The higher the complementarities, the larger the synergies, thus the higher the value of the strategic alliance
4. Advantages in terms of information, knowledge sharing and contextual adaptation (e.g. internationalization): One firm might have a very clear business model while the other a very sophisticated and efficient distribution channel. These advantages are capitalized especially whenever a company wants to stretch its geographical boundaries and creates an alliance with a local firm.
5. *Transitory* nature of the collaboration opportunity: An alliance is preferred whenever the nature of the collaboration is transitory

Drawbacks:

1. *Risk of losing control* over proprietary information → You are teaming up with another organization which might also become a competitor. To some extent, it could steal precious information since:
    - Intensive knowledge sharing necessary
    - Intellectual property protection mechanisms may be ineffective
2. Agency and influence costs (Costs a firm incurs when managers try to influence the allocation of resources. In different firms, different divisions do exist. The bulk of resources must be divided between these according to specific criteria. Takes time and there is the possibility to deviate from the optimal allocation if managers can influence it)
    - Coordination issues
    - No formal mechanisms for making decisions

- No formal mechanisms for resolving disputes

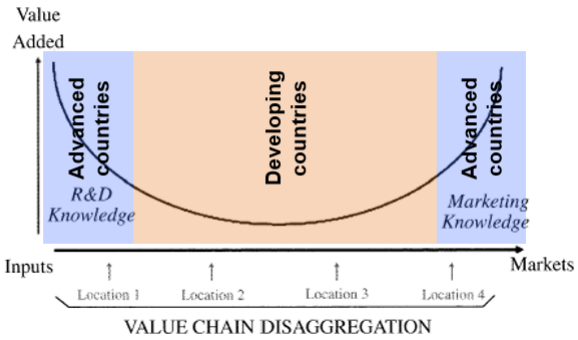
The internationalization process: stylized facts:



Internationalization has evolved enormously in the last decades. This process is about the movement of physical goods and information. **Offshoring** (moving production abroad) became widespread mainly due to abondance of resources and wages differentials. The most important aspect for companies to internationalize nowadays is about being able to reach locations that enables the company to absorb key knowledges (information) that can be spread all over the company business model.

(The phenomenon of offshoring consists of disintegration, relocation, and reintegration of a company from one country to another.)

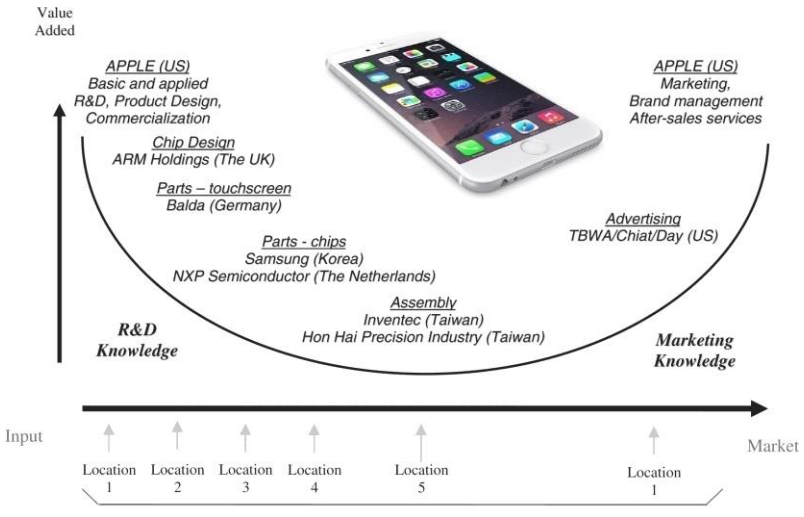
The smile of value creation (MNEs, knowledge and location)



VALUE CHAIN DISAGGREGATION

To this respect there many different activities that add more value than others. R&D, for instance, consists in the very foundation of a production process. Marketing, which is related to the perception of the product, is extremely important in adding value to the product itself. Both such activities are highly value adding and are both located in advanced countries (performed at the headquarter). The rest (packaging or production) not-so-value adding can be typically standardized and outsourced from developing countries.

Example: Apple



VALUE CHAIN DISAGGREGATION

After offshoring, some companies decide to *re-shore, backshore* and **near-shore** their activities

Why?

- Changes in priorities (e.g. cost vs differentiation) → Quality has become key → Costumer centricity
- Emergence of technological enablers (e.g. 3D printing) → New technologies allow cost efficient production → The contrary to modularity: 3D printing allows to literally print the entire product from scratch. There is no necessity in assembling different parts.
- Managerial mistakes

Internationalization: firm involvement beyond the home country boundaries

1. Export (simple trade flows between countries) → Weak form of internationalization
2. Intermediate forms granting indirect presence in other countries (e.g. some types of licensing agreements and strategic alliances)
3. Foreign Direct Investment (FDI): establishment of a controlling interest into another country → Strongest form of internationalization
    - Greenfield investments (new firm from scratch in another country to have direct interest)
    - Joint ventures (when the third entity being created is abroad)
    - Acquisitions (purchase of an existing local firm)

Firms engaging in Foreign Direct Investments (FDI) are called Multinational Corporations (MNCs) or Multinational Enterprises (MNEs)

15.4 The Economic Theories of MNCs

1. Hymer's seminal contribution:

Before 1960s (i.e., before *Hymer*'s seminal work) several economists investigated *international trade* (export)

These theories regarded *trade* and *capital movements* across borders (not FDI):

- the main determinant of movements of *goods* (trade) across borders is the difference in *factor endowments*
- the main determinant of movements of *funds* (capital) across frontiers is the difference in *interest rates*

Hymer (1960s) distinguishes between:

1. Portfolio Investments (purely financial investments)
2. Exports
3. *Foreign Direct Investments (FDI)* providing the firm *control* over the business activities abroad

FDI are the essence of the internationalization process: while exports are simple market transactions, FDIs allow firms to *grow internationally*. It does not consist solely in trade of goods and capital movements. FDI involves decisions pertaining extra costs and risks (what Zaheer (1995) has defined the "liability of foreignness"→ idea that by simply being foreigner you are going to incur into some direct and indirect costs):

- Costs of communication and acquisition of information in a different cultural, linguistic, legal, economic, and political context
- Costs of international coordination
- Costs due to less favorable treatment given by host countries' governments
- Risk of exchange rate fluctuations (purely macroeconomic risk)

→ The MNC's competitive advantage must be so strong as to overcome the liability of foreignness. Even though you are starting with an handicap you will be able to overcome difficulties and still enjoy the benefits from internationalization.

To sum up:

- Hymer is the first scholar defining the importance of *FDIs* as an *autonomous category*
- FDIs imply both potential to exploit multiple *advantages* and additional *costs*
- As always, FDI makes sense only if its benefits *outweigh* its costs

2. Dunning's eclectic paradigm (foundational theory in international business)

*Synthesis (*Eclectic = bringing multiple theories together) of previous approaches to study because firms become multinationals in a *comprehensive* framework. It is both prescriptive and descriptive.

The degree and structure of foreign activities depend on the existence of 3 types of advantages:

- **O**wnership advantages (firm-specific factors of competitive advantage)→ example might be a patent possessed by the firm or a particular skilled labor force. Why is it crucial to expand internationally? Since foreign markets are typically more hostile, you are required to have a wide set of ownership advantages to be able to compete.
- **L**ocation advantages (advantages inherent in the chosen location) → Distribution network or agglomeration of knowledge (e.g. Silicon Valley). Why is it crucial to expand internationally? If there were no location advantages (e.g. quality of transportation and communication, legal and commercial infrastructures, government policies, quality and price of inputs) the company would simply stay home. These consists thus basically in the reasons because of which the company is deciding to expand internationally.
- **I**nternalization (NOT INTERNATIONALIZATION) advantages (reasons why it is better to conduct the activity inside the firm than otherwise) → Why is it crucial to expand internationally? According to transaction costs theory, firms do exist because transaction costs are larger than coordination costs. But internalization advantages go a little bit beyond transaction costs' ones. It considers *knowledge* as well: It is better to internalize the knowledge required to do those tasks and the knowledge produce by those activities. First of all, I do not want another foreign firm doing given activity for me, because I do not want share the knowledge necessary to perform it. Secondly, whenever performing that activity, I could potentially acquire new skills and I want to internalize them.

→ *OLI* eclectic paradigm: In case the 3 aforementioned advantages are present,

Types of international expansion:

1. Resource seeking: search for cheap or productive resources
2. Market seeking: search for new markets
3. Efficiency seeking: international division of production aimed at increasing efficiency through selective exploitation
4. Asset seeking: development of international presence aimed at acquiring strategic assets

Problems:

- The original eclectic paradigm is static model
- It is well-suited to give a snapshot of the drivers of internationalization, but it fails to account for expansion dynamics
- It doesn't consider the evolution of resources and capabilities thanks to international presence
- Renewal of competitive advantage is neglected

3. Verbeke's dynamic framework

Main components:

- Internationally transferable firm-specific advantages (non-location-bound FSAs) → example might be company's brand
- Non-transferable firm-specific advantages (location-bound FSAs) → Although they belong to the firm they cannot be brought abroad
- Location advantages
- Resource recombination, and value creation through it
- Complementary resources of external actors
- Bounded rationality
- Bounded reliability

Non-location bound FSAs, location-bound FSAs and location advantages refer to the *bundle of resources* in the firm's possession. They can be distinguished based on

- *Mobility* (transferable vs non-transferable)
- *Availability* (firm-specific vs location-specific)

Location advantages can be exploited by any firm operating in that location. However, they do not benefit every firm in the same way. By expanding internationally, a MNE relinquishes its location-bound FSAs and its home country location advantages. However, it benefits from the host country location advantages

Recombination and complementary resources of external actors capture dynamic features such as innovation, adaptation, and evolution

- Complementary external resources may be needed for effective deployment of FSAs, especially in the initial phases of the expansion (dynamic adaptation)
- Resource recombination is essential to reinvigorate the MNE's competitive advantage, developing new resources and capabilities by selectively integrating existing ones with newly acquired ones
- Skillful resource recombination leads both to the upgrading of existing (non-location-bound and location-bound) FSAs and the development of entirely new ones

Bounded rationality and *bounded reliability* capture uncertainty

- It is necessary to account for the fact that human beings are boundedly rational, imperfectly informed, and inherently unreliable
- The concept of bounded reliability is very broad: besides including opportunism intended as "self-seeking interest with guile" (Williamson, 1979), it includes sources of benevolent preference reversal, such as reprioritization in good faith and failure to deliver on overcommitments

IMPORTANT examples in the message

Advantages of the framework

- It highlights the importance of the transferability of (some) FSAs abroad and their degree of complementarity with the host-country location advantages
- It accounts for environmental complexities (not only bounded rationality, but also bounded reliability)
- It captures international determinants of the evolution of competitive advantage over time

e.g. recombination of resources

To sum up

- Carefully consider the interplay between firm-specific (ownership) and location advantages before expanding internationally

- Be aware you may need to borrow complementary resources, especially in the first phases
- Modern internationalization is mostly about strategic positioning: do not adopt a short-term exploitation-oriented vision, but consider the possibility to tap into heterogeneous sources of knowledge and recombine them
- Do not forget about bounded rationality, bounded reliability, and the liability of foreignness

16. NETWORK ECONOMICS

Chart that shows the evolution of big B2C businesses in the online auction industry: after a given date, one firm (eBay) took the lead completely (the winner takes it all). This is something common in markets where *network economics* plays a crucial role – an inherent characteristic of the market scrutinized. This evolution in terms of competitive dynamics, indeed, couldn't be ascribed to eBay better performing its business compared to the others (i.e., better strategy or conduct). See the message box

16.1 An introductory example: *The story of the garbage bin.*

Suppose that in the town there are six guys that like playing videogames, Alan, Bud, Charlie, David, Eliah, and Frank.Their utility function is characterized by network externalities: their preferences depend on the intrinsic value of the game-console plus a bonus, which depends on the number of other guys owning the same type of console. Why? because they can exchange games, challenge friends, and have a greater variety of games. This is what it is meant with the term ***network externality***.

More formally: $U_i = X_k + w*N_k$, where X_k = intrinsic value, N_k = number of consumers owning type k console, and w = parameter (let us say, w = 0.2)

Let's suppose we have three different types of consoles, thus k = a, b, c. Each firm invested in *R&D* for its own console and managed to *patent* some product features. Each firm has also invested in *marketing and advertising* to win the consensus of the 6 guys. By doing so, a *standard war* occurs, i.e., a competition where firms develop and try to sell a product at the expenses of the competitors.

Note: R&D, marketing advertisement investments make sense *only if* the company can capture market shares, otherwise the value of the investment becomes null.

Yesterday, the guys received the catalogues, illustrating technical features. Therefore, tonight the guys will read the catalogues, assign their preferences (intrinsic value X_k), and make their consumption decisions. They will assign different preferences to different consoles since they value different characteristics by assumption. When they fall asleep, their structure of preferences is:

	A	B	C	Pref
Alan	0.4	0.6	0.5	B
Bud	0.6	0.5	0.4	A
Charlie	0.5	0.4	0.6	C
David	0.4	0.6	0.5	B
Eliah	0.6	0.5	0.4	A
Frank	0.5	0.4	0.6	C

So, the potential market when they all fall asleep, is shared in equal way between the companies (one third to A, B and C). What happens next?

Charlie's cat is very hungry on Monday morning and happens to turn the garbage bin upside down. Charlie wakes up suddenly at 8 am, while the other guys in town are still sleeping. Thus, Charlie thinks "since I am awake, I might go to the games shop and buy my console". He goes, and his purchase decision is obviously to buy C.

	A	B	C	Pref
Alan	0.4	0.6	0.7	C
Bud	0.6	0.5	0.6	A/C
Charlie	0.5	0.4	0.6	C
David	0.4	0.6	0.7	C
Eliah	0.6	0.5	0.6	A/C
Frank	0.5	0.4	0.8	C

From when Charlie makes his purchase, the utility of other customers is changed. The second customer is very likely to buy C. If he does so, the third will *certainly* buy C and the market will be locked in. So, C wins and takes all the pie.

Some questions might arise:

- What if Charlie was the boring guy in town? (C is the worst game console, thus he made the worst choice).

- What about company A & B? Indeed, it appears they have lost the war. There is little they can do to stop and revert the process according to which C will win the whole Pie. The only way is to come up with some radical innovation and open a new world (market) for competing.
- Are companies happy with such a high-risk profile of competition (winner takes it all)? We might argue that they are not so happy. To overcome this problem (avoid such high risky competition), they might come up with a two-stages strategy of *coopetition*: collaboration + competition, to come up with the same technological standards. So, ex-ante (before entering the market) they will try to collaborate and adhere the same technological standard. In our case, it would mean that the console companies would have come up with three consoles that are able to read the same software (game) – collaboration (ex-ante). If the first step occurs, it will be difficult that the competitive environment will turn up to a winner takes it all situation.

Further consideration: of course there are missing elements in the story, but you can already infer the role of past and casual events on today situation; how can we define business processes where a cat and a garbage bin are so important for determining the outcome of the market? There are several ways.

The most used one is called ***path dependence*** process (or 'history matters or 'ergodic' business processes). To understand where we are now in the market, we must analyze the initial conditions (the start, the early stages of the process) of the market.

"*The butterfly effect*" (chaos theory): small changes in initial conditions may lead to completely different outcomes to the ones we are in now.

16.2 Network markets' characteristics

1. Network externality
2. Strategic choice between compatibility and incompatibility
3. Switching costs & technology replacement
4. Significant economies of scale (in the production of hardware and software)

Network externality: a good exhibits a network externality when the positive change in the utility a consumer derives from it raises as the number of consumers that purchase the same good increases.

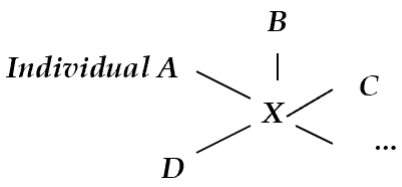
Examples: telephone (the more people exploit it, the more people you can reach and thus the more effective the solution will be), email, hardware-software, party (the more people there are, the more people will be attracted to participate. The value of the party depends on the number of people joining it), etc.

Direct externality: the value of the good increases *automatically* as the number of users increases (examples. phone or e-mail). Generally, the good has *no* intrinsic value.

Indirect externality: the value of the good increases as the number of users increases but (not automatically) only in the presence of *economies of scale in production* (of software) because of a greater offer of complementary products (or a better quality) of the good (admitting that consumers care for software variety). In this case, the good has often an intrinsic value. So, it is indirect because the increase in the value does not depend directly on the increment of users but from the existence of *complementary products*. Think of a smartphone. If no applications are developed for that specific software, users will be less inclined to purchase that can of product.

This difference between indirect and direct externality give birth to the distinction between a "two-way" and "one-way" network.

2-way network: all the networks where service AB is different from the service BA (they represent two different goods). Each knot of the network represents a user.



Direct network externality originates a 'two-way' network: The *value* of the network is a *function of the active links inside the network.*

In a 2-way network composed by *n* knots (representing an economic agent), there are *n*(n-1)* potential links. The entry of a new user produces a positive externality on existing users since she adds *2n* new potential links.

Metcalfe's Law: if a network is composed by *n users* and each user assigns a value to the network, which is proportional to the number of users (tool to estimate the value of a 2-way network):

- the value of the network $V = f(n^2-n)$
- If n is large, the squared variable becomes preponderant, thus: $V = f(n^2)$

Excursus:

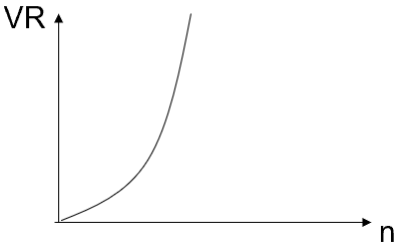
$$w = v(i)f(n) \quad i = 1, \ldots, N$$

$$f(n) = n$$

$$v(i) = c$$

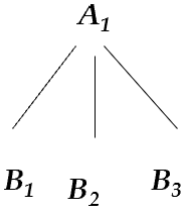
$$W = \int_0^n v(n)f(n)dn = \int_0^n cndn = \left(\frac{c}{2}\right)n^2$$

V(i) is the value each individual assigns to the network *a priori,* which is equals to a constant c. Metcalfe's law is the most used empirical rule to compute the value of a two-way network.

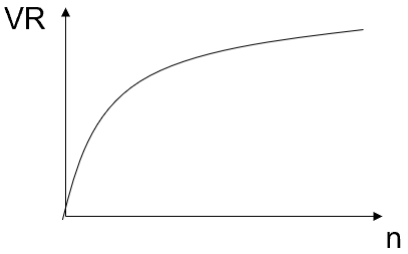


Relationship between the number of users and the value of the network implied by the law.

True relationship (concave relationship, second derivative which is negative rather than positive): *Metcalfe law could overestimate the value of a network.*



Indirect externality generates a *1-way network*: when one of AB or BA is unfeasible, or does not make any economic sense, or when there is no sense of direction in the network so that AB and BA are identical, then the network is called a one-way network.



The clearest example concerns the *hardware-software paradigm* (A is the hardware and B the software): the greater the usage of a hardware, the greater its attractiveness in terms of developing "software". Moreover, an increase in the number of software programs further raises the attractiveness of the hardware and increases the number of new adopters and so on ("*bandwagon effect*" or "*positive feedback*").

The same logic applies to the IBM – Intel – Microsoft vs. APPLE case. IBM has dominated the era of micro-computer, but has never believed in personal computers. It did not believe that everyone could own a portable computer, which was seen solely as a large processor that could be used only from every expert user. If the leader hence does not do the first step, start-ups have a free way to go. Komondor, Atari and Apple, entered in this new business of personal computer. IBM's judgement ("The size of the market for PCs is constituted by 5 people") changed after Apple 2 (1997) came out. The problem was that they had no in-house competence to come out rapidly with their own product. They decided to seek out for potential technology to be sourced out. At that time the mostly used operating system was the CPM produced by Digital Research and the microprocessors based on 8-bits by Intel. Time matured and microprocessors switched from 8 to 16-bit. This was recognized by IBM the window to enter the market and thus asked INTEL to provide them these new processors through an exclusive agreement which has been refused by INTEL (exclusivity has been refused, not the overall deal). IBM, which was able to force INTEL to let the microprocessor to another supplier, namely AMD. This gave start to the competition between INTEL and AMD**.** Precisely, in the summer of 1980 IBM managed to fly to Palo Alto trying to meet Gerry Kildall (CPM). The agreement was not stipulated; thus, IBM went to their second choice: Bill Gates. So, in 1977 Apple

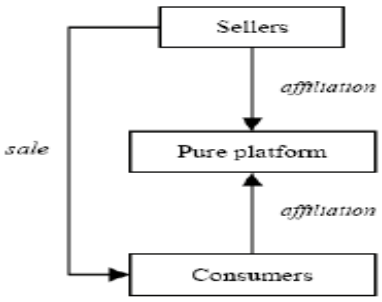
entered the market with Apple 2, whereas IBM in 1981 with its product, which was branded IBM but produced by many different players. The only fact that such a big player as IBM entered the market, pushed many software developers to write for IBM market rather than for alternative platforms. This, consequently, pushed consumers to opt for IBM products. More consumers, more incentives to develop further software and so on.

Some remarks on the story: see the message box to understand these 3 points

1. This dynamic was facilitated by the fact that IBM was forced to adopt an open standard architecture as opposed to Apple's one (closed, thus having full control of every piece produced by Apple).
2. Because of this open standard architecture, IBM-Intel-Microsoft platform became the standard. Who earned the most has been INTEL.
3. Strong non-ergodicity of business processes: at the very initial period, if the agreement between IBM and INTEL has been stipulated, we could have had a completely different story in the industry. We could not think of what we could had instead of a personal computer today if this happened.

Additional remarks:

1. **Bandwagon effect vs. negative feedback**. *If someone is experiencing positive feedback and has competitors, the latter are probably experiencing negative feedback.* The higher is the number of users of a platform, the higher is the incentive for the developers of complementary goods (SW) to create contents for such platform, the higher is the attractiveness of the platform, the higher is the number of users that use the platform (bandwagon effect or positive network feedback) and so on. The lower is the number of users of a platform, the lower is the incentive for the developers of complementary goods (SW) to create contents for such platform, the lower is the attractiveness of the platform, the lower is the number of users that use the platform (negative network feedback) and so on.



2. These dynamics also affect "*two-sided markets*": "markets where one enables interactions between end-users and try to get the two sides "on-board" → "Platform Businesses" (indirect network externalities or cross-side externalities). To some extent, the videogame industry can be seen as a 2-sided market as well, where the console is seen as the platform and the game developers and players as the two parties connected through the consoles. DEFINITELY SEE THE SLIDE (Network economics 1 , page 26)

3. Empirical Estimates of network externalities: there are multiple ways to check whether network externalities are present or not. We focus on the main one: the *complementary goods* approach (e.g., Gandal et al. 2000) derives a system of equations and uses the number of software available as a variable in hardware adoption regressions and vice versa.

Two equations thus:

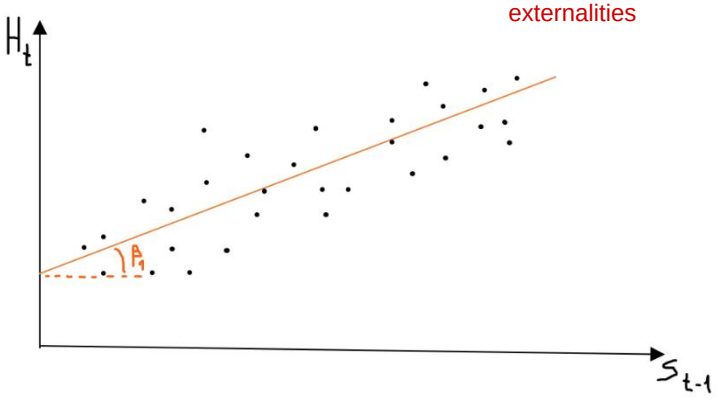
sales of hardware

$$H_t = \alpha + \beta_1 \cdot S_{t-1} + \gamma \cdot X_t + \varepsilon_t \rightarrow$$ Hardware sells depending on software developments

$$S_t = \alpha + \beta_1 \cdot H_{t-1} + \gamma \cdot X_t + \varepsilon_t \rightarrow$$ Software sells depending on hardware developments

sales of software — control terms — error terms

Both should be statistically significant and positive if there are indirect positive network externalities



16.3 Innovation in a network market: the start-up problem.

Having analyzed the previous cases, we can notice that the main turmoil is present at the very beginning of the business creation. Hence, to tackle it, we must understand how the *demand for network goods* works.

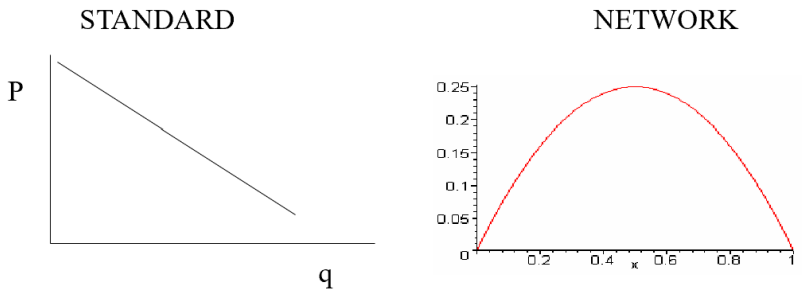
In a standard scenario, we have that the aggregate demand of a good depends simply on price. But, in case of network goods, a twist is added. Now, the demand depends on the *expectations* that consumers have on the penetration that the good will have. Why? Because the demand depends on what the other will do! This little twist has some implications, however.

1) Everybody in the economy is convinced that *nobody will buy* the network good: the network good will be effectively unsold (very pessimistic expectations) irrespective of a high or low price. Hence, this will be *"network failure"* → *self-fulfilling prophecy*: by simply thinking that something will happen, it will.

↓

Consumers want to buy a hardware if much software are available for that hardware. But software developers will write software *only if* a great number of people has already bought the hardware since they are potential buyers. Hence, if nobody makes the first move, the network good will be effectively unsold (again, irrespectively of a high or low price): *"network failure"* → *chicken-egg paradox* equilibria.

2) Since the value of what we consume depends on the quantity of how much others will consume, we *cannot assume* that the demand curve could not be upward-sloping instead of downward-sloping. The *aggregate demand* could be *upward sloping* (Rolfhs 1974, BJE):



If a firm can influence the expectations of potential consumers about the penetration that the good will have (and so its value in the future). So, it is totally plausible that as the expected size of the network increases, consumers are willing to pay a higher price for joining the network.

3) Demand presents a "*critical mass effect*", which stems from nuclear engineering and consists basically (very precise concept) in the following: under radioactive decay conditions, the amount of uranium necessary to start a self-sustaining process of production of neutrons that maintains unchanged its quantity. Any larger amount of uranium will cause an explosive nuclear chain reaction. Any smaller amount of uranium will cause nothing, and it will soon decompose.

for critical mass in sociology see the slides of network 2 page 7

Critical mass in economics: here, the concept is endogenous to the behavior of the firm rather than exogenous. The firm can choose the critical mass it can achieve. For any given price charged by the firm(s), it is the *minimal number of consumers that joins the network and is satisfied of this choice*. Any larger amount will trigger the bandwagon effect, any smaller amount will bring to a network failure. Thus, the critical mass is a *positive function of the price* (if you increase the price, you must convince more people).

Note that in these scenarios *no in-between cases* will occur. Think about the example of someone starting an applause: either he will be the only one clapping or if others start to join, at the very end, also the very reluctant to do so will clap a couple of times by inertia.

Stylized examples in network economics: firm A has invented (and potentially patented) a network good (exhibiting direct externalities) and is going to commercialize it under a monopolistic regime.

Trade-off

- High price and a high number of people to convince
- Low price and a small number of people to convince

Suppose:

- p = 1€; at that price, a representative agent may be willing to buy the good if it can be unable to communicate with (say) 100 people.
- p = 100 €; at that price one may be willing to buy the good if it can be unable to communicate with (say) 1.000.000 people.

If the choice of the firm is p = 1, the firm must convince 100 people. If the choice of the firm is p =100, it must convince 1.000.000 people (more investment in advertising and marketing expenditures). Whatever the choice, if the firm wants to sell the good, it must attract the critical mass (which is increasing in price). If p =1€ but the firm convinces only 99 individuals, someone of these individuals will be unhappy (someone who gives to the good a value of 1€ only if 100 individuals had joined the network); so, he will leave; the network size will shrink to 98 individuals, again someone of these individuals… If p =1€ but the firm convinces more than 100 individuals, e.g., 101, the value of the network good raises, and some more agents will want to buy the network good, network size becomes 102, so the value of the good raises…

Aim: attract critical mass so to trigger the bandwagon effect.

Aggregate demand for a network good: *a stylized model*.

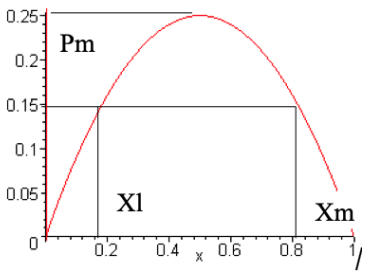
Consider a network that is of interest for N potential agents. They are indexed by *x*, which is uniformly distributed on the interval [0,1]. *x*, to some extent, gives the enthusiasm about a product. If the expected enthusiasm is high, *x* will be near to 0 and vice versa. In the former case, not many people will be required to trigger the bandwagon effect. These people will enter a priori (exaggerating the concept), irrespectively of the price. Each agent faces the binary decision of whether to buy the good or not. The good exhibits positive network externalities $U = (1 - x)n^e - p$ if she buys, and U = 0 if she does not buy. Moreover, note that the more the users, the larger the utility, and thus the positive relationship with *n*. With an *x* very close to 0, individual *i* cares a lot about the product, she is enthusiastic. Since a continuum of potential consumers exists in the interval [0,1], there will be therefore a particular consumer, indexed by *x**, such that she is indifferent between buying and not buying the good. This consumer is found by: $0 = (1 - x*)n^e - p$ *or rearranged* $\boldsymbol{x* = \frac{n^e - p}{n^e}}$

Hence all the consumers that have a higher willingness to pay for the service ($x \leq x*$) will buy the good, while all the agents that have a lower willingness to pay ($x > x*$) won't do so. Perfect foresight is assumed: self-fulfilling prophecy → $n^e = n = Nx*$ and then the inverse aggregate demand curve for a network good becomes $p = (1 - x*)Nx*$.

2 stable equilibria:

a) Network of zero size ("network failure" caused by pessimistic expectations)
b) Network of large size (critical mass is reached and positive feedback is triggered)

Critical mass: for any given price, it is the minimal number of consumers which join the network and are satisfied of this choice

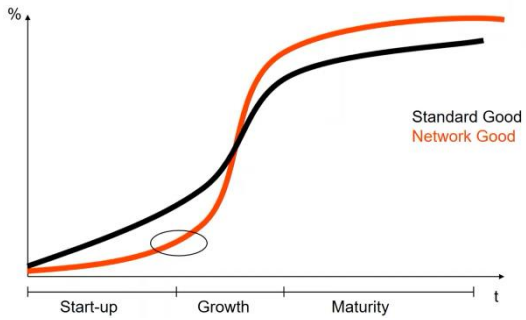


First consideration: whatever price is charged by the firm, a possibility that the network good remains unsold does exist. For people with a low X, we can say that the network effect prevails over the price effect. The converse is true as well: for people with a high X, we can say that the price effect prevails over the network effect (less enthusiastic people). In our case, given a price of 0.15, if the company can convince an amount of enthusiastic people equal to 0.2, critical mass is achieved.

1. Two stable equilibria: X = 0 and X = Xm
2. One instable equilibrium: X1

Market dynamics explained by Shapiro and Varian: "It is plausible to assume that when people are willing to pay more than the cost of the good, the size of the market expands and, when they are willing to pay less, the market contracts. Geometrically, this is saying that when the demand curve is above the supply curve, the quantity goes up and, when it is beneath the supply curve, the quantity goes down."

Typical path of a network good in a dynamic framework: the typical path of diffusion of a network good that has attracted the critical mass in the market → Successful network innovation diffusion. Typical logistic curve that describes the path mentioned before.



Comparison with the diffusion path of a standard innovation: what are the differences between network innovations and "usual" ones.

1. Reaching the critical mass can be more problematic: solving the chicken-egg paradox can be quite tricky. In general, the launch phase is more cumbersome.
2. Once the critical mass is reached, the rapid growth phase is more rapid because of the bandwagon effect.
3. In the maturity phase, the network goods can reach higher penetration than usual ones. These products become extremely important in the every-day life, indispensable.

From a managerial point of view, solving the start-up problem is of paramount importance to make the network good win. How to solve the start-up problem? *Strategic policies.*

- *Compatibility with competitors* (you are not alone): share the effort but be careful since interlinking can generate a free-riding problem. If both companies are trying to reach the same level of critical mass, one can let the other invest the most in advertising to reach the right number of followers and then exploit the already working network effect.
- *Promotional prices* (very often below costs): of course, if you choose a lower price, you have as reference a lower amount of people to convince to reach the effect.
- *Advertising* (size and nature): expectations are of paramount importance, and these can be influenced through advertising. The message to convey differ from industry to industry. Examples: In the car (BMW), you try to create the "snob" effect. In the Telecommunication industry (TIM), you try to create a bandwagon effect. It tries to convey the message that everyone has already joined the network and pushes consumers to join.

16.4 Further remarks on the possible *non-emergence* of "winner takes all (or most)" market equilibria

The first reason beside compatibility is basically the presence of *community of interest* in the market, i.e., the externalities are bounded within the community. Preferences in the market can be heterogeneous and there can be changes in interests within communities.

e.g., Suppose I use a social network that is used solely for communicating within my family. If the market is formed by this grouping, it is intuitive that each group can be served by another social network. Thus, we can have different providers coexisting without any *winner-takes-all* situation

The second reason why we do not observe a winner-takes-all market is due to ***multi-homing*** on both consumers' and producers' sides. Think about where we watch movies, for instance. Do we use solely Netflix, or do we sometimes use Disney+ or Amazon? Multi-homing is becoming more and more relevant and clearly lessen the differentiation for what concerns the hardware from the producers' perspective (more and more compatible).

Third reason concerns *local network externalities*. These resemble the situation where there will be many local dominant positions in different local markets, which does not necessarily mean to have a dominant position in another non-local one.

Of course, all three reasons can be applied simultaneously.

16.5 Strategic choice between compatibility and incompatibility

Compatibility: 2 products are defined as compatible if they "can work together", that is, the output of a given brand can be used by other brands. If this is the case, we say that different brands adopt the same standard. In the opposite case, we say that products are incompatible.

There is *downward compatibility* when a new release of a product is compatible with the old one, but it is not the other way around.

e.g., Windows Office. The updated version of Office software can read older versions, but the other way around is not true. This is often out in place by software producers to solve the trade-off and help consumers to transition from an old hardware to a new one. This incompatibility pushes consumers, who, if not fostered in such way, would probably avoid purchasing the newest version of a given software.

A product is *one-way compatible* if it can work together with the output of a rival brand, but it is not the other way around.

e.g., Linux-Windows. Linux, which is a small network of users offers compatibility towards windows-written products to survive. On the other hand, Microsoft has very little incentives to offer compatibility towards Linux products. It would be counterproductive as it would threaten Microsoft dominant position.

To better understand the consequences of this strategic choice (comp. vs non-comp.), let us focus on the characteristics of a network market. The success of a standard will depend on its capacity to solve the **start-up problem** (i.e., attract the critical mass) and trigger the bandwagon effect. The aim of the product will be to attract the critical mass. While, in case the firm choses the non-compatibility path, they must bear in mind three *stylized facts*:

1. A standard war may easily arise (winner takes-all can arise, i.e., monopoly position) keeping aside the three reasons aforementioned that explain possibilities of no standard war rise.
2. Choices of early adopters are fundamental and can determine the victory of a standard against the other ones (so early stages are extremely important for firms).
3. First mover advantage. This happens if the choices of early adopters are of such paramount importance. The firm that can launch its standard (product) will have huge advantages with respect to the competition.

Let's try to map into a more formal way the latter three aspects.

Suppose that there are in a market two *incompatible goods,* A and B. Population (N=1) is formed by agents who prefer A (a) and agents who prefer B (b) with $a + b = 1$.

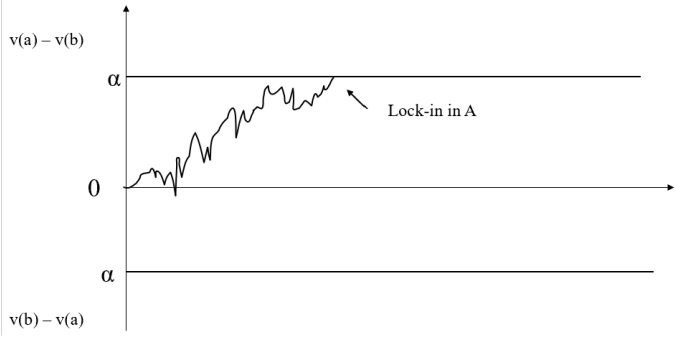
- Utility of a: v(x) if buy A and v(x) – α if buy B (alpha is a parameter that measures the *disutility* of buying a non-preferred good)
- Utility of b: v(x) if buy B and v(x) – α if buy A

The value of the good increases if the number of people using those specific good increases. Nevertheless, the *intrinsic characteristics* of the good play an important role in the choices. Let us suppose that market is perfectly competitive (PC), so firms charge the same price and suppose that being the only one to acquire the goodwill brings a zero benefit. Monopoly (M) will be an equilibrium if $v(1) - \alpha \geq 0$. If everybody has bought a good, even those who in principle do not prefer that good, will stay with that choice. Being the only one to choose a specific good in the market will bring 0 benefits.

Now suppose a *sequential entry* of consumers into the market.

Consumers *b* will buy A only if $v(a) - \alpha > v(b)$ and consumers *a* will buy B only if $v(b) - \alpha > v(a)$. If the number of consumers that are adopting already the non-preferred good is so large, the network effects will be so large to overcome the disutility of purchasing that good instead of the preferred one.

The achievement of a critical mass in a standard war



In the diagram above we observe the sequential entry of consumers in the market, where alpha (which measures the disutility) is a sort of technological absorbing barrier, i.e., once the technology touches it, we can say that the market is "locked-in". Hence, also those consumers that would prefer B would purchase A. And this is exactly what happens in standard wars.

1. We can characterize that "*tipping point*": we can either say that eBay was able to reach the critical mass before competitors or the first to achieve the "technological absorbing barrier" and lock the market.
2. *Early adopters* are more important (for reaching the critical mass): the higher are the network externalities (high v) and the lower is the love for variety (low α). A low alpha means that the goods are perceived to be extremely similar among each other. If it equals to 0, it means that the consumer is indifferent between two goods.
3. Same logics applies to the *first mover advantage*. If for some reason A is can enter the market before B does so, early adopters will have only one choice and thus A will have a higher chance to achieve the absorbing barriers.

Remark: the model also makes self-evident the risk that an inferior technology could emerge as winner of a standard war. This is a real possibility from a theoretical point of view, but (luckily enough) on an empirical ground is very difficult to affirm that such risk has ever become reality. The only controversial case is the one of QWERTY against DVORAK. Questionable whether there was path dependent inefficiency or not. Less questionable is the fact that the process has shown a higher degree of path dependency since we are adopting this system invented 150 years ago. Why is a keyboard a network good? And of what type? You may envisage this keyboard as a sort of platform that makes you communicate with computer users and computer producers and ally, we have cross-side externalities going on. The more are the users adopting the QWERTY technology, the more software and hardware developer will be keen to produce specific solutions for that type of keyboard.

2 main *forces* may spur (incentivize) companies to look for *compatibility* (standardization) and not for an incompatibility strategy.

1. Firms must bear high costs in R&D to develop a standard and in large measure they are sunk because highly specific.
2. Note also that marketing expenditure and coordination complementors costs can be extremely high when a firm choose incompatibility and opt for a go-it-alone strategy – strong non-redeploy-ability characteristic.

3 *formal ways* through compatibility among firms may arise:

1. Market standards are defined by international organizations (UL, ITU, NIST in the USA) or alliances between firms ("coopetition logic"). This is very much the case for the telephone industry.
2. Alliances: coopetition → collaboration in the upstream market and competition in the downstream. This way of conducting business will also end up in price competition since the products will be extremely similar since forced to be compatible with the same software (or hardware).
3. When a firm that has developed standards makes these standards available to other firms by making the patented technology very cheap (low licensing fees) → *Open standard policy*

It is possible to go a little more in depth for what concerns the open standard policy through a game-theory approach.

a) Battle of the sexes game. Classical in game theory, first presented when the two players were husband and wife battling whether to go to the theatre or to watch a football match

A\B	Lead	Follow
Lead	(3,3) standard war	(6,4)
Follow	(4,6)	(0,0)

→ 2 Nash equilibria will arise

→ Result: Both firms will prefer to cooperate rather than fight each other

b) The rather fight than switch game

A\B	Lead	Follow
Lead	(3,3)	(8,2)
Follow	(2,8)	(0,0)

→ 1 Nash Equilibrium

→ Both firms prefer to fight rather than cooperate. Note that here first mover advantage does not exist. We are assuming that players are moving simultaneously.

Note: a standard war leads to a worse result with respect to the cooperation situation. Therefore, this form of competition is costly. First, you must face upfront high costs to reach a critical mass. Secondly, standard war can originate a counterproductive type of inertia. Consumers may opt to wait to see which company wins before making their choice. Software will make the same reasoning as well. But, if everybody waits, it will be more difficult to escape from the network failure.

The bottom line is that in the second game, the leader is greedy for profits. If both leaders opt for that, a standard war will arise. The *lesson* is: this is an example in which firms are better off if they are cooperative or not too greedy. Firms can switch from the second to the first game if the leading firm:

- Set very low licensing fees (policy of *open standard*)
- Reduce *absorption costs* of the follower (in knowledge-intensive sectors these may be large, e.g. The Micronase case, in lecture on Transaction costs). The leader must help the follower to digest the technology that has not been developed by the follower, i.e., avoid the "not-produced ear syndrome".

More generally, the strategic choice "compatibility vs non-compatibility" implies the following *trade-off (2 strategies):*

1. No compatibility: leadership and closed standard. I developed the standard I want to keep it for myself. Possible result: a big slice of a small pie, high risk of network failure → lower probability to success: *inertia mechanisms play a crucial role; n*ow producers are producing incompatible (different) products, thus consumers will be most likely to wait and see which product will succeed → If a firm manages to reach the critical mass (success), it gets the entirety of the market (monopoly).

all the big pie

2. Compatibility: leadership and open standard. The leader sets at disposal to other firms the innovation at stake through low licensing fees or helping possible followers to in globate the technology within their products → possible result: a small slice of a big pie, more probabilities that the standard will reach the critical mass: high confidence by the agents, low risk of lock-in by potential consumers, low uncertainty by complementors → *higher probability* to succeed thus, BUT higher competitive pressures: other brands establish their position based on similar products, which have to share the same standards, i.e. very similar product. Competition will revolve around prices. *Note*: no inertia is involved neither on the software nor on the consumers' side → the bandwagon effect is easier to reach.

16.6 Cases on standard wars

3 case studies which offer a panoramic view of peculiar industry: VHS vs. BETAMAX; DVD vs. DIVX; Sony vs. Shyba.

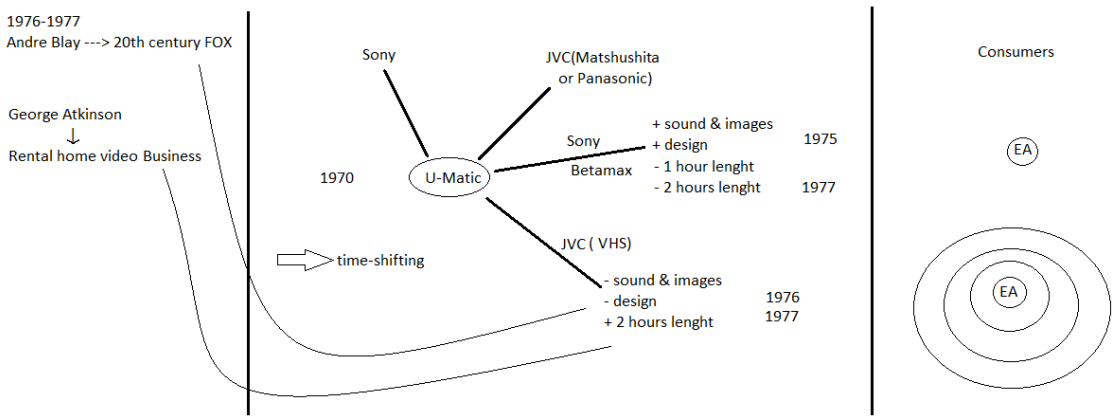
"Harvard Business Review: Increasing returns and the new world of business'"

This case will help us to understand what a standard war is.

It is *casino gambling*, where part of the game is to choose which games to play, as well as playing them with skill. We can imagine the top figures in high tech—the Gates of their industries—as milling in a large casino. Over at this table, a game is starting called multimedia. Over at that one, a game called Web services. In the corner is electronic banking. There are many such tables. You sit at one. How much to play? You ask. Three billion, the croupier replies. Who'll be playing? We won't know until they show up. What are the rules? Those'll emerge as the game unfolds. What are my odds of winning? We can't say. Do you still want to play? *High tech, pursued at this level, is not for the timid.* In fact, the art of playing the tables in the Casino of Technology is primarily a *psychological* one. What counts to some degree—but only to some degree—is technical expertise, deep pockets, will, and *courage*. Above all, the rewards go to the players who are first to make sense of the new games looming out of the technological fog, to see their shape, to cognize them. Bill Gates is not so much a wizard of technology as a wizard of precognition, of discerning the shape of the next game. Psychology is important: the most rational strategy may be to signal an irrational commitment to win the race at all costs. *Self-fulfilling prophecies* play a crucial role: If people think that you will effectively lose, you will eventually do so. *All in bet* = putting all the survival resources of your company at stake. This is analogous to a well-known result or oligopoly theory; the most rational strategy for an oligopoly may be to signal an irrational strategy commitment to punish price-cutters regardless of the ruinous cost that it may incur by doing so. The bottom line is that it is very difficult to predict how the standard war will end. Of course, there are some exogenous factors that can affect the outcome:

1. Financial resources of the firm
2. Marketing and distribution channels strength
3. Brands and reputation

1) Sony vs. JVC



These two firms developed together a technology called *U-Matics* for video recording. Very professional, sophisticated but quite cumbersome to use, thus not useful for the mass, rather suited for scientific research.

They owned the technology with a *patent sharing agreement*. However, they were not happy about their collaboration. Sony accused JVC of stealing its knowledge. JVC, on the other hand, accused Sony to exploit JVC solely for personal reasons.

Both understood that the technology, if developed correctly, could have been suitable for consumers. They developed it following two different trajectories: Sony with Betamax and JVC with VHS. In relative terms, *Betamax* had a superior quality with respect to sound and image and maybe also a better design (better cassette handling). Sony introduced Betamax in the US market in 1975, which allowed to record up to 1 hour on each videocassette. JVC entered in the market 1 year later (1976) with the *VHS*: lower sound and images quality, poorer design but larger capacity (2 hours video). Both allowed time-shifting (record some movies you could not watch streaming).

The only technical problem was that of a *blank cassette*. Is it a network good? NO. People do not care about what others do; they just care about recording they preferred tv series. We are still far away from a network market. Despite an early advantage by Sony, the early adopters' community was pending towards VHS. Why? Because the main usage of this technology was to record mainly *movies* which were lasting 2 hours. In 1977, Sony has been able to launch a new version of their product which enabled the fruition of the cassette for 2 hours. In the same period, JVC did not rest on its partial victory: they exerted their R&D investments to improve their product. VHS new product enabled the possibility to record up to 4 hours. Because of this, the community of early adopters continued to grow next to JVC's product. This is because the second usage was to record sports events (e.g., Super Bowl) which lasted more than 2 hours are still not talking about a network good, though. But such market became a network market around 1976-1977, thanks to an entrepreneur – Andre Bloy – who wrote 50 letters to the main movie producers to allow him to watch movies on these cassettes. The only one that answered has been 20th Century Fox. In this way, the video-movie industry started. Exploiting this new solution, another entrepreneur came up with the idea that video cassettes could be rented as well. These two actions changed the nature of the business making it a network market exhibiting indirect market externalities. Both entrepreneurs decided to target the larger customer community that was already existing (because less costly), the VHS community. In this way, the community of adopters of VHS continued to enlarge and a bandwagon effect occurred: entrepreneurs, seeing this increase in interest, were more convinced to invest and expand their businesses, launching thus new stores, new movies, etc.

Main *takeaways* from this story:

1. Difference between invention and innovation: it is always easy ex-post to understand which is the best technology usage, but it is extremely difficult ex-ante. The videocassette was thought for a given use that did not reveal to be the main one.

2. We see the power of software producers that determines which hardware producers will win. If the entrepreneur would have chosen to produce its movies for both technologies, we would have not seen a winner-takes-all market. Through a multi-homing strategy, this would have not happened.

3. We can see from the table depicted on the right that Panasonic (JVC), the leader of the market, whose shares remained between 10 and 20, decided to opt for an *open-standard* strategy, given its limited size, setting lower licensing fees → Higher chances to obtain the pie but a smaller share of it. But in war (as in love) everything is possible.

SALES MARKET SHARES FOR THE US VCR MARKET, 1978–1986

	Market Share (%)						
	By VCR Brand					By Format	
Year	RCA (VHS)	Sony (Beta)	Panasonic (VHS)	Fisher (VHS)	Magnavox (VHS)	VHS	Beta
1978	25.5	28.8	14.3	—	6.8	56.8	43.2
1979	34.1	38.1	4.3	—	3.8	55.0	45.0
1980	29.3	28.8	9.3	—	3.9	60.0	40.0
1981	28.0	14.2	15.3	—	4.5	72.4	27.6
1982	22.0	13.0	17.8	2.0	4.0	74.4	25.6
1983	16.0	7.0	15.0	5.0	4.7	81.9	18.1
1984	16.0	6.5	14.0	6.0	4.0	82.2	17.8
1985	13.8	4.8	12.1	7.7	4.0	87.8	12.3
1986	12.0	3.1	11.3	6.8	4.0	91.4	8.6

The five brands with highest average sales market shares in 1978–1986 are chosen in order. Thus the brands in the table are not necessarily in the top five in market share by year. Sources: *Annual Consumer Electronics Video Data Book*, 1st ed.–7th ed., and *Economic World* (February 1980).

2) The "fake" war between DVD 'Digital Versatile Disk' and DIVX: the use of vaporwave.

Digital Versatile Disk should have replaced the videocassette as technology. Sony-Toshiba-Panasonic participated to this follow along with major content creators (software) such as Columbia, Warner, MGM, and Polygon. The first *DVD* was launched in the US in September 1996. This DVD was heavily commercialized by one important home electronics retailer in US, namely 'estBuy''. The community of early adopter (for these specific DVDs) was quite satisfied with this technological solution. It was appreciating especially the great storage capacity. Some movie studios were pretty concerned about the raise of this new digital technology, which could have sparked a piracy attitude by consumers. *Macro-vision technology* was put in place to counter this problem, but it was not perceived as enough. Circuit city (other electronics retailers), since it did not want to develop the expansion of DVD, it launched an alternative to the DVD along with the aforementioned software producers and technology producer (2nd tier, Zenith): *DIVX* in 1998, one year after its announcement → *Vapourware*: a strategic declaration to control market's behaviour. ("Wait dear consumers, do not rush to buy DVDs, our technology, which is better, is coming soon"). DIVX could be played solely on DIVX players, while DVD could be played on IVX players. This technology came along an encryption code which had to be purchased every 48 hours. In 1999, the product has been retired from the market. What caused a real problem to the DVD market has not been the effective entry in the market, but rather the Vapourware.

$$\text{LSALES} = \beta_0 + \beta_1 \text{LPRICE} + \beta_2 \text{LSOFT} + \beta_3 \text{BOA} + \beta_4 \text{DIVX} + \beta_5 \text{ENTRY} + \beta_6 \text{DEMISE} + \beta_7 Q^2 + \beta_8 Q^3 + \beta_9 Q^4 + \varepsilon.$$

	Logarithmic specification (1)	
Independent Variable	Coefficient	*T*-Statistic
CONSTANT	11.71	1.69
LPRICE/PRICE	−1.70	−0.79
LSOFT/SOFT	0.25	1.18
BOA	5.55	3.74
q^2	0.31	1.82
q^3	0.47	3.36
q^4	0.61	5.75
DIVX	−0.23	−1.61
ENTRY	0.057	0.50
DEMISE	0.015	0.17
No. of observations	39	
Durbin-Watson	1.77	
Adjusted R^2	0.95	

The equation on the left depicts as a dependent variable the sales of the hardware and, as explanatory ones, its price, the software parts (movies) and the box-office return of the movies and three further three variables of interest: IVX, is a dummy variable that captures the moment in time when the vapouring has been made ENTRY, which shows when the technology has actually entered in the market, and DEMISE, which underlines when the latter has existed the market. The only statistically significant variable is in fact DIVX. Why is this possible in our context?

The strategic declaration occurred when the DVD technology had not yet achieved the critical mass, that is why it had an actual impact. So, vaporware can be intangible thing can have more tangible effects than tangible actions, such as the entry in the market from a product. The actual entry of DIVX in the market occurred when the critical mass was already reached, the bandwagon effect was already unstoppable.

Question: Was the *one-way compatibility* strategy as pursued by Circuit City a nice move? How could have been improved?

That probably was not a bad move, it was to some extent forced. The problem is that they did not come with exclusive agreements with software producers. They made agreements with big movie producers which were not forced to stop their relationship with DVD producers. Consumers had still the possibility to choose between the DVD and the DIVX, and they were just sticked to the solution previously used. Any breaks have been put on the consumers.

3) Toshiba & NEC (HD DVD) vs. Sony (Blue-Ray) for the technology replacing the DVD.

The Blue Ray was more a revolution (radical) in terms of the product, while the DVD was more an incremental one.

Note: before the creation of the video recording market (video cassettes), no one thought that it could have developed in a network market. Now, almost 30 years later (mid-2000s), they were aware of the market's nature, and they tried thus to avoid a possible standard war by stipulating an agreement:

August 2005: failure due to a) too different formats, b) too large sunk investments, c) too high absorption costs.

Launch and start-up phase: CES 2006: both formats are present and presented.

- April 2006: HD DVD commercialized at $499
- July 2006: Blu-ray commercialized at $999
- Fall 2006: Microsoft commercialize Xbox with HD DVD add-on at $200
- Fall 2006: Sony PS3 commercialized with Blu-ray at $499

Sony had some advantage in attracting moviemakers. Things became even more complicated when Paramount and DreamWorks abandoned Blue-Ray technology to produce movies solely in HD technology. Market is sluggish: "Despite tremendous efforts and investments by both sides to gain the upper hand, sales of both formats had been sluggish by any measure. The slow consumer adoption was generally attributed to their reluctance to upgrade while the format war was still ongoing. Uncertainty was also hurting the movie studios: in 2006 one analyst firm had projected that media companies stood to lose $16 billion in revenue over seven years if a single standard failed to emerge." (Takeaway, Inertia: Everybody wants to see who will win the standard war). In January 2008 Warner tipped the market towards Sony. Warner's reasons: 'The window of opportunity for high-definition DVD could be missed if format confusion continues to linger" "Within a month following Time Warner's move, electronics retailer Best Buy, movie rental company Netflix, and finally Wal-Mart, all of whom had been stocking movies in both formats, announced that they would drop HD-DVD, leading to Toshiba's surrender.

On February 19, 2008, Toshiba announced that it would discontinue production of DVD players, recorders, and parts in the HD-DVD format, which it had been championing since 2002 and in which it had invested large amounts of money to develop and bring into the market"

3 Takeaways:

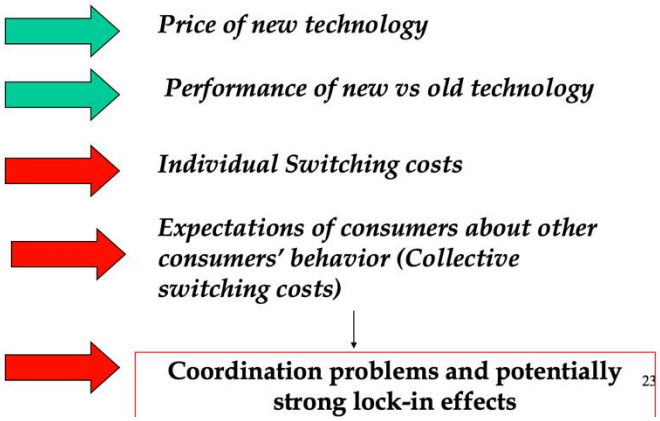
1. Since software producers are so important, hardware producers might be induced to pay them in such a way to push them towards their solutions
2. Inertia (on the consumers in particular) that the standard war may generate: Everybody wants to see what the producers/consumers will decide for
3. Precognition gift that Bill Gates has: *"[They should] understand that this is the last physical format there will ever be. Everything is going to be streamed".*

Nowadays, thanks to the phenomenon of multihoming on both sides we do not see any more one single standard (and thus company).

16.7 Third concept: Switching costs & technology replacement

Which factors determine the migration of consumers from an old to a new network technology? Switching costs. But why are switching costs so important?

When talking about *switching costs*, we refer to the cost faced by a user to migrate from the current used technology to another one. In network markets – i.e., often markets with high technology that require sophisticated knowledge – we have individual switching costs and *collective switching costs*, i.e., it would make sense for the consumer to migrate to the new good if others are doing so as well.



The presence of strong individual switching costs and collective ones makes the difference between the technology industry and the other ones. Think about a dishwasher: you will compare the price of the new technology to the residual value of the one you already have, as well as the performance of the latter against the former. These will be the only *two factors* considered. For what concerns the technology realm, strong coordination problems and potential lock-in effects.

Specifically, which are more these individual and collective switching costs?

1. Individual:

Learning & Non-transferability of complementary assets *(*i.e., sunk costs given by non-fully redeployable product-specific investments) – search costs, loyalty costs (psychological).

e.g., Think about a new operating system: you need to invest human capital (time) to learn how to use it properly. Suppose that there exist a new OS in the market. To adopt it, you will have to disregard the human capital invested to learn the old one and invest further to learn how to exploit the new one. These are very strong individual costs that are not present in other industries such as in the car industry. i.e., once you have learnt to drive on a Fiat, you will be able to drive a Mercedes as well. A further aspect that underlines high switching costs is potential non-transferability of complementary assets.

2. Collective switching costs:

the value of the new technology will depend on how many users will switch (because of network externalities). Thus, a new chicken-egg paradox to solve, which may prove to be more difficult than an ex-novo situation, since users already satisfy to some extent their needs with the existing technology.

Suppose there are 2 users that should decide if they want to shift from the old hardware platform to the new one.

Note: the payoff of each user depends on the decision of the other user because of network externalities.

Network externalities:

- $\alpha > \delta, \gamma$
- $\beta > \delta, \gamma$

1/2	New technology	Old technology
New technology	α, α	γ, δ
Old technology	δ, γ	β, β

Since we are dealing with network goods, the utility of both consumers receives making *the same choice* is greater than the one when the two consumers opt for different technologies. This is the reason why both alpha and beta, which represent the utilities given the same decision, are larger than both delta and gamma, which consist in the other two scenarios where both consumers opt for different choices.

Two Nash equilibria are present in this game: (new, new); (old; old)

- $\beta < \alpha$ and (old, old) is the market equilibrium: *excess inertia* (strong) → both typology of consumers stay with the old technology despite the fact that they would be better off switching to the new one. "They are waiting too much".
- $\beta > \alpha$ and (new, new) is the market equilibrium: *excess momentum* (strong) → despite the fact that both consumers would be better off in keeping the old technologies (no need to hurry up) they still have decided to migrate to the new one.

Things might be even more blurred than just the simple categorization of two extreme phenomena (strong inertia and momentum):

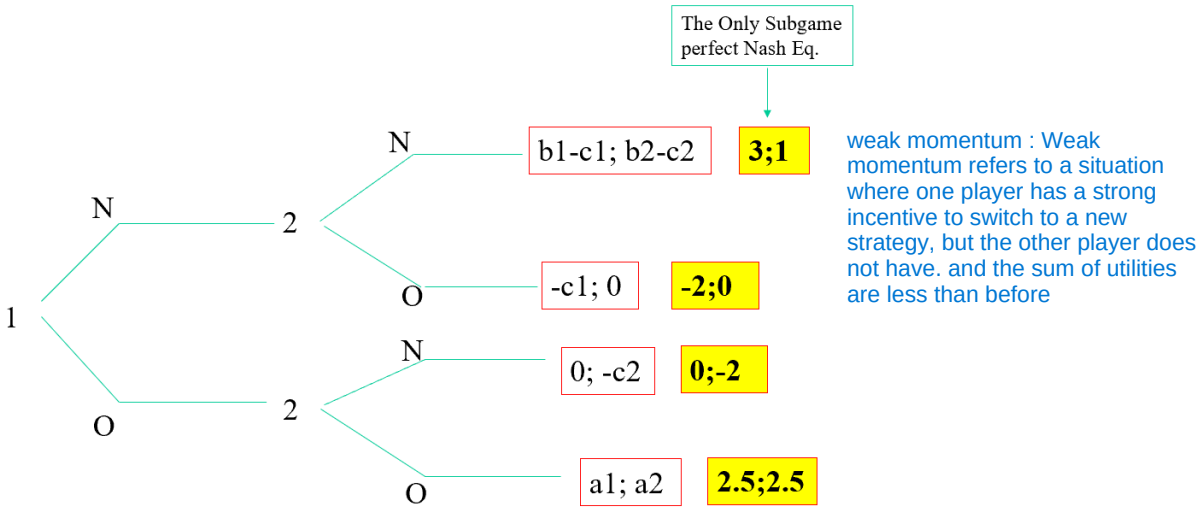
1/2	N	O
N	b1-c1; b2-c2	-c1; 0
O	0; -c2	a1; a2

If (b1-c1) = (b2-c2) > 0 [same analysis as before, 2 NE, potential excess inertia, or excess momentum, depending on actual values]. But (b1-c1) ≠ (b2-c2). Suppose (b1-c1) > a1 and (b2-c2) < a2. For example, b1 = 5, b2 = 3, c1= c2 = 2, a1 = a2 = 2.5

1/2	N	O
N	3; 1	-2; 0
O	0; -2	2.5; 2.5

2 NE as before (N; N) and (O; O) but now no Pareto dominates the other → excess (momentum or inertia) is limited to segments of population. Let's suppose the N, N equilibria comes out. Consumer 1 is better off, while consumer 2 is affected by excess momentum (high willingness to change).

Weak excess momentum in a sequential game:



Here, 1 consumer decides first and then she is followed by the other one. Since Consumer 1 is the first to move, consumer 2 will have no choice other than opting for the same new technology b1-c1 (slightly) > a1; b2-c2 (slightly) > 0 and < a2 → N, N

But (b1-c1) + (b2-c2) < a1+a2 [the gain of "1" type does not compensate for the loss of "2" type in changing]. The sum of the subgame perfect Nash equilibrium is 4, while there is globally another equilibrium (which will never happen, and does not Pareto dominate the other) in which we would have a larger sum of utilities. This underlines the *weak momentum*. If you would reverse the sequence letting player 2 decide first, we would have obtained a type of *weak inertia*.

Excess inertia example: AM vs. FM. FM is a better broadcasting technology → more stable and longer distance. Despite that, it struggled a lot to impose itself on the already existing technology, which was by far inferior. However, *we are only talking about inertia (as we know, in the end, in the long run, the best technology is likely to prevail).*

Excess momentum example: "Software updates are sometimes an example of excess momentum (the term *forced upgrades* is also used in this context). Many users of Office 95 were quite satisfied with their version of the software. However, when several key players adopted Office 97, owners of Office 95 had an incentive to switch to Office 97 for reasons of compatibility. This does not mean that, in the end, these users have been better off. They would probably prefer the initial situation when every user worked with Office 95. However, the new situation (everyone with Office 97) is better than sticking to Office 95 while the rest of the world switches to Office 97. Switching to Office 97 is then the lesser evil, as it was."

16.8 Final thoughts

Always keeping in mind caveats and exceptions (e.g., *community of interest, local nature of network externalities, multi-homing phenomena)*, we know that the choice of "non compatibility" ("standard wars") can frequently lead to global monopolies.

What can the losers do? Little can be done in a war for losers to revert the bandwagon effect. The only thing they can opt for, is opening a completely new game through a radical, disruptive technological innovation.

In network markets, competition and rivalry could not be *within* the market but rather be *for* the market. Monopoly can only be temporary, and it can always be interrupted by technological breakthroughs developed by startups (Schumpeterian competition, SM1). If this is the case, market shares can be a poor indicator of the degree of competition and rivalry that exist in a network market.

But network monopolies are – probably – becoming more and more stable. Why?

3 Reasons:

1) Network monopolies are now *conglomerates* with enormous advantages (financial resources but also data) and they are persistent top investors in R&D (SM2). Thus, the largest companies are top performers in R&D (those who invest the most to come up with new technologies). Therefore, these companies are still dominating the market they are competing in.

2) The narrative of a *successful entrepreneur* is now the one who "exits" by selling its business to an incumbent (lack of SM1). Recalling that one of the main reasons for being an entrepreneur was to establish "a private kingdom" achievable through entrepreneurship, now the narrative changes. Successful entrepreneur is the one who has managed to find a profitable startup but who at a given time wants to exit the market by selling its company to a larger competitor/incumbent.

3) *Killer acquisitions*: many of these large companies (GAFAM) have a strong tendency to acquire startups, to make "corporate shopping". Many observers fear that at least a considerable portion of these M&A are not for developing the technology, which stands at the basis of the startups, but rather to kill it once it is in the cradle because they do not want to risk that such technology will threaten their leader position.

→ data that corroborates these statements: Google has bought 1 firm per month for the last 17 years.

Incumbents cannot relax. They always must stay *on guard* to detect the threat for their positions. The most emblematic example is represented by the famous case study about the first internet browser which saw Microsoft vs. Netscape. Microsoft has been alleged to have implemented anti-competitive strategies.

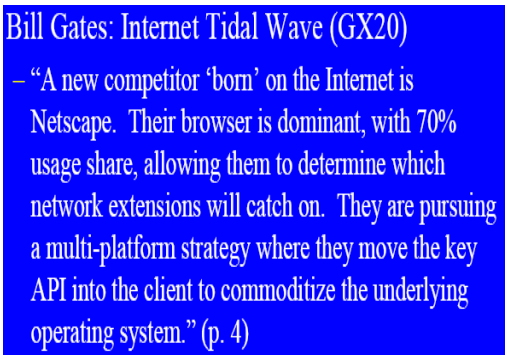
16.8.1 Microsoft vs Netscape case

Netscape has been launched in 1994 → Early story of internet. First good-quality solution to browse on the internet. It costed from 40$ to 60$. → "Middleware": something that is in-between an operating system and software. Netscape was a particular type of middleware, namely an API. It allowed to read many different software. At the very beginning, Netscape was not attached to any particular Operating System (OS).

Microsoft launched its first version of Internet Explorer in 1993, but due to its poor quality was far to be a competitor for Netscape. Why? Microsoft was not interested in competing in the browser market. However, the sole appearance of Netscape made sound the alarm for Microsoft. Why? Any computer very soon with Windows OS would have had inside Netscape Middleware. Fear came from the fact that software producers would have stopped to write programmes for Windows and start doing so for Netscape instead. By doing that, they knew they could have Windows users but others OS users as well. Microsoft understood this threat: From 1995 to 1999, the number of research devoted to the development for the browser passed from few units to thousands, and the budget increased consequently (up to 1.000.000$ a year). The first comparable version to Netscape was "Internet Explorer 30 / 40) accompanied by some strategies:

1. Commercialized at a price equal to 0. It was bundled with Windows. At the first it was just a bundle, but later it became what is called a "technological tying". If you would have deleted the product, this would have jeopardized the well-functioning of the entire OS.
2. Agreement with PC producers (first anti-competitive behaviors) which were forbidden to insert Netscape Navigator otherwise Microsoft would have reconsidered the license which was granting them.

Internet Explorer cut the air to Netscape, also because it was free.



Commoditization of OS → OS becomes a commodity; thus, buyers are indifferent between purchasing one or the other. Windows feared that this would have happened.

Bill Gates precognized the risk of OS commoditization →

New challenges & futuristic open questions:

1. Cloud computing and SaaS → Shift in how we consume software. Once it was something within a machine. Nowadays, thanks to the cloud, we are buying a service that allows to use the specific software. Many envisage tha OS will become a door to use all the software → Commoditized.

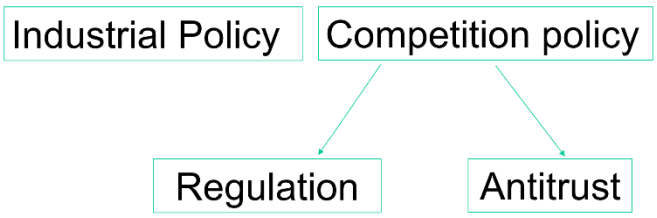
This further opens other questions:

- Will AR vs VR prevail?
- Will metaverse (or similar) completely replace desktops?
- How many metaverses we will have?

17. INDUSTRIAL AND COMPETITION POLICY

A necessary premise: We will mostly talk about micro and not macro interventions (e.g. monetary policy and/or general fiscal policy).

Industrial Policy vs. Competition policy, which can be subdivided into:



1. Regulation: rules set in advance, and they are specific to the markets which present natural monopolistic characteristics. Very vertical type of policies, tied to specific industries. "Ex-ante" regulation.
2. Antitrust: "Ex-post public intervention" → There to punish anti-competitive put in place by firms and it is more horizontal: it applies to all the firms in every sectors.

17.1 Industrial Policy

Industrial Policy: Definitions (among many others): Boundaries are not so needly defined. Slightly different definitions do exist.

"Industrial Policy is a nation's official total effort to influence sectoral development and thus, the national industrial portfolio" or

"Any type of selective intervention or government policy that attempts to alter the structure of production toward sectors that are expected to offer better prospects for economic growth than would occur in the absence of such intervention, i.e., in the market equilibrium."

Professor's own definition which tries to take the common parts of all the others: Bottom line that connects them → Industrial Policy implies the choice of the government to prefer some type of firm, activities and sectors rather than other:

"Industrial policy is any policy that affects a subset of firms, firms' activities and industries differentially from the remaining group of firms, firms' activities and industries. Any tax, subsidy, trade and other policy measure that affects only a limited and specific domain of a nation's production system can be considered as an industrial policy intervention."

Examples:

1) Support to young innovative companies (L. 221 Italy 2012, MISE)
2) (Some elements of) Italian PNRR (Italy's National Recovery and Resilience Plan).
3) Sustainment to specific sectors (e.g. Automobile industry) → Subsidy offered to consumers to purchase electric cars. Here, a specific industry is preferred and sustained.

We can classify **industrial policies** according to different dimensions:

1) **Offensive** (support to specific industries able to play a major role on the global value chains, e.g. "smart specialization") and **defensive** ("protectionism", "save losers" or "sunset industries") policies, where defensive in the short-term may become offensive in the long-run (successful "infant industries" policies → to compete afterwards internationally).
2) **Horizontal** (You sustain certain activity irrespective of the industry) vs. **vertical** industrial policies (Choosing a sector)

Different rationales behind its applications**:** economics (e.g. market failures), social justice (e.g. employment safeguard), national strategic aims (e.g. "national champions" policies→ we do not want that a foreign company purchases national firms because of several reasons), etc.

Different outcomes: Political and lobbying influence. **"Government failures"** → Even if the intervention aimed to increase social welfare fixing possible market failures, government might fail in doing so (benevolent view). Malevolent view: "politicians do not really want to increase social welfare but rather to make their own interest disregarding the community they are serving [apart from lobbying activities, i.e. often effectiveness of policies depends "on get the implementation details right", and this may be tricky details (Duflo, 2017)] .

Having said this, in order to make things more compelling, in dealing with this topic of industrial policies we will stick to the "main leg", namely innovation policy. Industrial Policy, at least in its "offensive" component is more and more coincident with its subset of Policy Innovation.

"*INNOVATION has become the industrial religion of the late 20th century. Business sees it as the key to increasing profits and market share. Governments automatically reach for it when trying to fix the economy."*

Normative arguments: how things should be.

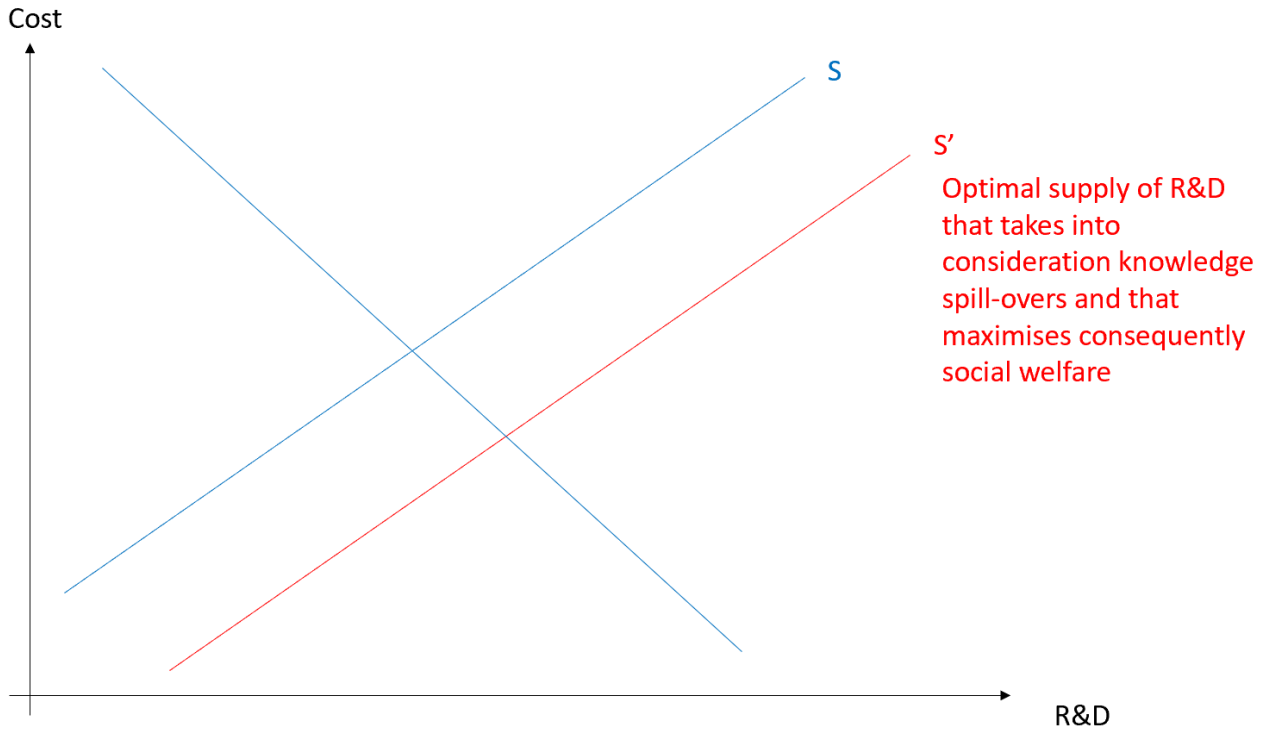
Social Welfare is basically made of three different efficiency concepts: Productive and Allocative efficiency. And the third one related to innovation, namely dynamic efficiency. Whenever implementing a policy, a trade-off between favouring one concept of efficiency among the others does exist. Usually these three concepts do not go hand in hand. For instance, whenever antitrust has to decide whether to allow two firms to merge by comparing the gains that can be gained in terms of productive efficiency (economies of scale and synergies) vis-à-vis with increase in market concentration→ Loss in allocative efficiency, though.

17.1.1 The role of Innovation Policy

Has public policy a role in the whole innovative process? Of course yes and at multiple levels! The innovative activity in the economy is strongly influenced by several forces, including competition policy, regulatory and law regimes, patent system (among others) all contribute to shape the interested dynamics. Without forgetting the institutional and cultural context. When we talk about innovation policies we are talking about direct interventions to ameliorate the situation rapidly.

Rationales for public sustainment to R&D activities: Private firms (may) invest less than the social optimum. There exist 2 rationale to justify market interventions.

1. Knowledge spill overs (positive production externalities)→ In the picture the firm supply is based solely on production costs whereas the right level in order to maximize Social Welfare should equal to S'.



2. Very large capital market imperfections, where asymmetric information (adverse selection and moral hazard problems) plays a crucial role.

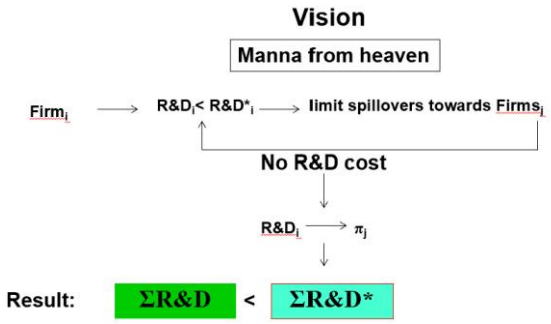
1) Knowledge Spill overs: "The primary output of R&D investments is the knowledge of how to make new goods and services, and this knowledge is not rival- use by one firm does not preclude its use by another. To the extent that knowledge cannot be kept secret, the returns to the investment in it cannot be appropriated by the firm undertaking the investment, and therefore such firms will be reluctant to invest, leading to the under provision of R&D investments in the economy". → Rationale: When investing in R&D you are producing knowledge. The problem is that you will not able to reap every single knowledge created, some of this will spill out to competitors. You do not want this to happen, thus you will invest less in R&D so that competitors will not exploit your R&D findings. Thus, the socially optimal levels of R&D will not be achieved.

Two annotations:

1.The argument rely on the stylized facts that any formal and/or informal mechanism to protect innovation is **only partly efficient** at the very best, whether they are patent or lead time, etc..

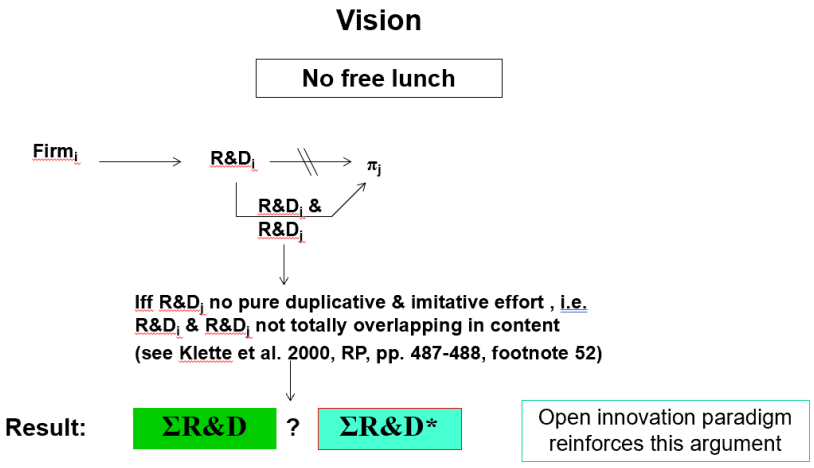
2. The under-provision of R&D is more severe for high-tech start-ups (the younger and the innovative, thus) due to the higher relative costs in protecting innovation (e.g. lack of complementary assets). Failing in patenting might be quite costly, thus the younger, the smaller the resources you have to devote to this scope. In tech industries complementary assets play a crucial role: the younger you are, the less resources you will have to devote to purchase complements.

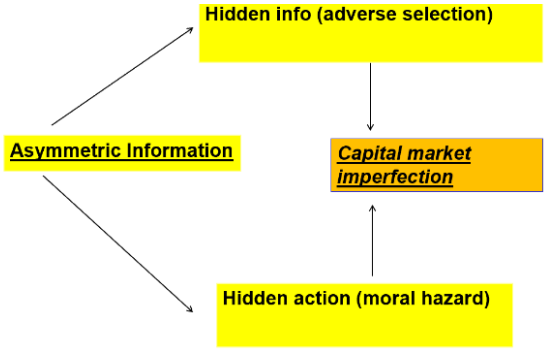
Critique to the spillover rationale: This argument very much rely something called "Manna from Heaven vision": Firm I invests less in R&D from a social point of view because it wants to limit spillover to firm J, which might exploit these latter to increase its profits without any costs in further R&D. Total R&D expenditure will be lower than the social optimum.



This cannot always be taken as a realistic vision. In reality, a more "No free lunch" vision exists.

Firm I investment in R&D does **not automatically** transform in knowledge for firm J, which has to combine the knowledges spilled over with its own R&D. In doing this, firm J will be able to increase profits. Thus, if firm j is not more pure duplicative & imitative, but rather complementary → the "Manna from heaven" vision does not apply. We are not sure anymore that the total R&D incurred in the system is actually lower than the social optimal one. From a more managerial point view, firms have to shar their knowledge and everybody will benefit from it, becoming more innovative.





But does the mere existence of spillovers justify policy intervention?

Capital markets heavily suffer from asymmetric information problems. Private actors might put in place some mechanism to solve the problems, which are unfortunately not always perceived as effective.

Empirical analyses generally confirm greater financial constraints suffered from R&D investments than other typology of investments. Even if not unanimously shared, the fact that R&D may suffer from capital markets inefficiencies. The problem is held to be more acute for young startups active in high tech markets. Why? Because (by definition) they do not have any prior performances to showcase the goodness of their business and high-tech because it Is for sure more risky than other typology of industries.

17.1.2 Policy Instruments to sustain R&D

3 main instruments that the government has at its disposal to alleviate these problems (they are not solely 3, but nevertheless we consider these 3 as the main important ones):

1. Fiscal incentives to R&D activities → State offer rebates of capital gains or income taxes
2. Grants to R&D
3. Innovative entrepreneurship policy

One important aspect that differentiate fiscal incentives (tax credits) from grants is related to the process: For grants, a firm shall participate in a contest in which only a winner wins the subsidy, i.e., the grant. On the other hand, fiscal incentives are automatically disbursed to all those firms that meet certain requirements set by the law.

What do we know about the performances of these schemes? Rather than to come out which context is better than another (which would be too difficult), let's get acquainted with the possible outcomes that might achieved:

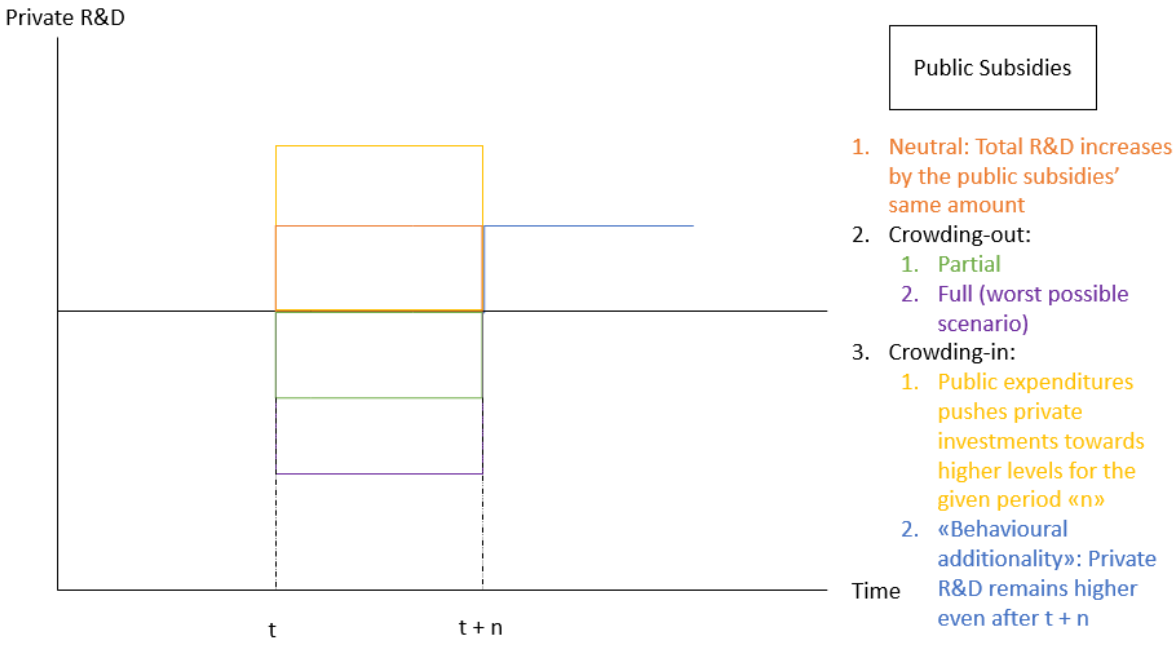
17.1.3. The "Crowding-in" or "Crowding-out" issue

Tons of literature on the crowding-in or crowding-out effects. 4 different outcomes:

→ 1) Crowding-in (additionality): public R&D subsidies increase private R&D expenditure. This might be solely during the given time span in which public subsidies are disbursed (limited effect) or continue even after.

→ 2) Neutral effect: public R&D subsidies do not stimulate neither depress private R&D expenditure. On one hand, public subsidy increases overall R&D budget of the firm so it is likely that R&D projects are undertaken that would have not materialize in the absence of the subsidy. On the other hand, subsidies must be financed (with taxes) so benefits are uncertain.

→ 3-4) (Full or partial) Crowding-out (substitution): public funding mere substitute private funding and there is no (or few) undertaking of new R&D project. Firms would have realized (many of) those projects even in the absence of the subsidy.



17.1.2.1 Grants vs. Tax credits (automatic fiscal incentives)

What are the problems with grants? **Selectivity**: Public operators might be incompetent (less competent than private agents) in finding which is the best project to finance than private. Public agents risk less compared to private ones. **Cherry Picking**: To avoid the public opinion to criticize how the government has invested the money (wasting them in a bad project), it will risk less and invest in the best projects that are already financed by private operators and thus do not need any finances. No one can criticize a possible waste of resources this way.

Major Pros with respect to tax credits:

1. No complications to the tax system
2. No room for recipient firms to re-label as R&D activities that are not.
3. Possible to send funds where most needed (from a social point of view)

Big implication: With automatic subsidies, the Matthew effect ('Piove sul bagnato') might be generated: Enabling all those companies that meet the requirements to receive the subsidies, even though they do not actually need them.

Note that Grants have in principle another main "indirect" advantage. In a context in which competition between firms is very high and the disbursement is decided by a **technical committee** (high expertise in the matter) a "Stamp of approval effect" (aka 'halo' or 'certification' effect) is generated. This will allow to open doors that were not accessible before to the firm/start-up. It signals how good the start-up is: the "stamp of approval" effect proves to be important also for high-tech start-ups in the European context. It **solves the asymmetric information problems**. Selective (grants) rather than automatic (fiscal incentives) proves to be beneficial

Policy intervention: 3rd Type of instrument that policy has on its quiver to increase the R&D activities of the economic system:

Important premise**:** Entrepreneurship policy ≠ Increase Entrepreneurship rate

1. Entrepreneurship policy in general does not mean aiming at increasing entrepreneurship rate but rather means to ***put in place the conditions that enable the most capable entrepreneur to become such*** → "Why encouraging more people to become entrepreneurs is bad public policy" → Why? 2 risks: **deadweight and substitution effects**:

→ "[…] If the selection process is the outcome of a Bayesian process of learning (recall "theta" parameter concerning the competitive selection model, where the company learns about it and adjust it once it has entered the market) a subsidy may be both useless (the more efficient entrepreneur does not need it, while the less efficient one leaves the market once the subsidy ceases to be in operation) and harmful (less efficient entrepreneurs are given an artificial seedbed, while market competition would have induced them to leave the market → because of the subsidy they will remain in the market although by looking to their theta they would have left). If the former situation is prevalent, the industrial policy supporting entry is affected by a ***deadweight*** component; if the latter is prevalent a ***substitution*** (staying in the market you may rob the market from more efficient entrepreneurs which will not enter because not finding space in the market) effect arises".

Several recent interventions after the Great Financial Crisis tried to stimulate "innovative" entrepreneurship with the tendency to overarch all these measures under a unique coherent umbrella (as a proper industrial policy should do). A policy for innovative entrepreneurship that does not look only at entry but which tries to sustain the growth of the firm and beside that looks at the possibility of failure (tries to reduce the burdens, legal, and cultural attached to the failing of firm) may be more efficient than those looking solely at the subsidies for entry.

→ The Italian Case:

- Regulatory change in Italy (2012) – "The Startup Act" intended to spark the national innovation ecosystem.
- Targeted Young Innovative Companies (YICs, aka "Innovative Startups").
- Requirements:
  - <5 years old,
  - <€ 5m annual sales,
  - Not listed,
  - No corporate spin-off,
  - Innovative:
    - Tangible IP rights (e.g. patent, license);
    - R&D investments >15% of the revenues;
    - >1/3 of employees/founders must hold a PhD or >2/3 must have a master degree.

According to the requirements met, several different incentives are put in place.

Failing in the US: Failure is very valuable experience (See. Picture)

The European attitude towards failure: From "The Buddenbrook" by Thomas Mann (1900):

"Grünlich is bankrupt?" Tony asked under her breath, half rising from the cushions and seizing the Consul's hand quickly. "Yes, my child," he said seriously.

"[….] "Oh, my God!" she suddenly uttered, and sank back on her seat. In that minute all that was involved in the word "bankrupt" rose clearly before her: all the vague and fearful hints which she had heard as a child. "Bankrupt" — that was more dreadful than death, that was catastrophe, ruin, shame, disgrace, misery, despair. "He is bankrupt" she repeated.

Note: **Success in Innovative Entrepreneurship** → "For a new, high-technology firm, the primary assets are the knowledge and skills of the founders. Any competitive advantage the new firm achieves is likely to be based upon what the founders can do better than others High importance of Human Capital

17.2 Competition policy

Competition authorities (Antitrust and Regulatory Agencies):

A. Anti-trust: general target & (mainly) **ex-post** → Anti-trust function:

Punish anti-competitive behavior put in place by firms <u>and</u> Prevent the formation of artificial dominant positions in markets

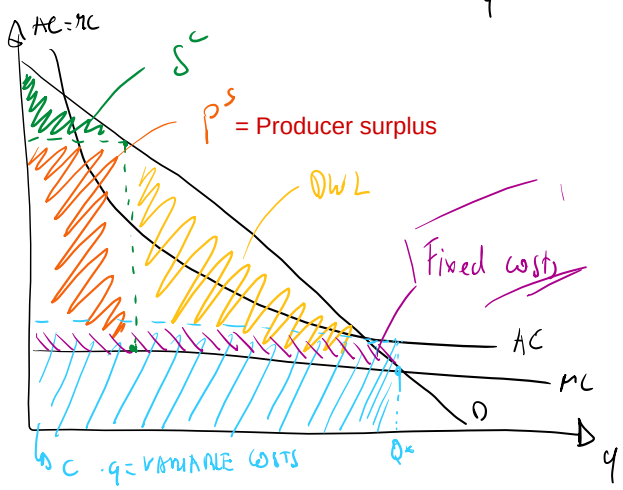
B. Regulation: specific target & (mainly) **ex-ante** → Regulation function

Identification of natural monopolies <u>and</u> Impose conducts to firms so that competition (and its results) can be mimicked

17.2.1 Regulation: One leg of competition policy

What is a natural monopoly and what the implications of a natural monopoly existence within a market.

If $TC = F + cq$ and $AC = \frac{F}{q} + c$ → AFFINE cost curve → $\frac{\partial AC}{\partial q} = -\frac{F}{q^2} < 0$



If an affine cost curve is present, we will be sure that a natural monopoly exists. Indeed, assigning production to a single firm is for sure the best choice in terms of productive efficiency but its behaviour must be taken into account! This will NOT be efficient in terms of social welfare (objective: to maximise OWN profits. Thus, the solution might be to establish a PUBLIC MONOPOLY which objective will be to **maximise social welfare**

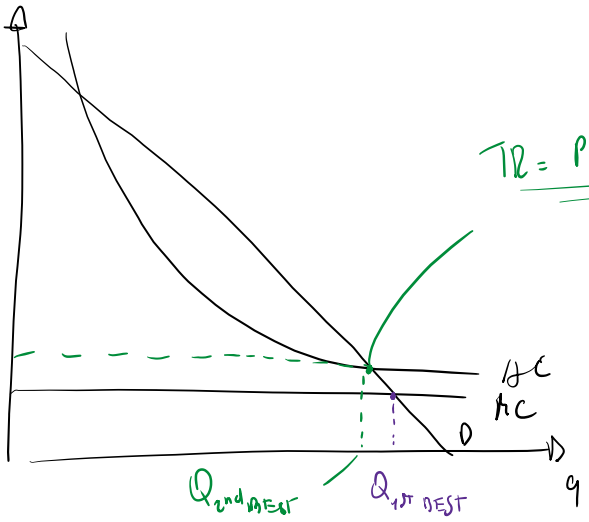
→ FORMALLY: $W = \int_0^q p(x)\,dx - TC(q)$

$\Rightarrow \frac{dW}{dq} = p(q) - MC(q) = 0$

$\Rightarrow p(q) = MC(q)$

First Best solution is NOT implementable for a private firm because of the losses that represent FIXED COSTS

2ND BEST SOLUTION:



$TR = P \cdot q = AC \cdot q = TC$

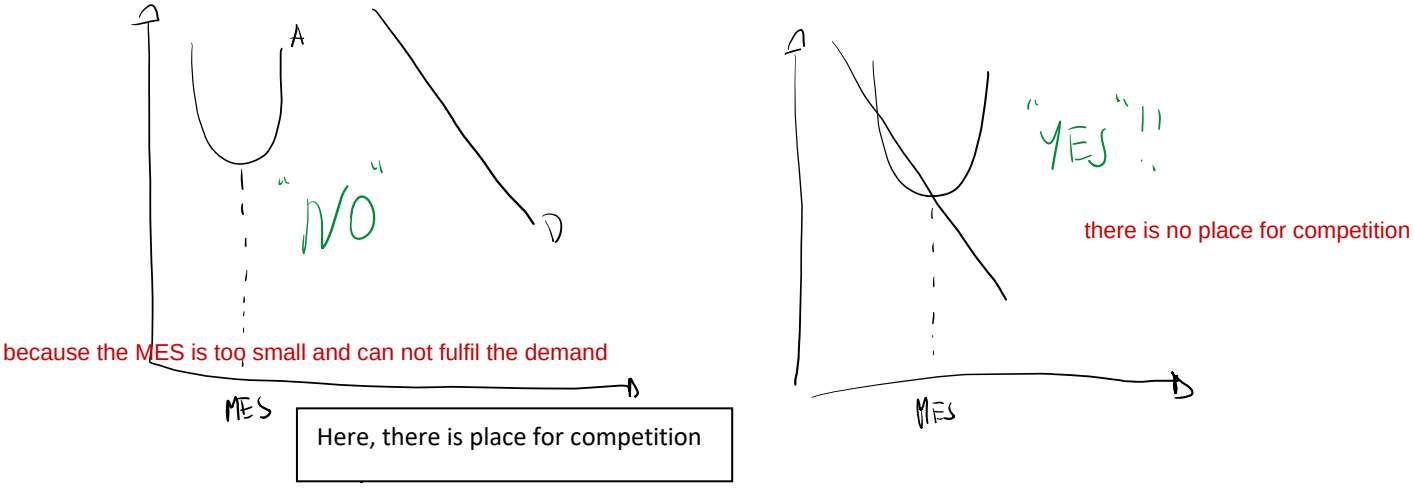
↳ NO LOSSES!

The second best solution allows the firm to still recover the fixed costs!

Given this framework we pose ourselves the following questions: How to regulate a sector that exhibits this characteristic of natural monopoly? → Vademecum ("roadmap to regulation") offered by **Braeutigam's** in 1989. These guidelines are presented by posing three different questions and from the answers of these we might understand how to regulate these sector.

17.2.1 **Braeutigam's** Vademecum

1. Is there a Natural Monopoly? Is there a MOS (Minimum Optimal Scale), aka MES, so that the competition within the market is not possible? If the answer to such question is NO, the best thing that the regulator can do is to avoid any regulation and let the market achieve the best condition. If the answer is YES, thus the MES is very close to the demand curve, we need to regulate the scrutinized market.



because the MES is too small and can not fulfil the demand

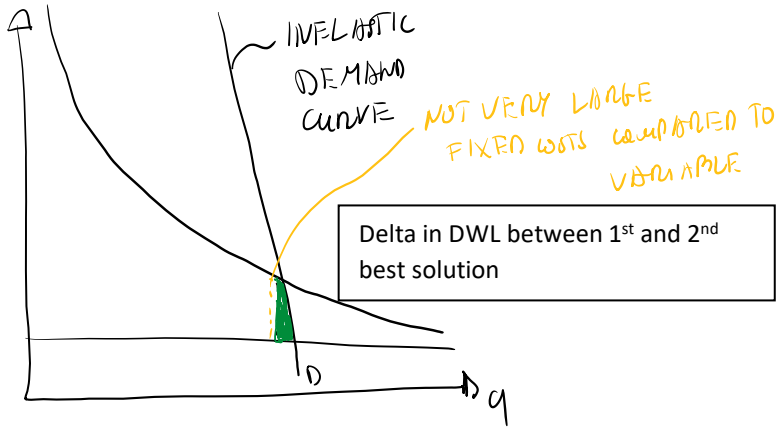
there is no place for competition

Here, there is place for competition

If we need regulation, the 2nd question will be the following:

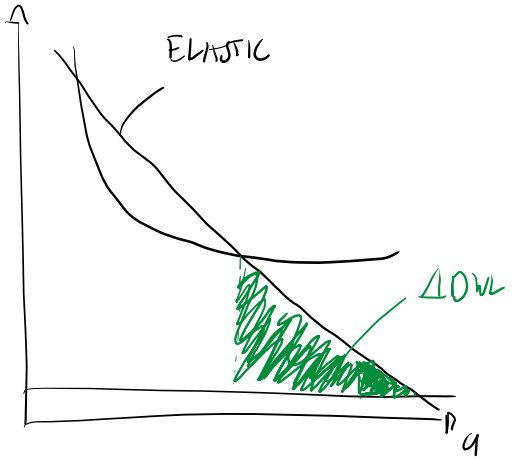
3. Is the second best solution in terms of Deadweight Loss tolerable? If NO, the regulator must try to achieve the first best outcome. If the answer is YES, we need to answer to third question.

Yes, tolerable: DWL is low



Delta in DWL between 1st and 2nd best solution

No, not tolerable: DWL is too large



1) The more elastic the demand curve, the larger the base of the triangle

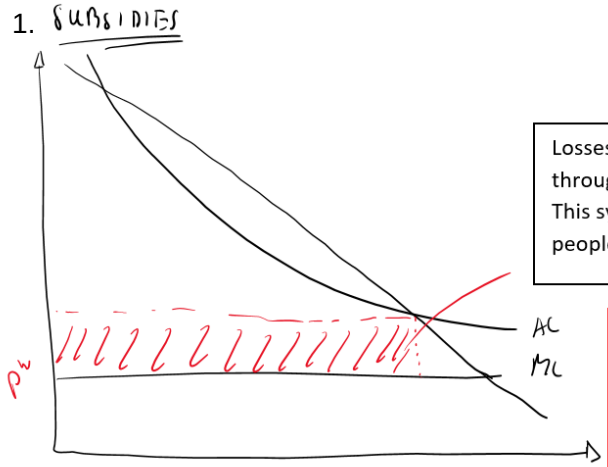
2) The larger the difference between fixed costs and variable ones, the taller the height of the triangle.

3.Can there be competition for the market? Is the market contestable? If the answer is NO, regulator will achieve the second best solution. If the answer is YES, introduce competition within the market

1. Densest competition
2. Contestable markets
3. Intermodal competition

Note: Fixing AC = P requires the regulator to estimate how large AC are, which might be quite cumbersome.

1. SUBSIDIES



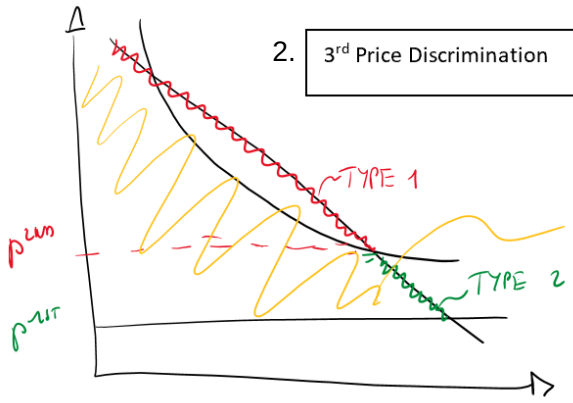
Losses are repaid by the government through general tax revenues…Problem? This system is also financed by those people that are NOT consuming the good…

Price Discrimination Schemes:

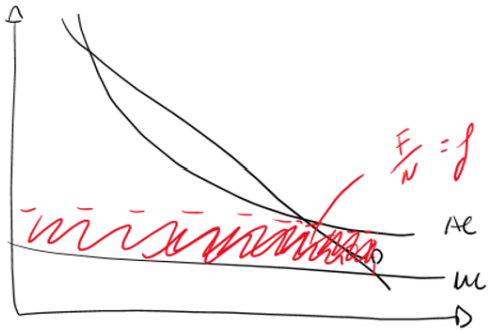
1. 3rd P.D: Consumers are divided into high and low willingness to pay and will be charged the amount they are willing to disburse. Consumers willing to pay high amounts will help the firm recover fixed costs while those willing to pay smaller amounts will be charged the 1st best solution price

2. 2nd P.D.: It can be divided into Two Part Tariffs (TT) or "Peak and Low Pricing"

2. 3rd Price Discrimination



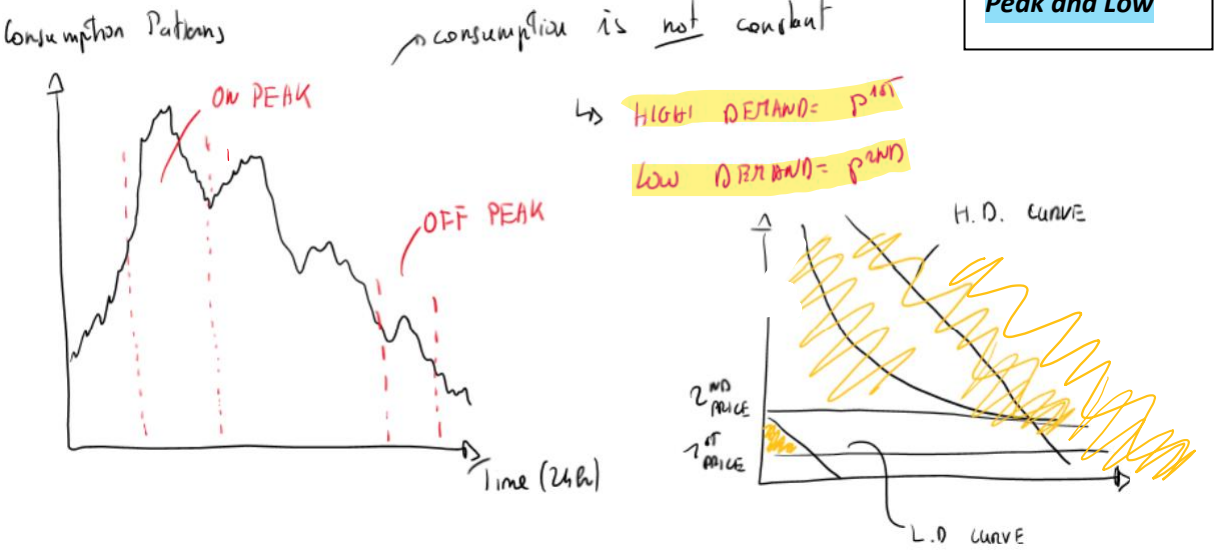
Problem: It is quite difficult if not impossible to identify these groups…



3. Two Part Tariffs: Consumers pay an Access fee in addition to a price for each good consumed.

Formally: TT = f + p*q, where f = F/N and p = the price charged at first best condition. This way social welfare is maximised….Implication: People who want to consume small quantities might refrain to pay a very high entry fee and won't consume at all

Peak and Low



The Logic here is the following: In periods of high demand (H.D.) the price offered is the 2nd best one. Those who find the price unaffordable may consume the good in period of low demand, where the price charged

will equal to the 1st best one. Also in this case, maximisation of social welfare occurs. The small triangle equals to the consumer's surplus in period of low demand. Moreover, recall that producers' surplus is given by the difference between Total Revenues and Variable Costs. This logic is used for instance for electricity market: The cost of it changes from day (high demand, thus higher price) to lower demand (lower demand, lower price).

Denset Competition: State lets companies to compete in the tender ("gara d'appalto"). The winner will consist in the company that offers the smallest price. Key characteristics:

1. High number of firms competing (necessary to avoid collusive behaviours. Why? See few pages below)

2. Technological symmetry: Production is similar: ACi = ACj (necessary condition to reach the 2nd best solution).

3. Bertrand competition is played.

Problem of Densest competition:

i) *Contractual incompleteness*: The winning firm commits itself to offer a given service according to a pre-defined price. But, since every contract is incomplete (what about quality, for instance) the regulator might exploit it to take the company in front of a judge. Mechanism, which was supposed to decrease the costs of regulation, increase becoming more invasive.

ii) *Technological symmetry* can exist in the beginning but once the firm has won the first coemption, it might exploit learning economies, reducing the average cost curves and create an advantage in subsequent tenders (more probabilities to prosecute the provision of the service in that area, even in for period ahead).

The other 2 possibilities are represented by

2) *Contestable markets*: Here the government tries to make the market more and more contestable. For instance, reducing entry barriers for newcomers, or reduce the time for potential entrants to enter the market. Think at the process of acquiring a license to enter a market to offer a service. Recall: If the incumbents have enough time to retaliate the new entrants by lowering prices making pointless to the new entrant to acquire the patent and thus enter the market.

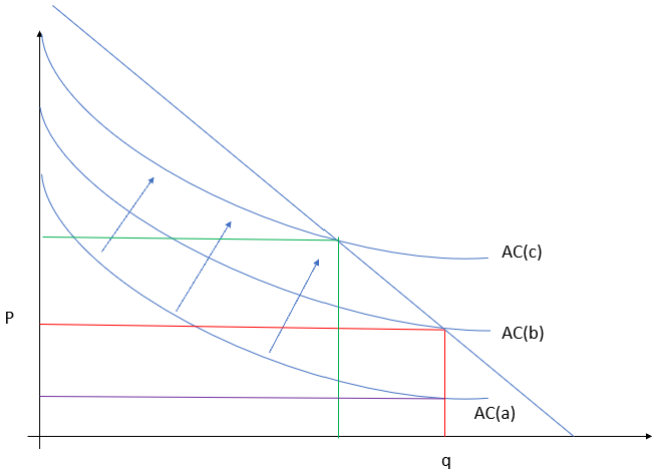
3) *Intermodal competition* which is another name (in this circumstance of regulated services) for monopolistic competition: Given a single firm (monopoly), this latter might face indirect competition stemming from other firms of a different but substitutable sector.

E.g., Flight company is limited to set prices because facing competition from other modality of transports. If set too high, consumers will prefer to take the train instead of the airplane. So, if strong enough, it can provide a solution that is to some extent like 2nd best price solution.

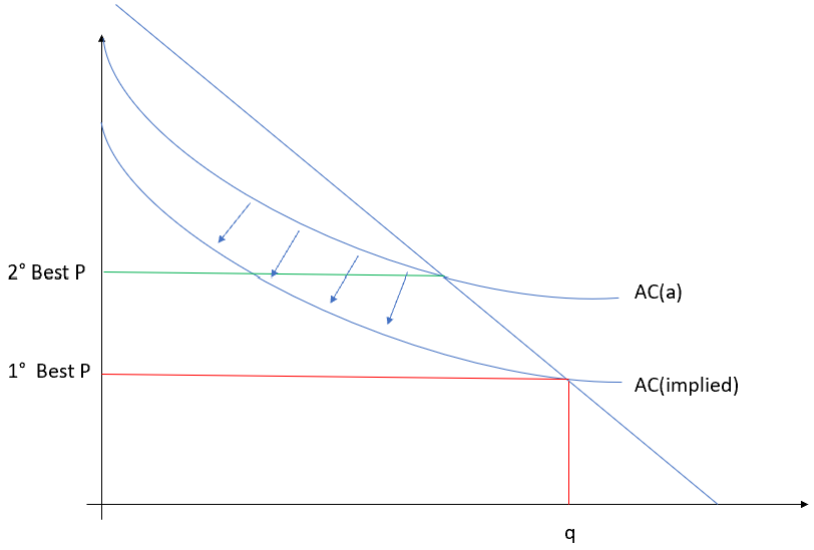
What are the difficulties in transforming these guidelines into practices?

Asymmetric information between the firm that shall be regulated and the regulator. This might give rise to adverse selection (hidden information) and moral hazard (hidden action).

Suppose the regulator wants to implement the 2nd best solution, so it is interested in the average cost curve of the firm it wants to regulate; but suffers from a problem of *hidden information*, which may lead to *adverse selection* problems. Suppose that the regulator is unsure about three possible levels of the cost curve of the firm that must be regulated (AC(a), AC(b), AC(c)). Which the relevant cost curve? Suppose that it chooses to set the price equal to AC(b) (In-between the other two). But suppose that the accrual cost curve is AC(c). The company has an immediate profit advantage. This because the real costs faced by the firm will be way smaller compared to the ones the regulator think. Why do we have adverse selection problem? Because this mechanism incentivize the firm to "fake" (not say the truth) the real costs, thus realizing greater profits. On the other hand, there is an incentive to make the regulator believe that the cost curve is even greater (larger costs) that what it is in reality. **Liars prosper**. How to alleviate this problem? Improving accountability procedures, i.e., the way they have to precisely estimate the real cost duffer by the firm.



The other problem concerns hidden actions, which might give rise to **moral hazard**, i.e., where the most informed party involved in the transaction does not behave in the most appropriate way with respect to the counterparty, which does not have as much information. For the sake of simplicity, let's suppose that the regulator is right and understands that AC(A) are the actual costs, but it does not know whether these are costs related to an efficient firm or not. The problem here is that, by fixing the 2nd best solution, the regulated firm it does not have any incentives in trying to become more efficient because the regulator will be able in this case to detect it and reap out any extra profits. In both cases the firm would achieve 0 extra profits. In this case we have moral hazard. The regulator would like the firm go searching for these more efficient ways but the firm will not because having no incentives.



These two illustrated problems are the biggest difficulties faced in translating the roadmap explained before into reality.

17.2.1.1 Regulation of tariff dynamics in practice

1. *Cost-based method* (cost plus "rate of return regulation") → Does little if not nothing to cope with hidden information problems. It means that the tariff closely adheres with the costs suffered by the firm.

2. *Incentive-based method,* which instead take more seriously into consideration the problem of hidden info and basically put in place ways to alleviate the latter problem (focus: price cap).

3. *Benchmarking-based method* (Yardstick competition), which in specific circumstances (those present with local monopolies) provides a solution by comparing the costs faced by monopolies locally distant.

17.2.1.1.1 Cost-based method: In particular, cost-plus:

The idea is very simple. Suppose you work for a company, and you have a boss that tells you that you must go in a mission and perform a given mission for the company, whatever spent during this mission will be refund by the company (salary is untouched). The cost-plus method basically follows this logic: The regulators establish total revenues (tariffs * q) that equals to the costs + r * RAB ("Regulatory Asset Base"→ Invested capital). "r" (WACC) should capture the opportunity costs of the firm in performing the regulated task instead of investing capital in other activities. Implications in terms of three different types of inefficiencies:

1. *Productive efficiency*: What is the incentive of the regulated firm to search for greater efficiency. Does it have any incentives? No, once the regulator has inferred the costs and decides the "r" if the company manages to lower costs, the regulator will next time see it and lower the tariffs. Of course, we are assuming that the regulator is quite capable of measuring all the costs. → High administrative costs.

every cost that can be described reasonably will be refunded so no incentive for the firm

costs, r , … should be estimated

practice

2. *Allocative efficiency*: In principle maximised, but in theory not.

because the tariff is the lowest possible but in reality due to lack of pro. eff. costs are for an inefficient firm

3. *Dynamic efficiency*: In this context means to substitute obsolete capital with new one to improve over time the quality of the service. The firm in this scenario may have strong incentives to invest news capital because included in RAB. The higher the invested capital (RAB), the larger the tariffs received. This might be seen also as a minus as well since companies might be incentives in overinvest above what necessary for the consumer → "Gold Plating Phenomenon".

So, wrapping up:

$$TR = tariff * q = costs + r * RAB$$

The regulator defines a tariff to allow a volume of revenues equal to the sum of incurred costs, which include a fair remuneration for the invested capital: TR= Costs + r * RAB, where TR: allowed total revenues. Costs: Operating costs (labour, materials, services) + annual capital depreciation (infrastructure, network, machineries); R: rate of return on capital; RAB: Regulatory Asset Base (net invested capital).

Procedure:

Based on firm accounting of the previous year, the regulator determines costs that can be recovered. Based on these costs and on expectation about future demand, total revenues are determined. Price structure is proposed to the firm, which approves it. Prices are implemented

For consumers:

- Allocative efficiency (minimum price but only with respect to observed costs)
- Possible productive inefficiency, due to information asymmetries →Price is based on accounting presented by the firm, not on an "efficient frontier" of costs → No incentives to reduce costs
- High administrative costs: Frequent revisions→ A lot of information needs to be elaborated

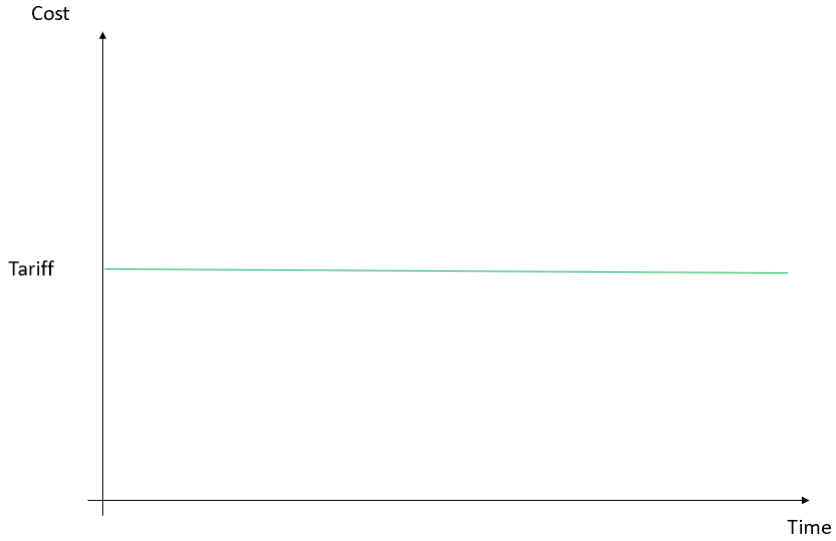
For the regulated firm:

- Financial-economic equilibrium: Costs are covered, and Investors and creditors are remunerated
- Reduced organizational and technological dynamics (low dynamic efficiencies)→ No incentives to implement best-practices and no incentives to adopt innovative systems
- High propensity to invest (it ensures remuneration)→ **Investments are not necessarily efficient**, in terms of localisation, technology, dimension, etc.

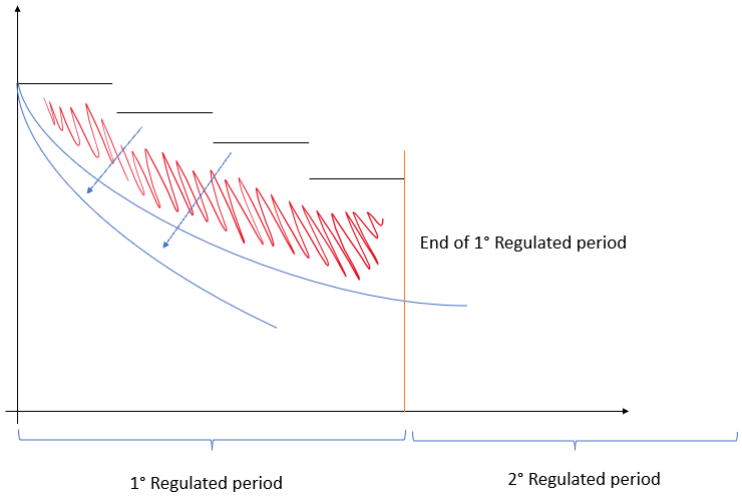
For this tendency to (over-) invest in physical capital generally dynamic efficiency results are not considered excessively bad with this regulatory scheme, but the risk of "gold plating" (Averch and Johnson, 1962) is high.

17.2.1.1.2 Price Cap

With cost-plus the best scenario you can have about the dynamics of costs you can have throughout time, is a flat line: the firm has no incentive to go search for productive efficiency, thus the best scenario is that the costs remain constant over time. The tariff will reflect these costs.



With price cap, tariffs are **lowered throughout time**. The company is pushed to lower costs to meet these declining requirements imposed by the regulator. It is even incentivized to reduce costs at higher rate than the tariffs to increase profits. So now we will have positive effects also in terms of allocative efficiency (prices decrease).



How does the regulator do that? Decides a regulatory lag, which may last from 3 to 5 years (Note: The length of the regulatory period allows the firm to keep gains from productive efficiency improvements and, hence, creates an incentive to adopt efficient behaviours). At the very beginning, the regulator may start either from a cots-plus tariffs or adopt an accounting analysis and costs' estimation. In the latter case, it can be assumed that the cap is equal to the most recent price, or to costs measured on a sample of similar firms (yardstick: see later); then, after a given period, it will decide to calculate the tariffs according to the *price cap formula:*

$$P_t^{MAX} = P_{t-1}^{MAX}(1 + RPI - X),\ t = 1,\ldots,n$$

[RPI = **Retail Price Index**, basically **inflation rate** occurring in the economic system], where absolute value |X| > RPI, since there are huge advantages to be reaped out in terms of productive efficiency in the first periods. **X gives the expected profits** that the company can reap in terms of productive efficiency. Due to this, during the regulatory period we will have this declining cap. So, the regulated firm maintains its natural objective of maximising profits but have only one instrument to do that, namely reducing costs since price is decided by the regulator. Productive efficiency is increased along with allocative efficiency.

Note: With cost tariff, the expected costs are set to a flat line. But these are costs of an inefficient firm since it has not incentives to search for higher **productive efficiency**. In terms of **allocative efficiency**, with respect to the latter, price cap is better since the firm is on the other hand forced to improve its production process.

Thus, from an allocative efficiency point of view, the consumer can also reap out the benefits from a lowered costs structure.

What about **dynamic efficiency**? Here the incentives to make new investments could be considered as mixed. Cost-reducing investments are indeed fostered: All those that enables the firm to lower its costs will tried to be implemented. The same cannot be said for quality enhancing investments. All those investments that do not imply a cost reduction but are aimed at improving per se the quality of the service are not strongly incentivised by the very basic price formula, because the regulated firm would not benefit in terms of revenues and profits in carrying out this type of investments. Therefore, in many regulated services, this basic price cap formula is augmented by the: $\mathbf{P_t^{MAX} = P_{t-1}^{MAX}(1 + RPI - X +/- Q), t = 1,....,n}$

Quality and other investment incentives:

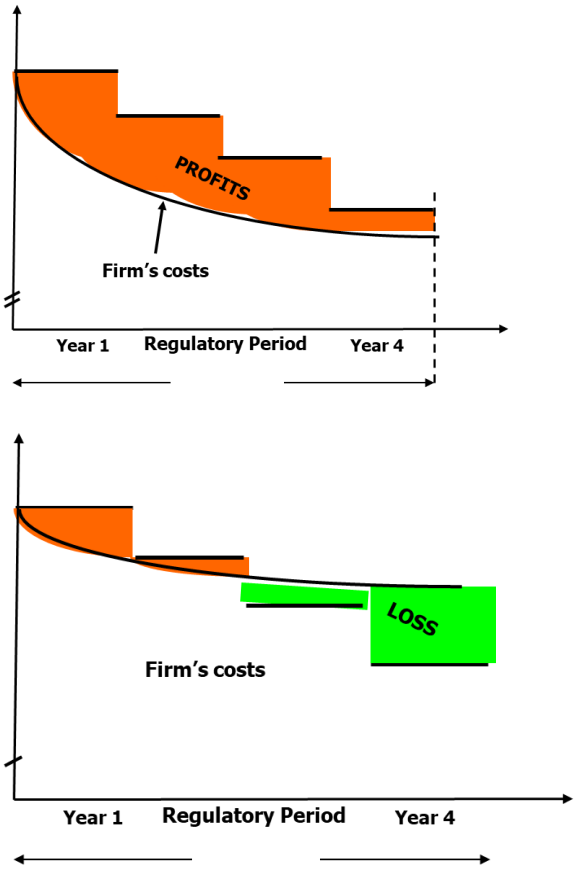
Q = Quality → For heavy improvements in quality the firm may charge a higher price to the consumers, thus enhancing revenues. If quality does not improve, a **malus is imposed**.

This augmented formula might be further augmented by what the professor calls the **"Pass-through factors"** → the regulator can modify the Price Cap formula to pass-through consumers those costs on which the firm has no direct control but do affect the cost structure. For instance, inputs' increasing costs. If this increase too much, this should be reflected in the maximum price that the regulated firm must charge to final consumers.

$\mathbf{P_t^{MAX} = P_{t-1}^{MAX}(1 + RPI - X +/- Q) + Pass\ Through\ Costs, t = 1,....,n}$ comment very important read it

17.2.1.1.3 Some Remarks

Some further considerations must be bore in mind with respect to the "X", which tremendously affects the overall result.



1. The lower it is, the better it is for the firm. This would allow the firm to reach higher level of profits since the decrease of the tariffs calculated by the regulatory agency are smaller than the decrease in costs for the firm. From the point of view of the regulator, the latter does not want to risk a situation in which the firms' costs reduce at a slower pace than expected, though. If the situation represented on the right occurs, the firm could also be willing to exit the market. Remember: From the point of view of the regulator it is better to put a slightly lower X and bear some "extra costs" instead of not having the service at all.

2. Because of that, the X factor is not dictated without any discussions, it does not fall out from sky. It Is decided through back-and-forth discussions with regulator, the regulated firm and possible 3rd parties' stakeholders that might have something at stake in the procedure.

3. In theory, this factor should be declining in absolute value throughout different regulatory periods. At the very beginning of the liberalization process where there were many productive efficiency profits to be reaped out. Once you start to become more and more efficient, it becomes tougher to continue at the same pace. There is a decreasing marginal rate of improvement.

Example for what concerns the electricity industry:

2° Regulatory period:

x = 3.5% for distribution (2.5% for transmission). In the 1°period it was higher (c.a. 4% for both)

3° Regulatory period (2008-2011)

x = 1.9% for distribution; x = 2.3% for transmission

4° Regulatory period (2012-2015)

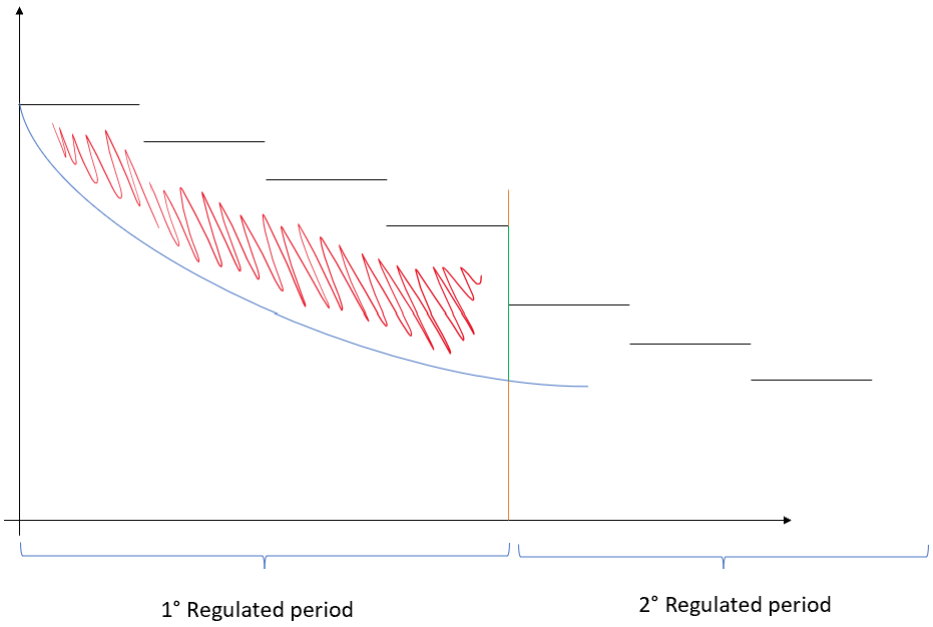
x = 2.8% for distribution; X = 3% for transmission

5° Regulatory period (2016-2023)

x = 1.9% for distribution; x = 1% for transmission

Note: If the people that must bargain about the X factor, once these bargaining occurs with the same parties, there is the risk that at the end they might become "friend", less neutral (*Regulatory Culture*).

Secondly, what happened when the regulatory period ends? In principle the procedure should repeat itself. The logic is applied iteratively, of course with different measures. The regulator should restart the procedures: Go to the regulated firm, estimate the costs the firm suffers of and set the initial level for the tariff which will change according to the price cap formula. Regulators prefer to adopt the "profits sharing rule" (50-50). 50-50 because at very start of the second period, the tariff will be set exactly in -between the costs borne by the firm and the last tariff, thus sharing the profits equally. The regulator still allows the firm to keep 50% of the profits stemming from having improved productive efficiency; instead of setting the new level of tariffs equal to its real costs thus cancelling any profits. This to alleviate the problem of "Rachet effect" ("Effetto inceppamento") of the price cap regulation: When the regulated firm is basically close to the end of the regulatory period, it could not be incentivized to look further improvements since at the beginning of the next period these improvements will "cancelled" out by the regulator (visiting the firm).



Concluding remarks on social welfare effects:

- Productive efficiency:
  - Incentives to cost reduction: beneficial to consumers
    - At the end of the regulatory period, part of benefits is passed to consumers, and part is a premium to the efficient behavior of the firm
- Possible allocative inefficiency:

- During the regulatory period costs are not measured
- Prices are normally higher than a firm's costs [but a likely win-win scenario at the end in comparison with cost-plus]

- Lower administrative costs with respect to Cost Plus: Why? Once the regulator establishes the regulatory period (at the very beginning performs the formula to establish the initial cap) what really regulates in the following period the firm and the price it can charge is the formula. During the regulatory period, basically, at least in principle if no revolutions occur, the regulator should not do anything. With cost-plus instead, the regulator should visit **year by year** the regulated firm, analyze the balance sheets and come out with new tariffs. There is a higher involvement of administrative tasks.

All in all, by correcting the formula and considering all the considerations mentioned before, the price cap is by far better than the cost plus. Therefore, it replaced the latter.

17.2.1.1.4 Yardstick Competition

Model of regulation based on the **principle of comparative competition** (based in geographical different locus) between firms

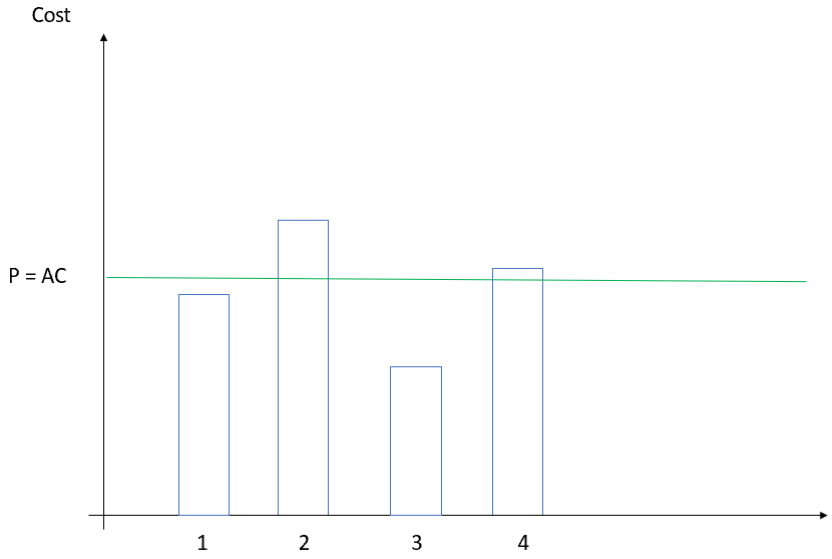
Assumption: if firms operate under similar conditions, they should, in principle, have similar costs

The model can be applied to price regulation:

- The regulator binds the price of a regulated firm to the **average costs** of other firms
- By imposing separation between price and cost, we obtain a mechanism with a high incentivizing power (analogous to price cap)

How does it work in practice?

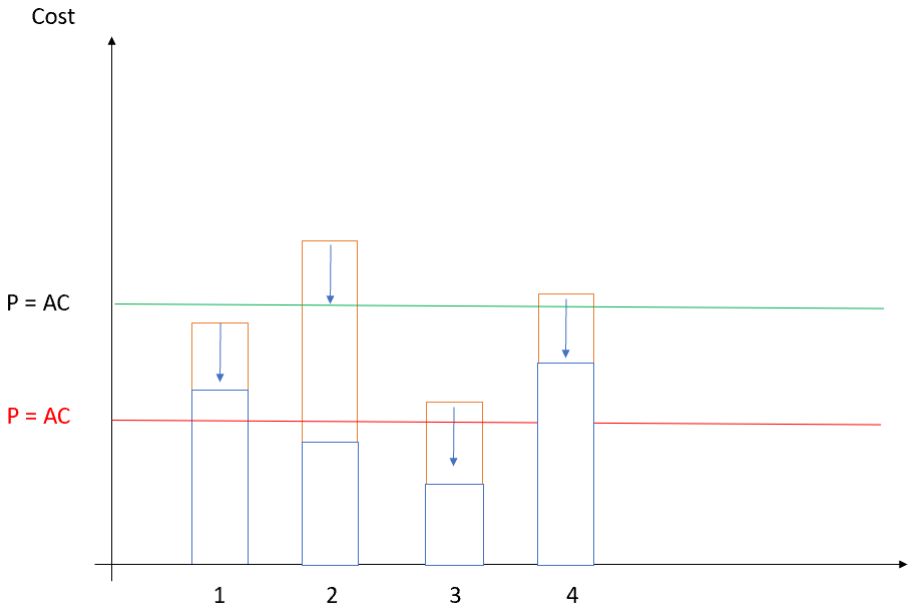
Average costs borne by each company is calculated and the average of these is calculated (grand mean of AC). So, the price set by the regulator will be an average between these different firms.



Note:

- Each firm has a great incentive to lower its costs. This allows to reach extra profits. The incentive to reduce costs is large since **efficiency gains do not translate into revenues reduction**. If the firm reduces costs at a level equal or lower than the average cost set by the regulator, the firm obtains an increase in profits. → This is a positive result in terms of **productive efficiency**.
- Like price-cap, in terms of allocative efficiency the result is not the best one since allowing the firm to make profits. Note that in the second period the regulator will be able to calculate the new average costs and set a price a lower level. This turns out to be positive from an allocative point of view.

- Dynamic efficiencies are not encouraged that much. How to face it? Set up a "**punishment**" mechanism as with the price-cap framework.



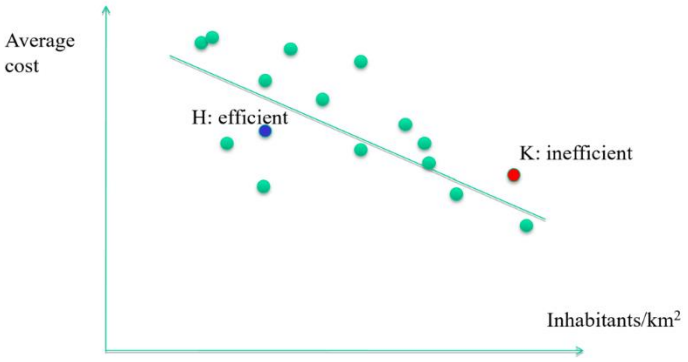
-

17.2.1.1.4.1 The problems of Yardstick competition.

1. Be able to compare apples with apples: Create this geographical area where in principle the costs to provide the service are similar.

In general, by applying this mechanism to all firms, we get an **overall cost reduction** and, consequently, a price reduction

- In theory, this is true for firms operating in identical contexts
- **In practice**, firms operate in **different contexts** (e.g., electricity distribution or water supply: environmental diversity → population density, climate, altitude), which generate different unit costs
- Partial **solution**: econometric/statistical methods to evaluate to what extent exogenous variables affect costs. This can alleviate the difficulties in assessing the right geographical regions.



In this case, the density of the population (exogenous variable) affects the overall average costs. In this circumstance, we can notice that H is more efficient than K. This because the former lies below the curve

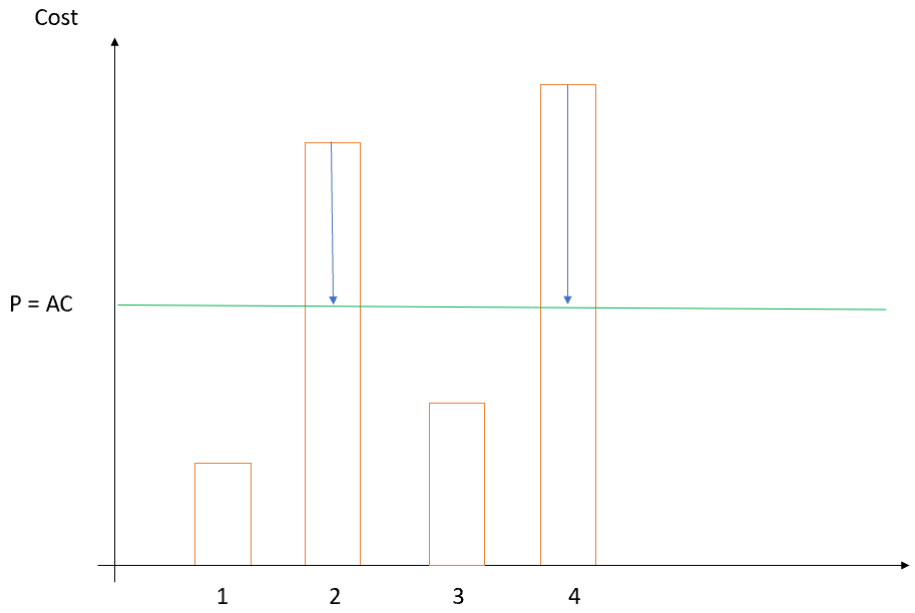
that represents the average costs curve which takes into consideration the relationship with the density of the population.

Implementation problems:

a) Inefficiency can be due to unobserved variables

b) what to do with inefficient firms that do not manage to reach the yardstick?

→ Think about the case in which the firms' costs chosen to differ by some amount. Thus, we can say that the variability of the costs structures is high. In this situation, if the firms that lie above the average cost curve cannot bear the costs and simultaneously reach higher efficiencies rapidly, they might be willing to **exit the market**.



In this case, the government might be willing to jump in and save them because they might be "too big to fail". This interventions tough is quite problematic since it might **incentivize** to some extent the firm **to not search for larger efficiencies**.

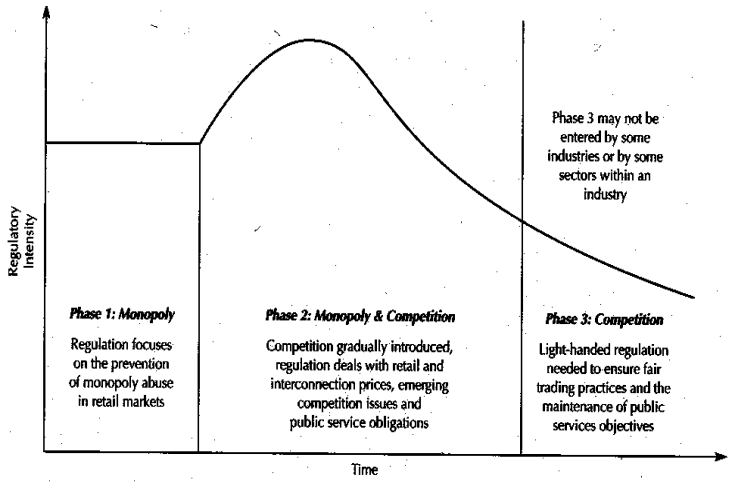
To sum up, yardstick regulation introduces an indirect form of competition, but its implementation is far from being trivial. In practice, regulators rarely use it in its "pure" form. At maximum, they use its logic when they build comparative measures relatively to some cost categories and use them to retrieve information within other regulatory mechanisms (e.g., Service distribution is usually regulated through Price Cap, but initial costs of the local monopolistic utilities are computed adopting a "yardstick" spirit).

Effects (analogous to Price cap):

- Possible allocative inefficiency
  - Costs of the regulated firm can be different from the price calculated on costs estimated on a sample of firms
- Productive efficiency
  - The firm benefits (temporarily) from cost reduction
  - During price revision, productivity gains enter average costs valuation and, hence, in the definition of the new price → producers' surplus extraction

17.2.1.2 Regulation intensity path overtime

Three phases can be distinguished regarding the regulatory intensity



Phase 1: (Pre-liberalization/privatization periods) Monopoly vertically integrated

Regulation of final tariff so to **limit market power** of the only incumbent firm in the market. This type of regulation could be implicit when the firm that needs to be regulated was governmentally owned (has a different objective function than a private firm) → Before privatization, no regulation is required.

Phase 2: Opening to competition

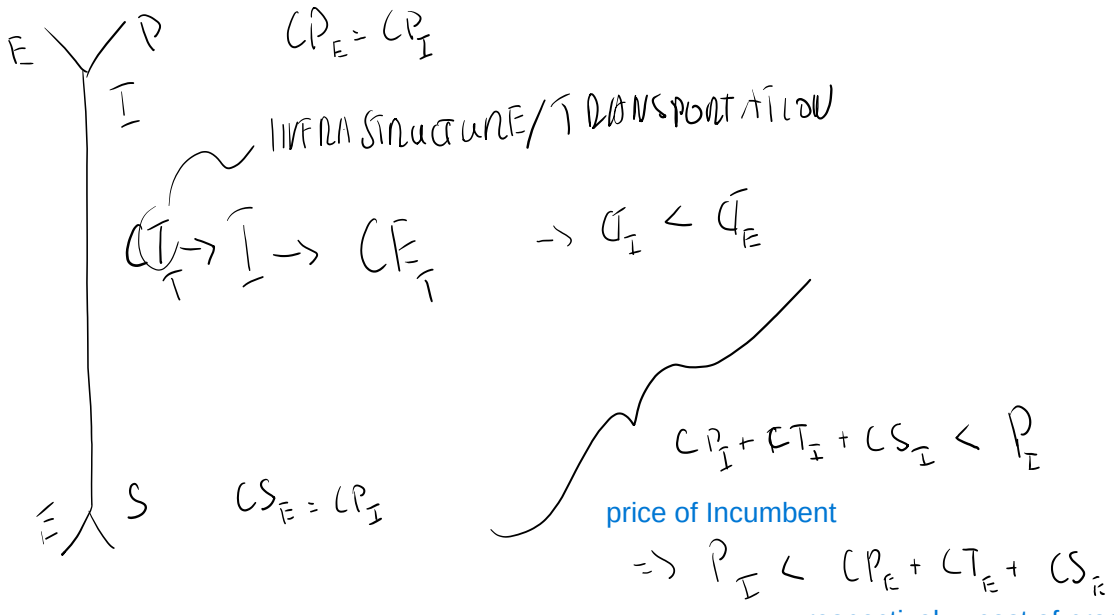
Regulation is aimed at **ensuring fair conditions of entry** to new firms but still final tariffs are regulated so to avoid exploitation of market power by the incumbent ex-monopolist which is now a dominant player in a market where competition is still at its infancy stage.

Phase 3: Access

Here new entrants have acquired the capabilities to pose a serious threats to the dominant incumbent (competition is more solid). Clearly, the focus of regulation can be relaxed a little bit since there will be **competition to drive the market to the optimal social result**.

Here, regulation focus is on ensuring that firms are not discriminated in **accessing essential facilities** still considered as natural monopolies.

Clearly if the incumbent is present in the production and sales stages and governs the infrastructure stage as well, the possibility that new entrants enter both the production and the selling stage is limited by the requirement of having access to the same infrastructure in order to be competitive



respectively : cost of production, cost of transportation, cost of sales of Entrant

Indeed, as it is depicted by the graph above, if the entrant's cost of production and the cost of sales are equal to the incumbent's ones, but the latter has full control on the price that can charge to the entrants regarding the costs of transport there cannot be any competition, since the incumbent will always able to find a price charged to the final consumers that will be lower than the costs faced by the new entrant. This is where the regulator must step in. But this is only one dimension (price) where the regulator might be involved. Quality might be another dimension: What if the incumbent does not allow the new entrant to have access to the same quality (e.g. timeliness) of the infrastructure.

What the regulator tries to do is "**unbundling**", i.e. inserting into the system some degree of separation. There are 4 different types of unbundling (from the strongest to the lightest).

1. Ownership: The state decides that the company that operates in the **infrastructure** shall not have any shares in the company's operating upward and downstream (e.g. ENI and SNAM, where the latter owns the infrastructure and the former the two extremes).

2. Legal: The company that operates in the infrastructure has to be a different company but still this firm can belong to the same group (i.e., there can be some sort of ownership linkages). They require to be legally separated entities.

3. Functional: There is no obligation anymore to be separate legal entities. The infrastructure has to be managed by a different department.

4. Accounting: (Lightest form). There is not even the obligation to carry the activities in different departments, but there is just the obligation to present different balance sheets.

Why these different forms? Because by implementing these, the regulator faces a trade-off. The more you opt for the lightest, the more you gain in terms of productive efficiency (economies from vertical integration can be reaped out à Saving resources, not obliged to duplicate resources) but you lose in terms of allocative efficiency (higher risk that the firm possess the stages discriminates some firms with respect to others).

Interesting is that this unbundling techniques are an open questions for digital markets.

17.2.2 Antitrust: The other leg of Competition Policy

Birth: Sherman Act and Clayton Act. → 3 types of interventions (considering the anticompetitive act):

<u>Ex-Post</u>

Agreements between companies that restrict competition → Cartels or other unfair arrangements in which companies agree to avoid competing with each other and try to set their own rules. → *Collusive agreements*

Abuse of a dominant position → A major player tries to squeeze competitors out of the market: *Predatory pricing* à Exploiting the dominant position to kick out competitors or at least to reduce competition

<u>Ex-Ante:</u>

Mergers & Acquisitions → Antitrust should approve or deny the undertaking of this acquisition if they danger allocative efficiency

17.2.2.1 Collusion

A firms, instead of competing, finds an agreement to behave as if they were a single one (large single monopolist). Firms may collude on many different dimensions (marketing, advertising, R&D, etc.)

Cornout example: Quantity produced is the same as the one that a monopolist would have produced:

STRATEGIC VARIABLE: Quantity

- $p = a - b(q1 + q2)$
- $\Pi 1 = pq1 - cq1 = [a - b(q1+q2)]q1 - cq1$
- $\Pi 2 = pq2 - cq2 = [a - b(q1+q2)]q2 - cq2$

- $d\Pi 1/dq1 = a - 2bq1 - bq2 - c = 0 \rightarrow q1^* = [(a - c)/2b] - (1/2)q2^*$
- $d\Pi 2/dq2 = a - 2bq2 - bq1 - c = 0 \rightarrow q2^* = [(a - c)/2b] - (1/2)q1^*$

The most frequent realm where firms usually collude is price.

Colluding, under very general conditions (if coordination costs are not extremely high) acting as a single monopolist realizing monopolist profits and splitting them among the firms participating within the agreement is always better than the profits that they would have obtained by competing (assuming similar costs structures and homogenous products). à $(1/2)\ \Pi^M > \Pi 1\ (\Pi 2)$

The other realm we have to understand is about how sustainable a collusive agreement is over time. How can the carte that has been established be persistent? If one of the member finds a better solution, collusion might be broken down.

3 facts that affects the sustainability of the collusive agreement:

1) n° of (identical/similar) firms

2) probability that no external factor will occur to change the game (e.g. entry of new firms, obsolescence of the focal product).

3) frequency of interactions between firms (i.e. the number of times firms play the same game)

Before doing this, a necessary premise is due:

Rohlfs, 2003, *Bandwagon effects in high-technology industries*, p. 45: [in the case firms choose NO COMP, the most rationale strategy] "may be to signal an irrational commitment to win the race (of the standard) at all costs. This is analogous to a well-known result of oligopoly theory; namely, the most rational strategy for an oligopolist may be to signal an irrational strategy commitment to punish price cutters-regardless of the ruinous cost that it may incur by doing so." → Which very irrational strategy can be a very rational one?

Suppose we have two firms that collude in order to obtain and then share monopolist profits. If these firms have no capacity constraints and consumers are quite responsive to price changes, both firms are incentivized to set a price an epsilon lower than monopolist price and take the entire pie (profits). If this is the framework, the cartel is not sustainable anymore. Both firms have an incentive to deviate from the agreement. But, let's suppose that the firms adopt a "grim-trigger strategy", i.e., "I will collude until you collude with me. Once you cheat, I will **never ever** collude with you again". And let's suppose that this grim strategy takes the very drastic form of: Fixing a price equal to the monopolist price but once the counterparty diverges to this, I will set the price equal to the marginal costs. This is totally irrational because the firm is not going to earn any extra profits anymore. But this is on the same time (if you are credible) the most rational one. It will enables the carte to be sustainable because preventing the other parties to diverge from the agreement. [End of the premise]

Getting back to the 3 facts affecting cartels…

1. n° of Firms: Of course, the more the firms that participate in the cartel, the less sustainable the cartel will be. This because of:

1. Increase in coordination costs.
2. Incentives to deviate. Why? Suppose that instead of 2 firms we have 4 firms. They all set their prices at monopoly prices and share monopolist profits in 4. In the previous scenario, the incentive to deviate from the monopolist profits equals to half to the monopolist profits. Now, it is ¾ .

2. Repeated interactions and no game changers (low probability that the game will change in the next future). To understand why these two dimensions are important, let's analyze the following game theoretic framework:

1/2	Cheat	Collude

Cheat	(1,1)	(3,0)
Collude	(0,3)	(2,2)

1. If the game is played only once, it is a prisoner's dilemma type of game: NE is (1,1). → Collusive agreement is not possible
2. Now suppose that the game is repeated an indefinite length of time and the probability the game will be played each time is 0 < p < 1. Also suppose that firms adopt a punishment strategy for deviation from the collusive agreement of a grim-type ("I collude until you collude, if you cheat once I will cheat forever").

3a. If a firm cheats its payoff will be $3 + p + p^2 + p^3 + \cdots p^n + \cdots$ that is equal to $3 + \frac{p}{1-p}$

3.b. if a firm colludes, its payoff is: that is equal to $2 + 2p + 2p^2 + 2p^3 + \cdots 2p^n + \cdots$ that is equal to $\frac{2}{1-p}$.

4. Both firms will collude if: : $\frac{2}{1-p} > 3 + \frac{p}{1-p}$, thus if $p > \frac{1}{2}$

5. If «p» is high enough (so «no game changer» is at the horizon) firms will likely collude (or to better say the collusive agreement is sustainable). → Multiple interactions are more likely to foster collusive agreements.

N.B. if |p| < 1, we have that $\sum_{n=0}^{n=\infty} p^n = 1 + p + p^2 + \cdots = \frac{1}{1-p}$

E-commerce increases frequency of interactions and, in doing so, enhance «collusion» possibilities. This temporal dynamics of a market can reveal that "big players" are colluding. Why? In internet, prices can be changed every instant. Of course, they can immediately observe what other firms are doing. Algorithmic logic ("if this, then do that") might lead to collusive behaviors.

"A super-game-theoretic model of price wars during booms" → INTUITION: Firms collude on price in periods of:

- High demand → High incentive to deviate from the cartel (i.e. higher profits at the expense of the other firms) → collusive price has to be reduced in order to reduce the incentive to deviate and make sustainable the cartel.
- Low demand → Low incentive to deviate from the cartel (i.e. lower profits at the expense of the other firms) → collusive price can be raised since incentive to deviate is low and the cartel is still sustainable.

When we observe this countercyclical behavior by firms, it can be a very important signal for understanding if a collusive agreement is present or not.

17.2.2.1.1 Tacit vs. Explicit Collusion

Despite being both socially detrimental, it is important to distinguish these two.

Explicit collusion: Both firms have agreed to collude and these latter has been corroborated by some tangible proof (contract, document) → Always illegal and even easily prosecutable by antitrust since provable through tangible documents.

Tacit collusion: May arise through the so called "conscious parallelism": This means that firms can find a way to understand their strategic interactions even without communicating one with another. **No interchange of communication** does exist. Because of this, it is very difficult to assert whether an agreement has been established or not. Under this scenario, collusion is """legal""", in the sense that it is impossible to actually prove that these two have any interactions.

In the middle of this spectrum we have a grey area:

"Concerted actions": Nothing written in any document but yet some prosecutable (provable) actions have been implemented (e.g. specific announcements).

Political matter: How much do political forces punish concerted actions?

Premise: Dominant position is never punished if not artificial

1. Suppose that demand is given by: $P = 120 - Q$ and all firms have constant marginal cost of $c = \$80$
2. Let one firm have innovation that lowers cost to $c_M = \$20$
3. This is a *Drastic* innovation. Why?
    - Marginal Revenue curve for monopolist is: $MR = 120 - 2Q$
    - If $c_M = \$20$, optimal monopoly output is: $Q_M = 50$ and $P_M = \$70$
    - Innovator can charge optimal monopoly price ($70) and still undercut rivals whose unit cost is $80

Should the innovator being blocked because it will turn out into a monopolist?

1. Ex-post allocative inefficiency but not compared to ex-ante: p now is 70$ and not 80$, 50 units traded rather than 40.

2. Ex-post productive efficiency gain: now producing 1 unit costs 20$ rather than 80$

3. Dynamic efficiency (a new technique available)

17.2.2.2 Relevant Market, Dominant Position and Abuse of Dominant Position

So what is punished is the abuse of the dominant position. The first question the regulator must ask itself should be: Is there a dominant position? The answer here depends on the definition of what a relevant market is: *What is the "relevant" market? (EU Commission definitions)*

(a) A relevant product market comprises all those products and/or services which are regarded as interchangeable or **substitutable** by the consumer by reason of the products' characteristics, their prices and their intended use.

(b) A relevant geographic market comprises the area in which the firms concerned are involved in the supply of products or services and in which the conditions of competition are sufficiently homogeneous.

It deals with:

1. **Demand side substitutability** (customers)
2. **Supply side substitutability** (suppliers)

Practical determination is far from obvious and quite complicated (e.g. survey data collection, analysis, econometrics, etc.). Several methods and techniques, all not-immune from critiques and problems do exist. No one of these are perfect (all have either practical or conceptual problems). The most used method is the so called "Small but significant non-transitory increase in price". How does it work?

17.2.2.2.1 Small but significant non-transitory increase in price test (SSNIP)

and ask whether five percent ...

1. Starts with smallest possible market guess and then questions whether 5% price increase would be profitable for a hypothetical monopolist in the market (market for bananas or market of concrete in the area x).
    a. If the answer yes, you have an estimate of the relevant market (bananas)
    b. If the answer is negative, then firm does not have sufficient market power to raise price.
2. Next closest substitute is added to the relevant market and test repeated keeping constant the price for the closest substitute (market for bananas & kiwi or market for concrete in x&y).

3. Process continues until the point is reached where a hypothetical monopolist could profitably impose a 5% price increase keeping constant the price for all the closest substitutes (fruit market or market for concrete in x&y&z).
4. Market then defined (fruit market/ market for concrete in x&y&z).

→ Problems with this process:

1. **Practical Construct**: Antitrust must understand how customers (in this case fruits consumers) are behaving → Questionnaire (statistics realm → Uncertainty is involved)

2. **Theoretical Construct**: *Cellophane fallacy:* Generally, the "prevailing" price is the starting point to apply the SSNIP test, but this is not necessarily a competitive one. If it is already the one of a monopolist (i.e. the optimal solution for the monopolist), a 5% increase by definition is not profitable, [the monopolist optimally sets a price in correspondence of an elastic part of the demand curve (remember the mark-up formula (Pm – MC)/Pm) = 1/|E|. By construction the term on the left is larger than one, thus the monopolist profits depend on how the demand is elastic]. In this case "an increase in the current level of prices might induce customers to switch to other products that would not necessarily be substitutes under competitive conditions (e.g. lower quality products)" (Willis 2005, Introduction to EU Competition Law, p. 28).

Once the antitrust is capable to define what relevant market is, the next question will be to understand whether there exists a dominant position in that specific market or not. How?

a. Market share (generally above 40% in EU; 50% in US)
b. Length in time of that market share → Transient or Permanent?
c. Differences in market shares with second competitors (high Herfindahl index)

And then, the third question will concern whether there is an **abuse of this dominant position** or not. What is an abuse? The company exploits this position to eliminate competition

This can take different forms:

a. depriving smaller competitors of customers by selling at *artificially* low prices they can't compete with (Predatory pricing)
b. obstructing competitors in the market (or in another related market) by forcing consumers to buy a product which is artificially related to a more popular, in-demand product
c. refusing to deal with certain customers or offering special discounts to customers who buy all or most of their supplies from the dominant company

17.2.2.3 Predatory Pricing

What makes a price a predatory one?
1.Areeda-Turner → Setting prices lower than its OWN marginal costs
2.*Recoupment loss* → Expectations to recoup the losses the firm is incurring
The logic: The firm is setting prices below marginal costs because it wants to induce other firms to exit, but of course by applying this pricing strategy it is incurring in negative extra profits (at least in the short-run). The goal is to push other firms to exit from the market. If this occurs. The market will be more concentrated. Thus, the dominant firm can pose a price much higher than marginal costs (extreme case: monopolist prices) so that in the long run, profits will be much higher than the negative profits in the short run. Thus, does the firm a reasonable expectation to apply such strategy?
3.Predatory intent → Find documents, documents that corroborate predatory pricing.
→ Of course, these are trials. You have to prove these in a court.
Note: The first condition is sufficient in Europe to decide for a predatory pricing behavior, whereas in the

US all three must be proved.
Final Thoughts:
Looking at social welfare maximization that should represent the aim of Antitrust, all this process turns out to be really problematic once put into practice. WHY?
→ Think to "predatory pricing": price down is the only sure thing of the whole process, which from a social welfare point of view is good thing. Thus there exists different views and different implementation philosophies about Antitrust policy
→ Chicago antitrust school: Antitrust legislation may end up harming rather than benefiting social welfare. Over time, the Chicago School has built a theoretically sounded framework aimed at diminishing the relevance of antitrust interventions.
See for example the "single monopoly" theory (Bork, 1978) to illustrate how monopolists do not have interest for monopolizing adjacent markets, since there's the possibility of only "one monopoly profit":
Homogenous consumers: each one has wtp = 10 for jars & lids.
Cost for producing a jar = 3
Cost for producing a lid = 3.

d. If a firm is a monopolist in the "jars & lids" market: Best unit profit = 4
e. If the two markets are separated, and jars offered at a competitive price (equal to cost, so price of jar = 3), the remaining wtp of consumers for lid is equal to 7, thus again the maximum unit profit is = 4.

No interest for the lid monopolist to monopolize even jars, because profits will remain the same.

The theory is valid, only under very specific assumptions, e.g. goods are consumed at fixed proportions, homogenous preferences, etc

1) Predatory Pricing is very unlikely:

- It is extremely expensive (price reduction may also mean higher output to produce at very high costs, if MC curve is increasing). (q)
- Acquisition can be a much less expensive way to eliminate a competitor.
- If the prey exit easily, this means that some others could also easily enter into the market à Lower exit barriers are present, which are correlated with lower entry barriers as well. This means that once an incumbent has kicked out a firm, it will be highly possible that a further one will join the market soon.

2) Firms may refrain from pricing aggressively so to avoid any risk of being accused of predation (MC has to be estimated and one never knows how things turn out in legal courts).

The other argument is about strategic use by inefficient competitors of antitrust infringement claims of efficient firms in a "Tonya Harding-style competition" → I am losing in this market because the other firm is more capable than me in doing things, instead of competing I ask for antitrust intervention alleging anti-competitive behaviors. Even if this is not true, the incumbent must deal with many bureaucratic issues (resources are drained from other more productive uses)

Different philosophies in implementing anti-trust legislation →2 families: **Rule of reason vs. Rule of law**

Rule of rule: There are some basic rules that if infringed, the firm will be automatically accused for predatory pricing. → Problems: risk of wrong decision in terms of welfare

Rule of reason: This does not contemplate violation per se but it has to be ascertained case by case. Problems: subjectivity, lack of transparency, little predictability to market players

Historically the US has stayed more towards the rule of law and the EU towards the rule of reason. Nowadays, the contrary might be said.

17.2.3 Merger regulation

3rd type of intervention by the antitrust regulation: Ex-ante.

What is the problem? If two companies that are competing in the same market will merge, the market's concentration (more on concentration later on) will rise. Their market power will as well, thus granting them the possibility to raise prices above competitive levels. The rational of M&A consists of gaining productive efficiency (reaping out economies of scale). The antitrust to this respect must check whether these do not translate also into allocative inefficiencies (detrimental for social welfare). It is an evaluation made on the basis of productive and allocative efficiency thus.

Considerations:

1) M&As that do not happen in the same relevant market, do not represent big concern for antitrust. The focus is on **horizontal mergers** (firms operating in the same relevant market)

2) Historically the third concept of efficiency (dynamic efficiency) did not enter very much into the consideration of the antitrust. It lied in the background. Firstly because innovation intrinsic nature regards uncertainty. Secondly, economic theory offers very little guidance with that respect. → This has always been true. More recently, the European realm has tried to take into consideration this third concept as well. The logic: If two firms are merging together and the result will be a higher R&D expenditures, this might **positive for innovation**. The contrary has to be considered as well.

A recent decision by EU antitrust has cleared the 130 Billion Euro merger between Dow/DuPont (agro-chemical sector) but only subject to structural remedies (i.e. the divestiture of major parts of DuPont's global pesticide business, including its global R&D organisation) on the augmented basis, then also formulated in the so-called "innovation theory of harm" (ITOH), that the eventual merger (without remedies) would have led (in the view of the EU Commission) to a significant decrease in the innovation rate among the involved parties that would exceed any possible gain in the other efficiency dimensions.

The ITOH has triggered an academic (surely not "ivory tower") debate on the merits of this (new) approach in antitrust merger control, with many opponents, see Jullien/Lefouilli – Denicolò/Polo.

From EU antitrust guidelines: "Horizontal mergers – i.e. those between competitors on a particular market – can lead to a loss of direct competition between the merging firms. By contrast, vertical and conglomerate mergers do not immediately change the number of competitors active in any given market. As a result, the main potential source of anti-competitive effects in horizontal mergers is absent from vertical and conglomerate mergers. They are thus generally less likely to create competition concerns than horizontal mergers. In addition, vertical and conglomerate mergers may also improve a company's efficiency by better coordinating their different production stages."

17.2.3.1 The relevant players

Having said this, mergers are scrutinized only if they involve sufficiently big actors (ex-ante or ex-post). The latter M&As shall not be communicated to the antitrust. Why is **"smallness" excluded**?

1. Antitrust has not infinite resources to analyze all the M&A activities happening in the market

2. The concern about the negative effect that the small M&A activities in terms of allocative efficiency are extremely reduced. When two small firms are merging, it is very likely that these will achieve **more efficient economies of scales**. This newly established firm could also become a serious threat for larger firms competing in the same market, establishing a more fierce competition (benefits the final consumers through lower prices) Thus, a potential win-win (productive and allocative efficiency) scenario might occur.

2) Is (ex-post) largeness automatically punished in horizontal mergers? Not really (rule of reason). What the regulator asks itself: **Synergies? Economies of scale?** If synergies do exist, can productive efficiency be transferred at least in part to consumers (also more allocative efficiency) or at least can we be sure that competition in the mkt will not decrease?

The procedure the Antitrust follows

1. M&A exceeding a certain threshold have to be communicated in advance

2. Antitrust investigates → Understanding the perimeter of the relevant market. If the M&A firms are not operating in the same market, the merge is allowed. In case they are operating in the same market, an horizontal M&A is in place. → Do productive efficiencies stem from the M&A activity?

3. Ascertained that productive efficiency gains do exist, is M&A leading to a dominant position or risking to seriously undermine competition? If not → Consent decree. If yes → Antitrust analyze whether allocative efficiency would be seriously compromised

4. Final decision → Decision: Denial (or Consent with remedies) if yes

Remedies might be put in place to allow the activity, such as:

1. Structural remedy

2. Behavioral remedy (for the next X years the price of a given product cannot increase more than Y%)

17.2.4 Extra Open Issues:

1. AI and collusion

"First, pricing algorithms had once been based on pricing rules set by programmers but now often rely on AI systems that learn autonomously through active experimentation. After the programmer has set a goal, such as profit maximization, algorithms are capable of autonomously learning rules of conduct that achieve the goal, possibly with no human intervention. The enhanced sophistication of learning algorithms makes it more likely that AI systems will discover profit-enhancing collusive pricing rules, just as they have succeeded in discovering winning strategies in complex board games such as chess" (Calvano et al. 2020, p. 1040, Protecting consumers from collusive prices due to AI, *Science, 370, 6520, pp. 1040-1042).*

"The field of discussion is the one of collusion between firms. Any textbook of microeconomics explains how collusion produces a distortion in the market, resulting in a deadweight loss of efficiency and a reduction in consumer welfare. […] Should algorithmic collusion emerge in a market, society lacks an effective defence to stop it. Indeed, firms are, or declare themselves, unaware of collusion, while AI does not have a mind, a will, a consciousness. Therefore, there can be no "meeting of minds" in the collusion between algorithms"

2. M&A control and killer acquisitions

Killer acquisitions may fall below thresholds (antitrust radar): Need of new ones? Not simply based on turnover (e.g. acquisition price)? Obligation to communicate (and scrutinize) every M&A for big tech companies? But how to assess the competitive threat of a start-up?

3. Is "digital" special and requires a specific attention?

See the recent Digital Market Act (DMA) proposal by the EU (expect to eventually become operative not before the end of 2022 – beginning of 2023, still needs formal passages by EU Parliament and Council. According to the European Union it is special, since it allows the owner ("gatekeeper") of digital platforms to discriminate participants.

4. Adequacy of sanctions?

It is an old question but which can be considered actual as well- The antitrust rarely opts for structural remedies. It never breaks up a company into divisions. It usually uses big fines, which, compared to the rewards achieved by behaving in anticompetitive manner can be still considered peanuts.

18. CONCENTRATION INDEXES

Why does industry concentration matters?

Concentration tells us how much the economic activity is conducted by few firms. The smaller concentration thus, the smaller the probability of having a monopoly or oligopoly.

- One of **the most studied areas** in the domain of market structure, particularly in the industrial organization literature.
- When **antitrust agencies** are evaluating a potential violation of competition laws, they will typically attempt to **measure** concentration within the relevant market. Recall that antitrust agencies are agency that try to ensure a fair competitive scenario, thus, they need to know how much the market is concentrated. Think of a M&A: They will have to grasp the level of concentration before and after to understand if this will be detrimental for the competitive landscape or not.

Concentration: The concept → Intuitively, concentration depends on the number and size of the firms belonging to the focal industry

- Concentrated industry:
  - Very few firms
  - Or many firms, but few of very large size
- Dispersed industry:
  - Many firms
  - None of the firms has a dominant position

In a nutshell, concentration

I. is a fundamental metric, especially for antitrust agencies
II. is useful to get a clear and immediate snapshot of the focal industry
III. affects firms' behavior and offers indications on competition mechanisms

- Analysis of the causal relationship: Is concentration leading to market power or vice versa (market power that leads to concentration)?

The relationship is bidirectional. When we have a very concentrated industry the economic activity is in a few firms. Does this mean that these firms have market power? Yes. Concentration leads to market power almost by definition: if they had no market power, these small number of firms could not do whatever they want with prices since other firms would join the market and reap out the latter profits. On the other hand, once you have market power you can raise entry barriers: you can invest expanding the firm's scope, ameliorating productive efficiency and attaining economies of scales à Market power leads to concentration.

because they don't need to worry about competition

18.1 Link to the market structure

Oligopolistic industries are characterized by high concentration. The maximum level of concentration is reached in the case of monopoly

Competitive industries are characterized by low concentration. The minimum level of concentration is reached in the case of perfect competition

→ The operationalization: The market share. It is the ratio between sales of firm I and total sales in industry j. Market share of firm i in industry j, where N firms operate:

$$s_{ij} = \frac{q_{ij}}{\sum_{i=1}^{N} q_{ij}}$$

Where q_{ij} is the sales of firm i in industry j

Intuitive representation of market shares in matrix form: N firms per M industries, where qij = sales of firm I in industry j.

Firm/Industry	...	...	j	....	M
...			q_{1j} Sales of firm 'i' in industry 'j'		
...			...		
i	...	...	q_{ij}	...	q_{iM}
...			...		
N			q_{Nj}		

Firm	Sales	Market share
1	q_1	$s_1 = q_1 / Q$
2	q_2	$s_2 = q_2 / Q$
...	...	...
i	q_i	$s_i = q_i / Q$
...	...	
N	q_N	$s_N = q_N / Q$
	$Q = \sum_{i=1}^{N} q_i$	$1 = \sum_{i=1}^{N} s_i$

Focusing on one industry, the market share would be:

➔ $s_i = q_i/Q$ with $Q=S_iq_i$

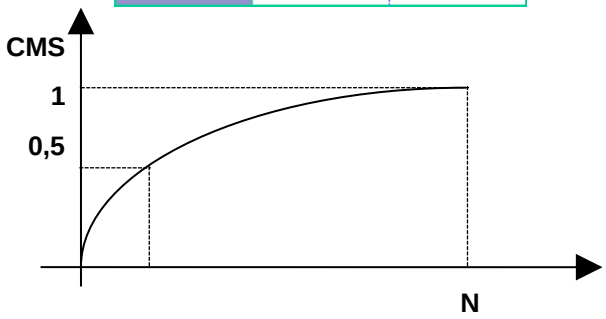
Firm	Output	Market share
1	500	0.5
2	300	0.3
3	100	0.1
4	70	0.07
5	25	0.025
6	5	0.005
	1000	**1**

The concentration vector lists market shares in descending order (it lists market shares starting from the highest market share and going down to the smallest one), in order to emphasize the presence of large firms in a given industry

A useful representation of **concentration** in the **Cartesian plane** (for a certain industry at time t) is given by the **concentration curve**

- **Horizontal axis**: firms ordered in **decreasing order by size** (from the largest to the smallest)
- **Vertical axis**: **Cumulative Market Shares** (CMS)

If the concentration curve grows very fast from the very first firms, it means that those firms virtually possess almost all of the market shares. On the opposite, if the growth is smooth, it means that even the market leaders do not own much of the market in relative terms.



18.2 A multitude of CIs

Concentration indices

- **synthesize** in one index the entire information contained in the concentration vector
- facilitate the **comparison** between the degree of concentration in different periods and industries

They can be:

I. Absolute: weighted sum of market shares. With different weights, different concentration indices are identified

$$I = \sum_{i=1}^{N} h_i(s_i)s_i \qquad with \quad h_i(s_i) = weight$$

→ $H(s_i)$ = weighting function which depends on the market shares.

II. Relative: capture the degree of inequality between firm sizes within a certain industry

Note: All indices facilitate analytical analysis while at the same time loose some information. Moreover relative concentration indices capture inequalities, but correlation is not perfect. In particular there are

some cases (symmetrical oligopoly) where relative indices underline perfect equality (thus no concentration) despite the fact we are in an oligopoly!

18.2.1 Absolute CIs

18.2.1.1 Concentration Ratio

Sum of market shares of the first k firms, in decreasing order $C_k = \sum_{i=1}^{k} s_i$

The weight =1 for the first k firms

The weight =0 for the remaining N-k firms → As rule of thumb k is usually equal to 3, 4, 8 or 20

Criticalities:

1) The choice of **k** is arbitrary → Loss of objectivity
2) It provides the **same result** for industries with **different concentration**
    - *Example:*
    - Industry A: S_A = (0.1, 0.1, 0.1, 0.1, …)
    - Industry B: S_B = (0.37, 0.01, 0.01, 0.01, …)
    - → C_{4A}=0.4 and C_{4B}=0.4. Despite completely different relationships (competitive structure), in industry A we have a symmetric oligopoly, whereas in industry B a monopoly.

→ While in A market shares are evenly spread, in B there is one firm that controls almost **40%** of the market: this crucial information **is lost**.

18.2.1.2 Herfindahl index

Very simple but very reliable → widely adopted

The weight of each share is given by the share itself: the HI is given by the **sum of the square** of the **market shares.**

$$HI = \sum_{i=1}^{N} s_i^2$$

Since we are calculating the square, the larger a firm, the more it contributes to the value of the index. So the contribution will be perfectly proportional to the market share itself. This gives us a quite accurate intuition of the concentration levels.

HI ranges from **0** to **1** → If there are N firms **with the same size,**

- In perfect competition HI=0 ($HI = \frac{1}{N}$)

- In monopoly HI=1 (**N** = 1)

Note: In a two sided oligopoly, the Herfindahl index is going to equal to 0.5 ($0.5^2 + 0.5^2$). If 4 firms are involved, 0.25, and so on.

US market of credit card suppliers – market shares in 2007 and 2008

Enterprise	QM 2007	QM 2008	2007	2008
JPMorgan Chase	17.74%	21.22%	**C_4**	
Bank of America	19.36%	19.25%	61.53%	63.01%
Citi	13.03%	12.35%		
American Express	11.40%	10.19%	**C_8**	
Capital One	6.95%	6.95%	80.85%	83.39%
Discover	5.65%	5.75%		
Wells Fargo	3.07%	4.21%	**HI**	
HSBC	3.65%	3.47%	11.00%	11.97%
U.S. Bank	1.84%	2.14%		
USA Savings	2.01%	2.02%		

18.2.1.3 Entropy Index

In the **Entropy index**, the weight of each share is equal to the logarithm of the inverse of the market share

$$EI = \sum_{i=1}^{N} s_i \cdot ln\left(\frac{1}{s_i}\right)$$

Smaller firms provide a greater contribution to the total amount of the index through the inverse of the logarithm. The smaller the firm, the higher the logarithm. The Entropy index measures the disorder of the market, thus if the value is pretty high, the market will be highly disordered and will represent the situation in which a huge number of firms is present in the market (perfect competition). The **entropy index** provides a measure of the **disorder**, ranging from 0 to infinity

- In an industry where there is only one firm, the entropy is minimal:

$$\boldsymbol{EI = \sum_{i=1}^{N} 1 \cdot \ln\left(\frac{1}{1}\right) = \sum_{i=1}^{N} 1 \cdot 0 = 0}$$

- In an industry with identical firms, the entropy increases together with the increase in N, theoretically approaching infinity in perfect competition

$$EI = \sum_{i=1}^{N} \frac{1}{N} \cdot \ln\left(\frac{\frac{1}{N}}{\frac{1}{N}}\right) = \frac{1}{N} \cdot N \ln N = \ln N$$

A remark on the role of the degree of disaggregation. In industrial economics we have different level of disaggregation (manufacturing sector → automotive industry)

Concentration indices are used to compare concentration:

- Between different industries at the same time
- In the same industry over time

The degree of disaggregation needs to be the same in order to meaningfully compare results. Indeed, the degree of disaggregation of the industry has an impact on concentration (typically, increasing the degree of disaggregation, concentration increases as well)

e.g. Biscuit industry (here we are aggregating many different firms that are competing in different industries) → specific segment: confectionery ind. → food ind.. The more specific the segment, the higher the likelihood that we will find a "leading" firm.

- The degree of disaggregation is defined relying on the classification of production activities
- There are several national and international classification of production activities that allow for a shared definition of the industries

18.2.2 Relative Indices

Relative concentration indices measure the degree of **size inequality** among firms

- Graphically → Lorenz Curve
- Numerically → Gini Index

These measures are traditionally adopted in labor economics in order to measure income inequality and income distribution.

18.2.2.1 Lorenz of curve

Firms are put in increasing order → It is to some extent the opposite of the concentration curve.

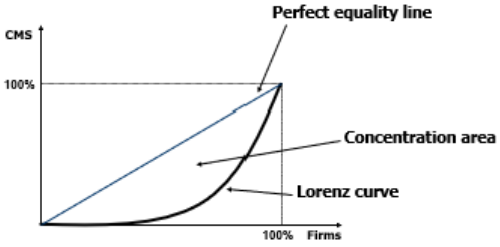
- On the horizontal axis: cumulated share of the number of firms

  e.g. The 10% smallest firms…
- On the vertical axis: cumulated market shares

e.g. …produce 20% of the industry output

The upper bound of the **Lorenz curve** is given by the **perfect equality line**

- The **perfect equality line** is the shape the Lorenz curve would get assuming perfect equality:
  - The first 10% of the firms produce 10% of the output
  - The second 20% of the firms produce 20% of the output
    …



Note: the portion of plane comprised between the Lorenz curve and the perfect equality line is called **concentration area.**

The more convex the shape of the curve, the larger the inequality. It can be seen as a bow. The more you stretched it towards the X-Axis, the concentration will be at it maximum. On the other hand, if concentration is not large at all (perfect competition), the Lorenz curve will equate to the perfect equality line.

18.2.2.2 GINI Index

Ratio between the concentration area and the area underlying the perfect equality line

The Gini index varies between 0 and 1:

- If all the firms have the same size
  - The absolute equality line and the Lorenz curve coincide
  - The concentration area is null
  - The Gini index is equal to 0
- If one firm produces for the entire industry
  - The concentration area coincides with area underlying the perfect equality line
  - The Gini index is equal to 1

Encava and Jacquemin qualitative classification

It synthesizes industry concentration in a finite number of categories:

1) Monopoly: there is one firm with a market share higher than 80%
2) Dominating firm: there is one firm with a market share between 50% and 80%, and the others are much smaller
3) Duopoly: 2 firms of similar size control at least 80% of the market
4) Asymmetric oligopoly: 3 or 4 firms control at least 80% of the market, the highest share being around 40%
5) Symmetric oligopoly: 3 or 4 firms equally control at least 80% of the market
6) Asymmetric competition: the largest firm holds a market share between 20% and 50%
7) Symmetric competition: the largest firm controls at most 20% of the market