No institution is a free lunch: 
a reconstruction of Ronald Coase

n. 633 – Febbraio 2012
Abstract - The two major contributions of Ronald Coase, written at distant points of his long life, have been often interpreted as different and, somehow, contradicting views of the merits of the market mechanism. We argue that the underlying point of the two articles is the same and it can be summarized by the statement that no institution is a free lunch. When the unity of the Coasian theory is properly understood, it offers a powerful challenge to standard neo-classical production theory and opens new analytical tools to understand and to compare the institutions of production.

Keywords: Coase, Returns to Scale, Transaction Costs, Production Entitlements.

Jel: B290, D020, D290, L230

The ideas, contained in this paper, were first presented during the ISLE 2010 meetings held in Bolzano at the workshop Deconstructing Coase. I am grateful to all participants for their comments and suggestions. Particular thanks are due to Antonio Nicita and to Matteo Rizzolli.

Ugo Pagano, Università di Siena and Central European University – ugo.pagano@unisi.it
1. Introduction.

Ronald Coase is now 100 years old. In his long life he has sharply criticized the economic for the neglect of the profession for its own subject matter. According to Coase, the economic approach lies in the rational choice explanations of the interactions which they examine and “while it may ultimately rejuvenate the study of law, political science and sociology had nonetheless……… serious adverse effects on economics itself” (Coase 1988 p. 3).

“What has been developed is - according to Coase – an approach divorced (or which can be divorced) from its subject-matter. Indeed since Man is not the only animal that chooses, it is to be expected that the same approach can be applied to the rat, cat and octopus in the same way in which a man does”. (Coase 1988, p.3). However, in spite of its imperialism outside its traditional subject-matter, economics is dramatically losing its own field. Indeed, “one result of this divorce of theory from its subject-matter has been that the entities whose decisions economists are engaged in analyzing have not been made subject of study and in consequence lack any substance.” (Coase, 1988 p. 3) The undesirable consequence of this state of affairs is that in current economic theory we have “consumers without humanity, firms without organization, and even exchange without markets” (Coase 1988 p. 3).

In his long life, Coase has tried to re-address this situation. However, while his contributions have been very much appreciated, the overall harmony of his theoretical building has often been missed. Coase has even appeared to show some inclination for schizophrenia - a disease that could forgiven only considering that his two major contributions are located at distant points of his long life. In the words of Guido Calabresi (1991, p. 1211), the trouble with the Coasian personality could be described as follows:

Some fifty-five years ago, in a seminal article called The Nature of the Firm, a young socialist named Ronald Coase sought to explain the existence of firms, of organizations within which markets were replaced by hierarchy and command. Twenty-five years later, in The Problem of Social Cost,
Ronald Coase, by then a middle-aged libertarian, indicated how markets could replace hierarchy and command structures to the perceived benefit of those who organized them.

However, in spite of this superficial impression of divided personality, Calabresi (1991, p. 1212) points out that:

\textit{As Professor Coase recognized, the underlying point is the same.}

In this short paper we will try to show that this latter statement of Calabresi is correct.

However, we will take a different route and will not only claim that Ronald’s deconstruction into “Young Ronald” and “Old Ronald” is unjustified but also that re-constructing the unity of his self allows us to understand the power of his theory and the strength of his challenge to orthodox economics. In the next section we will examine the motivations underlying the de-construction hypothesis of Ronald Coase. In the third section we will show how Ronald can be properly re-constructed following his own self-analysis. In section 4, we will claim that some fundamental propositions of orthodox economics cannot meet the challenge of a properly re-constructed Ronald. His fundamental message may sound obvious. It can be even expressed in a short sentence: \textit{no institution is a free lunch!} However, its implications are really far-reaching and – we argue in the concluding section - they have major implications not only for the ontology of economics (or its subject-matter) but also for its methodology.

\textbf{2. A three-step deconstruction of Ronald.}

Ronald’s de-construction can be made in three steps:

I) 	extbf{Young Ronald (1937) claims:} In a world of zero transaction costs firms would not exist. Firms exist only when market transaction costs are positive.

II) 	extbf{Old Ronald (1960) claims:} In a world of zero transaction costs (and well defined property
rights) all externalities (harmful and beneficial effects) are internalized (taken into account) by market transactions.

**III)** Old Ronald (1960) also claims: That the first statement does not contradict the second. He does realize that in that in (I) he is emphasizing the cost of using the market mechanism and, in (II), with the so-called Coase theorem, he is claiming that all sorts of externalities can be handled by costless market transactions.

Let us consider each one of these three steps:

**I)** In his famous (1937) article Coase observes that the allocation of resources within the firm is not governed by the price mechanism which characterizes a market economy. In a market economy "the direction of resources is dependent directly on the price mechanism" (p. 34) and "the allocation of factors of production between different uses is determined by the price mechanism" (p. 35). If in a market economy the price of factor A becomes higher in X than in Y, then A moves from Y to X until the prices in X and Y become equal. By contrast within a firm "if a workman moves from department Y to department X, he does not go because of a change in relative prices, but because he is ordered to do so". Thus, Coase defines firms as islands of central planning or, quoting D. Robertson, as "islands of conscious power" existing in "the ocean of unconscious cooperation defining the market economy". So he can rephrase the question "Why do firms exist" in the following way. "But in view of the fact that it is usually argued that co-ordination will be done by the price mechanism, why is such co-ordination necessary? Why are there these islands of conscious power?" (1937, p. 35)

Coase's answer is that the existence of firms can only be explained by admitting that the use of the price mechanism is costly and the allocation system used within the firm can be relatively cheaper than market transactions. Discovering the relevant prices, negotiating and enforcing contracts are all costly activities which are required by the use of the price mechanism and which can be greatly reduced if firm-type co-ordination replaces the market system.

For instance, under firm-type co-ordination, a factor of production should not negotiate with all the other factors of production the terms under which their co-operation is going to take place. These many contracts replaced by the typical contract under which a factor is employed in a firm. This is one whereby "the factor of production for a certain remuneration (which may be fixed or fluctuating),
agrees to obey the directions of an entrepreneur within certain limits". Such conditions define the employer and employee or the master servant relationship. "It is this right of control or interference, of being entitled to tell the servant when to work within (the hours of service) and when not to work, and what work to do and how to do it (within the terms of such service) which is the dominant characteristics in this relation and marks off the servant from the independent contractor, or from one employed merely to give to this employer the fruits of his labour." (1937, p. 49) The existence of the firm or of employment contract, which defines its limits, does not only imply that each person negotiates with one person (the employer) instead than with many persons (the other factors used in the firm). It makes also the contract simpler because it can leave to a future date the exact specification of the workers' activities. When one of the many ex-ante possible states of the world will occur, then the employer will ask the worker to perform a particular activity Thus, the existence of the firm implies a considerable saving of market transaction costs. One simple contract substitutes for many complex market transactions.

But if planning and firm type co-ordination imply such a considerable saving of market transaction costs, the economy should become a single-firm economy, or in other words a centrally planned economy. Coase’s solution of this problem is again grounded on the observation that no institution is a free lunch and that one must always compare the costs of alternative institutions.

When expanding and internalizing additional market transactions, firms face increasing organizational costs. According to Coase, there are decreasing returns to management or to the entrepreneur function. As the size of the organization increases the entrepreneur is more likely to fail "to place the factors of production in the uses where their value is greatest, that is, fails to make the best use of resources" and smaller firms can compete him out of the market. "Naturally a point must be reached where the costs of organizing an extra transaction within the firm are equal to the costs involved in carrying out the transaction in the open market or to the costs of organizing by another entrepreneur." (1937, p. 43). In a competitive system the expansion of the firm will stop at this point where organizational costs are minimized. The central planning occurring within the firm and the market activities existing outside it will therefore be combined in the optimal way. "In a competitive system, there is an "optimum" amount of planning" (1937, p. 37).

In Coase's view the optimal mixture of firm type and market type organization which is
achieved by the competitive system will change over time because technological innovation is likely to change the relative costs of these two ways of organizing economic activity. An increase (decrease) of the size of the firm will result if a new invention makes firm type organization cheaper (more expensive) than market type organization. "For instance, if the telephone reduces the costs of using the price mechanism more than it reduces the costs of organizing, then it will have the effect of reducing the size of the firm" (1937, p. 46). The mix of planning and markets must be recalculated by the competitive system, each time that this is required by technological change. The analysis must always be comparative as Coase stated in a paper, written more than 50 years later, that had the purpose of clarifying the meaning of his contribution (Coase 1991 p. 59):

..... the main transaction costs that are saved are those which would otherwise have been incurred in market transactions between the factors now cooperating within the firm. It is the comparison of these costs with those which would have to be incurred to operate a firm which determines whether it would be possible to establish a firm.

Thus, it is true that in his 1937 article Young Ronald claims that markets are costly and firms exist for this reason. However, his claim is more general: all institutions are costly and an institutional mix of markets and firms is going to characterize a competitive system. The relative costs change in different times, productive sectors and in all sorts of other circumstances. The basic message of the article is that no institution is a free lunch and economists must deal with a complex comparative institutional analysis.

II) The so-called Coase theorem, contained in the 1960 paper on The Problem of Social Cost, seems to contradict the claim that markets are costly. Standard Pigouvian theories had associated externalities with the need of state intervention. Coase argued that, also in these cases, (costless) markets could solve the problem. What impressed most economists was that, in these circumstances, the initial distribution of property rights was largely irrelevant. Externalities would have been internalized even if the pollution rights were granted to the polluters. Introducing markets for externalities implied that polluters faced a similar cost in terms of forgone income. In a world of zero transaction costs, the rights to resources would always flow to (or be detained by) the individuals who valued them the most and, in the process, they would be subdivided and combined in such a way to
yield the maximum value\textsuperscript{1}.

Coase challenged the view that the externality was a one-way-interaction from individuals causing a damage to some victims and re-addressing the damage required necessarily interventions such as Pigouvian taxes, regulations or liability rules. Coase (1960 p. 96) pointed out the reciprocal nature of the problem. Reducing the pollution in a river would increase the production of fish but it would decrease other forms of production. “If we assume that the harmful effect of pollution is that it kills the fish, the question to be decided is, Is the value of the fish lost greater or less than the value of the product which the contamination of the stream makes possible? It goes almost without saying that this problem has to be looked at in total and at the margin.” The individuals who are to carry out the most valuable production processes will be willing to offer a price to compensate the reduction of production of the other party. However, in order for these transactions to take place, “it is necessary to know whether the damaging business is liable or not for damage caused, since without the establishment of this initial system of rights there can be no market transactions to transfer and recombine them. But the ultimate result (which maximizes the value of production) is independent of the legal position if the pricing system is assumed to work without cost.” (Coase 1960 p. 104).

Thus, it is true that, in the 1960 Old Ronald argued that markets could solve problems that according to established theory required forms of state intervention. However his claims were subject to the assumption that the pricing system was assumed to work without cost.

\textbf{III)} One can de-construct Coase into a young socialist 1937 Ronald emphasizing the cost of using the market transactions and a 1960 old conservative Ronald focusing on the virtues of costless market transactions. According to a well-established wisdom this schizophrenic approach could be justified by this sentence (with some doubts attributed to Churchill or Clemenceau):

\textit{If a man is not a socialist by the time he is 20, he has no heart. If he is not a conservative by the time he is 40, he has no brain.}

\textsuperscript{1} This proposition was never stated as theorem by Coase. Stigler (1966) re-formulated it as the Coase theorem. The proposition may sound tautological and has, indeed, been accused of being so (see for instance Usher 1998). However, its purpose was simply to show the paradoxical implications of the world without positive transaction costs which was implicitly assumed by most economists.
However, if somebody was going to live for more than one 100 years, he could afford to be a socialist at the age of 27 and a pro-market conservative at the age of 50.

3. Ronald re-constructing Ronald.

In spite of its consistency with the accepted wisdom on political psychology, Ronald’s deconstruction is not tenable. In the introduction to his essays Coase (1988 p. 14) claims:

*I showed in “The Nature of the firm” that in the absence of transaction costs, there is no basis for the existence of the firm. What I showed in the “Problem of Social Cost” was that, in the absence of transaction costs, it does not matter what the law is, since people can always negotiate without cost to acquire, subdivide and combine rights whenever this would increase the value of production. In such a world the institutions which make up the economic system have neither substance nor purpose."

Indeed, Ronald’s deconstruction is only tenable if we stop reading his article at the beginning of Section 6. In that section (with the title “The Cost of Market Transactions Taken into Account”) Coase points out that the market is and cannot be the only mechanism by which the economy is organized:

"It is clear that an alternative form of economic organization which could achieve the same result at less cost that would be incurred by using the market would enable the value of production to be raised. As I explained many years ago the firm represents such an alternative to organizing market transactions” (Coase 1960, p. 115).

The 1960 article extends the market vs. firms issue and generalizes the analysis to other institutions.

"But the firm is not the only possible answer to this problem…. an alternative solution is direct governmental regulation” (Coase 1960 p. 116).

Thus, Ronald shows the way to the reconstruction of his own un-divided self. The so-called
“Coase theorem” (but there is no theorem according to Coase!) is perfectly consistent with the 1937 article and they add up to one consistent theory. According the “Coase theorem”, in a world of zero transaction costs, markets can internalize all externalities. State regulation or Pigouvian taxes, as well as firms, are unnecessary in that framework. Unsurprisingly, if one institution (the market) is assumed to be costless, only two things can happen:

(1) If the other institutions are costly, they disappear.

(2) If the other institutions are also costless, they may co-exist but their mix becomes an irrelevant economic issue.

The consequences of the analysis not only continue to apply to the comparative analysis of markets. They can also be generalized to any number of institutions.

4. Marginal cost and the challenge of Re-constructed Ronald.

In the standard building of microeconomics the existence and the size of firms are explained by U-shaped cost curves. An initial phase of increasing returns explain why people cluster in firms while the subsequent phase of increasing costs explains their finite dimension. A re-construction of Ronald, integrating the 1937 and the 1960 articles, challenges this technological approach which relies only on technology to determine the main features of the organization of production.

In his 1946 article on the “Marginal Cost Controversy”, Coase struggled with the paradoxes of marginal cost pricing observing that, under a regime of increasing returns, this rule involved a loss relatively to the average cost, sustained to produce all the units. He observed how when this pricing rule is applied by public utilities then

Consumers who buy products which are produced under conditions of decreasing average costs will therefore obtain products for any given expenditure embodying a greater value of factors than those who not. There is a redistribution of income in favour of consumers of goods produced under conditions of decreasing average costs. (1946, p. 85).
And in the (1988, p. 19) introduction to his essay he restated that “Marginal Cost Pricing is largely a policy without merit”.

In this section we will develop the (1946) Coasian critique by referring to the theories included in the 1960 article.

Let us start by observing that (Dis)economies to scale can be seen as the effects that production of a certain number of units has on the provision of units of the same kind while (dis)economies of scope refer to the interactions among different production processes. From this perspective, the positive and negative effects (increasing and decreasing returns) which the production of some units has on units of the same kind can be viewed as a particular type of externality that can be internalized by some institution such as markets or firms.

Consider now a world in which one institution –the market- works at zero cost. This is the world of the Coase theorem where all externalities could be internalized if the relevant property rights were properly defined. In this perspective, the issue becomes which are the property rights and the related markets that are missing for this type of externality.

Economies to scale can be due to the fact that the costs of production can be decreased by sharing a common facility such as, a larger and a cheaper plant. In this case a public o private central agent can define production entitlements (rights/duties to produce) and auction them to the different agents. The central agent can auction production entitlements at prices corresponding to different plant sizes. All production entitlements will have the same price for each unit. If a greater facility entails lower entitlement productions price (and the facility is fully utilized), it will be convenient to expand facility size. The supply and demand of production entitlements will be equalized at a price where the shared facility (for instance a plant) has an optimal size. In this situation each agent will have constant returns to scale in the sense that unit costs are independent of the amount produced.

In other cases, the units can be sequentially produced and it is cheaper to produce the first units relatively to the subsequent units. Earlier units act as a facility for the later units. This is the case when innovation and learning characterize the production process. Unlike the case of production units sharing a common facility, the units have different production costs. Suppose again, that it is possible to define and auction production entitlements before production. In this case, also this other
externality, taking the form of economies to scale, can be internalized. The producers of the subsequent units will be willing to pay a price for producing their units which is greater than the price to be paid to produce the preceding units. The equilibrium prices will again be such that we have constant returns in the sense that unit costs, being equal to the sum of production costs and production entitlements, are independent of the number of units being produced.

A similar argument holds for the case of diseconomies to scale.

Also in this case, diseconomies may depend on a single facility whose use is characterized by the fact that increasing production involves higher unitary costs. Again a facility of optimal size could be shared and producers will be charged the same price for the production entitlement of each unit. Also in this situation each agent will face constant returns to scale in the sense that unit costs are independent of the amount produced.

Similarly to the case of economies of scale, also in the case of diseconomies of scale the units can be sequentially produced but, in this case, it is cheaper to produce the subsequent units relatively to the early units. Again, production entitlements with different prices can compensate for these negative externalities and equalize unit costs so that they become independent of the quantity supplied by an individual producer.

Let us now consider in more detail the second (more complex) case, in which the production of the preceding units has positive or negative effects on the production of the subsequent units and different production entitlements command different prices.\(^2\)

Suppose that there are two units of a good x. Call \(x_1\) the first unit and \(x_2\) the second unit of x to be produced. Define new commodities \(y_1\) and \(y_2\) as the rights (or the "duties" when the amounts paid are negative) to produce respectively the first unit \(x_1\) and the second unit \(x_2\) of x. Suppose now that we have economies to scale in the production of x so that producing \(x_2\) is cheaper than producing \(x_1\). Producers will be willing to pay a price for \(y_2\) higher than the price for \(y_1\). In particular, under a competitive regime, producers will be willing to pay a higher price for \(y_2\) which equals the difference in costs between the price of \(x_1\) and the price of \(x_2\).

---

\(^2\) A formalization of the argument can be found in Chichilnisky G., Heal G., Pagano U. (1994). I wish to thank Kenneth Arrow and Frank Hahn who gave us very useful suggestions.
In other words, if we denote by \( c_1 \) and \( c_2 \) the costs of producing respectively \( x_1 \) and \( x_2 \) by \( p_1 \) and \( p_2 \) the prices of the production entitlements \( y_1 \) and \( y_2 \), then, in a competitive equilibrium, we must have that:

\[
p_1 y_1 + c_1 x_1 = p_2 y_2 + c_2 x_2
\]

A competitive equilibrium, with complete markets for production entitlements, is characterized by constant returns because the costs of buying the rights of producing a certain unit must in equilibrium offset the cost differentials (due to (dis)economies to scale) among the units of product.

For the single producer, complete markets imply constant returns to scale under any regime of (dis)economies to scale. Increasing or decreasing returns are due to market failure in the sense that markets for \( y_1 \) and \( y_2 \) fail to exist. In other words increasing returns are not the cause of market failure; rather both increasing or decreasing returns are the consequences of market failure. When the markets for positive production externalities are lacking, a single producer who is internalizing them faces increasing returns to scale. Indeed, positive or negative externalities, which cannot be internalized by markets (or internalized at comparatively higher transaction costs) are, consistently with the 1937 Coasian approach, the reason why firms exist.

By contrast under a regime of constant returns there are no positive or negative production scale externalities and firms have no reason to exist. Thus, an evident route links together Old Ronald (Coase, 1960) and Young Ronald (Coase, 1937) and leads inevitably to the conclusion that firms cannot exist only for technological reasons. The standard neoclassical explanation is based on the assumption that a single producer faces a U-shaped cost curve (characterized by an initial phase of decreasing costs and a subsequent phase of increasing costs). However, in a world of well defined property rights and zero transaction costs, only constant costs are possible and the U-shaped curve does not make sense. In a world of well-defined property rights and complete markets, non-constant returns to scale are wiped out by the markets internalizing the relevant externalities and firms have no reason to exist. Under these assumptions, we are in the so-called world of the Coase theorem where all externalities, including those arising from (dis)economies to scale are internalized. In such a world,
firms, Pigouvian taxation, regulations and other (costly) institutions have no reason to exist because there is no cost involved in the use of the market mechanism.

Ronald's reconstruction poses serious challenges to orthodox economic theory.

In the first place, the non-existence results which are traditionally attributed to increasing returns must rather be ascribed to incomplete markets and can be solved by completing the markets introducing the appropriate property rights.

Secondly, returns to scale cannot anymore be seen simply as technological phenomena but must be looked at as a joint consequence of the technology and of the property right system.

In the third place both increasing and decreasing returns must be seen as a sign of market failure to internalize production externalities. Thus the firm can be seen as an institution which achieves that internalization when the market transaction cannot achieve this result or can only achieve it at higher costs. Observe that both types of market failures (decreasing and increasing returns) can be the cause for the existence of the firm. In both cases, the firm can emerge as the appropriate local governance structure to internalize the externalities that central governments and markets fail to internalize\(^3\). The firm is not simply a hole of market incompleteness. It is a centralization of market transactions that is made possible by a decentralization of some powers of the public ordering to a private ordering. Firms' private governance structures may be particularly precious in the internalization of the externalities arising from the production of homogeneous units. When homogeneous units require different levels of effort, the contributions of different individuals cannot be inferred from the quality and quantity of their products and must be directly monitored. Market transactions are likely be characterized by costly court litigations, to be adjudicated by public judges who have made little specific investments in the understanding of the internal life of the organization. According to Alfred Sloan (1963 p. 458) the role of courts may be performed by management and the Executive Committee "which views the corporation as a whole and at the same time is closely familiar with operating problems, has a

---

\(^3\) Unlike Hart's (1985) new property rights approach, the firm is seen by Coase as a system of private governance. This view is developed by Williamson O. E. (1985) and Coase's (2000) analysis of the Fisher-Body merger. Pagano (2000) observes the symmetry between the Coasian theory of the firm (as a centralization of market transactions) and Fuller's (1969) view of the firm (as one possible decentralization of a public ordering). Pagano (2010) shows that the two views can be usefully joined in the "Cathedral Framework" developed by Calabresi and Douglas Melamed (1972).
somewhat judicial function”. In some cases, the private orderings, introduced by an appropriate corporate governance, can do better than public markets (Pagano 2000).

Finally when neither markets nor firms can internalize externalities arising from (dis)economies to scale there is some room for State intervention. Observe that our analysis implies that the existing level of production is a public good (bad) for all producers and consumers of the good. Thus some possible advantages of aggregate demand management follow quite naturally from our analysis. Aggregate demand management may be useful for the welfare of the individuals for the simple reason that, in certain circumstances, the level of production activity is a public good that markets and firms may fail to supply.

5. Conclusion: methodology and ontology after Coase.

The re-construction of Ronald Coase can be synthesized in one trivial single sentence: no institution is a free lunch. However, this simple point involves a genuine revolution in economic theory. All institutions have to be analyzed and assessed in a comparative perspective. The complexity of economic problems increases not only because all institutions are costly but also because their costs will be different for different technologies. For different technologies, adopted in different productive sectors, different institutional mixes will prevail in the economy.

Even the least costly institutional mix is still costly and the advantages of the mix have to be compared with its costs. In some cases, it will not worthwhile to have any institution by which individuals can internalize in their behavior the effects that their actions have on the other individuals. In these cases externalities should remain as such: internalizing them does not make any economic sense. Moreover, while there will be a tendency to move to less costly institutional mixes, one should not forget that changing the mix is itself costly. Thus path dependency and importance of past history are a necessary consequence of the Coasian approach. These consequences of Ronald’s reconstruction become even more evident when one considers that costly entities, such as goods and institutions, are not only substitutes but also complements. Institutional complementarity can severely limit the possibilities for efficient institutional substitution. Institutional change is necessarily path-dependent.
and must be framed in the evolutionary dynamics of complex systems.4

The no-free-lunch hypothesis is a well-known foundation of orthodox economics and it may sound paradoxical that re-constructing Ronald along these lines could be such a challenge for the standard approach. Indeed, the no-free-lunch hypothesis seems to imply that individuals, faced with scarcity, have to make rational choices on the basis of market prices, and the standard methodology seems to follow quite naturally from the ontology of universal scarcity. However, in some respects, this impression is misleading: it involves restricting scarcity to material scarcity and, at the same time, assuming free rationality and at least one free institution (the market). Or, in other words, it is to assume that we have no cognitive or institutional scarcity. In this way, in the orthodox approach there is a divorce between the subject-matter of economics (or its ontology which should consider all forms of scarcity) and its methodology (that is implicitly a world without cognitive and institutional scarcity). By making it clear that no institution is a free lunch, Coase has not only shown the contradictions and the limitations of the orthodox approach but has also re-aligned economic methodology with its ontology, opening many fruitful ways to study real-life economies.

4 Institutional complementarities imply that any analogy between evolutionary biology and economics must take into account the degree of complexity of the units that are being analyzed (Pagano, 2011).
References.


