Journal of Economic Issues

ISSN: 0021-3624 (Print) 1946-326X (Online) Journal homepage: www.tandfonline.com/journals/mjei20

Transaction Costs and the Historical Evolution of the Capitalist Firm

C. N. Pitelis

To cite this article: C. N. Pitelis (1998) Transaction Costs and the Historical Evolution of the Capitalist Firm, Journal of Economic Issues, 32:4, 999-1017, DOI: 10.1080/00213624.1998.11506106

To link to this article: https://doi.org/10.1080/00213624.1998.11506106

JOURNAL OF ECONOMIC ISSUES
Vol. XXXII No. 4 December 1998

Transaction Costs and the Historical Evolution of the Capitalist Firm

C. N. Pitelis

The aim of this paper is to critically assess the contribution of transaction costs economics (TCE) in explaining the nature of the firm. TCE suffers from conceptual problems and lacks a historical-evolutionary basis. The pursuit of a historically informed evolutionary perspective exposes the limits of TCE and points to the need for a synthetic, integrative framework.

Transaction costs economics dates back to Ronald Coase's 1937 classic article. There, Coase claimed that given markets, firms should not exist in the absence of inherent market failures related to, what we call today, transaction costs. Such are costs of measurement, information, bargaining, contracting, enforcing, and policing agreements. Following O.E. Williamson's [1975] now classic book, these costs can be attributed to bounded rationality, opportunism, and asset specificity. The internalization of (imperfect) markets by firms (hierarchies) is said to increase overall efficiency. Firms are the result of a contractual arrangement between those involved in transactions. Removing market inefficiency is seen as both the reason for and effect of firms. Given costs of hierarchy (organization, management costs), the boundary of the firm is defined as being where an extra transaction can take place equally efficiently by organizing this transaction by the market or another firm. Accordingly, the observed market-hierarchy mix enhances overall efficiency. Market internalization, due to transaction costs, can also explain the evolution and strategies of firms, notably vertical integration, the M-form organization, the conglomerate,

The author is Director of the Center of International Business and Management at the Judge Institute of Management Studies, University of Cambridge. *Earlier versions of this paper have been presented in numerous conferences and workshops. I am grateful to their participants for comments and discussion, in particular Gregory Dow, John Groenewegen, Geoff Hodgson, Steve Masten, and Oliver Williamson. For comments on earlier versions, I am grateful to Keith Cowling, Douglass North, Edith Penrose, A. Pseiridis, Roger Sugden, and two anonymous referees of this journal. Errors are mine.*

and the transnational corporation (TNC) [Williamson 1981]. Hybrid forms of organization, such as strategic alliances, networks, equity joint ventures, etc., can also be attributed to the quest for efficiency, a progression from markets to hybrids to hierarchies [Williamson 1993]. In short, the quest for efficiency explains why firms, hybrids, and markets exist. Efficiency is also the outcome of the process.

Numerous papers and books, some summarized in Y. Eggertson [1990], C.N. Pitelis [1991, 1993], and John Groenewegen [1996], have debated and critiqued the TCE perspective. In this paper, I focus on some conceptual problems of TCE, starting from the conception of the market and moving critically through the main claims of TCE.

Transaction Costs Economics: Some Problems

Our starting point is the important claim of TCE that "in the beginning there were markets." This is the explicit assumption of Coase and can be seen in a quotation from Williamson [1975, 20]. However, this claim need not be true. It depends on the definition of the market and the firm and also on the *context and timing*, whether we refer to a particular historical context or to the first emergence of firms (and markets). Let us clarify by using Coase as an example.

For Coase, markets differ from firms in that the former rely on voluntary transactions between individuals, while the latter count on the authority of one party over the other, namely, the entrepreneur over the laborer. This is one (arguably narrow) definition of a firm. One can define a firm as a single producer or as a family or friends (partners) cooperatively involved in production. The firm need not be a multiple person hierarchy, which the Coasean firm is.[1] Logically, one could argue that firms (or organizations) precede the market as firms produce while markets may not. Moreover, it could be questioned whether, historically, markets or hierarchies (families, states) appeared first. All of these criticisms are standard in the literature and are by no means original [Pitelis 1993]. They highlight the point that the Coasean firm is only one of many possible definitions of the firm.

Suppose, however, that we accept the Coasean definition. The question then becomes the *context and timing* one chooses for the beginning. Is it the beginning of a particular historical period or of life in general? This is important. If, for example, we are talking about the emergence of Coasean firms in capitalism (as, for example, Williamson's 1985 title implies), then one could be more willing to accept the claim of pre-existing markets. However, if we start from the very beginning, then surely this claim is false. Following the emergence of the human race, there was self-sufficiency of the individual or a (extended) "family" unit. Transactions, production, markets, and such other starting points adopted by various authors surely came later. Accordingly, we need to distinguish between the generic firm and the various possible (definitions of) types of firms, including the Coasean firm. Starting from self-sufficiency need not modify TCE-type analysis, but it may.[2] At the moment, suffice it to state the obvious: if we start from self-sufficiency, then we need to explain why social units, "families," or individuals (decide to) exchange a world of no exchange (costs) for one with (positive costs of) exchange. In my view, this is the question. We are back to square one; we need an explanation of exchange, production, production for exchange, markets, the firm, the Coasean firm, the whole lot!

Another tenet of the TCE perspective is that *firms are not markets*. This seems self-evident, but it is not. Once again, it depends on our definitions. If firms differ from markets, it is important to know the nature of this difference. This difference may (not) be the existence of Coase's hierarchy (authority, fiat) of one agent over another, or it could be that distinction only in certain cases or contexts-historical periods. For example, in the case of a single producer (unitary firms for P.J. McNulty [1984]), Coase's distinction breaks down by definition. This, moreover, is true for the case of a cooperative production unit (firm?). Authority and fiat may exist in a family who operates as a production unit (firm?), but this would not be a Coasean firm, as the latter presupposed an employment contract, namely, one agent working (voluntarily) under the authority of another. In the case of the family, the employment contract may only be implicit, there may not be a single source of authority as such, and/or the agents (children, wife) may not obey the principal's (father, husband) authority out of choice. In brief, we need to define our terms.

Even accepting that the firm is a multi-person unit, it does not necessarily follow that the principal has authority over an agent. This will depend on the alternatives the agent has. If, for example, market conditions are such that alternative employment is easy to find, the agent can sack the principal, not just vice versa. If so, the relationship will be symmetrical, at least *ex post*, as A. Alchian and H. Demsetz [1972] have claimed.[3] The question then becomes, why an employment contract to start with? If, for example, agents work for principals because they simply cannot do otherwise (i.e., have a principal or no employment contract at all), then the relationship is asymmetrical to start with, and even market-type relationships become hierarchical.[4] In this case, asymmetry is endemic in the (market) relationship; it is part and parcel of the system, or the capitalist market is a hierarchy [see Pitelis 1991]. The interesting question, of course, is when, why, and what is the case in each real life situation.

A third tenet of TCE is that *firms result from transaction costs market failure*. However, if markets need not pre-exist and/or need not differ from firms, this point need not follow either. Importantly, the point at hand need not follow even if the previous two points are true. Markets could operate perfectly, and yet firms (organizations) exist because they perform different functions (for example, production, as F. Fourie [1993] has claimed) and/or because they complement the market by virtue, for example, of being an integral part of the (market) system, or because they are better at doing some things than markets are. In modern jargon, it would be said that they have some distinct capabilities [Penrose 1959; Foss 1993, 1997]. In fact, when seen from a dynamic perspective, firms can be said to create (and/or close) markets, as a host of theorists on the transnational corporation have argued [Hymer 1976].[5]

A further reason why the existence of firms need not presuppose transactional market failures is simply because firms may be able to increase productive efficiency through characteristics described above as part of the non-Coasean perspective and by sheer virtue of authority. Such authority may allow them to achieve a better (from the point of view of the principal) use of specialization and the division of labor [see Pitelis 1993]. Versions of this idea go back to Karl Marx and more recently to S. Marglin [1974].[6]

In sum, (the existence of) organizations need not presuppose transactional market failure. It all depends on the nature of the organizations (and markets), which, however, is the issue under consideration.

A fourth tenet of TCE is that *the employment contract is (the result of) a contractual process*. In the Coasean assertion, agents agree to work under the authority of principals voluntarily, because they expect to gain from the increased (transaction costs-related) efficiency gains resulting from emergent hierarchies.[7] This need not be true. The alternative is that agents are obliged to do so by the principals [Marglin 1974], perhaps by employers, and/or by the state, in mutual support of each other and in the pursuit of mutual gain [Hymer 1979; Galbraith 1987]. Or they may accept such work because they simply have no better alternative, i.e., the alternative is to work for another principal or not to work at all (be unemployed). In the first two cases, the process is clearly predatory, while in the third it is contractual in a rather limited, almost meaningless sense.

It is interesting to note that the predatory theory of hierarchies receives support from Douglass North [1981, 1991]. It is notable here that this support follows his earlier adoption of an efficiency perspective. He rejected that view in favor of the predatory view after his analysis of institutions and institutional change in history. On this particular issue, namely, whether firms emerge for contractual or predatory reasons, there is enough evidence to persuade North that predation is more likely to have been the case, at least historically.

A fifth tenet of TCE is that *firms are the result of efficiency (not power)*. This need not be true. It depends on the definition of efficiency, on whether we are concerned with *ex ante* or *ex post* efficiency, and on the vantage point taken (efficiency from whose point of view, for whom?). In brief, the Coasean argument need not follow if the first two points we considered are not true. In the case of the third point, efficiency may result but for different reasons, including the possibility of coercion. In this last case, the result will not be efficient in the Pareto sense. The same is true in the case of the fourth point, if the process is not contractual. In Marglin's [1974] perspective, firms (the factory system) may ensue even if they result in inefficiency, as long as firms enhance the power and control of capitalists over workers, so as to allow the former to gain, even at the expense of the latter.

Whether large firms emerge for efficiency or power-related reasons cannot be answered on a priori grounds. It requires detailed empirical work. However, despite appeals to evidence, there is simply no corroborating support to the TCE perspective on this particular issue, i.e., that firms cannot be the result of power and (thus) can potentially lead to inefficiency. In fact, the only existing historical evidence—provided by K. Stone [1974], Marglin [1974], and Hymer [1979]—seems to support the power perspective. Although TCE supporters have marshalled evidence to its favor, they have not considered historical evidence on the issue.

Williamson's approach is interesting and illustrative. Consider his [1985] reinterpretation of Stone's [1974] paper on the transformation of the steel industry in the United States in the late nineteenth and early twentieth centuries. Stone suggests that in the late nineteenth century, a strong, skilled, steel workers' union undermined capital's power over labor and led to apparent inefficiencies and stifled labor-saving innovations. Following a challenge to the union by A. Carnegie and H.C. Frick at Carnegie's Homestead mill, and with the support of the state and federal governments, the union was defeated. That was followed by further union defeats by other companies. For Stone, "the decade that followed the Homestead defeat brought unprecedented developments in every stage of steelmaking" [1974, 66].

Stone attributes these changes to the class struggle between capital and labor, and he goes on to suggest that alternative work organizations could achieve similar efficiency gains without hierarchical imposition.

Williamson [1985] doubts this. He focused on the efficiency gains observed and proposed the possibility that

> The incentive to challenge the union in the first place and the efforts to organize labor subsequently were principally geared to achieving efficiency [1985, 235].

He further observes (rather surprisingly) that

> Given . . . the large efficiency gains that Stone reports, the efficiency hypothesis or *a combined efficiency-power hypothesis cannot be rejected!* [1985, 236; emphasis and exclamation mark added].

Another interesting case is Williamson's [1996] interpretation of the Central Selling Organisation (CSO) of De Beers' practices on sales of gem-quality, uncut diamonds. De Beers has market dominance (an 80-85 percent market share), and it organizes the sale of diamonds as follows. Stones are sorted into imperfectly homogeneous categories and are offered for sale to prospective invited buyers in preselected blocks (sights) at non-negotiable prices. In the event a buyer rejects the sight, she or he is offered no alternative box. In addition, buyers who reject the diamonds offered are deleted from the list of invited customers! Williamson suggests that such apparently blatantly power-based practices have potential efficiency gains. The all-or-none purchase rule is said to economize on oversearching, while "what the in or out trading rule does is to introduce an intertemporal dimension that further reduces search costs and brings reputation effects to bear more effectively" [Williamson 1996, 27].

The result for Williamson is efficiency increases through reduced transaction costs. However, who has ever disputed that? Indeed, this is exactly the alleged *raison d'être* of the firm in the very Coasean perspective: that authority can increase efficiency!

In considering the last point, it is important to ask how can we distinguish between authority (the *sine qua non* of the Coasean firm) and power, which Coase and Williamson reject. One answer could be that market power is acceptable, but Marglin-type power is not. However, it is not obvious how one can distinguish between bad power and good (efficiency-enhancing) power and/or whether the distinction is relevant or useful.[8]

Another issue arises concerning the alleged superior efficiency properties of hierarchies and the process through which an efficient institution, or organization, supersedes an inefficient one. I do not intend to summarize here the huge debate on the efficiency of selection, path dependency, and related topics. Suffice it to say that Williamson himself observes that the efficacy of selection remains controversial [1996, 3]. He also treats path dependency as one of the unresolved tensions between TCE and organization theory in Williamson [1993].

All these issues point to the possibility that efficiency and (monopoly) power need not be seen as polar opposites. For example, they could co-exist; one could lead to the other, and/or they could even be different sides of the same coin. At the very least, these observations suggest that this fifth point need not be true.

A sixth tenet of TCE is that the *objective of firms is to reduce market transaction costs*. This relates to the previous point and can be seen as a stronger version of it. It need not be true either. First, firms' objectives may be to increase their power over rivals, markets, labor, and/or the state, as numerous authors have argued. Interestingly, that firms *could* do so also follows from the very neoclassical tenet of profit maximization. That firms *should* do so is at the very heart of the competitive strategy analysis of M. Porter [1980], which in effect is an application of industrial organization ideas to strategic management. Could firms be paying millions of dollars in consultant's fees to receive advice on how to increase their power over rivals, potential entrants, customers, and suppliers (in Porter's well-known competitive forces model) and at the same time have as their major goal the reduction of market transaction costs? Now that would appear rather irrational, and rationality (even bounded) is part of the TCE agenda.

It is interesting that Williamson does not reject the idea that the goal of firms is to maximize profits. If so, as noted above, it is simply unclear why firms should not use all available methods at their disposal to do so, namely, reduce transaction costs *and* increase their power over markets, labor, etc., if they can get away with it, and for as long as they can. Not to do so simply defies self-interested behavior (let alone opportunism), the very basis of Williamson-type analysis. Moreover, firms can pursue non-transaction costs reducing means of increasing profits, even if their original success was due to efficiency gains. This is confirmed by a host of everyday experiences, notable being the case of Microsoft.

> In a telling judgement upon the methods Microsoft used to maintain its predominance in the software industry, the US Justice Department said: "While the company fairly and legally climbed to the top of the industry ladder, it used unfair and illegal practices to maintain its dominant position" [Kehoe 1994].

Which is to say, that efficiency can lead to market power, which may then be used to further market power, as one would expect!

The last tenet of TCE we examine here is that *firms evolve in order to reduce market transaction costs* (or at least by reducing transaction costs). Williamson's [1981] explanation of vertical integration, the M-form organization, conglomeration, and transnationalization is in terms of firms pursuing such strategies or evolving-growing in order to reduce market transaction costs. This relates closely to the previous point, so I will not belabor it. Suffice it to note again that this flies in the face of all early literature in industrial economics to the effect that (market) power is the main reason for the growth of firms and/or for the pursuit of their strategies. Relatedly, it flies in the face of firms themselves paying consultants to advise them on how to increase their market power! It stretches credibility to its limits not to accept what is supported by logic, previous theory, and firms' own behavior, that in order to make profits, efficiency increases and (market) power increases (can) go hand in hand.

As already noted, Williamson [1985] could not reject a combined efficiency-power hypothesis. However, he supports his preference for efficiency in terms of the concept of power being hard to operationalize, a point he has reiterated recently [Williamson 1996]. There are various problems with this claim. The idea that power has not been operationalized is not fully accurate. For example, in the ownership versus control debate, sociologists, notably S. Nyman and A. Silbertson [1978] and A. Francis [1980], have gone some way toward providing operational measures of owners' control by identifying the existence of interlocking directorships, kinship networks, background of the chairperson of the board, etc. In the political science literature, moreover, there have been numerous attempts to operationalize power, with varied success. S. Clegg [1989] provides a critical survey. Last, but not least, if power has theoretical merit, then it cannot be rejected simply because it cannot be made operational, as Williamson defines this term. The proper way of rejecting power is to test against TC-type efficiency, having first operationalized both transaction cost efficiency and power and then finding power not to have any explanatory power at all. At the moment, we are simply very far from being able to do anything like this.

All these bring us to the evidence in favor of TCE. There is indeed substantial evidence to support the claim that TC-type theorizing has merit [e.g., see Caves 1982 for an early survey and Masten 1996]. There are also problems, notably that estimated equations are mostly of the reduced form type and do not adequately test the hypotheses in hand [Masten 1996]. I do not wish to enter this debate here, other than to note that the evidence to which TCE proponents appeal is almost exclusively concerned at this stage with the make or buy decision in general and with vertical integration and/or the internalization of intangible assets by conglomerates and/or TNCs in particular [Hennart 1991]. It is not surprising that TCE-type theorizing has explanatory power here [Foss 1997]. However, there is simply no evidence in favor of what are arguably the main assertions of TCE—that markets differ from firms, and that firms' existence presupposes market failure; there is no evidence regarding the TCE claims on the nature of the firm, namely, that the employment relation is the result of a contractual process. The possibility of testing these has not even been addressed. The evidence, moreover, on transaction costs efficiency versus power being the reason for and/or result of the emergence and evolution of firms is indecisive. In my view, the evidence clearly points to the combined efficiency-power hypothesis to which Williamson has hinted.

To summarize, TCE explanations of *the nature of the firm* are based on very shaky foundations and offer no evidence in support of their main claims concerning these foundations.

Reading the above, one might get the feeling that TCE has nothing to offer. This could not be further from the truth. In my view, TCE has been an important development in the economics of organizations and political economy more generally. The main reason is because it has relaxed the hitherto prevalent assumption that one can analyze markets and organizations while ignoring exchange costs. Once such costs are allowed for, interesting implications follow, notably why exchange often fails to take place, as well as what explains the development or non-development of organizations, institutions, and economies. In brief, we are suggesting that the main scope of TCE is in explaining the conditions that nurture an institutional framework conducive to efficient production and exchange.

The other application of TCE is the make or buy decision, notably the case of vertical integration. In this case, in the presence of intangible assets, transaction costs-related arguments for internalization are both logically persuasive and (not surprisingly and notwithstanding problems of operationalization and interpretation) have led to successful empirical investigations [Williamson 1998].

An important question is whether we need all the paraphernalia of TCE. Do we need to assume pre-existing markets that fail due to transaction costs, which are then replaced by more efficient firms whose only aim and result in this world is to increase efficiency, where conflicting interests do not exist and everything is a happy contractual process, etc.? I believe we do not. It is excess baggage. There is no better proof of this than North's work. There, all of the above have been rejected, yet the relevance of exchange costs for institutional change and of economic (non) development is profound. In this light, I suggest that Coase has indeed come up with a fundamental insight that costs of exchange do matter and should be (part of) the explanation of the nature of the capitalist firm.

Coase's fundamental insight provides one explanation for the remarkable influence of TCE. Another is that TCE explanations of the emergence of the employment relation seem to be the only ones still available in the mainstream tradition. The pure neoclassical view of Alchian and Demsetz, for example, according to which there is no authority involved in the capitalist employment relationship, has been abandoned by its very proponents [Demsetz 1988; Williamson and Winter 1993]. On the other hand, in the radical tradition, Marglin was not concerned with the issue of why there is an employment relation to start with, but rather with why there was a transition from one form (putting-out) to another (factory system) [Pitelis 1991].

An explanation of why there is a capitalist employment relation to start with (in line with Marglin's perspective) is the Marxist one as developed by, among others, Hymer [1979]. According to this, the employment relation was imposed on workers by capitalists through coercion, often acting in concert with the capitalist state. There is historical support for such an argument by, among others, J. K. Galbraith [1987].

The Marxist view focuses on the power-control related distributional benefits for capitalists of the employment relation. As already noted, this predatory view has also been adopted by North [1981], who, however, observes the transaction costs-related savings of this development, savings important enough for him to justify the efficiency properties of capitalism as a system. Both the Marxist and North's views, however, do not consider the possibility of other (than transaction costs related) dynamic benefits resulting from the capitalist firm. This is Edith Penrose's agenda.

For Penrose [1959], the firm is a bundle of human and non-human resources under an administrative structure. This administrative structure provides for the cohesive character of the firm, in contrast to the market. Moreover, the coexistence and cooperation of the human resources within this cohesive structure allow firms to outperform markets in preserving and creating new knowledge. In this sense, firms are different from and better than markets in generating knowledge-related dynamic benefits [Pitelis and Wahl 1998]. Although Penrose did not explicitly address the issue of why there was employment relation to start with, her theory provides a rationale for it that is different from, yet consistent with, the TCE perspective. According to this, the employment relation can be said to arise because of the dynamic benefits of firms vis-à-vis markets. These are beyond any transaction costs savings resulting from this relationship.

There is now a huge literature on Penrose's contribution, including discussions of complementarities and differences with TCE [Foss 1993; Foss and Knudsen 1996; Pitelis and Wahl 1998]. Despite significant differences, Penrose's perspective can be seen to complement TCE on this issue: the potential benefits of the capitalist employment relation. However, Penrose fully recognizes the possibility, indeed the probability, that firms will make use of their power. She recognizes, for example, that vertical integration can serve as a barrier to entry, protecting a monopolistic position. She also recognizes the potentially detrimental effects to social welfare of vertical integration. She observes that "the power to keep others out may rest on legal (or illegal) control over the relevant technology, raw materials, or even essential producers' goods" [1959, 230].

The competences perspective of Penrose does not address the issue of whether the employment relationship is contractual or predatory. On this, our previous observations still hold. Last, but not least, it is worth mentioning that the potential complementarities between TCE and the Penrosean perspective hold at the conceptual level. It could be argued that an integrative framework is required, bringing together useful elements from all perspectives. Ideally, this should not be done eclectically (a collection of good points), but rather it should derive such ideas endogenously. Moreover, given that the emergence of the (Coasean) firm (the employment relation) is an actual (thus) historical phenomenon, what is needed is a historical analysis. As history involves time, such analysis would also be evolutionary and dynamic. The starting point for such an approach could be the emergence of capitalism and the capitalist firm (the main focus of Coase, Williamson, Marglin, and Penrose), and it is this beginning that is adopted in the next section.

A Historically Informed, Evolutionary Perspective

In his *Economic Institutions of Capitalism*, Williamson [1985] does not define the term capitalism! Here we define capitalism as a system of commodity production, where labor service or labor power is itself a commodity. This labor power can be purchased or sold in the market for a compensation, the wage (rate), by owners (controllers and/or coordinators) of productive means such as physical capital, who become residual claimants of any profit generated by the transaction by virtue of their coordinating, owning, and/or controlling function. These are invariably referred to as employers, capitalists, principals, or entrepreneurs [see also Pitelis 1991].

The attractive feature of this definition is that by avoiding attribution of profits to ownership and/or control by employers/capitalists, it can encompass both mainstream and radical (including Marxist) perspectives. It still suffices to make the point, however, that under capitalism there are agents who sell their labor power and principals who buy it [Braverman 1974]. Also, for as long as the agents are contracted for, they have to obey the directions (order, authority) of the principal. Why does this division of labor come about?

Coase [1937] raised a similar question. He observed that it may be that some people prefer to work under someone but dismissed this idea in favor of the view that people tend to prefer quite the opposite, i.e., being one's own master [1937, 390].[9] An alternative would be that some people realize the (transaction costs- or productive resources-type) benefits from cooperation and the need for supervision during this process. In order to achieve these benefits, they are prepared to sacrifice part of their autonomy for the duration of the productive process. This is the idea underlying Coase's own perspective as seen in the previous sections. A third possibility is that people (agents) are coerced to do so either legally and/or because of their inability to survive (live what they perceive as a satisfactory life) if they do not do so. This view is favored by the Marxist tradition. In Hymer's view, under capitalism:

> People became unencumbered by property in the twofold sense: They were free of feudal claims on their time and had not property of their own, and therefore had no alternative to working for others [Hymer, in Cohen et al. 1979, 30].

It is the need to survive, given the absence of property, that has led people to work for others in this Marxist scenario.

Whatever the relative merits may be of these conflicting scenaria, a consensus emerges that when we focus on the process of capitalist production there is an asymmetry between agents and principals in that the former have to obey the latter. Given this, the question a historical approach has to address is why historically the emerging capitalist class—the merchants—might have chosen to seek to replace market-type (spot contracts or putting-out) relationships with their suppliers-employees with the firm-like, multiple person, hierarchical employment relationship and why the (putting-outers) laborers have accepted this. Below I try to address this issue in a historically informed way, drawing on both conceptual issues and historical evidence as available.

Historically, the early capitalist institutional device through which division of labor was performed was the price mechanism.[10] Through it, the emerging capitalists—the merchants—were able to profit by the simple act of buying and selling, by turning products into commodities, or by changing use values into exchange values. Product availability in often remote markets and (thus) demand and supply considerations were the means through which a source of revenue was generated for the merchants in the process of exchange.

The use of the price mechanism was an inherently uncertain and (thus) imperfect means of satisfying the merchants' interests: profit and (thus) power. Neither the availability of producers nor the price of products and commodities-to-be could be guaranteed by the merchants. Alternatively, such availability could only be guaranteed at the price of excessive market transaction costs, that is, the costs of finding producers, negotiating with them, contracting with them and policing such agreements. Such costs could endanger the very existence of a satisfactory profit. In this sense, the price mechanism failed the merchants. It its earliest form, the putting-out system, the merchant-manufacturer, replaced the price mechanism with an employment relation, where laborers worked for the merchant-manufacturer for a wage, initially at their own place (and pace) and by making use of their own capital equipment.[11]

Within the capitalist logic of profit generation for the principals, the employment relation is transactionally superior to the price mechanism and is thus a more efficient means of the organization and division of labor. In this sense, the transactional properties of firms are not per se the reason for market failure; the reason is the transactional properties needed for capitalist control to be established. Such control is easier under the employment relation, which renders the latter transactionally superior from the capitalist point of view. Given this, (transactional) efficiency cannot in itself explain the existence of firms; the principals' objectives (thus power and distribution) also need to be considered. Efficiency and power-distribution are inseparable, as are existence and objectives. This need not necessarily imply that firms are *not* Pareto-inefficient *ex post*. Given efficiency gains, from tighter control and reduced transaction costs, the opposite is as likely, at least from a pure pecuniary point of view. In a historical context, moreover, whether would-be laborers preferred the status of laborers to (the uncertainty of) being free peasants is a thorny question, leaving unanswered the question of whether the firm was Pareto-efficient or not *ex ante* as well. Important, however, in our context is that the driving force has been efficiency for profit (power-distribution) and not efficiency for efficiency's sake (economizing) as the Coase-Williamson neoclassical new institutionalism suggests. All that matters is that merchants, by virtue alone of their being merchants (knowledge of markets, willingness and ability to sell), were able to persuade (or coerce, directly or through the state) some single producers to work for them. In this sense, the capitalist was born as was the multi-person firm (first in the form of putting-out) (see Landes [1966] and Pitelis [1991] for support of the view that the putting-out system involved hierarchy).[12] Following this, organizational knowledge of production within the firm was obtained in the process. The move to the factory system and the growth of factories increased the relative significance of the capital ownership barrier by removing capital ownership from the putting-outers and thus consolidating the victory of the merchant-capitalist and the capitalist firm. The state facilitated the process (see, e.g., Galbraith [1987] and Hymer and Resnick [1979] concerning the farmers [land enclosures]).

The emergence of the employment relation does not remove the market *in toto*. It replaces the use of the price mechanism with a hierarchical device, for a particular type of transaction relating to the employment of labor. Outside this relation, the price mechanism remains and grows more important. Going further, it can be said that the firm emerges *for* (the production of commodities for sale in) the market. The replacement of putting-out by the factory system, which, initially at least, brings together laborers on one site under the direct control of the capitalist, who now owns the means of production too, further enhances efficiency for profit by solving two problems of imperfect control (and thus transaction costs) associated with putting-out: imperfect supervision of labor and imperfect control over organizational knowledge, which would allow some laborers to join the ranks of capitalists, given their ownership of capital. The factory system both solves these problems and fully frees labor from property. It also brings about the benefits of teamwork. From the point of view of the capitalist, the factory system represents a further efficiency gain. Whether it is Pareto-efficient is doubtful. Vis-à-vis the putting-out system, laborers lose control over their lives, for an uncertain increase in their wage, as their share in the efficiency gain.

In the above framework, the price mechanism and the firm complement each other in the principals' efforts to further their profits by achieving a more efficient organization and division of labor. The price mechanism in itself is an imperfect and insufficient means of achieving this aim because of the excessive transaction costs associated with its inherent reliance on voluntary exchanges. The private hierarchy—the firm—solves some of these problems.

It follows from this short account that, historically, it can be claimed that improvements in the exploitation (of the fruits of the division) of labor and, relatedly, the productive benefits related to teamwork and knowledge enhancement in the framework of the cohesive shell of an administrative organization could be a driving force behind the emergence and evolution of the (Coasean) firm. Further, that independent of the *ex post* Pareto-efficiency (or inefficiency) properties of such evolutionary changes, their initiators (and major beneficiaries) were the principals (to-be). This is broadly in line with, and provides a basis for, a startling consensus on the part of a widely divergent array of contributors to the definition of institutions, markets, and hierarchies. For example, the neoclassical economist North [1981] observes that "institutions are sets of rules, compliance procedures, and moral and ethical behavioural norms designed to constrain the behaviour of individuals *in the interests of maximizing the wealth or utility of principals*" [North 1981, 202; emphasis added]. The Marxist economist Hymer [1979], on the other hand, regards the market and the factory as "the two different methods of *co-ordinating the division of labor*. In the factory, entrepreneurs consciously plan and organize cooperation, and the relationships are hierarchical and authoritarian, in the market coordination is achieved through a decentralised, unconscious competitive process" [Hymer 1979, 57; emphasis added]. The Austrian economist and philosopher Friedrich von Hayek observes that "the commands as well as the rules which govern an organisation *serve particular results aimed at by those who are in command of the organisation*" [quoted in Clegg 1989, 33; emphasis added].

To summarize, my historically informed account suggests that the transition from the market to the firm (e.g., putting-out first, factory after) can be claimed to have been, at least in part, the result of the merchants' desire to further their interests by effecting a more efficient division of labor. The process seems to have involved consensus and coercion, potentially including state intervention. *Ex ante*, it was likely to have been undesirable. *Ex post*, often Pareto efficient!

The observation that the employment relation (the Coasean firm) is in part the result of (the merchants') attempts to effect a more efficient division of labor, and the related idea that the transition from putting-out to the factory system was simply a choice between two hierarchies, renders inseparable the notions of the existence and objectives of firms. In this scenario, firms exist because of their principals' objectives; objectives lead to existence, and existence is due to objectives. It is neither necessary nor helpful to attempt to explain existence first and discuss objectives or vice versa. Existence and objectives are the two sides of the same coin.

As suggested in my historically informed account, an aim of the merchants in using the market (and underlying their successful attempt to replace the latter by an employment relationship) was to further their profits.[13] Similarly, the transition from one type of employment relation (putting-out) to another (factory) was due to the merchant-manufacturers' desire to further their profits through a more efficient exploitation of the division of labor. From this it follows that both Williamson and Marglin are simultaneously right and wrong. The drive for profits by merchant-manufacturers would lead them to seek more efficient means of exploiting (the division of) labor. Through such efficiency improvements, profitability and power would also increase. Power and efficiency are linked inseparably. Concerning, in particular, transaction costs, it is exactly through their reduction that profit, thus potentially market power, can increase, as observed by Malcolmson [1984]. To further profits, principals would try to reduce all types of costs, including transaction and labor-production costs. They would achieve the latter by increasing their control over the labor process, as indeed both Marglin and Williamson emphasize. In this sense, involved here are not two different explanations of the existence of firms, rather, differences in emphasis concerning two sides of the more general objective/reason for existence of firms, the furtherance of profits for the principals.[14]

To the extent that transaction costs and resource virtues could further enhance profits, through cost saving and/or teamwork and knowledge enhancement benefits à la Penrose, further support for Coase-type hierarchy can be provided. Again, this suggests that separating efficiency from power and distribution is both hard and not useful.[15]

Summary and Concluding Remarks

To summarize, the conceptual contribution of TCE to the issue of the nature of the firm is based on shaky theoretical grounds and offers no evidence to substantiate its core proposition that pre-existing markets fail due to transaction costs, which leads through a contractual process to emerging efficient firms, which (thus) exist for, and result in, efficiency. TCE's main contributions are in the make or buy decision of firms and in the analysis of economic development. Its main limitations are that it lacks history, dynamics, and a historically informed evolutionary perspective.

In this paper, I have attempted to explain the emergence, evolution, and failures of the firm, the market, and the state in a framework that focuses on principals' attempts to further their interests by trying to remove constraints to the realization of these objectives through attempting to initiate suitable organizational and institutional changes. In this framework, objectives and existence are inseparable, part of a dialectical process, and markets in capitalism are seen to contain elements of hierarchy (endemic in the system), which cast doubt on the markets-hierarchies dichotomy. Firms are also seen to complement and create markets, while internalizing-replacing markets at the same time.

This approach goes well beyond TCE and also integrates these with power and competence virtues perspectives.[16]

Notes

1. Coase [1993] now recognizes this and refers to the need to examine *the essense* of firms, which is running a business. This is taken to include direction of others but also of the entrepreneur's own faculties and of non-human resources.
2. It is worth noting that starting with an individual in a state of nature is in line with the general neoclassical focus on the individual as the unit of analysis. An alternative is to assume that in the beginning there were parental units (families) and/or other socio-institutional arrangements. We assume here self-sufficiency in order to pursue a critique of TCE from within its own perspective but also because we consider self-sufficiency in its more general form (e.g., parental unit) as a plausible starting point.
3. Williamson [1993] accepts Masten's [1991] distinction between firms and markets in that disputes in the former case are solved inside the firms (an inside court of appeal), which is not true in the market. The problem with this distinction is that its adoption presupposes the existence of firms and markets, which is the very issue in question!
4. An example is building companies in Athens. They employ some individual workmen as long-term employees (a firm-type transaction) and some others for one day's work in the spot market in Omonia Square in Athens early every morning. In the first instance, we have a hierarchical relationship, in the second a market relationship. However, to claim that the latter situation is more symmetrical would be preposterous. Workers looking for a day's work incur more than just transaction costs to get to the meeting point, and they may well end up unemployed for the day.
5. It is worth observing that from the host of prominent authors dealing with this question, few share the Coasean perspective. For example, Demsetz [1988] views the *differentia*

specifica of firms to be that they involve specialization, continuity of association, and direction, which incidentally need not imply authority over others, as it could be limited to the direction of one's own faculties and/or non-human resources. See also Coase [1993]. Simon [1991] stresses the coordination and motivational aspect of organizations. Others still see the firm as a bundle of competencies such as tacit knowledge, skills, etc., which are part and parcel of organizations, not markets. This perspective comes from Knight [1921], Penrose [1959], Hymer [1979], Chandler [1962], Demsetz [1973], Nelson and Winter [1982], and originally from the founding fathers, Adam Smith and Karl Marx.

6. For Marglin, the emergence of the factory system from the putting-out system was not due to any efficiency attributes of the former, but rather to the enhanced control-power it conferred on capitalists over workers. This perspective has been recently supported by Simon who suggests that "co-ordination mechanisms in organisations taken in combination with the motivational mechanisms discussed earlier, create possibilities for enhancing productivity and efficiency through the division of labour and specialisation" [1991, 42].
7. Importantly, TCE proponents fail to examine other efficiency-related reasons why employees could voluntarily accept working for others, notably that firms are more efficient in production than markets. This is implied by Penrose's [1959] theory.
8. Continuing with the De Beers case, for example, Williamson observes that "this efficiency interpretation takes initial conditions—namely the De Beers monopoly—as given. Conceivably that is too passive, in that, as a matter of good public policy, the De Beers monopoly should be broken up. That, however, is another story . . ." [1996, 27].

   However, that is not another story. First, in the Coase-Williamson perspective, monopoly could (should) be the result of transaction costs-reducing internalization of markets, i.e., efficiency. Second, and relatedly, breaking monopolies is the very opposite of what the TCE perspective implies. Third, the state (which implements good public policy) may be endogenous to the system, which could lead to a combined state-monopoly. See Dunning [1992] for this case for the transnationals, as well, of course, as all relevant Marxist literature, such as Baran and Sweezy [1967] and Pitelis [1990] for a critical assessment of alternative perspectives.
9. Bowles and Edwards [1993] offer historical evidence in support of this observation. They found that in early nineteenth century America, working for a boss was seen as slavery. They also provide evidence on how much things have since changed.
10. For this to happen, merchants should be able to sell something for more than they paid, i.e., they should be able to buy a product below its value (the socially necessary labor time of average skill embodied in the product as the labor theory of value would have it), and/or because of the supply/demand and relative scarcity conditions of such products in the markets they were sold. I will not enter the value debate here, as it is inconsequential for this analysis. The *existence* of a pool of merchants, by virtue of being merchants, suffices to make the point that some surplus (however derived) was present. For different theories of value, see Dobb [1973]. A consoling (albeit not uncontroversial) view is Marglin [1984], who suggests that "the essense of capitalist profit does not lie in exchange—a point of view that is now common to virtually all economic theory" [1984, 462].
11. This is not to deny the potentially important role of technological change in driving production of cloth from small peasant homes into factories. Rather, my aim is to emphasize the often neglected issue of control.
12. Landes, for example, observes that although most domestic weavers owned their looms and nailers their forges, "they were not, however, independent entrepreneurs selling their products in the open market; rather they were hirelings, *generally tied to a particular employer*, to whom they agreed to furnish a given amount of work at a price stipulated in advance" [1966, 12; emphasis added].
13. This is not to deny the importance of other motives, such as control over others, power, prestige, etc. In a capitalist economy, however, these can plausibly be suggested to be

positively and strongly related to profits. For discussions, see Schumpeter [1987], Penrose [1959], Eichner [1976], and Pitelis [1987].

14. The similarity between the two models is also observed in Putterman [1986], who, however, attributes the apparent difference in their assumptions concerning the conditions of the labor market. "Once market conditions are made explicit . . . the radical theory . . . translates sensibly into the 'mainstream' terminology" [Putterman 1986, 27].
15. In the light of the above, the Alchian and Demsetz argument also appears correct, given its exclusive focus on exchange. From a legal point of view, the relationship between principals and agents is one of symmetry. Labor market conditions (affecting the agents' alternatives), as well as actions by principals and agents affecting the exit conditions of agents in particular, can put limits (constraints) on the relative power of the two parties but cannot change the fundamental equality in the face of the law between the two. Once, however, we focus on the production side (once it is recognized, that is, that the agents' only alternative to an employment relation may be another employment relation or none at all), the relationship becomes one of asymmetry (hierarchy). In the production process, the agent has to obey the principal (or else choose another principal to obey).
16. Interestingly, the solution of transaction costs-related failure by firms (can) represent(s) a firm competence, which moreover could vary across the firms' population and (thus) be firm-specific. In this context, transaction costs reductions could be seen as part of the specific competences arsenal of the firm(s).

References

Alchian, A., and H. Demsetz. "Production, Information Costs and Economic Organization." *American Economic Review* 62, no. 5 (December 1972): 777-95.

Baran, P., and P. Sweezy. *Monopoly Capital*. London: Penguin, 1967.

Bowles, S., and R. Edwards. *Understanding Capitalism*. New York: HarperCollins, 1993.

Braverman, H. *Labour and Monopoly Capital: The Organisation of Work in the Twentieth Century*. New York: Monthly Review Press, 1974.

Caves, R. *Multinational Enterprise and Economic Analysis*. Cambridge: Cambridge University Press, 1982.

Chandler, A. D. *Strategy and Structure*. Cambridge, Mass.: MIT Press, 1962.

Clegg, S. *Modern Organisations*. London: Sage, 1989.

Coase, R. H. "The Nature of the Firm." *Economica* 4 (1937): 386-405.

________. "The Nature of the Firm: Meaning," "The Nature of the Firm: Influence," and 1991 Nobel Lecture: "The Institutional Structure of Production." In *The Nature of the Firm: Origins, Evolution and Development*, edited by O. E. Williamson and S. G. Winter. Oxford: Oxford University Press, 1993.

Demsetz, H. "Industry Structure, Market Rivalry and Public Policy." *Journal of Law and Economics* 16 (1973): 1-9.

________. "The Theory of the Firm Revisited." In *Ownership, Control and the Firm: The Organization of Economic Activity*, vol. 1. Oxford: Basil Blackwell, 1988.

Dobb, M. *Theories of Value and Distribution since Adam Smith*. Cambridge: Cambridge University Press, 1973.

Dunning, J. H. "The Competitive Advantage of Countries and the Activities of Transnational Corporations." *Transnational Corporations* 1, no. 2 (1992): 135-68.

Eggertson, Y. *Economic Behaviour and Institutions*. Cambridge: Cambridge University Press, 1990.

Eichner, A. S. *The Megacorp and Oligopoly*. Cambridge: Cambridge University Press, 1976.

Foss, N. J. "Theories of the Firm: Contractual and Competence Perspectives." *Journal of Evolutionary Economics* 3, no. 2 (1993): 127-144.

_______, ed. *Resources, Firms and Strategies*. Oxford: Oxford University Press, 1997.

Foss, N. J., and C. Knudsen, eds. *Towards a Competence Theory of the Firm*. London: Routledge, 1996.

Fourie, F. C. V. N. "The Nature of Firms and Markets: Do Transaction Approaches Help?" *The South African Journal of Economics* 157, no. 2 (1993): 142-160.

Francis, A. "Company Objectives, Managerial Motivations and the Behaviour of Large Firms." *Cambridge Journal of Economics* 4 (1980).

Galbraith, J. K. *A History of Economics: The Past as the Present*. London: Penguin Books, 1987.

Groenewegen, J., ed. *Transaction Costs Economics and Beyond*. Dordrecht: Kluwer, 1996.

Hennart, J. F. "The Transaction Cost Theory of the Multinational Enterprise." In *The Nature of the Multinational Firm*, edited by C. N. Pitelis and R. Sugden. London: Routledge, 1991.

Hymer, S. *The International Operations of National Firms: A Study of Foreign Direct Investment*. Cambridge, Mass.: MIT Press, 1976.

_______. "The Multinational Corporation and the International Division of Labour." In *The Multinational Corporation: A Radical Approach, Papers by Stephen Herbert Hymer*, edited by R. B. Cohen, N. Felton, J. Van Liere, and M. Nkosi. 1971. Reprint. Cambridge: Cambridge University Press, 1979.

_______. *The Multinational Corporation: A Radical Approach, Papers by Stephen Herbert Hymer*, edited by R. B. Cohen, N. Felton, J. Van Liere, and M. Nkosi. 1971. Reprint. Cambridge: Cambridge University Press, 1979.

Hymer, S., and S. A. Resnick. "International Trade and Uneven Development." In *The Multinational Corporation: A Radical Approach, Papers by Stephen Herbert Hymer*, edited by R. Cohen, N. Felton, J. Van Liere, and M. Nkosi. 1971. Reprint. Cambridge: Cambridge University Press, 1979.

Kehoe, L. "Microsoft Deal Settles Antitrust Legislation." *Financial Times* (July 18, 1994).

Knight, F. *Risk, Uncertainty and Profit*. New York: Houghton Mifflin, 1921.

Landes, D. S., ed. *The Rise of Capitalism*. New York: Macmillan, 1966.

Malcolmson, G. "Efficient Labour Organisation: Incentives, Power and the Transaction Costs Approach." In *Firm Organisations and Labour*, edited by F. Stephen. London: Macmillan, 1984.

Marglin, S. "What Do Bosses Do? The Origins and Functions of Hierarchy in Capitalist Production." *Review of Radical Political Economics* 6 (Winter 1974): 60-112.

_______. *Growth, Distribution and Prices*. Cambridge, Mass.: Harvard Economic Studies, 1984.

Masten, S. "A Legal Basis for the Firm." *Journal of Law, Economics and Organisation* 4 (1991): 181-198.

_______. "Empirical Research in Transaction Cost Economics." In *Transaction Costs Economics and Beyond*, edited by J. Groenewegen. Dordrecht: Kluwer, 1996.

McNulty, P. J. "On the Nature and Theory of Economic Organization: The Role of the Firm Reconsidered." *History of Political Economy* 16, no. 2 (1984).

Nelson, R. R., and S. G. Winter. *An Evolutionary Theory of Economic Change*. Cambridge, Mass.: Harvard University Press, 1982.

North, D. C. *Structure and Change in Economic History*. London and New York: Norton, 1981.

_______. "Institutions." *Journal of Economic Perspectives* 5, no. 1 (1991): 97-112.

Nyman, S., and A. Silbertson. "The Ownership and Control of Industry." *Oxford Economic Papers* 30 (1978).

Penrose, E. T. *The Theory of the Growth of the Firm*. Oxford: Basil Blackwell, 1959.

Pitelis, C. N. *Corporate Capital: Control Ownership, Saving and Crisis*. Cambridge: Cambridge University Press, 1987.

_______. "Neoclassical Models of Industrial Organization." In *Perspectives in Industrial Organization*, edited by B. Dankbaar, J. Groenewegen, and H. Schenk. Kingston-upon-Thames, UK: Kluwer, 1990.

_______. *Market and Non-Market Hierarchies*. Oxford: Basil Blackwell, 1991.

_______, ed. *Transaction Costs, Markets and Hierarchies*. Oxford: Basil Blackwell, 1993.

Pitelis, C. N., and M. Wahl. "Edith Penrose: Pioneer of Stakeholder Theory." *Long Range Planning* 31, no. 2 (1998).

Porter, M. *Competitive Strategy*. New York: Free Press, 1980.

Putterman, L. *The Economic Nature of the Firm: A Reader*. Cambridge: Cambridge University Press, 1986.

Scherer, F. M. *Industrial Market Structure and Economic Performance*. 2d ed. Chicago: Rand McNally, 1980.

Schumpeter, J. *Capitalism, Socialism and Democracy*. 5th ed. 1942. Reprint. London: Unwin Hyman, 1987.

Simon, H. A. "Organizations and Markets." *Journal of Economic Perspectives* 5, no. 2 (1991): 25-44.

Stone, K. "The Origins of Job Structures in the Steel Industry." *Review of Radical Political Economics* 6 (1974): 61-97.

Williamson, O. E. *Markets and Hierarchies*. New York: Free Press, 1975.

________. "The Modern Corporation: Origins, Evolution, Attributes." *Journal of Economic Literature* 19, no. 4 (1981): 1537-68.

________. *The Economic Institutions of Capitalism*. New York: Free Press, 1985.

________. "Transaction Costs Economics and Organization Theory." *Industrial and Corporate Change* 2, no. 2 (1993): 107-156.

________. "Efficiency, Power, Authority and Economic Organization." In *Transaction Costs Economics and Beyond*, edited by J. Groenewegen. Dordrecht: Kluwer, 1996.

________. "The Institutions of Governance." *American Economic Review*, Papers and Proceedings, 88, no. 2 (May 1998).

Williamson, O. E., and S. G. Winter, eds. *The Nature of the Firm*. Oxford: Oxford University Press, 1993.