EXPANDING THE BOUNDARIES OF THE FIRM
LESSONS FROM DIGITALISATION

Jörg J. Dötsch

1Associate professor, Chair of Economic Policy, Andrássy University, Hungary
https://doi.org/10.47833/2024.1.ECO.008

Keywords:
theory of the firm
digitalisation
transaction costs

Abstract
The theory of the firm provides different approaches to explain the existence and boundaries of organisations. The approach of Coase starts from the underlying principle of coordination and distinguishes between market and organisation with regard to transaction costs. However, the optimisation of organisations can reach the limits of feasibility, which Williamson treated under the term "economics of atmosphere". This contribution first aims to clarify the relation of both theoretical approaches, then to narrow their gaps by placing them into a complexity theory perspective, and finally applies this new point of view exploratively with attention to the conditions of digitalisation.

1 Introduction

Why do firms exist? Ronald Coase provided a concise and prominent answer: “the distinguishing mark of the firm is the supersession of the price mechanism" [9]. His transaction cost theory provides the crucial argument: The coordination of activity for a given economic goal can be solved cheaper by an organisation than by the market. This is a calculative solution for the aim of making a profit. Accordingly, this organisation is built and led by calculative principles. However, the question arises to which extent coordination can be maintained on the basis of purely calculative principles. Oliver Williamson has drawn attention to the fact that social coordination guided solely by "calculativeness" becomes dysfunctional beyond a certain point, but without defining a critical threshold. He called this fuzzy area “economics of atmosphere". It is still a debated question what this term means exactly [see, e.g. 8].†

The two perspectives focus the same problem from different angles. While Coase’s approach allows for a precise distinction, Williamson’s approach is quite open. The first provides a mechanism for the question either/or, the latter questions the scope of it. Both approaches can be understood as treating problems of the boundaries of the firm. While Coase focuses on the ‘outer boundary’ – market or organisation –, Williamson rather focusses the limits of its operation, its ‘inner boundary’.‡

Calculativeness is an essential condition for the firm to exist, but not sufficient for its maintenance.

† Oliver Williamson received his “Nobel Prize” – correctly: The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel – together with Elenor Ostrom in 2009 “for his analysis of economic governance, especially the boundaries of the firm” [28]; Ronald H. Coase was awarded with the same prize in 1991 “for his discovery and clarification of the significance of transaction costs and property rights for the institutional structure and functioning of the economy” [27]. I owe the reference to Williamson’s peculiar category and its possible connections to complexity theory approaches to a stimulating discussion with Stefan Okruch.

‡ These are spatial metaphors that may entail their own problems, but are sufficient as heuristic expressions for the following considerations.
This depends on what motivates people to cooperate with each other, to win trust [see 19], to be innovative, mindful, and so on [see 29]. Of course, all this is for achieving the best possible calculative goal.

The problem is that the inner boundary may not be clearly defined by economic instruments only. Furthermore, both the outer boundary and the inner have come under new conditions in the age of digitalisation. This is the subject of this contribution. It pursues two aims. Firstly, it provides a new perspective on a problem of the theory of the firm by using principles of emergence theory. Secondly, it sheds light on the fact that due to digitalisation the role of people and firms is currently changing decisively.

The remainder of this paper is structured as follows. The second section briefly outlines the theoretical access and highlights the concept of “economics of atmosphere”. The third section clarifies this concept by using the approach of emergence theory. The fourth section takes this as a starting point to illustrate the extent to which the internal and external boundaries of the firm have come under new conditions in the age of digitalisation and are tending to shift. A brief conclusion brings the perspectives together and points out possible challenges for society and the regulator.

2 Theoretical principles

2.1 Concepts of boundaries

According to Coase, scientific disciplines can be distinguished by “(...) one or more of the following: common techniques of analysis, a common theory or approach to a subject, or a common subject matter” [9]. He defines the subject area of economics as “(...) the working of the social institutions which bind together the economic system: firms, markets for goods and services, labour markets, capital markets, the banking system, international trade, and so on” [ibid.].

Williamson attached greater weight to the “approach” rather than the subject matter: “Calculativeness is the general condition that I associate with the economic approach and with the progressive extension of economics into the related social sciences” [32]. His contribution on Calculativeness, Trust, and Economic Organisation (1993) demonstrated the extent to which “calculativeness” is worthwhile regarding the organisation of transactions. This obviously firstly sheds light on the question of the existence of firms.

Accordingly, Coase uses this concept of “calculativeness” to determine the outer boundaries of firms: “the distinguishing mark of the firm is the supersession of the price mechanism” [9]. Whether organisation is worthwhile instead of the market thus can be understood as a transaction cost problem.

Using the conception of calculativeness as distinctive principle is illuminating with regard to the emergence and differentiation of firms. A theoretical problem arises when this perspective is applied to the maintenance of the organisation, because for this aim economic stringency alone is not sufficient. For transactions to be carried out optimally and to make an organisation sustainable and resilient, something beyond mere calculativeness is needed. In case, quite the opposite, “(...) piecemeal calculativeness can easily be dysfunctional” [32]. The question arises as to how to determine the threshold where this dysfunctionality begins. Or, asked differently, how and under what circumstances does this threshold come about? For this obviously ambiguous area Williamson had coined a term which appears in Markets and Hierarchies (1975) and The Economic Institutions of Capitalism (1985): “economics of atmosphere”.

2.2 The role of atmosphere

From the strictly economic point of view this area is diffuse. Asked about it by Geoffrey Hodgson and David Gindis, three years before his Nobel Prize, Williamson stated that “(o)n the one hand it seemed an important concept and on the other it was hard to give a specific content” [34]. That is where it has remained. Chassagnon (2021) points out that “Williamson’s organisational atmosphere has barely been exploited in the economic literature” (p. 3) and emphasises the interdisciplinary potential of this concept, which “(...) gives the opportunity to question quantitative and calculative neoclassical economics” (ibid, 4). Indeed, “atmosphere is a concept that non-economists
are apt to find less objectionable than economists (…) much of what economists do is rationalise how human behaviour is to be understood in familiar net gain terms and debunk such relatively loose concepts as atmosphere” [31].

Do the “economics of atmosphere” fall into the lap of sociology [see e.g. 26]? They only partially do. For institutional arrangements evoke specific environments. Transactions, too, are not only “embedded” [12], but they evoke possibly new and specific environments. According to Williamson, organisations differ in “nontrivial atmospheric respects” [31] and “(...) atmosphere refers to interactions between transactions that are technologically separable but are joined attitudinally and have systems consequences” [33].

More theoretical clarity about the quite elusive concept of “economics of atmosphere” can be gained by revealing the systematicity of this relationship. From Williamson’s perspective, the marginal utility of calculativeness becomes negative at some point, because a non-calculative “atmosphere” is the prerequisite for carrying out calculative actions. The other way round: an organisation gets unfeasible as soon as the marginal benefit of the atmosphere is higher than the marginal costs of calculativeness. For organisational theorists the challenge is that the organisation must generate such an environment and/or confirm such an environment, which itself stabilises the calculative transactions. Williamson himself probably saw it this way, because “…it may be more accurate, and sometimes even essential, to regard the exchange process itself as an object of value…” [31]. To see something “itself” as an object of value obviously means here a non-calculative value [see e.g. 29]. In another perspective this can be seen as an externality which can’t be internalized. The term “economics of atmosphere” thus captures a special class of external effects that consist of self-referentially ascribing an intrinsic value to the transaction outside its explicit objective.

However, transactions have external, non-calculative effects that can have systemic consequences. These consequences are systematic in the sense that they create and/or reproduce an environment – an “atmosphere” – that is of its own quality. That is, it is a phenomenon of emergence. By means of system theory, we will illuminate the connection that is diffuse in Williamson’s work because he cannot appropriately capture this dimension. Emergence theory considerations provide a conclusive and elegant complement to Williamson’s position, whose “(...) concept of atmosphere, (...) was originally intended to convey the idea that the firm as a whole is more than the sum of the parts” [34] – i.e., a phenomenon of emergence.

3 Emergent effects

3.1 Basic distinctions

The well-known metaphor of Granovetter, who also explicitly referred to problems of Williamson’s approach in his seminal 1985 paper, sheds light on Williamson’s point of view. “Economic action” [12] or “economic behaviour” [ibid] are “embedded” in a social reality. Accordingly, Granovetter distinguishes economic action from non-economic action according to a majority of “(…) sociologists, anthropologists, political scientists, and historians” [ibid] at the time, very similarly to Williamson. He “(…) sees the economy as an increasingly separate, differentiated sphere (…) with economic transactions defined (…) by rational calculations of individual gain” (ibid). Let us focus on the difference between these spheres.

Granovetter’s metaphor of embeddedness is a residual category for the prerequisites of transactions which is similar to Williamson’s proposition. In the following, we highlight four core arguments on which Williamson’s metaphor is based:

* Hannah Arendt’s critique of capitalism offers a connecting point, as she sees the mechanism of capitalism in the appropriation and accumulation of what is in principle exchangeable. Everything is exchangeable because it only acquires its value through a (monetary) transaction. In the end, the owner of these goods also becomes exchangeable too, since he obtains his value through their accumulation – or simply through the accumulation of money. See [2] (esp. chapter III and IV); on this problem see e.g. [3].
a. The terms “transaction” and “atmosphere” can be interpreted as the equivalents for an inside-outside dichotomy.

b. Transactions represent a form of social action through which “atmosphere” is simultaneously. In other words, the dichotomy in question is created through action.

c. The created atmosphere has a quality of its own. Otherwise, it would be quite pointless to provide an own designation. We claim that this is an emergence effect.

d. The emergent atmosphere has, reversely, an effect on the transaction. In other words, there is a feedback effect.

On this basis, two gaps in Williamson’s argument become clear. Firstly that the importance of the feedback effect is noted by Williamson. However, it is not theorised further. Secondly that the existence of an emergence effect is an implicit prerequisite of the argument. However, it is not described what this actually consists of. Admittedly, this needs some theoretical sharpening.

3.2 Emergence in organisations

Phenomena of emergence show properties that cannot be explained only by properties of their elements or by the interaction of their elements. A common term for this is irreducibility [see 21]. Early on, there has been quite a wide range of concepts (ibid.) and descriptions of the preconditions for this phenomenon [see also 23].

The differences in these concepts, e.g. with regard to the probability of the occurrence of emergence, result partly from the great diversity of the subject areas in which this phenomenon is observable, such as e.g. biology or the social sphere. In any case, however, an interaction of elements in a kind of network is required. It should be noted that “network” is also a central concept for Granovetter’s view [11].

Even if we do not get involved in the jungle of studies on questions of culture (which, after all, provides the embeddedness) in management literature [e.g. 25; 14] we are faced with a fundamental problem here. In any case, the working atmosphere that arises when people cooperate is neither sufficiently predictable nor reducible to the individuals. They have too many characteristics anyway. Nevertheless, this “atmosphere” is there and it is intersubjectively relevant.

Within organisations there can be differences concerning the significance of emergent effects of social action. Forms of emergence are less likely to occur in firmly ritualised forms of encounters. Strict forms of ritualization leave nothing to chance and are there precisely to exclude emergence effects. Emergence effects become particularly likely when groups will have to solve new problems. It is evident that even if we know everything about the actors, we cannot predict the outcome: neither the “atmosphere”, nor the way leading to a ‘solution’, nor the ‘solution’ found. There are simply too many possibilities for that among the cooperating (or covertly defecting) actions. This non-calculability is even necessary for the solution of some problems and is desirable from the organisation’s point of view. It is a calculated non-calculability, a basic problem of innovation research and its variations e.g. with regard to the role of diversity of teams [e.g. 14; 21; 19].

This means for our focus: one cannot explain or predict emergence on the basis of a theory that deals with individuals and their interaction. This is what the calculative, the economic approach tries to do.

4 The organisation under new conditions

4.1 The impact of digitalisation

Up to this point, the boundaries of the organisation have been dealt with theoretically, without considering specific technological options in the narrower sense. This did not even seem necessary as technological development had not caused any categorical change in the internal or external boundaries of the firm until recently. However, this seems to be changing under the conditions of digitalisation. Digitalisation is a very broad umbrella term, under which all new phenomena made possible by digital technology can in principle be summed up. This article takes the view that this is a new general purpose technology [5], the centrepiece of which is the ability to process large
amounts of data – big data – automatically. The diverse phenomena that digital technology enables are based on this phenomenon. Their overall effect and appearance create a process of change that encompasses all areas of society and can be described as digitalisation. The following two sections briefly outline what the possibility of digitally processing data means for the object area of the boundary of the firm focussed on here.

4.2 The shift of inner boundaries

The considerations in the previous sections first of all shed light on the question of the criteria where the inner boundary of organisations is. Admittedly, an emergence-theoretical approach makes it clear that the “atmosphere” obeys an own, non-calculative operating principle and comes about in leaps. However, digitalisation seems to offer opportunities to extend the boundary of the calculative principle further. Calculative action is much further applicable through developed possibilities of (unnoticed) monitoring. Fernández-Macías [11] refers to sensory control in production processes stating that “the Internet of Things (…) promises a big leap in the efficiency of industrial processes, but it can also transform a factory into an invasive surveillance system.” First of all, this does not mean that the “atmosphere” would be affected by these possibilities, which can also be realised unnoticed. But it can be, under certain circumstances, a shift of the organisation into the private sphere. It can subvert the atmosphere. The organisational challenge is known as “digital trust”) [1].

These possibilities potentially go beyond the boundaries of the workplace whenever an employee uses a network-enabled device. “Remote sensors, cheaper and smaller cameras, and geo-localisation, make it easier than ever for firms to monitor work, whether inside or outside the boundaries of the firm” [17, with reference to 25]. Zuboff (2019) has prominently discussed the fundamental problem of the omnipresent potential for surveillance through digital technology and offers an instructive overview of a new problem area [see also the critique by 16]. But the actual extension of “calculativeness” is well illustrated by the fact that it is driving the differentiation process of the division of labour and increases the volume of outsourcing [20].

According to the increasing scope of calculativeness, the inner boundary of the firm shifts. However, the original problem has of course not been abolished. For whether an employee who has gained new skills through digital technology must also be prepared to use them meaningfully and creatively for his company. This will vary considerably depending on the task, his/her involvement in teams and thus also with regard to the degree to which the “atmosphere” is important. Under the conditions of digitalisation, it has become much easier for companies to differentiate on a small scale. The global availability and accessibility of workers, for example via platforms, has provided new opportunities for this. Employers have greater incentives to keep their commitments as low as possible, depending on the skills required.

This tendency can properly be explained in terms of the approach of Coase. The existence of the firm as an arrangement of contracts with workers occurs if and only if this arrangement is cheaper than procuring labour services not within the organisation but through the market. Blix (2017, 13) has correctly observed that “(...) (w)ith gig platforms, the cost of hiring temporary staff on a needs-only basis is much smaller than in the past”. New forms of work such as “crowdworking” or “clickworking” [10] lead to the fact that “(...) much work seems to be carried out in small units and irregularly” [21].

So while the possibilities of calculative organisation grow, precisely because they seem to rely less on an emergent phenomenon like “atmosphere”, the possibilities of workers to defend their interests become less, because their representation can no longer be organised cheaply enough.

4.3 The shift of outer boundaries

If “(...) the distinguishing mark of the firm is the supersession of the price mechanism” [9], then it is worth taking a closer look at the function of prices. Prices have the task of encoding scarcity and, like a language, enabling coordinative relationships between all actors in markets [6; 15]. But
can the use of digital technology relying on big data not encode scarcity and its conditions even better than the price system?

Friedrich August von Hayek put that “(...) (it) is useful to recall that wherever we make use of competition, this can only be justified by our not knowing the essential circumstances that determine the behavior of the competitors” [13]. From this the pre-digital perspective, the price system is relevant mainly because of its superiority in the discovery, generation and distribution of knowledge. However, the possibilities of digital technology to generate and process large amounts of data can absorb parts of the price system. Digital technology precisely codes “(...) the essential circumstances that determine the behavior of the competitors” – and (potential) customers. The outer boundary of the firm is expanding.

This expansion is empirically well illustrated by the fact that market shares currently obviously shift in favour of those who have the best command of digital possibilities. The price system is becoming less important in the digital age for the discovery, processing and dissemination of knowledge. So the “discovery procedure” of competition [13] is certainly working; no one could e.g. claim that modern technology companies are not innovative. However, the role of price signals pushed back by the use of big data.

This leads to a somewhat paradoxical situation. On the one hand, the higher degree of transparency provided by the internet seems to lead to a reduction in transaction costs and thus enables welfare gains. On the other hand, the reduced role of the price system creates a new form of information asymmetry. This pays off in monetary units for some organisations. At the same time, the capacities of such organisations go even beyond this by e.g. influencing consumer behaviour through techniques such as endorphin triggering and the like.

5 Consequences

New questions arise for both economic theory and regulation. On the one hand, a more differentiated theory is needed for the question of how to describe the inner boundary of the firm more adequately. While this paper briefly showed that the boundary, which for Williamson began where the marginal benefit of the atmosphere is higher than the marginal costs of calculativeness, is shifting further in favour of calculativeness, it would be worthwhile to consider how far this shift can actually extend.

However, one must distinguish two different aspects. One is the ethical and at the same time legal aspect, namely how extensively employees can actually be monitored and under what conditions. Digitalisation creates a new degree of complexity, where the employer’s legitimate interest in optimisation must be weighed against the employee’s legitimate interest in protecting his or her privacy. On the one hand, this poses a challenge for the regulator, who must determine which rules must be complied with in which sector and under which conditions. This is relevant for the ‘outer boundary’ of the company and is no easy task given the dynamic nature of the changes that digitalisation in particular has made possible. On side of the firms, this not only poses the legal challenge of compliance, but also – and this now concerns the ‘inner boundary’ – the leadership challenge of how digital trust can be cultivated among employees. Ultimately, this is a question of atmosphere!

Furthermore, competition policy issues become relevant. A “discovery procedure” based on the capacity to deal with big data instead of relying on the price system can lead to monopolisation tendencies. Its outcome may be more efficient on the one hand, but can have a negative impact on (economic) freedom. Provocatively formulated one could state that the relatively higher transaction costs of using the price system are the opportunity costs of protecting individual freedom. Without being able to go into detail, the boundary shift described here explains why traditional competition law thresholds are no longer sufficient in the digital age, but instead face a new problem of complexity.

For both parties, the regulator and the companies, this new level of complexity makes it much more difficult, if not impossible, to establish general rules about when which kind of use of data is actually problematic. Regulation and attentive leadership will only be possible on a smaller scale and
will have to consider more aspects at the same time, such as data protection, privacy, property rights and incentives for innovation.

References

[1] Abraham, C.: How digital trust drives culture change: privacy is a right, not a privilege: but organizations and leadership often struggle with it comes to adapting their culture toward digital trust and stewardship, MIT Sloan Management Review.


