

Six big ideas

Coase's theory of the firm

If markets are so good at directing resources, why do companies exist? The first in our series on big economic ideas

Jul 27th 2017

ONE morning, an economist went to buy a shirt. The one he chose was a marvel of global production. It was made in Malaysia using German machines. The cloth was woven from Indian cotton grown from seeds developed in America. The collar lining came from Brazil; the artificial fibre from Portugal. Millions of shirts of every size and colour are sold every day, writes Paul Seabright, the shirt-buying economist, in his 2004 book, "The Company of Strangers". No authority is in charge. The firms that make up the many links in the chain that supplied his shirt had merely obeyed market prices.

Throwing light on the magic of market co-ordination was a mainstay of the "classical" economics of the late-18th and 19th centuries. Then, in 1937, a paper published by Ronald Coase, a British economist, pointed out a glaring omission. The standard model of economics did not fit with what goes on within companies. When an employee switches from one division to another, for instance, he does not do so in response to higher wages, but because he is ordered to. The question posed by Coase was profound, if awkward for economics: why are some activities directed by market forces and others by firms?

His answer was that firms are a response to the high cost of using markets. It is often cheaper to direct tasks by fiat than to negotiate and enforce separate contracts for every transaction. Such "exchange costs" are low in markets for standardised goods, wrote Coase. A well-defined task can easily be put out to the market, where a contractor is paid a fixed sum for doing it. The firm comes into its own when simple contracts of this kind will not suffice. Instead, an employee agrees to follow varied and changing instructions, up to agreed limits, for a fixed salary.

Coase had first set out his theory while working as a lecturer in Dundee, in 1932, having spent the prior academic year in America, visiting factories and businesses. "The nature of the firm", his paper, did not appear for another five years, in part because he was reluctant to rush into print. Though widely cited today, it went largely unread at first. But a second paper, "The problem of social cost", published in 1960, by which time he had moved to America, brought him to prominence. It argued that private bargaining could resolve social problems, such as pollution, as long as property rights are well defined and transaction costs are low (they rarely are). He had been asked to expound his new theory earlier that year to a sceptical audience of University of Chicago economists. By the end of the evening, he had won everyone around. Coase was invited to join the university's faculty in 1964; and there he remained until his death in 2013 at the age of 102.

In 1991 Coase was awarded the Nobel prize for economics, largely on the strength of these two papers. But as late as 1972, he lamented that “The nature of the firm” had been “much cited and little used”. In a strange way, Coase himself was partly to blame. The idea of transaction costs was such a good catch-all explanation for tricky subjects that it was used to close down further inquiry. In fact, Coase’s paper raised as many difficult questions as it answered. If firms exist to reduce transaction costs, why have market transactions at all? Why not further extend the firm’s boundaries? In short, what decides how the economy as a whole is organised?

Almost as soon as Coase had wished for it, a body of more rigorous research on such questions began to flourish. Central to it was the idea that it is difficult to specify all that is required of a business relationship, so some contracts are necessarily “incomplete”. Important figures in this field include Oliver Williamson, winner of the Nobel prize in economics in 2009, and Oliver Hart and Bengt Holmstrom, who shared the prize in 2016. These and other Coase apostles drew on the work of legal theorists in distinguishing between spot transactions and business relations that require longer-term or flexible contracts.

Spot markets cover most transactions. Once money is exchanged for goods, the deal is completed. The transaction is simple: one party wants, another supplies. There is little scope for dispute, so a written contract can be dispensed with. If one party is unhappy, he will take his business elsewhere next time. Spot markets are thus largely self-policing. They are well suited to simple, low-value transactions, such as buying a newspaper or taking a taxi.

Things become trickier when the parties are locked into a deal that is costly to get out of. Take a property lease, for instance. A business that is evicted from its premises might not quickly find a building with similar features. Equally, if a tenant suddenly quit, the landlord might not find a replacement straight away. Each could threaten the other in a bid for a better rent. The answer is a long-term contract that specifies the rent, the tenure and use of the property. Both parties benefit.

But for many business arrangements, it is difficult to set down all that is required of each party in all circumstances. In such cases, formal contracts are by necessity “incomplete” and sustained largely by trust. An employment contract is of this type. It has a few formal terms: job title, work hours, initial pay and so on, but many of the most important duties and obligations are not written down. It is thus like a “mini-society with a vast array of norms beyond those centred on the exchange and its immediate processes”, wrote Mr Williamson. Such a contract stays in force mostly because its breakdown would hurt both parties. And because market forces are softened in such a contract, it calls for an alternative form of governance: the firm.

One of the first papers to elucidate these ideas was published in 1972 by Armen Alchian and Harold Demsetz. They defined the firm as the central contractor in a team-production process. When output is the result of a team effort, it is hard to put the necessary tasks out to the market. That is because it is tricky to measure the contribution of each member to the finished work and to then allocate their rewards accordingly. So the firm is needed to act as both co-ordinator and monitor of a team.

Chain tale

If a team of workers requires a firm as monitor, might that also be true for teams of suppliers? In some cases, firms are indeed vertically integrated, meaning that suppliers of inputs and producers of final goods are under the same ownership. But in other cases, suppliers and their customers are separate entities. When is one set-up right and not the other?

A paper published in 1986 by Sanford Grossman and Mr Hart sharpened the thinking on this. They distinguished between two types of rights over a firm's assets (its plant, machinery, brands, client lists and so on): specific rights, which can be contracted out, and residual rights, which come with ownership. Where it becomes costly for a company to specify all that it wants from a supplier, it might make sense to acquire it in order to claim the residual rights (and the profits) from ownership. But, as Messrs Grossman and Hart noted, something is also lost through the merger. The supplier's incentive to innovate and to control costs vanishes, because he no longer owns the residual rights.

To illustrate this kind of relationship, they used the example of an insurance firm that pays a commission to an agent for selling policies. To encourage the agent to find high-quality clients, which are more likely to renew a policy, the firm defers some portion of the agent's pay and ties it to the rate of policy renewals. The agent is thus induced to work hard to find good clients. But there is a drawback. The insurance firm now has an incentive of its own to shirk. While the agent is busting a gut to find the right sort of customers, the firm can take advantage by, say, cutting its spending on advertising its policies, raising their price or lowering their quality.

There is no set-up in which the incentives of firm and agent can be perfectly aligned. But Messrs Grossman and Hart identified a next-best solution: the party that brings the most to any venture in terms of "non-contractible" effort should own the key assets, which in this case is the client list. So the agent ought to own the list wherever policy renewals are sensitive to sales effort, as in the case of car insurance, for which people tend to shop around more. The agent would keep the residual rights and be rewarded for the effort to find the right sort of client. If the insurance firm shirks, the agent can simply sell the policies of a rival firm to his clients. But in cases where the firm brings more to the party than the sales agent—for example, when clients are "stickier" and the first sale is crucial, as with life insurance—a merger would make more sense.

This framework helps to address one of the questions raised by Coase's original paper: when should a firm "make" and when should it "buy"? It can be applied to vertical business ties of all kinds. For instance, franchises have to abide by a few rules that can be set down in a contract, but get to keep the residual profits in exchange for a royalty fee paid to the parent firm. That is because the important efforts that the parent requires of a franchisee are not easy to put in a contract or to enforce.

The management of ties between a firm and its "stakeholders" (its customers, suppliers, employees and investors) is another variation on this theme. A firm often wants to put restraints on the parties it does business with. Luxury-goods firms or makers of fancy sound equipment may ban retailers from discounting their goods as a

way to spur them to compete with rivals on the quality of their shops, service and advice.

Inside the cubicle

If one of the challenges set by Coase was to explain where the boundary between firms and markets lies, another was for economic analysis not to cease once it reached the factory gate or office lobby. A key issue is how agreements are structured. Why, for instance, do employment contracts have so few formal obligations? One insight from the literature is that a tightly specified contract can have perverse outcomes. If teachers are paid according to test results, they will “teach to the test” and pay less regard to other tasks, such as inspiring pupils to think independently. If chief executives are paid to boost the firm’s short-term share price, they will cut investment projects that may benefit shareholders in the long run.

Mr Holmstrom and Paul Milgrom established that where important tasks are hard to monitor, and where a balance of activities is needed, then a contract should shun strong incentives tied to any one task. The best approach is to pay a fixed salary and to leave the balance of tasks unspecified. A related idea developed by Mr Hart and John Moore is of a job contract as a “reference point” rather than as a detailed map. Another insight is that deferred forms of pay, such as company pension schemes and promotions based on seniority, help cement long-term ties with employees and reward them for investing in skills specific to the relationship.

Coase noted in 1937 that the degree to which the mechanism of price is superseded by the firm varies with the circumstances. Eighty years on, the boundary between the two might appear to be dissolving altogether. The share of self-employed contractors in the labour force has risen. The “gig economy” exemplified by Uber drivers is mushrooming.

Yet firms are unlikely to wither away. Prior to Uber, most taxi drivers were already self-employed. Spot-like job contracts are becoming more common, but flexibility comes at a cost. Workers have little incentive to invest in firm-specific skills, so productivity suffers. And even if Mr Seabright’s shirt was delivered by a set of market-based transactions, the supply chains for complex goods, such as an iPhone or an Airbus A380 superjumbo, rely on long-term contracts that are often “incomplete”. Coase was the first to spot an enduring truth. Successful economies need both the benign dictatorship of the firm and the invisible hand of the market.

Coase’s theory of the firm: a reading list

1 “The Nature of the Firm” by R H Coase, *Economica*, 1937

2 “The Problem of Social Cost” by R H Coase, *Journal of Law and Economics*, 1960

3 “Industrial Organisation: A Proposal for Research” by R H Coase, NBER, 1972

4 “Production, Information Costs and Economic Organisation” by Armen A Alchian and Harold Demsetz, *American Economic Review*, 1972

5 “Transaction-Cost Economics: The Governance of Contractual Relations” by Oliver E Williamson, Journal of Law and Economics, 1979

6 “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration” by Sanford Grossman and Oliver Hart, Journal of Political Economy, 1986

7 “Multitask Principal-Agent Analysis: Incentive Contracts, Asset Ownership and Job Design” by Bengt Holmstrom and Paul Milgrom, Journal of Law, Economics and Organisation, 1991

8 “The Firm as Sub-economy” by Bengt Holmstrom, Journal of Law Economics & Organisation, 1999

9 “The Theory of the the Firm as Governance Structure: From Choice to Contract” by Oliver E Williamson, 2002

10 “Contracts as Reference Points” by Oliver Hart and John Moore, Quarterly Journal of Economics, 2008

This article appeared in the Economics brief section of the print edition under the headline "Coase call"

Copyright © The Economist Newspaper Limited 2017. All rights reserved.