

Chapter 2

Why Do Firms Exist?

[T]he operation of a market costs something and by forming an organization and allowing some authority (an “entrepreneur”) to direct the resources, certain marketing costs are saved.

—Coase (1937), p. 392

Consider the operation of an ice cream shop. The owners decide what inputs to use and how to organize them. This set of decisions will have significant implications for what the shop does and how it operates as a firm. The shop requires ice cream, workers, and other inputs (such as cones, cups, and electricity) that vary depending on how much ice cream the shop sells, its operating hours, and so on. It also requires freezers, a place of business, and other pieces of capital to work complementarily with the other inputs to produce ice cream. An economic analysis of the ice cream shop as a firm could examine production costs, and it could investigate the market for ice cream to determine the most profitable prices to charge. It could also explore the firm’s operational details to see how it goes about organizing its productive activities. Coase emphasized the latter.

As a 21-year old student at the London School of Economics, Coase won a scholarship to spend a year in the United States learning about and analyzing how firms organize production. This question remains one of the most profound and fundamental in economics—why do firms exist? Coase observed the contrast between markets, where individual actions and decisions are coordinated by the decentralized price system, and firms, where actions and decisions are coordinated by internal hierarchy and central planning. If spot markets using the price system to coordinate production can

maximize economic welfare, why is not all production done through spot market transactions? These two options present alternative institutional structures for the organization of production.

In the neoclassical economic theory Coase was learning in the 1930s, called the cost-based theory of the firm, economists modeled firms based on their input costs, with inputs falling into two categories: labour and capital. Having decided what to produce, firms choose combinations of inputs that maximize their profits. The neoclassical theory of the firm focuses on how firms allocate resources to their highest-valued uses and how they make profit-maximizing investment decisions over time. This theory does not explore what determines the use of hierarchies over markets, or which transactions firms choose to perform internally. It also says little about how firms as organizational structures enable innovation, or how entrepreneurship is expressed in the forms or the actions that firms take.

Coase took the opportunity to explore how firms organize production, which did not negate the cost-based theory of the firm, but rather built upon it to examine how to determine which functions should be performed within firms and which should be performed through contracts with independent suppliers, as well as how internal organizational decisions within firms are made. The article based on this research, “The Nature of the Firm” (1937), was one of the two most influential works cited by the Nobel Committee in awarding the Nobel Prize to Coase in 1991.

In response to the question “why do firms exist?” Coase answered that they exist in order to address—specifically, to keep to a minimum—transaction costs. Coase’s answer unleashed a stream of influential research that is still generating new ideas today (although he did not use that phrase in his 1937 article, calling them “marketing costs” instead). Coase defined transaction costs as “the cost of using the price system” (1937: 390). A more general definition is the cost of establishing and maintaining property rights (Allen 1999: 898). As examples of transaction costs, Coase included the task of discovering what market prices are and the cost of negotiating a separate contract for each transaction. Institutions emerge to reduce those costs, but they can never be eliminated entirely. Firms still use contracts, but they are of longer duration and of a different nature:

It is true that contracts are not eliminated when there is a firm but they are greatly reduced. A factor of production (or the owner thereof) does not have to make a series of contracts with the factors with whom he is co-operating within the firm, as would be necessary, of course, if this co-operation were as a direct result of the working of the price mechanism. For this series of contracts is substituted one. (1937: 391)

Organizing and using managerial hierarchy within the firm has costs, so the decision of what transactions to perform internally involves weighing the trade-off between transaction costs and organization costs. That was Coase's fundamental insight.

Let's return to the economic analysis of the ice cream shop as a firm employing labour and capital to produce output. So far this model of the ice cream shop fits with the neoclassical view of the firm. Coase's insight gives us deeper understanding, by prompting questions in several dimensions around how the owners organize production. Do the shop owners make the ice cream on-site, or buy in ice cream from a supplier? If they buy ice cream, do they contract for standard flavours, or do they have the supplier make custom flavours that are unique to that shop but (perhaps) might be sold to other shops? Do they have a long-term contract with a single supplier, or do they place orders with any one of a number of multiple suppliers in a spot market when needed? When the shop hires workers, do they contract with them on a day-to-day basis, or do they enter into longer-term employment contracts?

Consider how costly it would be to have to settle on a new contract each day for each worker who comes to the shop, and for that contract to specify the tasks to be performed. Longer-term employment contracts that make the employee part of the firm typically economize on transaction costs, enabling the shop owners to schedule and plan production and the workers to schedule tasks based on more stable expectations and routines. Longer-term employment contracts also encourage firms to invest in worker training, making them more productive. But shop owners may decide not to bring all of the relevant transactions into the firm. It may be cheaper for them to specify the quality of ice cream they want and contract with a private label ice cream manufacturer

(or lease an “industrial kitchen” and hire specialized ice cream “chefs”) than to buy all of the equipment and inputs to make the ice cream in their retail shop. The firm, simply responding to profit incentives, tends to discover and implement the lowest transaction costs solution, and thereby deliver quality ice cream to customers at the lowest possible price.

The basic idea is deceptively simple: transaction costs determine what a firm does in house and what inputs it buys, so firms perform functions internally that are cheaper (given a specific level of quality) for them to accomplish than through independent contracts in markets. The firms contract with others for functions that are cheaper to accomplish through markets than by organizing internally. This paradigm may seem basic, but it has sparked a wide range of research and created new fields of inquiry in economics, management, and political science.

If a firm is successful and faces sufficient demand to expand, it can expand by increasing the amount of its production, by expanding into related product lines (product differentiation), or by merging with a competitor (horizontal integration). It can also integrate backward by producing its own inputs, or forward into more finished goods and marketing (vertical integration). Coase argued that the comparison between transaction costs and organization costs determine the size and boundaries of the firm as well as the extent of vertical integration.

[A] firm will tend to expand until the costs of organising an extra transaction within the firm become equal to the costs of carrying out the same transaction by means of an exchange on the open market or the costs of organising in another firm. (1937: 395)

Note how this theory uses the fundamental economic idea of evaluating opportunity costs at the margin, which Coase embeds in all of his work. As an alternative to integration into a firm, Coase pointed out that long-term contracts can avoid some transaction costs and can be attractive to risk-averse parties—but they still have the risks associated with imperfect foresight and forecasting. Thus, contracts are necessarily incomplete and cannot cover every possible circumstance that could arise in a production relationship.

Throughout his career Coase viewed Adam Smith's pioneering ideas as important insights. In the idea that transaction costs determine a firm's boundaries, Coase built on Smith's foundations. Smith grounded his economics in the division of labour, and the idea that by specializing in a task and working with others who are specialized in complementary tasks, people can be more productive, be wealthier, and create economic growth. In "The Nature of the Firm" Coase takes this idea of specialization and asks where and how specialization occurs, and how specialization affects which functions are best accomplished within the firm and which are best accomplished through contracts in markets. Specialization and organization are two dimensions of the same question of how best to organize production.

Coase's introduction of transaction costs and organization to the theory of the firm initiated new work in industrial organization, leading to new fields of transaction cost economics (TCE), organizational economics, and new institutional economics, now broadly called institutional and organizational economics (IOE). IOE focuses on governance institutions and their diversity in organizing production relationships, enabling organizations to adapt to unknown and changing conditions, to protect their investments in assets specific to that relationship, and to harmonize the interests of the parties in the relationship.

Figure 2, adapted from Shelanski and Klein (1995), shows the continuum of diverse governance institutions for organizing production, from open spot markets to fully integrated firms. IOE research building on Coase (1937) has expanded the analysis of hybrid methods of organization beyond long-term contracts to include relational contracts that are informal relationships held together by the expectation of future value, as well as other forms of hybrid ownership and control.

from additional control. Such governance can provide benefits in two general categories—it can enable people in production relationships to adapt better in the face of imperfect foresight, and it can enable them to make longer-term commitments of relationship-specific assets that they might otherwise not make. Analyzing governance institutions takes the idea that incomplete contracts are pervasive and unavoidable and the analysis examines how people manage that incompleteness. Even market contracts can be complex, with short-term and long-term contracts achieving different objectives, so the study of governance institutions applies to market relationships as well.

Relationship-specific assets, also called asset specificity, play a large role in TCE research. Suppose, for example, that our ice cream shop owners want a particular shape of cone to be their signature way of serving ice cream, embossed with their logo. Their make-or-buy decision is whether to make the cones or to contract out cone production to a supplier. If they contract out, do they provide the machinery to make the logo-embossed cones, or does the supplier purchase the asset? If the supplier purchases the machinery, and it is expensive and can only be used to make cones for that single firm, the supplier will want a long-term contract to ensure that it earns what it considers to be a sufficient return on its investment since it cannot use that machinery in any other production relationship. The ice cream shop owners, though, are concerned that the supplier, knowing how essential its work is to the firm, might hold out for a larger share of the economic pie that arises from cone sales (also known as “the holdup problem”). Given that contracts are incomplete, writing a long-term contract that negotiates a mutually beneficial split of those rents might be difficult (in other words, transaction costs exist). Hence, the more profitable arrangement may be for the ice cream shop owners to buy the machinery, hire employees from the supplier, and make the signature cones themselves.

More generally, a cooperative production relationship generates value that the parties did not (or could not) allocate in advance in their contract and which they have to divide between them. The contractual incompleteness gives each one an opportunity to try to get a bigger share of the pie, and they exert effort to do so, so it may be that vertical integration proves to be less costly

because it eliminates the incentive to behave opportunistically (Monteverde and Teece, 1982).

The most commonly cited application of this idea is the analysis of the relationship between General Motors and Fisher Body in the 1920s from Klein, Crawford, and Alchian (1978). General Motors had a 60 percent ownership stake in Fisher Body, which made closed car bodies for GM and other manufacturers and had considerable autonomy in decision-making in its relationship with GM. To accommodate Fisher's production process, GM had to make some very costly investments in production machinery and processes that would have become obsolete had GM switched to another body supplier. Klein, Crawford, and Alchian used this relationship as an example of the potential cost of the holdup problem that GM would bear. The holdup problem arises when party A depends upon party B to perform some action, but party B—knowing that party A has become dependent on B to carry through with the action—threatens not to complete the action unless party A pays more than was originally agreed to by party B.

This specific case ultimately resulted in GM acquiring Fisher Body in 1926 and vertically integrating into auto body production. With the producer of automobile engines and chassis now also owning the maker of automobile bodies—that is, with both operations owned by GM—there was obviously no incentive for one “division” to try to hold up the other. However, the Klein-Crawford-Alchian interpretation of this history as resulting from a holdup problem remains controversial, with a lively debate resurfacing in 2000 that included further research from Coase. As Peter Klein notes,

Klein, Crawford, and Alchian (1978) and Klein (1988) cite the case as a classic example of vertical integration designed to mitigate holdup in the presence of asset specificity. Fisher refused to locate its plants near G.M. assembly plants and to change its production technology in the face of an unanticipated increase in the demand for car bodies, leading G.M. to terminate its existing ten-year supply contract with Fisher and acquire full ownership. Coase (2000), revisiting the original documents, argues instead that the contract performed well, and was gradually replaced with full ownership

only to get Fisher's top managers (the Fisher brothers) more closely involved in G.M.'s other operation.... In short, G.M. did not acquire the remaining 40 percent of Fisher's stock in response to an inappropriate alignment between transactional attributes and an existing governance structure. Rather, the long-term contract signed in 1919 was adequate for mitigating holdup in the face of asset specificity and uncertainty, and was replaced by vertical integration for secondary reasons. (2005: 446)