HARVARD | BUSINESS | SCHOOL

Design Rules, Volume 2: How Technology Shapes Organizations

Chapter 5 Complementarity

Carliss Y. Baldwin

Working Paper 19-036

Design Rules, Volume 2: How Technology Shapes Organizations

Chapter 5   Complementarity

Carliss Y. Baldwin
Harvard Business School

Working Paper 19-036

Design Rules, Volume 2: How Technology Shapes Organizations

Chapter 5 Complementarity

By Carliss Y. Baldwin

Note to Readers: This is a draft of Chapter 5 of *Design Rules, Volume 2: How Technology Shapes Organizations*. It builds on prior chapters, but I believe it is possible to read this chapter on a stand-alone basis. The chapter may be cited as:

Baldwin, C. Y. (2018) "Complementarity" Harvard Business School Working Paper (October 2018).

I would be most grateful for your comments on any aspect of this chapter! Thank you in advance, Carliss.

Abstract

The purpose of this chapter is to relate the theory of task networks and technology set forth in previous chapters to theories of firm boundaries from economics and management. Complementary goods have more value when used together than separately. Complementarity may be strong or weak. Strong complements are *specific* and *unique* goods that have no value (or greatly diminished value) unless all are present in use. In the task network, dense technical interdependencies create strong complementarity, but it can arise for other reasons as well.

Transaction cost economics and property rights theory advise that strong complements should be placed under unified governance, for example, through common ownership. Agency theory suggests that weak complementarity can be handled via arms-length transactions and contracts. Furthermore, strong or weak complementarity are not innate properties of tasks and assets, but can be the result of choices regarding task networks, incentives and job design.

Supermodular complementarity exists when more of one input makes more of another input more valuable. *Distributed supermodular complementarity (DSMC)* exists when two or more independent actors can create complementary value by pursuing their own interests, and will not find it advantageous to combine in order to coordinate their actions. I derive formal conditions under which DSMC holds as a consistent pattern in a dynamic equilibrium. Given DSMC, clusters of firms making different complementary goods, including open platforms with surrounding ecosystems, can survive and compete effectively against integrated firms that control all complementary inputs.

Introduction

"The time is ripe for a fresh, modern look at complementarity." Paul Samuelson[1]

In Chapter 2, I argued that densely interdependent sectors of a task network cannot be split across two firms without incurring high transaction costs. Such sectors are more efficiently structured as transaction free zones. However, there is nothing that stands in the way of placing several transaction free zones under the governance of a single corporation. A single legal entity thus may own and control multiple transaction free zones.

In economics, theories about what assets and activities should be grouped together under common ownership and unified governance go by the titles "theory of the firm" or "theories of vertical integration." All three branches of organizational economics—transaction cost economics, property rights theory, and agency theory—have something to say on these issues. Although the theories are broadly consistent, there is no single theory that integrates all three approaches. Management scholars have also studied these questions and offered their own theories of how best to set the boundaries of a firm.

The purpose of this chapter is to relate the theory of task networks and technology set forth in previous chapters to theories of firm boundaries from economics and management. I begin by reviewing prior theories of the firm. We shall see that several prior theories can be boiled down to the principle: "assets, skills and activities that are *strong complements* should be owned and governed by a single person or corporation," i.e., should be subject to unified governance. Strong complementarity exists when an asset, skill or activity has no value (or greatly diminished value) except in the presence of other *unique* assets, skills or activities.

I shall argue that task interdependence gives rise to strong complementarity between the assets and people performing the tasks. Thus the argument (in Chapter 3) that it is economical to encapsulate interdependent tasks within the boundaries of a single corporation is consistent with economic theories based on strong complementarity.

However it is also possible for strong complementarity to exist without task interdependence. This in turn can lead to a corporate structure made up of multiple multiple divisions governed by a central headquarters. Such organizations generally set up multiple internal transaction free zones, corresponding to business units, and create formal or quasi-formal internal transactions between them. Such transactions are not enforceable by law, but are generally recognized and upheld by the corporate hierarchy. Oliver Williamson labels this type of organization the M-form of corporation.[2]

Weak complementarity exists when the value of an asset, skill or activity is enhanced by the presence of other, non-unique assets, skills and activities. Existing

[1] Samuelson, 1974, p. 1255.

[2] Williamson (1985).

theories of the firm are mostly silent on the impact of weak complementarity on firm boundaries. However, recent studies of platforms and their ecosystems indicate that complementary assets and activities may be distributed over many autonomous organizations, including corporations, commons organizations, and open collaborative projects.

When is it advantageous to spread complementary assets, skills and activities over many separate organizations vs. uniting them within the boundaries of a single firm? Agency theory suggests that there may be a tradeoff between unified governance on the one hand and an agent's effort on the other. Freeland and Zuckerman (discussed in Chapter 3) also argue that hierarchical authority can lead to disaffection and reduced effort.[3] Thus just as it sometimes makes sense to hire an outside contractor instead of an employee, it sometimes makes sense to distribute complementary assets, skills, and activities across separate organizations. Agency theory and sociology both teach us that the benefits of autonomy may exceed the costs.

To investigate when and how this happens, I introduce another concept of complementarity from economics—*supermodular complementarity* as defined by Francis Edgeworth, Donald Topkis, Paul Milgrom, and John Roberts.[4] Supermodular complementarity arises when increasing one input to a technical system increases the positive impact of some other input. More of one makes more of the other more valuable.

Supermodular complements are not necessarily strong complements. It follows that some forms of supermodular complementarity can be managed effectively by autonomous actors each pursuing separate interests. I describe the mathematical conditions that support *distributed governance* of supermodular complements. When these conditions are satisfied, a group of independent firms will *perform as well or better than a single firm* with unified governance. Organizational forms characterized by distributed governance are the principal focus of Parts 3 and 4 of this book.[5]

5.1 Strong vs. Weak Complementarity

Complementary goods are used together. Examples include right and left shoes; razors and blades; cake and icing; horses and carriages; cars and highways; TV sets and TV shows; computer hardware and software; and tea, hot water and a cup. The value of a group of complements in joint use is *super-additive*, that is, the things used together are more valuable (to someone) than the sum of their values in separate use:

$$V(\text{Tea, Hot Water, Cup}) > V(\text{Tea}) + V(\text{Hot Water}) + V(\text{Cup}) \quad . \tag{1}$$

[3] Jensen-Meckling (1976); Baker Gibbons Murphy (2002); Freeland and Zuckerman (2016).

[4] Edgeworth (1881); Topkis (1998); Milgrom and Roberts (1990; 1995).

[5] Single firms with unified governance include proprietorships and partnerships as well as corporations. Commons organizations, open collaborative projects, and clusters of firms are characterized by distributed governance.

Goods are *strong complements* if they are unproductive (or have greatly diminished value) unless used together. Strong complementarity applies to *specific and unique* goods that through some technological transformation have been tied together. Examples are the windows of a house and the house itself, a lock and its key, the engine and chassis of a truck after the truck is built, and the two ends of a pipeline.[6]

Tea, hot water and a cup are not strong complements because one can drink tea from a glass, use hot water to make coffee, and serve other liquids in the cup. Furthermore, if there are several providers of each component, then each component will retain its value even if a *specific* complement is withdrawn. If Twinings tea is not available, a tea drinker can use Lipton's instead. For most tea drinkers, the value of the bundle is not (much) diminished by using another brand of tea.

Hence if there are substitutes for all components in a group of complements, withdrawal of one component does not destroy the value of the others. Such components are *weak complements*.

5.2 One-way vs. Two-way Complementarity

In some cases, one or more components in a group of complements may be unique while the other components have substitutes. In such cases, the withdrawal of any of the unique components will destroy the value of the system. Conversely, any of the non-unique components can be withdrawn without inflicting much harm.

The presence of unique and non-unique complements in a technical system creates an asymmetric economic relationship. The non-unique components *depend on* the unique ones for their value. Hence the owner of a unique component can threaten to harm the owners of non-unique components by excluding them from the system. In such cases, strong complementarity operates in one direction, hence is "one-way." In effect, the owners of unique components have the ability to "hold up" the owners of non-unique components. By virtue of this threat, they can demand a disproportionate share of the surplus value of the system.

If there are several, separately owned unique complements in a technical system, their owners are in a symmetric relationship. Each can independently destroy the value of the system by withdrawing their piece of it. Strong complementarity operates in both (or all) directions, hence is "two-way." Each owner of unique complement has equal bargaining power.

In transaction cost economics, Oliver Williamson defines "asset specificity" as arising when the value of a given asset is high when it is used in conjunction with another

[6] Hart and Moore (1990) p. 1135.

(unique) asset, but much lower when the other asset is not present.[7] Asset specificity is a form of strong complementarity.[8]

In property rights theory, investments by workers in firm-specific skills are termed "relationship-specific investments." [9] Relationship-specific investments create strong two-way complementarity between a particular firm and workers who have made relationship-specific investments.

In the management literature, David Teece has argued that complementary assets may be generic, specialized or co-specialized. "Specialized assets are those for which there is unilateral dependence between the innovation and the complementary asset. Co-specialized assets are those for which there is a bilateral dependency." Specialized assets are one-way strong complements, while co-specialized assets are two-way strong complements. According to Teece, if an innovation is specialized or co-specialized to a complementary asset owned by someone else, the innovating firm can expect to split the value created by the innovation with the other party.[10]

Asset specificity, relationship-specific investments and specialized innovations all tie an asset or activity to some *unique* other asset or activity. Thus all three concepts entail strong complementarity. The strong complementarity locks in one or both parties and makes them vulnerable to an *ex post* threat of holdup by the other. (Throughout this book, I will use the term "strong" complementarity to refer to complementarity between specific and unique elements. This type of complementarity is also called "strict", "perfect", and "unique" complementarity.[11])

5.3 Task Interdependence Creates Strong Complementarity

In Chapter 2 we saw that tasks within a module of a task network are interdependent by virtue of the many uncodified and contingent transfers (of material, energy and information) that are necessary to perform the tasks adequately. Problem-solving within a module consists of tracing the cause-and-effect relationships among tasks and then modifying *specific* tasks and transfers to achieve a better outcome. Over time, the *specific* tasks and transfers become co-specialized through an ongoing process of mutual adaptation.[12]

---

[7] Williamson (1985), p. 55.

[8] *Ibid.* p. 53.

[9] Hart and Moore (1990).

[10] Teece (1986).

[11] Jacobides, Cennamo, and Gawer (2018).

[12] My definition of task interdependency differs in its details from that of Puranam (2019). In Puranam's usage "interdependency" is derives from value: tasks are interdependent if their inputs or outputs are either substitutes or complements. In my usage, an individual task may not have separable outputs. Each task, however, can have multiple material effects on a production system or a design. These material effects in turn may require modification (adaptation) of other tasks. If Task A has an impact on

In Chapter 4, I argued that people performing tasks within a module will generate a pattern of organizational ties that works to solve frequently-arising problems. Assets will also be configured to support the most common work flows. As the tasks are co-adapted, the assets and people performing the tasks generally become co-adapted as well.

In this fashion, task interdependency gives rise to strong two-way complementarity between tasks, skills and assets. Task interdependency within modules is thus a *technological cause* of strong complementarity. Assets used within a module have specialized roles and depend on other assets in a specific configuration. People working within a module must invest in organizational ties and procedures that work only in that specific context. And innovations within a module are generally not transferrable to other parts of the system without significant modification.

5.4 Strong Complementarity and Theories of the Firm

Two branches of organizational economics—transaction cost economics and property rights theory—place strong complementarity at the center of theories of firm boundaries and vertical integration. Theories in the management literature based on the concept of "synergy" also rest on the existence of strong complementarities between business units within the same corporation. Although they are based on different assumptions, these theories arrive at the same conclusion: strong complements should ideally be placed under unified governance.[13] In this section, I review these theoretical arguments with the aim of highlighting their similarities.

Transaction Cost Economics

In transaction costs economics, a market is an institution where voluntary exchanges (transactions) take place between autonomous agents. In contrast, a firm is a hierarchical institution in which a single actor has the authority to make decisions for the benefit of the entire enterprise.

The main problem with bilateral exchange (markets) is holdup, that is, the opportunistic withdrawal of a unique complementary asset from a technical system. Within a hierarchy, i.e., within a firm, holdup is controlled through the exercise of authority (fiat). It follows that, in a given technical system, a single firm should own or otherwise control all unique complementary assets.[14]

Task B, then B depends on A, and vice versa. If A depends on B and B depends on A (directly or indirectly), then A and B are interdependent. My use of the term complementarity is consistent with Puranam's: assets, skills, and activities are complementary if their joint use exceeds the sum of the value of their separate uses, i.e, the value function of the inputs is super-additive.

[13] Unified governance is normally achieved by having one entity own the assets and employ the people needed to perform the tasks specified by a technical recipe.

[14] Williamson (1985); Argyres and Zenger (2012).

Property Rights Theory

In property rights theory, a firm is viewed a coalition of suppliers, some of whom provide of physical and intellectual capital (assets) and while others provide skills and effort (workers).[15]

Initially the actors form coalitions (firms) that include owners of assets and workers. There is no hierarchy or authority: all participants act autonomously in their own best interest. Once a coalition has formed, workers make relationship-specific investments in skills and other forms of human capital. These investments improve the productivity of the coalition, but are worthless outside it. As a result the assets and skills provided by different members of the coalition become strong two-way complements. Their joint value can only be realized if all are present.

At the next stage, a specific product design is revealed.[16] Members of the coalition then bargain over the terms of their participation. By assumption, there are always gains to maintaining the coalition, and as a result, the coalition always survives and joint production always occurs. However, ownership of non-human assets (property rights) affects the way the joint surplus is divided, because (again by assumption) an owner can threaten to withdraw her assets, thus damaging workers and other asset owners.

The main problems arising in this context are under-investment in relationship-specific skills by workers and *ex ante* defensive investments by asset owners aimed at improving their *ex post* bargaining positions. Rather than maximizing the value of their joint output, workers and asset owners will instead invest to improve their outside options.

Within this framework, it can be shown that strongly complementary assets should be owned by a single actor in order to maximize workers' returns on relation-specific investments. In this fashion, a "second-best" outcome can be achieved. (First-best can be attained only if a single entity controls both physical assets and human capital.[17])

[15] Holmstrom and Roberts (1998), p. 77 say that "a firm is exactly a collection of assets under common ownership." However, the assets must be used in conjunction with suppliers of human capital (workers) who in turn make relationship-specific investments. Thus the productive entity includes both the assets and the workers, working as a voluntary coalition.

[16] The *ex post* revelation of the design prevents the parties from reaching an *ex ante* contract specifying their separate roles. The work that will be done is "non-contractible."

[17] Worker ownership of physical assets is ruled out by assumption. The assumption is justified by the fact that many technologies require indivisible physical assets or intellectual property, which are beyond the means of most individuals in society. Workers must then contract with entities, such as corporations, that do have the means to acquire and manage the complementary assets. The possibility that commons organizations may manage the assets as common pool resources is not discussed in this literature.

Management Theories of Corporate Boundaries and Vertical Integration

Management theories of corporate boundaries are mainly based on the notion of synergies. The argument is that business units should be integrated within the same corporation when the value of the combination is substantially greater than the stand-alone value of the separate businesses.[18] Synergies are in effect another name for strong complementarities between business units.

There are many sources of business synergies. Synergies can arise from the opportunity to create denser organizational ties, hence greater coordination between the business units. These synergies derive from the strong complementarity of interdependent tasks.

In other cases, synergies arise through of the reduction of a threat. Generally a threat (to one organization) exists if another organization controls a strong complement, for example, a unique source of raw materials or a blocking patent. The threat disappears if the strong complements are controlled by a single organization. Resource dependency theory maintains that unified governance (achieved through vertical and/or horizontal integration) eliminates the threat. Thus in common with transaction cost economics and property rights theory, resource dependency theory recommends that strong complements be placed under unified governance.

Finally, synergies between organizations can arise if one firm has developed superior knowledge or competence that can be used to improve another firm's performance. If the transfers from the first firm to the second are well-defined, countable and easy to value, then the two firms can structure a transactional relationship, such as a consulting engagement. However, if the first firm's contributions are hard to define, measure, and value then its best course of action may be to acquire the second firm, as long as it can do so at a favorable price.

In these circumstances, the first firm's distinctive capabilities are strong one-way complements of the second firm's business. If the first firm's capabilities are not unique, then firms with similar capabilities will bid up the price of the target. If the complementarity is symmetric, i.e. the target is unique, then the first firm may need to split the value created with the shareholders of the target in order to obtain control.[19]

As indicated, strong complementarity between specific assets and/or human skills and synergies between business units can arise because of interdependency of tasks.

[18] Puranam and Vanneste (2016) Chapter 2.

[19] Management theory also recognizes the possibility of negative synergies. In these cases, the presence of different businesses under one roof detracts from overall performance. For example, pay inequalities between divisions of a corporation may cause disaffection resulting in low performance in the inferior division. Feldman and Gartenberg (2018).

However, the prescriptions of transaction cost economics, property rights theory, and the management literature rest on strong complementarity alone, and make no mention of task interdependence. Can people, assets, and activities be strong complements, yet *not* linked by dense transfers of material, energy and information? The next section gives examples of cases of strong complementarity without task interdependence.

5.5 Strong Complementarity without Task Interdependence

Separate modules of a task network, separated by thin crossing points, can be strong complements if one or or more modules are unique and essential to the functioning of the system. The subsections below describe two situations in which strong two-way complementarity can exist without task interdependence.

Example # 1: Coal to Philadelphia

Margaret Blair described a technical system to transport coal via the Lehigh River from the Lehigh Valley to Philadelphia.[20] Getting the coal from underground to the point of use required three separate investments: (1) a coal mine; (2) roads from the mine to the river; (3) a series of dams and walls in the river so that "a flat boat loaded with coal could float down the river without running aground on the rocks." [21]

The three parts of the project were clearly strong complements: all were necessary to derive any revenue from the coal. They did not, however, display a high degree of task interdependence. As with the smiths and the cooks in Chapter 2, the transfers of coal from the mine to wagons and wagons to boats were fairly standard and easy to define, measure and value. In principle, such transfers could easily be converted into transactions.

The problem was that—at the time the investments were needed—there were no other potential suppliers of these services nor customers for them. Thus each stage, if built, would be a unique and essential step in the process of getting coal to Philadelphia. The promoters attempted to finance the three investments separately, but they were unable to attract outside investors. In the end, concerns about transactional hazards led the three stages to be united within the boundaries of a single corporation.[22]

Task interdependence is thus *sufficient but not necessary* for strong two-way complementarity between activities. Even if they are divided into separate modules, different stages of a single production process can be strong complements, if all stages are essential and no alternative sources exist.

[20] Feldman and Gartenberg (2018).

[21] *Ibid.* p. 400.

[22] This example falls under Williamson's heading of "dedicated asset specificity," compounded by "site specificity." Williamson (1985) pp. 55.

Example # 2: The Anticommons

Michael Heller defined an *anticommons* as "a property regime in which multiple owners hold effective rights of exclusion in a scarce resource."[23] Examples of anticommons include separate titles to very small parcels of land and separate patents on multiple subparts of a technical process. In cases like these, permissions must be assembled from *all* the owners before the land can be put to use or the technical process can be implemented.

Heller cited the example of land in Kobe, Japan, where "there are thousands of parcels the size of a U.S. garage" and a building "can be based on a plot that is actually dozens of smaller parcels thrown together by developers."[24] During the Kobe earthquake of 1994, many buildings were destroyed. However, the consent of all claimants to the land had to be obtained before redevelopment could take place. Several years after the quake, rebuilding plans throughout the city were blocked by various claimants, and more than half the housing stock remained "damaged or in rubble."[25]

Land titles are separable components that are not part of an interdependent task structure. However, each title is unique and essential to the final product (a building that straddles the subdivisions). The titles are thus strong one-way complements of the building.[26] To avoid the hazard of hold up, all titles should be held by the developer before construction begins.[27]

Similarly a complex technical process may be based on many pieces of individual knowledge about different techniques and mechanisms. Patents create property rights in such knowledge: the patent holder can demand that individuals and firms pay to use the knowledge, or may exclude them altogether.

Each patent is a separate asset. Obtaining diverse patents may not involve interdependent tasks. However, patent rights are unique and may be essential, if there is no other way to carry out a technical process other than by using the knowledge covered by the patent. In such cases patent rights are strong one-way complements of any process that uses the patents. As in the case of land titles, to minimize the risk of holdup, licenses

---

[23] Heller (1998); Heller and Eisenberg (1998).

[24] Sapsford, J. *Wall Street Journal* (December 12, 1996), quoted by Heller (1998) p. 684.

[25] *Ibid*.

[26] If the "highest and best use" of the land is a specific building, then the titles and the building are strong two-way complements.

[27] Transaction costs can be high in the pre-construction phase, but they may be astronomical after the building is built and construction costs are sunk.

to all essential patents should be held a single firm before work on the process begins. (In Chapter 7 below, I show how functional components that are unique and essential to a technical system give rise to "bottlenecks" that can be used to capture value within the system.)

5.6 Vertical Integration

In summary, there are several reasons to bring different parts of a technical system under *unified governance* within the boundaries of a single firm. The placement of different complementary activities within a single firm is commonly known as "vertical integration."

The first reason for vertical integration might be called "technological." As we saw in Chapter 3, unified governance makes it relatively easy to place technologically interdependent tasks within a single transaction free zone that is encapsulated by transactions. There unprogrammed transfers of material, energy and information can take place freely. As people within the zone solve problems posed by the technology, the tasks, assets, and organizational ties will be co-adapted, over time becoming strong two-way complements.

A second reason for vertical integration might be termed "defensive." Here the aim is to prevent holdup, defined as the withdrawal of a unique and essential input by an opportunistic counterparty. This type of vertical integration is the main focus of transaction cost economics and property rights theory. Defensive vertical integration unites assets and activities that are strong two-way complements in the sense that if one part is withdrawn the other parts will suffer a significant degradation in value.

A third reason for vertical integration might be called "strategic." In these cases, the focal firm believes it has special resources or capabilities, that can be applied in a variety of settings. If the resources and capabilities are unique, then they are strong one-way complements of the external opportunities.

When pursuing strategic vertical integration it is important to gain control of the target assets at the right price. Because the complementarity is one way, failure to gain control of a specific target will not threaten the acquirer's business. In contrast, failing to gain control of strong two-way complements creates a potential for holdup that may jeopardize the entire enterprise.

Technology affects all forms of complementarity. For example, task interdependence can be reduced by creating modules separated by thin crossing points. The threat of opportunistic holdup can be decreased by finding substitutes for unique and essential inputs. (In fact, the search for technical substitutes, such as patent work-arounds, is a classic form of defensive investment.) Finally, advantageous one-way strategic complementarity may be obtained by developing a core product or process that is valuable in a wide range of settings.

In contrast to defensive and strategic forms of vertical integration, vertical integration aimed at increasing technical interdependency has received comparatively little attention in the management and economics literatures. In most theoretical models, technologies are considered to have clear dividing lines established by natural laws. The advantages and disadvantages of integrating or separating different parts of a technical recipe are not usually a subject of analysis in economics or management.

Is it possible to have complementary assets and activities that create more value under distributed governance than under unified governance? Agency theory shows how this is possible.

5.7 Agents Within and Outside the Firm

The third branch of organizational economics—agency theory—does not focus on strong complementarity between assets and activities or the attendant risk of holdup. Firms are characterized by principals (owners) and agents (workers). A particular firm is a nexus of contracts designed by the principal to attract agents and elicit productive effort from them. Agents, however, have their own utility functions and there are technological barriers to paying each agent his or her marginal product to elicit the "right amount" of effort.

Importantly, principals and agents are not uniquely matched nor essential to one another. Agents always have an "outside option." Principals do not need to attract a unique or specific agent, but can work with whoever finds their offer attractive. Thus principals and agents are complementary—they create more value together than either could separately. But they are weak, not strong, complements.

The main problems encountered in agency theory are effort directed away from unmeasured or unrewarded activities, the diversion of resources to perquisites, and excessive or insufficient risk-taking. Asset ownership and governance can be used to change incentives in ways that reduce these *agency costs*.

The internal structure of the firm is assumed to be hierarchical in the sense that some actors have the ability to establish incentive compensation for others. There is a difference in the incentives offered to outside agents vs. employees, *but, consistent with the nexus of contracts view of firms, these are simply names given to two generic contracts.* Both outside contractors and employees are viewed as part of the firm's technical system.[28]

Technologies are assumed to differ in terms of (1) how easy it is to measure outputs and inputs of various stages; (2) the degree to which the principal can observe directly ("monitor") what the agent does; (3) the amount of risk inherent in the process (this affects the reliability of measurement); and (4) the degree of "multitasking" that is

[28] Alchian and Demsetz (1972); Jensen and Meckling (1976); Holmstrom and Milgrom (1991); Holmstrom and Roberts (1998); Baker, Gibbons and Murphy (2002); Gibbons (2005).

desirable. These properties are generally taken as fixed, although risk is sometimes assumed to be under the agent's control.

One determinant of whether an activity should be carried out by a (generic) contractor or a (generic) employee is the degree of "multi-tasking" involved in the job. In an influential paper, Bengt Holmstrom and Paul Milgrom present a model in which the agent's job consists of multiple tasks.[29] The agent's performance can be measured with precision for some tasks and only imprecisely (or not at all) for others. For example a sales agent may be tasked with making current sales (measurable); transmitting information about customer needs (less measurable); and keeping up an ongoing customer relationship (even less measurable).[30] All of these tasks demand effort and attention, hence are costly to the agent.

If the agent's pay increases in proportion her measurable output (current sales), then she will focus her effort and attention on that activity to the detriment the other two.[31] Thus, if customer intelligence and long-term customer satisfaction are important to the firm, giving sales agents so-called "high-powered incentives" based on current sales leads to a suboptimal allocation of their effort and attention.

An alternative contract would give the agent higher base pay with "low-powered incentives," that is a low commission rate. The agent will then spend more time on the unmeasured tasks. But if she is an outside contractor who sells the goods of many different companies, she may direct her effort to selling those goods with the highest commission rates. *Thus low-powered incentives conducive to "multi-tasking" are not effective unless the agent agrees to work for the principal exclusively, in effect, becoming an employee.*

In these examples, the boundaries of a firm are permeable and the degree of co-specialization of assets and activities becomes a matter of choice given the constraints of the underlying technology. When the agent's performance has a small number of dimensions and can be measured objectively, then the optimal incentive contract is to give him control of all inputs and set rewards based on measured performance. In effect, the agent becomes the owner of a separate firm supplying goods and services to the principal.

Outside agents have incentives to maintain the measurability of their products. A small number of performance dimensions consistent with measurability requires a thin crossing point to exist between the supplier's product and the customer's process. Also without long-term assurances, outside contractors will generally avoid being overly dependent on a single customer, and customers will seek alternate sources of supply. Other things equal, both sides have reason to avoid co-specialization, hence their

[29] Holmstrom and Milgrom (1991).

[30] This example is taken from Holmstrom and Roberts (1998).

[31] See Steven Kerr's classic article "On the Folly of Rewarding A, While Hoping for B." (Kerr, 1975.)

complementarity will be weak. (Of course, as we saw in Chapter 4, firms can use formal and relational contracts to increase their interdependence.)

Conversely, when the job involves several discretionary activities whose output is difficult to measure, the optimal contract is low-powered. To make her commitment to multi-tasking credible, the agent may need to work for one principal exclusively. In effect, she becomes an employee, working under the principal's authority. Having committed to an exclusive relationship, the employee would like to be embedded in as many aspects of the technical system as she can manage.

Thus the technological conditions favoring employment vs. outside contracting are very similar to conditions prevailing within vs. across modules in a task network. Within a module, we have seen, transfers occur in many directions and are difficult to define, measure, and value. A given agent's contribution to the whole depends on transient circumstances as well as what other agents do. She must respond appropriately in different situations and most of her specific actions will be uncounted and unmeasured. This is very close to the definition of multi-tasking.

In contrast, across modules, transfers are simple and few, hence relatively easy to define, measure and value. Transactions at module boundaries have low mundane transaction costs, and measurement of what passes between modules limits opportunistic behavior on both sides (see Chapter 2). Finally, if an agent is paid for every transfer, he will have incentives to increase his effort in order to increase his pay. This is close to the definition of single-tasking.

It follows from this line of reasoning that strong or weak complementarity are not necessarily an innate quality of tasks and assets, as is usually assumed in the transaction cost and property rights literature. Instead strong or weak complementarity may be the result of choices made about the modular structure of the task network, incentives, and job design. Work within modules rewards co-specialization and strong complementarity, while work across modules rewards arms-length relations and weak complementarity.

Furthermore, since task networks, incentives and job design are all malleable (to some extent), there is room for variation across companies. Managers may make different strategic bets as to the level of task interdependency, multi-tasking, and high vs. low powered incentives needed to succeed in their product markets.

Agency theory teaches us that some forms of weak complementarity can be handled via arms-length transactions and contracts. I now apply that lesson to the case of platforms and their complementary ecosystems.

5.8 The Conundrum at the Heart of Platform Research

As discussed in Chapter 1, the last two decades have seen a growth in the phenomena of *open platforms*, *business ecosystems*, and *clusters of complements*. From a technological perspective, a platform system consists of a technical core and optional complements that use the platform's functionality.[32] The platform generally has little or no value except in conjunction with some set of complements and complements have no value except in conjunction with the platform.

Platforms in turn can be divided into product platforms and exchange platforms. *Product platforms* support the design and/or production of goods. The platform sets standards that allow the complements to interoperate with the platform and with each other. Examples include a computer operating system and applications, video game consoles and games, a microprocessor and compatible hardware, and TV sets and TV shows. An important sub-type of product platform is a *logistical or supply chain platform*, which facilitates the movement of products through a network of producers, distributors, wholesalers and retailers.[33]

Exchange platforms are physical or virtual spaces that facilitate valuable exchanges between diverse agents. Exchange platforms are also called two-sided (or N-sided) markets.[34] We will examine both types of platforms in greater detail in later chapters of this book.

A critical difference between product and exchange platforms is that in exchange platforms, the autonomy of actors is generally taken as given. For example, if the platform is a marketplace, the underlying task networks of buyers and sellers are already separate modules (as was the case for the smiths and the cooks). It makes no sense to merge the two task blocks for the purpose of conveying simple good, such as a pothook from one establishment to the other.

Similarly, with most communication platforms, such as the telephone system or email, the communicators are independent actors. Again, it does not make sense to merge them for the purpose of sending messages and receiving replies. Instead, an exchange platform allows users to make transient connections and then go their separate ways.

With product platforms (including logistical platforms), the situation is different. From the user's perspective, the platform and complements form a system. Moreover, the user needs both platform and complements for his system to have value. Finally, the modular structure of the system is not given *a priori* as it is in an exchange platform. The platform and complements are separate modules because the platform sponsor decided that a modular architecture was advantageous. In principle, the platform and complements could be produced interdependently and sold as a bundle.

In fact, we do see cases where a platform and complements are supplied by the same firm. But we also observe many cases where the platform and complements are

[32] Baldwin and Woodard (2011).

[33] Bresnahan and Greenstein (1998); Gawer and Cusumano (2002).

[34] Rochet and Tirole (2003); Parker, Van Alstyne and Choudary (2016).

supplied by different firms. Often an *ecosystem* of autonomous firms surrounds the platform.

Virtually all theoretical research on platform systems takes the existence of an ecosystem as given and proceeds from there. But this common premise begs the question: What allows the platform and complements to be supplied by autonomous entities not subject to unified governance and central control? Paraphrasing Ronald Coase's question, why are the platform and complements not produced by one big firm?[35]

In the rest of this chapter, I investigate the technological conditions that make the platform-cum-ecosystem an advantageous way to organize complementary activities. In this effort, I make use of the concept of supermodular complementarity, described in the next section.

5.9 Supermodular Complementarity

Strong complementarity, we have seen, arises when two or more *unique* assets or tasks have no value except in joint use. If assets or activities are strong complements, then property rights theorists and transaction cost economists alike recommend that they be owned and managed by a single entity to avoid holdup and opportunistic transaction costs.

Another type of complementarity exists, however. The concept was defined by Francis Edgeworth in the 19th Century, and refined by Donald Topkis, Paul Milgrom, and John Roberts in the 1990s.[36] It goes by the impressive name "supermodularity on a sub-lattice." In what follows, I will refer to it as "supermodular complementarity."

Intuitively, two items are supermodular complements if *more of one makes more of the other more valuable in relation to some desirable end result*. Examples of supermodular complements include sunny days and sunscreen, free time and leisure activities, smart phones and apps, and cars and highways. So-called direct and indirect network effects are also forms of supermodular complementarity.

Formally, two input variables are supermodular complements if an increase in one increases the positive effect of the other. Thus let x and y be variable inputs to a technical system that creates value $V(x,y)$. Let Δx and Δy denote positive changes in x and y respectively.[37] Then x and y are (strong) supermodular complements if:

$$V(x+\Delta x, y+\Delta y) - V(x+\Delta x, y) > V(x, y+\Delta y) - V(x,y) \qquad (2)$$

[35] Coase (1937) p. 394.

[36] Edgeworth (1881); Milgrom and Roberts (1990;1995); Topkis (1998).

[37] Technically, supermodular functions can be decreasing in the input variables. Then the property of supermodularity implies that moving from low to mixed will destroy less value than moving from mixed to high. This reversal can be cured by redefining the input variables so that the value increments are positive.

Here, the impact of increasing *y* is greater if *x* is increased as well. This is known as the property of "increasing differences." In effect, Δx provides Δy with an "extra kick."[38] The definition can be applied pairwise to any group of input variables.

The variables *x* and *y* may be continuous variables, discrete variables, or binary (yes/no) variables. Indeed, one of the strengths of the analysis is that many different kinds of variables, including material quantities, decisions, endogenous choices, and exogenous parameters, can all be incorporated within a single framework.[39]

Strong complements as defined earlier in this chapter are also supermodular complements. Recall that strong complements are unique (i.e., have no substitutes) and withdrawing any strong complement destroys the value created by the system of complements. Let $x=1$ if complement *X* is present, and $x=0$ if it is not present. Similarly let $y=1$, if complement *Y* is present, and $y=0$ if it is not.

The strong complementarity of *X* and *Y* implies that:

$$V(0,0) = V(1,0) = V(0,1) = 0 \text{ ; and}$$

$$V(1,1) > 0 \ .$$

Substituting these values into equation (2) above we have the test:

$$V(1,1) - V(1,0) > V(0,1) - V(0,0) \ .$$

All terms in the expression are zero except $V(1,1)$, which is greater than zero, thus the strong complements *X* and *Y* are perforce (strong) supermodular complements.

Figure 5-2 presents a graph of a generic value function for strong complements. The function only has value if both inputs are present, i.e., $x = 1$ and $y = 1$. Otherwise no

---

[38] Supermodular complementarity is often defined using a weak inequality (≥) instead of a strong inequality (>). Although this generalizes the concept and the proofs, it makes the exposition more complex and the concept less intuitive. My arguments depend on the existence of strong supermodular complementarity, hence I will take the strong inequality as given.

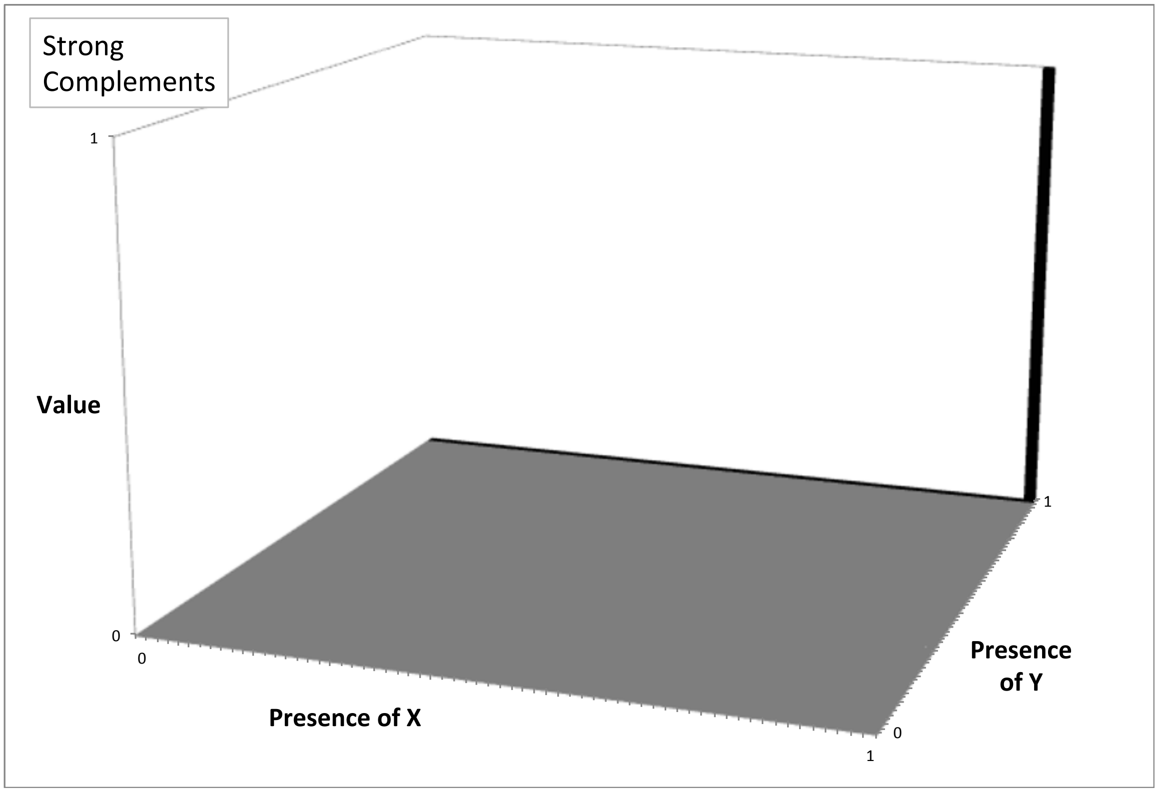
[39] However, the input variables must form a sub-lattice. On each input dimension of a sub-lattice, the quantity of the input can be measured and ranked. Furthermore, an increase in one variable never causes a decrease in another variable:

> Constraining the choice *x* to lie in a sublattice… says that increasing the value of some variables never prevents one from increasing others as well … and that decreasing some variables never prevents decreasing others. Milgrom and Roberts (1995) p. 182.

Variables that can be quantified (thus ranked) and chosen independently of one another form a sub-lattice. At the beginning of this chapter, I noted that tea, hot water and a cup are complements, but not strong complements because alternatives exist for each input variable. However, there is also no natural ordering relationship between a tea and coffee, hot water and other liquids, or a cup and a glass. Thus the input variables do not form a sub-lattice, and the function, although super-additive, is not supermodular.

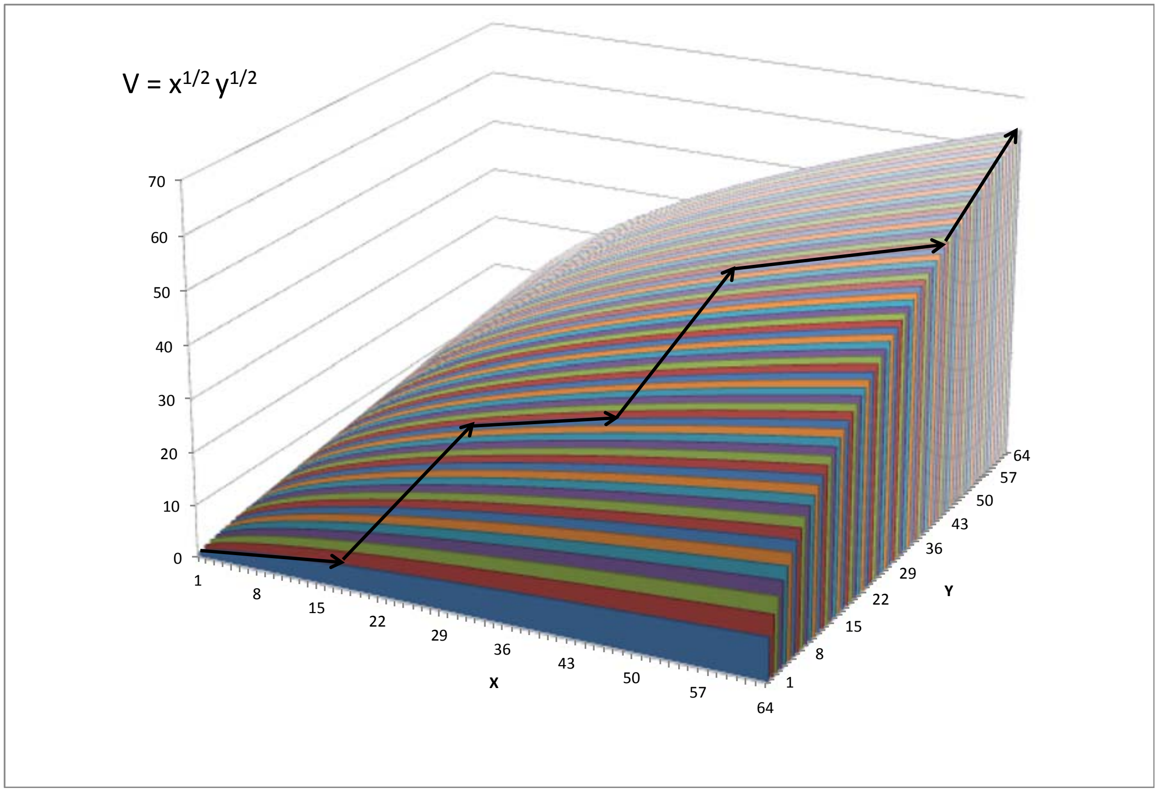
value is created. Obviously, this argument can be generalized to any number of strong complements.

Figure 5-2 Value Function for Strong Complements X and Y



However, other patterns are also possible. For example, Figure 5-3 shows the graph of a supermodular function of two variables that displays diminishing marginal returns in both inputs. (The function $V(x,y) = x^{1/2}y^{1/2}$ is in the family of Cobb-Douglas production functions.) The function passes the test of increasing differences: each time the input variable *x* rises, the slope of the function with respect to *y* rises, and similarly for *y*. Thus investments in *x* create a positive externality for investments in *y*, and vice versa. This pattern is indicated by the zigzag arrows along the surface of the graph.

Figure 5-3 Supermodular Value Function with Diminishing Marginal Returns to Each Input



Supermodular value functions have the property of *monotone comparative statics*, meaning that the choice variables (optimizers) will be non-decreasing in the parameters. In other words, the choice variables will tend to move up or down systematically in response to changes in environmental variables or actions taken by complementors.[40]

The fact that one choice can positively (or negatively) affect the payoff to another choice has led some authors to argue that choices exhibiting supermodular compementarity should be placed under unified governance.[41] It is true that if two (or more) choices are controlled by different agents, each may not receive compensation for the "extra kick" their choice gives to the other agent's payoff. The agents may seriously underestimate the joint value created and thus underinvest.

This is certainly the case when the decisions/actions are strong complements (see Figure 5-2): the owner/controller of *X* obtains no value unless the owner/controller of *Y*

[40] The test for supermodularity implies that the value function V cannot decline in any input variable. Thus any function with an interior maximum is not a supermodular function.

[41] See, for example, Milgrom and Roberts (1990; 1995).

takes symmetric action. If *X* and *Y* are owned and controlled by the same person, he or she will have incentives to act appropriately. Autonomous agents have no incentives act independently.[42]

However, looking again at Figure 5-3, it is possible to imagine Agent *X* acting to increase *X* and capturing the marginal returns from his action. Agent *Y* will then have greater incentives to increase *Y* in order to capture her marginal returns. Two (or more) separate agents, acting independently, can in this fashion move up the supermodular value function, as indicated by the arrows. Each would not even have to know of the other's existence, as long as both could correctly assess and capture the marginal value of their separate actions.

The requirement for consistent action by independent agents places constraints on a supermodular value function beyond the test of increasing differences. Not all supermodular value functions support consistent, independent action. As we just saw, the value function for strong complements fails this test.

Let me define the property of enabling consistent, independent action in pursuit of complementary goals as *distributed supermodular complementarity*. Distributed supermodular complementarity means that the group of complementary assets or activities need not be owned or controlled by a single entity. Two or more independent actors can create complementary value by pursuing their own incentives at any point in time. The actions of each will redound favorably on the rest. *But the actors and their separate inputs need only to be joined in the minds of users who perceive the complementary value of the goods.*

5.10 Direct and Indirect Network Effects

So-called direct and indirect network effects are forms of supermodular complementarity.

Direct network effects arise when additional users of a good increase the value of the good to other users. For example, as more telephones are added to a network, users can reach an increasing number of counterparties. The number of possible connections goes up as the square of the number of installations.

In the telephone (or any other) communication network, senders and receivers of messages are complements, though not unique complements. A completed message requires a sender and receiver, but the identities of senders and receivers vary, and they may switch roles from one message to the next. In principle, the n^{th} member to join the network increases the number of potential connections by $2(n-1)$, since any of the existing users may now connect with the new user and vice versa.

[42] If the inputs are costly and moves are simultaneous, the actors face a classic Prisoners' Dilemma game. The equilibrium of the game is for both to do nothing.

If each user values the network in proportion to the number of connections they can make, then the total value of the network of n users is proportional to $n(n-1)$. The difference between a network of n users and $n-1$ users is $2(n-1)$.[43] This difference is increasing in the number of users, hence the value function is supermodular. More users make the next user more valuable.

Indirect network effects arise when one set of participants in a networked system values the presence of another set. For example, users of computers value hardware and software that works on their platform, while software and hardware developers value platforms that offer many potential customers. Thus users will tend to migrate to platforms with more developers while developers will migrate to platforms with more users.[44]

In cases of indirect network effects, each defined group of participants (sometimes called a "side") values the presence of members of other groups (other "sides"). As the number of participants on one side increases, the value of the platform to the other sides increases *pari passus*. Again differences in value increase as numbers increase on each side. The value function is supermodular.

Network effects are common in all kinds of platform systems. In product platforms, more complements to the platform increase its option value to users, and symmetrically, more users of the platform increase the potential revenues of each complement. In logistical platforms, more nodes in the network increase the number of potential paths, thus increasing the number of goods that can be produced and transported. In exchange platforms, more buyers and sellers increase the potential number and variety of exchanges.

As indicated, in most exchange platforms, the autonomy of platform participants is take for granted. Transactors generally seek to exchange goods, services and/or information, and do not wish to merge their separate task modules into a larger organizational entity. However in product platforms (including logistical platforms), the boundaries of modules are endogenous. Participants thus need to consider whether complementary activities should take place within a single firm (unified governance) or in separate, autonomous organizations (distributed governance).

The next section defines mathematical conditions under which distributed governance of supermodular complements can be sustained in equilibrium when unified governance not ruled out *a priori*.

---

[43] Calculation: The difference in value between network of n members and $n–1$ members is

$$n(n-1) – (n–1)(n–2) = (n–1)(n – n + 2) = 2(n–1) \quad .$$

[44] Church and Gandal (1992; 1993).

5.11 Distributed Supermodular Complementarity

The purpose of this section is to describe necessary and sufficient conditions for distributed supermodular complementarity (DSMC) to exist in equilibrium. The opposite of distributed supermodular complementarity is *unified supermodular complementarity*. With unified supermodular complementarity, complementary inputs are subject to unified governance, i.e. they are chosen by a single agent to optimize the value of the entire system. This section derives conditions under which DSMC can be sustained in equilibrium.

The equilibrium envisioned in this analysis is dynamic. In essence, it is a pattern of interaction that is sustained over many cycles of action, reward, and reaction by a group of rational actors. At the end of each cycle, the actors' beliefs about the system are confirmed, hence the pattern of interaction continues in the next round.[45]

Separable Supermodular Values

The first necessary condition for DSMC is that the joint value function $V(\cdot)$ must be divisible into *N* separable supermodular value functions, one for each autonomous actor. For example, if there are two inputs, *x* and *y*, there must exist two value functions, $V_X(x;y_0)$ and $V_Y(y;x_0)$, representing the value that can be captured by actor *X* and actor *Y* respectively. Each actor's own value is a function of his (her) own input variable and the "starting value" of the other input. Each actor views the other input as environmental parameter, not subject to his (her) control.

Inputs are costly and their costs are not included in the value functions.

As indicated, the individual value functions (excluding costs) are supermodular in the input variables.[46] The joint value function $V(x,y)$ is defined as the sum of the values realized by the individual actors:

$$V_X(x;y) + V_Y(y;x) = V(x,y) \quad . \tag{3}$$

Sums of supermodular functions are supermodular, thus the joint value function satisfies Equation (2) above.[47]

Separable value functions for complementary goods arise in modular technical systems when the user (or a systems integrator) assembles the system out of modular

[45] Masahiko Aoki (2000) calls these stable patterns of interaction "institutions." He analyzes them as repeated games with "self-confirming" beliefs.

[46] Thus the actors are engaged in a supermodular game. Formally, in a specific supermodular game, DSMC obtains when the equilibria of the game (including all costs) are also Pareto-efficient. In such cases, centralizing decision-making (collapsing the game) and distributing the joint output cannot make all players better off. On general supermodular games, see Topkis (1998) and Amir (2005).

[47] Milgrom and Roberts (1995).

components, none of which is unique and essential.[48] In a modular system, each module (or upgrade) is a distinct object that can be mixed and matched with other objects to make a system.[49] Users can thus ask and answer the question: how much am I willing to pay to add this object or upgrade to my present system? For example, given my present computer system, how much is this new application worth to me? How much is an upgrade of the computer worth to me? Given what is already in my kitchen, how much is a new appliance worth to me? Or, given my present factory layout, how much is an upgraded machine worth to me?

When the incremental value of one module or upgrade is positively affected by the presence of another module or upgrade, the modules and upgrades are supermodular complements. By estimating the willingness-to-pay of current users, independent producers of separate modules can construct demand functions for their modules in the usual fashion. $V_X(x; y_0)$ and $V_Y(y; x_0)$ are then the revenues each producer obtains by maximizing profit under their respective demand functions.[50]

However, as each module provider makes better modules, the users' systems will change. If the modules are supermodular complements, the value of a given *x* (or change in x) will increase, as *y* increases. The separate demand functions for each module will then shift outward as a result of improvements in the other modules.

Costs of Integration

A second necessary condition for DSMC is that the value obtained under unified goverance (first-best decision-making) exceeds the value obtained under distributed governance by less than the cost of integrating the separate enterprises.

Define $V_{FB}(x^*, y^*)$ as the "first best" value obtained if a single actor solves the joint maximization problem and chose optimal values (x^*, y^*) for the input variables. To make the comparison fair, I assume that the single actor faces the same cost of inputs as the separate actors, however, she does not have to make the same choices. For example, if the inputs are strong complements, and each actor receives his or her marginal value, separate actors would not supply any inputs. In contrast, if the sum of costs is below

48 If a particular module is unique and essential, then it is a strong complement to all other modules, and its value is no longer separable from the rest of the system. A basic platform may be unique and essential to the user, but upgrades of the platform are neither unique nor essential unless the platform sponsor withdraws support for earlier models.

49 Mixtures of input variables make up the sublattice on which the supermodular value function is defined. See Topkis (1998); Milgrom and Roberts (1990; 1995).

50 It is straightforward to show that, under most reasonable conditions, supermodular complements will exhibit traditional demand complementarity. That is, for continuous prices and quantities, the quantity demanded of *x*, Q_x will be decreasing in the price of y: $\partial Q_x / \partial p_y < 0$. Reducing the price of *y*, will push out the demand curve for *x* and vice versa. Also *x* and *y* are "strategic complements" as defined by Bulow, Geankopoulos and Klemperer (1985).

$V_{FB}(x^*, y^*)$, a single actor would supply both.

Defining *I* as the cost of integration, the second necessary condition can be written as:

$$V_{FB}(x^*, y^*) - I < V(x_0 + \Delta x, y_0 + \Delta y) \ . \tag{4}$$

In equation (4), $V_{FB}(x^*, y^*)$ is the value achievable under unified governance. $V(x_0 + \Delta x, y_0 + \Delta y)$ is the sum of the values achieved by independent actors *X* and *Y* under conditions of distributed governance.

If a unified actor has the same range of choice as the distributed actors, unified control is always better (or at least as good). At a minimum the single actor can simply choose the same input values as the separate actors.

However, a single actor does not always have the same range of choice as distributed actors. As we saw earlier in this chapter, separate enterprises may have advantages over unified enterprises in terms of the technologies they can implement, their access to human capital, and high-powered incentives. If two enterprises start out separate, there are also generally transaction costs that stand in the way of combining them. Finally enterprises making complementary modules may not be aware of the other's existence and would incur search costs trying to find each other. (In contrast, suppliers of strong complements cannot help but be aware of each other's existence.) Loss of opportunities, constraints on incentives, transaction costs of integration, and search costs are all subsumed in the "cost of integration" variable, *I*.

Under distributed governance, separate actors will make independent adjustments Δx and Δy in the input variables *x* and *y*. Because their inputs are supermodular complements, the joint value they create will exceed the sum of values each would create if the other failed to act. The joint value under distributed governance is Pareto-efficient if it exceeds the value of unified governance net of the costs of integration. In this case, it is impossible to make all actors better off by moving from separate to unified control.

Given Equation (4), if the actors start out in separate firms, DSMC will survive in equilibrium even though unified governance exists as a possibility (i.e., the actors could elect to combine their enterprises). *An ecosystem of complementary suppliers is then the predicted organizational outcome.*[51]

Equation (4) sheds light on why the autonomy of actors is "taken for granted" on exchange platforms. In these cases, the exchange itself is designed to have very low transaction costs. *A priori*, the transactors can set up modular task networks whose boundaries mirror the scope of their particular technical recipes. Through the exchange platform they can then transfer goods and provide compensation at low cost. The

[51] The outcome may be path dependent, i.e., if the actors start out under unified control, there may costs of breaking the organization apart.

exchange platform generally has features that make it easy to find good matches; define, count and pay for goods; and build trust between counterparties.

When search costs, mundane transaction costs, and opportunistic transaction costs are low, the difference between values obtained under distributed and unified governance will be low. At the same time, specialized knowledge about production processes (the smith knows how to make a pot hook; the cook does not) and high-powered incentives (the smith gets paid by the piece) can make unified governance uneconomic. The head cook does not want to run the smithy, she simply wants a pot hook and is willing to pay make it worth the smiths' effort!

Cost Alignment

The last condition needed to support DSMC is the alignment of each actor's separate costs with their separate rewards. The statement of the condition is slightly different depending on whether we assume the actors move simultaneously or sequentially and whether side payments are allowed. Each assumption is reasonable in some settings, thus I will examine each in turn.

Sequential Action

Let us assume the current state of the modules is x_0 and y_0, and that all relevant users have already have these modules in their systems. Actor *X* contemplates an upgrade Δx to the module she supplies; and Actor *Y* simultaneously contemplates an upgrade Δy to his module. The upgrades will be available separately from each provider.

The investment cost of the upgrades is $c_{\Delta x}$ and $c_{\Delta y}$ and will be borne by the respective module makers. For simplicity, I assume that the upgrades and their costs are discrete, i.e., the module makers face take-it-or-leave it investment decisions. (This is consistent with the fact that investments in new technologies are generally discrete projects and thus not infinitely divisible.)

The timing of events is as follows: at *t=0*, each module maker decides whether to invest in an upgrade or not, based on his or her marginal returns. At *t=1*, the module upgrades are revealed and users' willingness to pay changes to reflect the new state of technology. Because of their supermodular complementarity, the demand curves for both modules shift outward. Each actor chooses an optimal price and quantity based on the new demand conditions. (I assume neither is capacity constrained.)

For distributed supermodular complementarity to hold, each module maker must invest in his or her upgrade based on the state of users' systems at *t=0* and his or her own cost:

$$V_X(x_0 + \Delta x; y_0) - V_X(x_0; y_0) > c_{\Delta x} \quad ; \tag{5a}$$

$$V_Y(y_0 + \Delta y; x_0) - V_Y(y_0; x_0) > c_{\Delta y} \quad . \tag{5b}$$

Equations (5a) and (5b) together with equations (3) and (4) are sufficient conditions for simultaneous distributed supermodular complementarity. Each module maker will invest on his own account, *and then receive a positive surprise due to the other's investment*. It is as if "the environment" became more favorable post-investment.

However, under this scenario, the positive surprises are surprises, thus the module makers will not include the positive impact of the other's investment in their calculations. Also, if there are many modules in the system, some investments may clear the cost hurdles, while others do not. The latter group will not invest, even though their investments might be profitable after the fact. For example, if

$$V_Y(y_0+\Delta y; x_0+\Delta x) - V_Y(y_0; x_0) > c_{\Delta y} > V_Y(y_0+\Delta y; x_0) - V_Y(y_0; x_0) \qquad (6)$$

then *Y* will not invest *ex ante*, even though *ex post* the investment would have been profitable.

Sequential action

The problem of underinvestment can be partially addressed by staging the module makers' investments. To see this, let us assume that investments proceed in rounds. In the first round, investment in module *X* clears hurdle (5a) but investment in module *Y* does not clear hurdle (5b). Thus Actor *X* invests, but Actor *Y* does not.

In the second round, however, the rewards to investment in *Y* will increase because of the property of increasing differences:

$$V_Y(y_0+\Delta y; x_0+\Delta x) - V_Y(y_0; x_0+\Delta x) > V_Y(y_0+\Delta y; x_0) - V_Y(y_0; x_0)$$

As shown in equation (6), the investment in *Y* may now clear the cost hurdle $c_{\Delta y}$.

Thus, sufficient conditions for sequential distributed supermodular complementarity are:

$$V_X(x_0+\Delta x; y_0) - V_X(x_0; y_0) > c_{\Delta x} \quad ; \qquad (7a)$$

$$V_Y(y_0+\Delta y; x_0+\Delta x) - V_Y(y_0; x_0+\Delta x) > c_{\Delta y} \quad . \qquad (7b)$$

Equation (7b) is a looser constraint than Equation (5b). Value-enhancing investments that do not take place simultaneously may occur sequentially.

Sequential distributed supermodular complementarity may lead each actor to undertake a series of investments. Investments in new technology are not infinitely divisible, but they can often be broken up and implemented in stages. Actor *X* may find it worthwhile to invest in Δx_1 at *t=0*, before Actor *Y* makes his move. But the next increment, Δx_2, may not be worthwhile because of diminishing marginal returns. (See

Figure 5-3.) However, *after Y moves*, the marginal returns to Δx_2 will be higher, and that investment may clear its cost hurdle. Figure 5-3 depicts this alternating investment pattern.

Actors in both the simultaneous and sequential scenarios do not have to coordinate their actions. They do not even have to be aware of the others' existence. Each actor only needs to be able to assess *the marginal value of his or her own investment* given the state of the world at the time he or she makes the investment.

Thus distributed supermodular complementarity, whether simultaneous or sequential does not create the same transactional hazards or need for close coordination as strong complementarity. Supermodular complementarity between two or more input variables does not *in itself* create a need for unified governance.

In summary, sufficient conditions for distributed supermodular complementarity (DSMC) to hold in equilibrium are as follows:

- The actors' value functions must be separable;
- There must be some costs of integration;
- The actors' costs must be aligned with the value he or she can capture.

Modular components and upgrades that are sold separately and can be mixed and matched perforce have separable value functions. If the other conditions also hold, actors in different firms can obtain the benefits of complementarity while independently pursuing their own self-interest. This means that a *cluster* of separately owned, independently governed firms can create a functional and evolving technical system. Each can pursue the development of one or more modules and each will benefit from investments by others.

However, even with sequential investments, distributed supermodular complementarity does not necessarily cause *all* jointly attractive investments to be undertaken. For example, suppose part of the value created by an investment in Δx is directly captured by *Y*, but the investment cost is entirely borne by *X* and exceeds the value *X* can capture:

$$V_Y(y_0;x_0+\Delta x)-V_Y(y_0;x_0)>0 \ ;$$

$$V_X(x_0+\Delta x;y_0)-V_X(x_0;y_0)<c_{\Delta x} \ .$$

In this hypothetical case, *X* will not invest, even if the total value captured by *X* and *Y* exceeds $c_{\Delta x}$. This pattern can arise in systems with direct network effects such as social networks. When a new person joins the network, she gains the direct benefits of linking to others, but the others also gain the benefits of linking to her. The benefits gained by others may exceed the benefits to the new member by a large margin. If all costs are borne by the new member, she may decide not to join, even though the value for all exceeds her cost.

The issue in this case is not with supermodular complementarity per se, but with any form of externality. If one party can help (or do harm) to another, then individuals acting independently will not necessarily arrive at a jointly optimal outcome.

Side Payments

Supermodular complementarity is a specific type of externality. Given externalities, in a world with two actors, each individual's action (Δx or Δy) potentially has two effects: one on V_X and one on V_Y. For example, from a given starting point (x_0, y_0), the total impact of a change Δx on the entire system is, by definition:

$$V(x_0 + \Delta x, y_0) \equiv V_X(x_0 + \Delta x, y_0) + V_Y(x_0 + \Delta x, y_0) \quad .$$

If V_Y in this expression is positive, and yet the entire cost $c_{\Delta x}$ is borne by Actor *X*, then there may be underinvestment in Δx. Symmetrically, if V_Y is negative, there may be overinvestment.

We are now in the familiar world of the Coase Theorem, wherein a side payment from Actor *Y* to Actor *X* can achieve the first-best joint outcome.[52] Side payments involve a stronger form of coordination than independent pursuit of value. At a minimum, *X* and *Y* must know of each other's existence, and have some conception of the other's individual value function in order to structure the exchange between them. Side payments are also transactions and thus subject to transaction costs, both mundane and opportunistic.[53]

This argument can be generalized from two to *N* actors in a straightforward fashion. As Ronald Coase famously demonstrated, given externalities, there is always a system of side payments that can achieve a Pareto-efficient outcome.

The side payments will be subject to transaction costs, which can be non-trivial and even prohibitive. Notwithstanding this fact, if the complements are *modules* in the technical system, the side payments will take place across modules. Module boundaries are thin crossing points in the task network, *where transaction costs are generally low*.

In practice this means that the complementors of a given module do not need to know specifically how a module is made, but only what it does. To encourage a change Δx in the design of Module *X*, Actor *Y* does not need to know in great detail *how* Δx will be accomplished but only *what* Δx is worth to him. Because *X* is a module with

---

[52] Coase (1960).

[53] One can think of any purchase of a good as a side payment from its user to its provider: Let Δx be the act of making a pothook (see Chapter 2). The value of the pothook to its maker, $V_X(x_0 + \Delta x, y_0)$, is virtually zero, while its value to a cook, $V_Y(x_0 + \Delta x, y_0)$, may be considerable. The price paid by the cook to the smith must exceed the smith's cost, or the smith would not make it. And it must be less than the cook's value, otherwise the cook would not buy it.

measurable performance (by assumption), it may be possible for Actor *Y* to contract with Actor *X* to supply Δx in return for compensation. Of course, the more Actor *Y* knows about *X's* internal processes, the more favorable will be the terms of its contract. This is why platform sponsors and systems integrators, who must manage a distributed ecosystem of complementors, generally "know more than they make."[54]

In summary, I have shown that it is possible for independent actors to achieve complementary outcomes without placing the input variables under unified governance. I have labeled this state of affairs *distributed supermodular complementarity* (DSMC). The first necessary condition for DSMC is value separability, that is, the existence of individual supermodular value functions representing the value to a particular actor of a change in his or her input variable, holding the other variables constant. The second necessary condition is that there are costs of integration that diminish the value of "first-best" decisions.

Beyond these necessary conditions, one of the following sets of conditions must hold: (1) Equations (5a) and (5b) are sufficient for simultaneous DSMC; (2) Equations (7a) and (7b) are sufficient for sequential DSMC. If either set of conditions holds, the individual actors can achieve coordinated action without communication and even without knowledge of the others' existence. Complementarity in such cases will appear in the form of positive surprises in the market and/or a positive trend in buyers' willingness to pay.

Side payments (transactions) between actors can be used to make either set of equations true. Side payments are transactions, therefore subject to transaction costs, both mundane and opportunistic. However, *side payments between module providers occur at thin crossing points in the task network where both types of transaction costs are low*. Therefore, even when side payments are needed, they can take the form of a goods purchase or a service contract between independent agents.

Strong complements do not have separable values, by definition, since all are needed to achieve any value at all. Thus strong complementarity does not satisfy Equation (3), the first necessary condition for distributed supermodular complementarity. As both transaction cost economists and property rights theorists have observed, strong complementarity requires unified governance of assets and activities to guarantee the presence of all necessary inputs.

5.12 Conclusion: How Technology Shapes Organizations

Complementary goods have more value when used together than separately. Complementarity may be strong or weak. Strong complements are *specific* and *unique* goods that have no value (or greatly diminished value) unless all are present in use. In the

[54] Brusoni, Prencipe and Pavitt (2001).

task network, dense technical interdependencies create strong complementarity, but it can arise for other reasons as well.

Transaction cost economics and property rights theory advise that strong complements should be placed under unified governance, for example, through common ownership. Agency theory suggests that weak complementarity can be handled via arms-length transactions and contracts. Furthermore, strong or weak complementarity are not innate properties of tasks and assets, but can be the result of choices regarding task networks, incentives and job design.

Supermodular complementarity exists when more of one input makes more of another input more valuable. *Distributed supermodular complementarity (DSMC)* exists when two or more independent actors can create complementary value by pursuing their own interests, and will not find it advantageous to combine in order to coordinate their actions. I derive formal conditions under which DSMC holds as a consistent pattern in a dynamic equilibrium. Given DSMC, clusters of firms making different complementary goods, including open platforms with surrounding ecosystems, can survive and compete effectively against integrated firms that control all complementary inputs.

References

Alchian, Armen A. and Harold Demsetz (1972) “Production, Information Costs, and Economic Organization,” *American Economic Review* 52:777-795.

Amir, R., 2005. Supermodularity and complementarity in economics: An elementary survey. *Southern Economic Journal*, 71(3):636-660.

Aoki, Masahiko (2001). *Towards a Comparative Institutional Analysis*, Cambridge, MA: MIT Press.

Argyres, N.S. and Zenger, T.R., 2012. Capabilities, transaction costs, and firm boundaries. *Organization Science*, *23*(6), pp.1643-1657.

Baker George, Robert Gibbons and Kevin J. Murphy (2002) “Relational Contracts and the Theory of the Firm,” *Quarterly Journal of Economics* 117:39-84.

Baldwin, Carliss Y., and C. Jason Woodard (2011) "The Architecture of Platforms: A Unified View." In *Platforms, Markets and Innovation*, Annabelle Gawer, ed. (London: Edward Elgar).

Brusoni, Stefano, Andrea Prencipe and Keith Pavitt (2001) "Knowledge Specialization, Organizational Coupling and the Boundaries of the Firm: Why Do Firms Know More Than They Make?" *Administrative Science Quarterly*, 46(4):597-621.

Bulow, J.I., Geanakoplos, J.D. and Klemperer, P.D., 1985. Multimarket oligopoly: Strategic substitutes and complements. *Journal of Political economy*, *93*(3), pp.488-511.

Church, J., & Gandal, N. (1992). Network effects, software provision, and standardization. *The Journal of Industrial Economics*, 85-103.

Church, J., & Gandal, N. (1993). Complementary network externalities and technological adoption. *International Journal of Industrial Organization*, 11(2), 239-260.

Coase, Ronald H. (1937) “The Nature of the Firm,” *Economica*, 4(4):386-405, reprinted in *The Firm, the Market, and the Law*, University of Chicago Press, Chicago, IL, 33-55.

Coase, Ronald H. (1960) “The Problem of Social Cost,” *Journal of Law and Economics* 3:1-44, reprinted in *The Firm, the Market, and the Law*, University of Chicago Press, Chicago, IL, 95-156.

Edgeworth, F. Y., 1881, *Mathematical Psychics*, London: C. Kegan Paul.

Feldman, E.R., Gartenberg, C. and Wulf, J., 2018. *Pay inequality and corporate divestiture*s. Strategic Management Journal.

Freeland, R.F., and Zuckerman, E., 2018. The Problems and Promise of Hierarchy: Voice Rights and the Firm. *Sociological Science*, *5*, pp.143-181.

Gawer, Annabelle and Michael A. Cusumano (2002) *Platform Leadership: How Intel, Microsoft and Cisco Drive Industry Innovation*, Harvard Business School Press, Boston, MA.

Gibbons, R., 2005. Four formal (izable) theories of the firm, *Journal of Economic Behavior & Organization*, *58*(2), pp.200-245.

Hart, Oliver and John Moore (1990) "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98(6):1119-1158.

Heller, M.A. (1998) "The Tragedy of the Anticommons: Property in the Transition from Marx to Markets" *Harvard Law Review* 111(3): 621-688.

Heller, M.A. and R.S. Eisenberg (1998) "Can Patents Deter Innovation? The Anticommons in Biomedical Research," *Science* 280: 698-701.

Holmstrom, B. and Milgrom, P., 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics and Organization*, 7: 24-52.

Holmstrom, B. and Roberts, J., 1998. The boundaries of the firm revisited. *Journal of Economic perspectives*, *12*(4), pp.73-94.

Jacobides, M.G., Cennamo, C. and Gawer, A., 2018. Towards a theory of ecosystems. *Strategic Management Journal*. 39:2255-2276.

Jensen, Michael C. and William H. Meckling (1976) "Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure," *Journal of Financial Economics*, 3(4):305-360.

Kerr, Steven (1975) "On the Folly of Rewarding A, While Hoping for B," *Academy of Management Journal*, 18:769-783.

Milgrom, Paul and John Roberts (1990) "The Economics of Manufacturing: Technology, Strategy and Organization," *American Economic Review* 80 (3): 511-28.

Milgrom, P., and Roberts, J. (1995). Complementarities and fit strategy, structure, and organizational change in manufacturing. *Journal of accounting and economics*, 19(2), 179-208.

Parker, G.G., Van Alstyne, M.W. and Choudary, S.P. 2016. *Platform revolution: How networked markets are transforming the economy--and how to make them work for you.* WW Norton & Company.

Puranam, P. (forthcoming) *The Microstructure of Organizations*, Oxford University Press.

Puranam, P. and Vanneste, B., 2016. *Corporate strategy: Tools for analysis and decision-making*. Cambridge University Press.

Rochet, J-C and J. Tirole (2003) "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association*, 1(4):990-1029.

Samuelson, P.A., 1974. Complementarity: An essay on the 40th anniversary of the Hicks-Allen revolution in demand theory. *Journal of Economic literature*, *12*(4), pp.1255-1289.

Teece, David J. (1986)."Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy," *Research Policy*, 15(6): 285-

305.

Topkis, Donald M. (1998) *Supermodularity and Complementarity*, Princeton University Press, Princeton, NJ.

Williamson, Oliver E. (1985). *The Economic Institutions of Capitalism*, New York, NY: Free Press.