Coordinating Supply Chains

Robert Akerlof and Richard Holden*

December 12, 2023

Abstract

Many goods have complex supply chains and a variety of specialized production inputs. Notable examples include airplanes, lasers, and photolithography machines. Manufacturers of such goods often face the problem of coordinating suppliers (i.e. convincing them to produce specialized inputs). We build a model to analyze this problem. A manufacturer endowed with capital needs a specialized input from each of n suppliers. The manufacturer can pay a markup to overcome a supplier's reluctance to produce. Alternatively, the supplier's reluctance can be overcome by integrating them; but integration inflates costs when there is a lack of congruence between the manufacturer and supplier. We derive sharp predictions about firm structure and apply our model to a number of applications including international trade and industry policy.

Keywords: Coordination, Supply Chains, Theory of the Firm.

*Akerlof: University of Warwick and CEPR, r.akerlof@warwick.ac.uk. Holden: UNSW Business School, richard.holden@unsw.edu.au. We are grateful to Steve Callander, Oliver Hart, Fuhito Kojima, Simon Loertscher and Hongyi Li for helpful discussions, and also to seminar participants at Keio University, University of Melbourne, and the Australasian Meetings of the Econometric Society.

1 Introduction

Many goods have complex supply chains and a variety of specialized production inputs. A crucial yet often overlooked problem of producing such goods is coordinating suppliers—that is, convincing them to produce the needed inputs.

To get a sense of the importance of coordination in supply chains, consider the following example from the semiconductor industry. Dutch company ASML is the world's leading manufacturer of photolithography machines and the sole manufacturer of extreme ultraviolet (EUV) lithography machines, which are instrumental to producing the most advanced chips. Initially, lithography machines utilized visible light with a wavelength of several hundred nanometers, limiting their ability to carve small transistors. To overcome this limitation, manufacturers turned to ultraviolet light with wavelengths as small as 193nm, and eventually to extreme ultraviolet light with a wavelength of 13.5nm.[1]

ASML's EUV machines require a vast array of specialized parts. In all, ASML uses more than 4,700 individual suppliers. As Miller (2022) notes, one "can't simply buy an EUV lightbulb." Producing EUV light at scale is incredibly complex. ASML identified Cymer, a San Diego based company, as a potential supplier of EUV light and decided to acquire Cymer in a deal announced in October 2012.[2] Cymer engineers devised a method for generating EUV light involving the propulsion of a 30nm-wide ball of tin through a vacuum at a speed of 200 miles per hour. The tin is then subjected to two laser blasts, the first to warm it and the second to transform it into a plasma with a temperature 40-50 times hotter than the sun's surface. This entire procedure is repeated an astounding 50,000 times

[1] The interested reader may wonder how 7nm, 5nm, and even 3nm transistors can be produced with EUV light having a wavelength of 13.5nm. In 1873 Ernst Abbe showed that the smallest resolvable distance between two features of a sample–the *optical resolution* of a microscope–is proportional to the wavelength of light used to illuminate the sample and inversely proportional to the refractive index of the medium between the lens and the sample. Denoting the wavelength as λ and NA as the *numerical aperture* given by the sine of the half angle multiplied by the refractive index of the medium, Abbe showed that $d = \lambda/2NA$. This principle applies to any light-based projection system, including semiconductor photolithography.

[2] https://www.asml.com/en/news/press-releases/2013/asml-completes-acquisition-of-cymer.

per second.

The lasers required for this process did not exist at the time. Cymer and ASML approached German precision tooling company Trumpf to develop the necessary laser. One key challenge was preventing the tin droplets' light from reflecting back into the laser. Each laser consists of 457,329 individual components, including ultra-pure diamonds—developed in collaboration with partners—which serve as "windows" for the laser beam to exit the chamber.

Another hurdle was directing the EUV light towards a chip. Due to its short wavelength, EUV light behaves more like an X-ray than visible light, with many materials absorbing it rather than reflecting it. To surmount this challenge, ASML turned to German lens manufacturer Zeiss. Zeiss engineered mirrors consisting of one hundred alternating layers, each measuring a mere 2nm in thickness, composed of molybdenum and silicon. According to Miller (2022), "building such a mirror with nanoscale precision proved almost impossible. Ultimately, Zeiss created mirrors that were the smoothest objects ever made."

There are a number of salient features of the ASML example. First, many suppliers needed to develop entirely new products. Second, multiple components were essential. Trumpf's laser and Cymer's light generation method would have been futile without Zeiss's groundbreaking mirrors. Third, ASML acquired Cymer but not Trumpf or Zeiss.[3] Across its supply chain, ASML is integrated with about 15% of its suppliers but contracts with the remainder (Miller (2022)). Fourth, by being the first to successfully coordinate these suppliers and build the requisite supply chain, ASML has—at least for now—a complete monopoly on EUV machines. In fact, ASML's stock price has risen from \$154 in January 2019 to \$720 toward the end of 2023, or more than a quadrupling.

We analyze a model of the supply-chain coordination problem. There is a manufacturer (M), endowed with capital, who needs an input from each of n suppliers to produce a final good. M makes a take-it-leave it offer to each supplier consisting of a price for the input and a fraction of M's capital that is payable if the supplier

[3] Our model provides a rationale for why ASML bought Cymer but not Trumpf or Zeiss. See Section 3.2.3 for a discussion.

produces and M refuses to buy. Suppliers decide whether to accept this contract and produce. Finally M decides whether to buy the inputs produced or pay the penalty for not purchasing. If M buys all of the inputs, M produces the final good.

We characterize the set of equilibria of this game using *introspective equilibrium* as our solution concept (Kets and Sandroni (2021)). Introspective equilibrium is based upon level-k thinking. Each player has an exogenously-given "impulse" which determines how they react at level 0; at level $k > 0$, players react according to the belief that other players are acting at level $k - 1$. Introspective equilibrium is defined as the limit of this process as $k \to \infty$.

We show that the manufacturer uses their capital to insure the suppliers who are most reticent to produce. These suppliers are paid marginal cost, while the remaining suppliers receive a markup. We offer an explicit expression for the markup as a function of marginal costs and the suppliers' impulses. If the size of the aggregate markup exceeds the surplus generated from final good production, coordination is not possible.

We next consider M's decision to make input i in-house or buy it from an external supplier. We use an approach in the spirit of Aghion and Tirole (1997) where M and supplier i decide ex ante on the allocation of formal authority. If M has formal authority, we describe them as "integrated." M faces the following basic tradeoff: under non-integration, M must pay a markup to induce the supplier to produce; under integration, M has the authority to produce so there is no need to pay a markup, but the cost of production is greater due to a lack of congruence of preferences between M and the supplier. In the general case with n suppliers, we show that supplier i is more likely to be integrated when they have a lower impulse to produce or when their preferences are more congruent with M's.

An important point of comparison to our paper is Property Rights Theory (Grossman and Hart (1986), Hart and Moore (1990)). One contrast with PRT is that hold-up problems do not arise in our model. Additionally, our model makes two starkly different predictions. First, since investments are made by suppliers, PRT predicts that suppliers should own manufacturers.[4] In our model, it is the

[4]If suppliers own manufacturers, this protects their investments from hold up.

other way around: it is only ever optimal for manufacturers to own suppliers (or for there to be non-integration). Second, integration is only optimal in our model with two or more suppliers since, with one supplier, there is no coordination problem.

Scholars of international trade have focused on "global value chains" (GVCs) since the 1980s. This literature concerns the size of trade flows within GVCs and how different countries and industries are positioned within GVCs. For an excellent survey, see Antràs and Chor (2022). Within this literature, it is now customary to distinguish two broad types of value chains: "snakes" and "spiders."[5] Snakes involve a sequential production process where, at each stage along the value chain, a supplier produces an input that is used by a supplier at the next stage. This process proceeds stage-by-stage until a manufacturer produces a final good. Models of snakes that speak to the organization of the production process typically take a PRT approach, following the earlier trade literature beginning with Antràs (2003).

In a spider, by contrast, each supplier's input flows directly to the final-good manufacturer. As Gereffi (1999) puts it: "producer-driven commodity chains [(i.e. spiders)] are those in which large, usually transnational, manufacturers play the central roles in coordinating production networks...This is characteristic of...industries such as automobiles, aircraft, computers, semiconductors and heavy machinery." For instance, an aircraft consists of many components (fuselage, wings, engine, and so on) that are assembled by a manufacturer such as Boeing or Airbus.

Our paper focuses squarely on spiders as we are principally concerned with the coordination-aspect of supply chains. We do not address sequential production processes, which are undoubtedly important in many industries, and we do not model the final goods market and thus cannot speak to issues driven by elasticities, which is the focus of much of the international trade literature. On the other hand, we put coordination of suppliers at the heart of the analysis. In contrast to PRT, we take a complete contracts approach. This is consistent with our motivating example, and our results on supplier integration are consistent with

[5] Baldwin and Venables (2013).

the empirical evidence on spiders (see, for instance, Gereffi (1999)).

To get a more concrete sense of how incomplete-contracting models of snakes work, consider the pioneering work of Antràs and Chor (2013). They build on the Property Rights Theory (PRT) and follow a now venerable tradition of applying PRT to international trade, which includes Antràs (2003), Antràs (2005), Antràs and Helpman (2004), Acemoglu et al. (2007), and Antràs and Helpman (2008).

Antràs and Chor (2013) analyze a model in which final goods production requires a continuum of production stages, each performed by a different supplier. Each supplier makes a relationship-specific investment which affects compatibility with other components, but contracts cannot be written on compatibility. Importantly, there is a specific technologically-determined sequence of production stages.

The authors show that if final-good demand is sufficiently elastic, early stages of the production process are not integrated while later stages are integrated. Conversely, when demand in sufficiently inelastic, early stages are integrated while later stages are not. The authors find empirical support for this striking result using data on the intrafirm share of total imports in the US.[6]

Another strand of the literature on value chains concerns spiders. A complete-contracting model of spiders is offered by Antras et al. (2017) and an illuminating incomplete contracting approach is offered by Chor and Ma (2021). Antras et al. (2017) interpret international trade flows as the "legs" of a spider. Firms that assemble inputs from the legs decide from which countries they will source inputs. A key ingredient in this framework is that countries differ in their ability to reduce marginal costs for the assembling firm and also in the fixed costs associated with importing. This implies that—in contrast to standard models of exports—whether it is optimal to import from one country depends on the other countries from which a firm is sourcing inputs. That is, there is interdependence across suppliers. In fact, depending on elasticities of demand in the final goods market,

[6] Alfaro et al. (2019) consider an environment where choices of organizational form have spillovers along the value chain. In particular, upstream relationship-specific investments affect the incentives to make relationship-specific investments downstream. The authors show that the central results of Antràs and Chor (2013) go through in this richer setting.

different input-source countries can be complements or substitutes. The authors show how, despite this complexity, such a model can be structurally estimated. They show how a change in the benefits of sourcing from a country like China affects the sourcing decisions of US firms generally and how optimal sourcing varies depending on firm productivity and the other countries from which a firm initially sources.

Chor and Ma (2021) analyze a state-of-the-art multi-country, multi-industry model of sourcing inputs. Final good producers source a continuum of inputs from different suppliers. Both suppliers and the headquarters make relationship-specific investments, leading to the possibility of bilateral holdup. The key contribution of the paper is that, for each input variety, the firm chooses two things: from which country to source the input and whether to vertically integrate with the supplier. In this environment, integration can be the optimal organizational form because of hold-up between firms and suppliers. These integration decisions can be made on a dyad-by-dyad basis since there are no relevant interactions among suppliers that affect integration decisions. The authors show how these firm-level decisions can be aggregated into a gravity equation by industry and organizational form. This has direct implications for the share of intra-firm trade. The authors also provide an expression for the welfare gains from trade.

Finally, a number of papers have emphasized the importance of coordination in investment decisions. Among these, Akerlof and Holden (2016) and Akerlof and Holden (2019) focus on the role of coordination in investment decisions, as does the literature on "catalytic finance" such as Corsetti et al. (2006) and Morris and Shin (2006).

The remainder of the paper proceeds as follows. Section 2 states the model of supply-chain coordination and characterizes its equilibrium. Section 3 considers the manufacturer's make-or-buy decision and analyzes which suppliers will optimally be integrated. Section 4 considers some applications. Section 5 concludes by putting the "coordination-based theory of the firm" analyzed here into context. Proofs not offered in the main text are contained in the appendix.

2 The Model

A manufacturer of widgets (M) is endowed with capital k. Selling widgets generates revenue R. There are n suppliers of production inputs. To make widgets, M needs an input from each supplier. M has no production costs aside from the cost of obtaining inputs. It costs supplier i (S_i for short) c_i to produce the input needed by M. The surplus from making widgets is strictly positive: $S \equiv R - \sum_i c_i > 0$.

The contracting environment is one where M makes take-it-or-leave-it offers to suppliers (and thus has all of the bargaining power). A contract between M and S_i takes the form of a price p_i that M offers to pay for input i, plus a claim over a fraction θ_i of M's capital if S_i produces the input and M fails to pay (with $\sum_i \theta_i \leq 1$). We assume that the suppliers have limited liability, so in the event they are unable to deliver inputs, it is not possible to penalize them.[7]

The timing of events is as follows:

1. M makes take-it-or-leave-it offers to each supplier i of the form (p_i, θ_i).
2. Suppliers simultaneously decide whether to accept the offers and produce inputs.
3. For each supplier i who produces, M decides whether to buy their input or pay the penalty $\theta_i k$. If M buys all inputs, M produces widgets and generates revenue R. Otherwise, M produces no widgets and generates zero revenue.[8]

2.1 Analysis

For ease of exposition, we will assume that suppliers accept M's offer when otherwise indifferent, and both M and the suppliers produce when otherwise indif-

[7] If liability was not limited, the problem of coordinating suppliers could be solved by offering contracts where suppliers receive a penalty for failing to produce if and only if all other suppliers sign a contract.

[8] We assume that M contracts with suppliers before suppliers decide whether to produce. Notice that if M contracted with suppliers after they produce inputs—rather than before—suppliers would receive zero for their inputs since M has all of the bargaining power. Suppliers would, consequently, choose not to produce in the first place.

ferent.

We can analyze the game by backward induction. At time 3, there are two possibilities: either all suppliers produced inputs, in which case M can produce widgets, or M is unable to produce widgets. In the former case, we can assume that M will pay p_i to each supplier and produce widgets (if not, M would have offered different contracts at time 1). M's payoff is $R - \sum_i p_i$ and S_i's payoff is $p_i - c_i$. In the event that M cannot produce widgets, M obtains zero revenue and pays each supplier i that produced either p_i or $\theta_i k$ (whichever is smaller). Notice that the penalty $\theta_i k$ is only used if $\theta_i k \leq p_i$. It is weakly preferable to choose θ_i ex ante such that $\theta_i k \leq p_i$, since θ_i can always be set such that $\theta_i k = p_i$. Thus, we will assume that $\theta_i k \leq p_i$ and the penalty is paid rather than p_i.

Now, consider time 2. Suppliers are weakly better off if they accept M's offer since, if they accept, they can always choose not to produce (there is no penalty for failing to do so). Thus, suppliers will always accept M's offer. Let a_i denote S_i's decision whether to produce. Based on our analysis of time 3, we conclude that S_i's payoff is:

$$\pi_{S_i}(a_1, ...a_n) = \begin{cases} 0, & \text{if } a_i = 0 \\ p_i - c_i, & \text{if } a_1 = a_2 = ... = a_n = 1 \\ \theta_i k - c_i, & \text{otherwise} \end{cases}$$

Supplier i's payoff is clearly zero if they do not produce ($a_i = 0$). If all suppliers produce inputs, M produces widgets and pays p_i to supplier i. Consequently, S_i's payoff is $p_i - c_i$. If supplier i produces but some other supplier does not, M is unable to produce widgets and hence cannot pay p_i to supplier i. Consequently, supplier i receives $\theta_i k - c_i$ instead.

In order to get the suppliers to produce, M must set $p_i \geq c_i$ for all i, so let us assume that this is what M does. First, suppose $k \geq \sum_i c_i - c_{\max}$, where $c_{\max}$ denotes the maximum of the c_i's. M can set $\theta_i \geq \frac{c_i}{k}$ for $n-1$ of the suppliers, giving them full insurance. As a result, these $n-1$ suppliers will produce for sure. Moreover, since $n-1$ suppliers produce for sure, the nth supplier is also effectively insured and so will produce for sure as well. Notice that M can set $p_i = c_i$, which gives M all of the surplus ($R - \sum_i c_i$).

If instead $k < \sum_i c_i - c_{\max}$, there will be at least two suppliers who are not fully insured. Consequently, the suppliers face a coordination game with multiple Nash equilibria: there is a Nash equilibrium in which all suppliers produce ($a_1 = a_2 = ... = a_n = 1$) and a Nash equilibrium in which at least some suppliers do not produce. Going forward, we will focus on this case.

Assumption 1. *We assume that* $k < \sum_i c_i - c_{\max}$.

In order to refine the set of time-2 equilibria, we will apply an equilibrium concept developed by Kets and Sandroni (2021) called introspective equilibrium. Introspective equilibrium is based upon level-k thinking (see Crawford et al. (2013) for a review). Each player has an exogenously-given "impulse" which determines how they react at level 0; at level $k > 0$, players react according to the belief that other players are acting at level $k - 1$. Introspective equilibrium is defined as the limit of this process as $k \to \infty$.

As we will see, the introspective equilibrium depends critically upon players' impulses, or level-0 thinking. We make the following assumption regarding suppliers' impulses.

Assumption 2. *Let* $x_i^0 \in \{0, 1\}$ *denote the impulse of supplier* i*:* $x_i^0 = 1$ *means that supplier* i*'s impulse is to produce and* $x_i^0 = 0$ *means that supplier* i*'s impulse is to not produce. We assume that* $x_i^0 = 1$ *with probability* γ_i*. The* x_i^0*'s are mutually independent.* x_i^0 *is private information of supplier* i*.*

The parameter γ_i captures supplier i's perceived willingness to produce. This willingness could arise from a number of sources. These include, but are not limited to: a past relationship between supplier and manufacturer (e.g. past sales), cultural affinity between the supplier and manufacturer, or behavioral characteristics of the supplier.[9]

[9] One question is whether it is possible, in practice, to measure γ_i. While it is presumably difficult to exactly measure γ_i, it may be straightforward to identify factors (such as cultural affinity or past sales) that cause γ_i to be higher or lower.

We formally define introspective equilibrium in our context as follows.

Definition 1 (Introspective Equilibrium for Supply Chains)**.** *An introspective equilibrium $(a_1^*, a_2^*, ..., a_n^*)$ is constructed as follows:*

1. *Supplier i's choice at level $k > 0$, denoted x_i^k, is obtained by letting each supplier best respond to the belief that other suppliers are at level $k-1$:*

$$x_i^k = \begin{cases} 1, & \text{if } \mathbb{E}_i[(\Pi_{j \neq i} x_j^{k-1})(p_i - c_i) + (1 - \Pi_{j \neq i} x_j^{k-1})(\theta_i k - c_i)] \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

2. *An introspective equilibrium is the limit as $k \to \infty$:*

$$a_i^* = \lim_{k \to \infty} x_i^k.$$

To better understand equation 1, observe that at level k, if supplier i produces ($a_i = 1$), they believe that M will obtain all inputs if $x_j^{k-1} = 1$ for all $j \neq i$. Thus, if $a_i = 1$, supplier i expects to receive $p_i - c_i$ if $\Pi_{j \neq i} x_j^{k-1} = 1$ and $\theta_i k - c_i$ otherwise.

The following lemma elucidates when M is able to obtain inputs from all of the suppliers.

Lemma 1. *An introspective equilibrium exists in which all suppliers produce if and only if condition 2 holds for all suppliers:*

$$p_i \Gamma_{-i} + \theta_i k (1 - \Gamma_{-i}) \geq c_i, \tag{2}$$

where $\Gamma_{-i} = \Pi_{j \neq i} \gamma_j$. Moreover, the introspective equilibrium is unique.

Proof. First, consider supplier i's level-1 choice (x_i^1). At level 1, supplier i assumes that other suppliers are at level 0. The expected return to supplier i from producing is $p_i \Gamma_{-i} + \theta_i k (1 - \Gamma_{-i})$ (since Γ_{-i} is the probability that all of the other suppliers produce). Hence, supplier i produces at level 1 ($x_i^1 = 1$) if and only if equation 2 holds.

If equation 2 holds for all suppliers, so that all suppliers produce at level 1, all suppliers produce at level 2 (since it is optimal to produce when all other suppliers produce). By the same logic, all suppliers produce at level 3 and higher levels. Hence, when equation 2 holds for all suppliers, there is a unique introspective equilibrium in which all suppliers produce.

Now suppose equation 2 is violated for some supplier j. Notice that equation 2 holds for any supplier who is fully insured ($\theta_i k \geq c_i$), so supplier j must be a supplier who is not fully insured. By assumption 1, there are at least two suppliers who are not fully insured. Let supplier k be a second supplier who is not fully insured. Notice that, at level 2, supplier k does not produce since it is not optimal to produce under the assumption that supplier j is not producing. Consequently, at level 2, at least one supplier does not produce. By a similar logic, at least one supplier does not produce at level 3 and higher levels. Hence, an introspective equilibrium does not exist in which all suppliers produce. This establishes the lemma. □

Suppose the manufacturer wishes to produce widgets rather than not. The manufacturer's problem is to minimize costs ($\sum_i p_i$) subject to the incentive compatibility constraint of the suppliers (equation 2) and the incentive compatibility constraint of the manufacturer ($\theta_i k \leq p_i$). It clearly makes sense to choose p_i such that the suppliers' incentive compatibility constraints bind, which implies that $p_i = \frac{c_i - \theta_i k(1-\Gamma_{-i})}{\Gamma_{-i}}$. Substituting for p_i, we can rewrite the incentive compatibility constraint for the manufacturer as $\theta_i k \leq c_i$. We can thus rewrite the manufacturer's problem as:

$$\min_{\theta_i} \sum_i \frac{c_i - \theta_i k(1 - \Gamma_{-i})}{\Gamma_{-i}} \tag{3}$$

$$\text{subject to: } \theta_i k \leq c_i \text{ for all } i \text{ and } \sum_i \theta_i \leq 1$$

Without loss of generality, assume that $\Gamma_{-1} \leq \Gamma_{-2} \leq ... \leq \Gamma_{-n}$. This means that it is hardest to convince supplier 1 to produce and easiest to convince supplier n. Note that, with this indexing, supplier 1 also has the greatest impulse to produce:

$\gamma_1 \geq \gamma_2 \geq ... \geq \gamma_n$. Intuitively, despite supplier 1's high impulse, it is hard to convince supplier 1 to produce because of the low impulse of other suppliers.

From equation 3, we see that it is optimal for the manufacturer to use capital k to insure the suppliers for whom $\frac{1-\Gamma_{-i}}{\Gamma_{-i}}$ is greatest, or equivalently, the suppliers for whom Γ_{-i} is lowest. In other words, it makes sense for M to insure the lowest-index suppliers who are the most reticent to produce.

Let us define m to be the number of suppliers who can be insured using capital k. More precisely, let m be the integer such that $\sum_{i=1}^{m} c_i \leq k < \sum_{i=1}^{m+1} c_i$. Note that there might be some capital remaining after paying off the first m suppliers that could be used to partially pay off the $m+1$-st supplier. For the sake of simplifying exposition, we will assume that there is zero capital remaining after paying off the first m suppliers.

Assumption 3. *There is no capital remaining after paying off the first m suppliers: $\sum_{i=1}^{m} c_i = k$.*

The following proposition characterizes the solution to the manufacturer's problem.

Proposition 1. *If M produces widgets:*

1. *M fully insures the first m suppliers and does not insure other suppliers:*
   - $\theta_i = \frac{c_i}{k}$ *for* $i \leq m$.
   - $\theta_i = 0$ *for* $i > m$.
2. *M pays the first m suppliers marginal cost and pays a markup to other suppliers:*
   - $p_i = c_i$ *for* $i \leq m$.
   - $p_i = \frac{c_i}{\Gamma_{-i}} = c_i + \underbrace{c_i \cdot \frac{1-\Gamma_{-i}}{\Gamma_{-i}}}_{\text{markup paid to i}}$ *for* $i > m$.

M produces widgets if the total markup does not exceed the surplus:

$$\underbrace{\sum_{i>m} c_i \cdot \frac{1-\Gamma_{-i}}{\Gamma_{-i}}}_{\textit{total markup}} \leq S \tag{4}$$

Otherwise, M does not produce widgets and offers suppliers nothing for their inputs ($p_i = 0$ for all i).

According to Proposition 1, it is optimal for M to use capital k to fully insure the most reticent suppliers ($\theta_i = \frac{c_i}{k}$). Because these suppliers are fully insured, they do not need to be paid a markup. To convince the remaining suppliers to produce, M must pay a markup. The size of supplier i's markup is $c_i \cdot \frac{1-\Gamma_{-i}}{\Gamma_{-i}}$, which is increasing in supplier i's reticence (i.e. decreasing in Γ_{-i}).

3 Make or Buy

We now extend the model to consider whether it makes sense for M to make input i in-house or buy the input from another firm. Our approach to the make-or-buy problem is based on Aghion and Tirole (1997). As in Aghion and Tirole (1997), M and S_i decide ex ante on an allocation of formal authority. If M has formal authority, we will say that M and S_i are integrated; otherwise, we will say that they are separate firms. The party with formal authority has the right to make two decisions: (1) whether input i is produced and (2) which production methods are used. The timing is similar to the baseline model:

1. M makes take-it-or-leave-it offers to each supplier of the form (p_i, θ_i, A_i), where $A_i \in \{M, S_i\}$ is an allocation of formal authority.
2. Suppliers decide whether to accept the offers.
3. Suppliers exert effort at generating cost-cutting production methods. The party with authority then chooses among the production methods.

4. The parties with authority simultaneously decide whether to produce inputs.

5. For each input i that is produced, M decides whether to use the input and pay p_i to S_i, or pay a penalty $\theta_i k$ to S_i. If M uses all of the inputs, M produces widgets and generates revenue R. Otherwise, M produces no widgets and generates zero revenue.

In Section 3.1, we will carefully spell out what happens at time 3. As a preview, the main finding is that M-authority raises the cost of producing input i to $(1 + \beta_i)c_i$. In Section 3.2, we will show that, while M-authority increases *production costs*, it also lowers the *costs of coordination*. This is the main tradeoff governing the make-or-buy decision.

3.1 Cost cutting

Let us consider how production methods are chosen at time 3. We will focus on input i and we will drop i subscripts to reduce notation.

There are P aspects of the production process and a production method for each aspect. Let m_p denote the production method used for aspect p and let $\frac{c(m_p)}{P}$ denote the associated cost. The overall cost of production is $\frac{1}{P}\sum_{p=1}^{P} c(m_p)$.

For each aspect of production, there is a default method with a cost c_H and two cheaper methods (m1 and m2) that might be uncovered. One of the cheaper methods has an associated cost $c_L < c_H$ and the other method has an associated cost $c_L - \Delta < c_L$. The cost of production is borne by the supplier, so the supplier prefers the cheapest possible method. The manufacturer receives a private benefit of negligible size from one of the two cheaper methods. The manufacturer prefers whichever of the cheaper methods generates this negligible private benefit. The supplier's preferred method is the same as the manufacturer's preferred method with probability α. Parameter α captures the degree of preference congruence of M and the supplier.

The supplier uncovers the cheaper methods for process p with probability $\eta_S > 0$ and the manufacturer uncovers the cheaper methods with probability $\eta_M > 0$. The chance of uncovering cheaper methods is independent across players and processes. After methods have been uncovered, the party with formal authority chooses among the production methods they know or, alternatively, delegates the choice to the other party.

Analysis

Let us now examine how the allocation of authority affects the cost of production.

M-authority. If the manufacturer has formal authority, they choose the method themselves whenever they uncover the cheap methods and they delegate the choice to the supplier otherwise. The expected cost to the supplier of producing the input is:

$$C^{M-auth} = c_H - \underbrace{\eta_M(c_H - c_L + \alpha \cdot \Delta)}_{\text{M uncovers the cheap methods}} - \underbrace{(1-\eta_M)\eta_S(c_H - c_L + \Delta)}_{\text{M delegates and S uncovers the cheap methods}} .$$

$$= c_H - (\eta_M + \eta_S + \eta_M\eta_S)(c_H - c_L) - (\eta_S + \eta_M(\alpha - \eta_S))\Delta$$

Supplier-authority. If the supplier has formal authority, they choose the method themselves whenever they uncover the cheap methods and they delegate to the manufacturer otherwise. The expected cost to the supplier of producing the input is:

$$C^{S-auth} = c_H - \underbrace{\eta_S(c_H - c_L + \Delta)}_{\text{S uncovers the cheap methods}} - \underbrace{(1-\eta_S)\eta_M(c_H - c_L + \alpha \cdot \Delta)}_{\text{S delegates and M uncovers the cheap methods}} .$$

$$= c_H - (\eta_M + \eta_S + \eta_M\eta_S)(c_H - c_L) - (\eta_S + \eta_M(1 - \eta_S)\alpha)\Delta$$

Notice that $C^{S-auth} < C^{M-auth}$ when $\alpha < 1$. Intuitively, $\alpha < 1$ means that the manufacturer is not sufficiently concerned with cost-cutting and allocating formal authority to them leads to inflated costs.

Because it is probabilistic whether the manufacturer and the supplier uncover the cheaper methods, the cost may differ from the expected cost. However, as the number of aspects of production P becomes large, the cost converges to the expected cost (due to the law of large numbers). Since it simplifies the exposition to deal with a fixed cost rather than a random cost, we will focus on the case where $P \to \infty$. Moreover, we will let $c_i \equiv C^{S-auth}$ and $\beta_i \equiv \frac{C^{M-auth}-c_i}{c_i}$, so that $(1+\beta_i)c_i = C^{M-auth}$. Parameter β_i, which is increasing in preference incongruence, captures the extent to which costs increase under M-authority.

3.2 Authority allocation

Let us now examine how authority will be allocated at time 1. To aid intuition, let us begin by considering a simple case where there are just two suppliers and the manufacturer has zero capital ($k = 0$). We will then examine the general case.

3.2.1 Two suppliers and $k = 0$

If neither supplier is integrated, the problem is the same as in Section 2. According to Proposition 1, to produce widgets, the manufacturer must pay supplier i a markup of $c_i \cdot \frac{1-\gamma_j}{\gamma_j}$. The total amount the manufacturer pays on top of marginal cost is:

$$\Phi^{NI} = c_1 \cdot \frac{1-\gamma_2}{\gamma_2} + c_2 \cdot \frac{1-\gamma_1}{\gamma_1}$$

Now suppose that both suppliers are integrated. At time 4, there is no longer a coordination problem since M has the authority to produce both types of inputs. Hence, there is no need to pay a markup because of the coordination problem. However, there is cost inflation in this case arising from the added production costs of M authority. At time 1, M must offer supplier i a contract where $p_i = (1 + \beta_i)c_i$. Consequently, there is an added cost of $\beta_i \cdot c_i$ on input i. The total amount the manufacturer pays on top of the original marginal cost is:

$$\Phi^{12} = c_1 \cdot \beta_1 + c_2 \cdot \beta_2$$

Finally, consider the case where only supplier i is integrated. At time 4, the parties deciding whether to produce inputs are M and supplier j. The coordination game is played, in other words, between M and j rather than i and j. To analyze this case, we will assume that M has an impulse just as suppliers do.

Assumption 4. *The manufacturer has an impulse to produce, denoted $x_M^0 \in \{0, 1\}$, just like suppliers. We assume that $x_M^0 = 1$ with probability γ_M. x_M^0 is private information of the manufacturer.*

In order to get supplier j to produce at time 4, the supplier must pay a markup of $c_j \cdot \frac{1-\gamma_M}{\gamma_M}$. Because M is the second party in the game, the only party that must be paid a markup is supplier j; however, there is an *incentive compatibility constraint* that must be satisfied in order for M to choose to produce input i at time 4. The incentive compatibility constraint for M is analogous to equation 2 in Lemma 1:

$$\underbrace{(S - \Phi^i)}_{\text{M's payoff if widgets are produced}} \cdot \gamma_j + \underbrace{0}_{\text{M's payoff if widgets are not produced}} \cdot (1 - \gamma_j) \geq \underbrace{c_i \cdot (1 + \beta_i)}_{\text{the cost of producing input i}}$$

Rearranging terms, the incentive compatibility constraint becomes:

$$S - \Phi^i \geq \frac{1}{\gamma_j} c_i (1 + \beta_i) \tag{5}$$

While supplier i is not part of the time-4 game, and thus does not need to be paid a coordination markup, the allocation of authority to M inflates the cost of producing input i by $c_i \cdot \beta_i$. Thus, the total cost inflation in this case is:

$$\Phi^i = c_i \cdot \beta_i + c_j \cdot \frac{1 - \gamma_M}{\gamma_M}$$

To summarize, the manufacturer will choose among the possible authority structures based upon which minimizes the production cost; however, the authority structure where supplier i is integrated but j is not is only available if the

IC constraint for the manufacturer (condition 5) is satisfied. The following proposition immediately follows.

Proposition 2. *Suppose there are two suppliers and $k = 0$.*

1. *Integrating supplier i is preferred to no integration ($\Phi^i \leq \Phi^{NI}$) if and only if:*

$$c_j \cdot \Big(\frac{\gamma_M - \gamma_i}{\gamma_M \cdot \gamma_i}\Big) \geq c_i\Big(\beta_i - \frac{1-\gamma_j}{\gamma_j}\Big) \tag{6}$$

2. *Integrating both suppliers is preferred to integrating i only ($\Phi^{12} \leq \Phi^i$) if and only if:*

$$\frac{1-\gamma_M}{\gamma_M} \geq \beta_j \tag{7}$$

Equation 6 shows what makes it is preferable to integrate supplier i rather than have no integration. Whether it is preferable depends in part upon how taking supplier i's authority away affects the cost of obtaining input i ($\beta_i - \frac{1-\gamma_j}{\gamma_j}$). The tradeoff here is between supplier i's reticence to produce ($\frac{1-\gamma_j}{\gamma_j}$), which pushes towards integration, and supplier i's preference incongruence with M (β_i), which pushes against integration. Additionally, whether it is preferable to integrate supplier i depends upon whether supplier i has a low impulse to produce compared to the manufacturer ($\gamma_M - \gamma_i$).

Equation 7 shows what makes it is preferable to integrate both suppliers rather than supplier i only. What matters in this case is how taking away supplier j's authority affects the cost of obtaining input j ($\beta_j - \frac{1-\gamma_M}{\gamma_M}$). The tradeoff is between supplier j's reticence to produce ($\frac{1-\gamma_M}{\gamma_M}$), which pushes towards integration, and supplier j's preference incongruence with M (β_j), which pushes against integration. Notice that supplier j's reticence in this case depends upon M's impulse.

In the special case where there is a uniform impulse to produce ($\gamma_1 = \gamma_2 = \gamma_M = \gamma$) and uniform preference incongruence ($\beta_1 = \beta_2 = \beta$), full integration is preferred when $\frac{1-\gamma}{\gamma} \geq \beta$ and non-integration is preferred otherwise. In other words, integration occurs when the cost of giving authority to M (β) is low relative to the costs of coordinating suppliers (γ).

3.2.2 The General Case

Let us turn now to the general case where there are n suppliers and the manufacturer has capital k. The following lemma characterizes the cost of producing widgets when M integrates a set of suppliers I and insures a set of suppliers K. (For ease of exposition, we will assume that, after fully insuring set K, no additional capital remains to partially insure other suppliers.)

Lemma 2. *Suppose the manufacturer integrates a set of suppliers I and uses capital k to fully insure a set of suppliers K (i.e. $\theta_i = \frac{c_i}{k}$ for $i \in K$ and $\theta_i = 0$ for $i \notin K$). When at least one supplier is integrated,*

1. *The amount the manufacturer pays to produce widgets on top of the original marginal cost is:*

$$\Phi^I(K) = \underbrace{\sum_{i \in I} c_i \cdot \beta_i}_{\text{non-congruence costs}} + \underbrace{\sum_{i \notin I \cup K} c_i \cdot \frac{1 - \Gamma_{-i-I} \cdot \gamma_M}{\Gamma_{-i-I} \cdot \gamma_M}}_{\text{coordination costs}}, \tag{8}$$

   where $\Gamma_{-i-I} = \prod_{j \notin \{i\} \cup I} \gamma_j$.

2. *The incentive compatibility constraint for the manufacturer is:*

$$S - \Phi^I(K) \geq \frac{1}{\Gamma_{-I}} \sum_{i \in I} c_i(1 + \beta_i),$$

   where $\Gamma_{-I} = \prod_{j \notin I} \gamma_j$.

The first term of equation 8 is the added cost of producing inputs when suppliers are integrated. This cost is lower when the integrated suppliers' preferences are more congruent with those of M. The second term of equation 2 is the cost of coordinating the non-integrated suppliers. The difference relative to the markups suppliers receive in Proposition 1 (equation 4) is that manufacturer M takes the place of the integrated suppliers. As in the two-supplier case, there is an incentive compatibility constraint for the manufacturer since the manufacturer must be properly incentivized to produce at time 4. This incentive compatibility constraint is analogous to the incentive compatibility constraint in the two-supplier case (condition 5).

Lemma 2 characterizes the cost of producing widgets as a function of who is integrated (I) and who is insured (K). The following corollary follows directly.

Corollary 1.

1. *The manufacturer does not insure suppliers who are integrated ($\theta_i = 0$ for $i \in I$).*
2. *Among the non-integrated suppliers, the manufacturer insures those who are most reluctant to produce (i.e. those of lowest index).*

Point 1 of the corollary can be seen from equation 8. The reason to insure a supplier is to reduce the coordination markup; but there is zero coordination markup for integrated suppliers, and hence no reason to insure them. Point 2 can also be seen from equation 8, and the logic is analogous to Proposition 1. The low-index suppliers are the most reluctant to invest. Insuring these suppliers has the greatest bang for the buck in terms of reducing the total markup.

The next proposition also follows from Lemma 2. It examines when a given supplier is more or less likely to be integrated. The proof is given in the appendix.

Proposition 3. *Lowering supplier i's congruence with M (β_i) or supplier i's impulse to produce (γ_i) makes it more likely supplier i will be integrated. More precisely, let $I^*(\beta_i, \gamma_i)$ denote the firms that are integrated as a function of β_i and γ_i. If $i \in I^*(\beta, \gamma)$:*

1. *$i \in I^*(\beta', \gamma)$ for all $\beta' < \beta$.*
2. *$i \in I^*(\beta, \gamma')$ for all $\gamma' < \gamma$.*

Intuitively, lowering β_i reduces the cost of integrating supplier i, and hence raises the chance that supplier i will be integrated. When γ_i is low and supplier i is not integrated, it is hard to coordinate the other suppliers. Thus, lowering γ_i raises the benefit of integrating supplier i, and hence raises the chance that supplier i will be integrated.

Proposition 4, which gives some additional results, follows from Lemma 2 as well. The proof is given in the appendix.

Proposition 4.

1. *If $\beta_i = 0$ for all suppliers, all suppliers are integrated.*

2. *If suppliers differ only in their β's ($c_i = c$ and $\gamma_i = \gamma$ for all i), the suppliers who are integrated are those for whom β_i is lowest.*

3. *If suppliers differ only in their γ's ($c_i = c$ and $\beta_i = \beta$ for all i), the suppliers who are integrated are those for whom γ_i is lowest.*

4. *If suppliers are homogeneous ($c_i = c$, $\beta_i = \beta$, and $\gamma_i = \gamma$ for all i), it is possible that some suppliers are integrated and some are not.*

Point 1 follows from the observation that there is no cost to integration—only benefit—when $\beta_i = 0$ for all i. Point 2 follows from the observation that the cost of integration is lowest for the most congruent suppliers (i.e. the suppliers with the lowest β_i's). Point 3 follows from the observation that the coordination benefits of integration are greatest for the suppliers with the lowest impulses (i.e. the suppliers with the lowest γ_i's).

Point 4 considers the case where all suppliers are homogeneous. In this case, the cost of integrating an additional firm is β. The benefit of integrating an additional firm is decreasing, however, in the number of integrated firms. Consequently, the point where the marginal cost of integration equates with the marginal benefit may be interior. Point 4 has the interesting implication that initially identical suppliers may be treated differently by the manufacturer; moreover, their ultimate costs of production may be different (c vs. $(1 + \beta)c$). This result speaks to the growing literature on persistent performance differences between firms (see, for instance, Gibbons and Henderson (2012), Chassang (2010), Ellison and Holden (2014)).

3.2.3 Discussion

Returning to the case of ASML, our model provides a rationale for ASML's acquisition of Cymer but not Trumpf or Zeiss. The value of Cymer's business was largely tied to the successful development of EUV lithography, and ASML had made a large bet on that prospect as well. In terms of the model, they had high "congruence" (low β). By contrast Trumpf made lasers for many other purposes and the lens Zeiss developed for ASML was a small part of its business (i.e. β was higher). The two other acquisitions ASML made in this period were also companies whose products were strongly tied to the success of EUV lithography. Wijdeven Motion (acquired in September 2012) develops systems which help in achieving nanometer accuracy; and Hermes Microvision (acquired June 16, 2016) tests the functioning of advanced chips.

Note also that Zeiss, Trumpf, and ASML are all Northern European firms. This likely created a degree of trust between them which, in terms of the model, might correspond to a high γ. Cymer, based in San Diego, may by contrast have had less trust of ASML. This may be another reason why Cymer was integrated.

4 Applications

Supplier Investments

Suppose the inputs the manufacturer needs to produce a widget are generic and have a low marginal cost of production. In such a case, the manufacturer is likely to have sufficient capital to purchase all of the inputs ($k \geq \sum_i c_i - c_{\max}$), in which case no coordination problem arises.

There are two circumstances where coordination problems are more likely to arise: (i) the inputs are generic but have a high marginal cost of production, or (ii) the inputs are non-generic. In both cases, a coordination problem is likely because the cost of producing inputs (c_i) is high. In the latter case, the cost of producing inputs is high because suppliers need to make investments to produce them. This was a key reason for ASML's coordination problem: its main suppliers (e.g.

Cymer, Trumpf, and Zeiss) needed to make expensive, years-long investments.

The idea that coordination problems arise when suppliers make expensive investments suggests an interesting extension to the model. Consider a dynamic version of the model where there are several potential manufacturers of widgets. At time 1, it is hard to produce widgets because input suppliers have not made the requisite investments. If suppliers make investments at time 1, the coordination problem disappears at time 2. Such a model suggests that a given manufacturer M might *prime the pump* at time 1. If M solves the coordination problem at time 1, they create a supply chain that other manufacturers can exploit at time 2.

One consideration that arises here is akin to the classic dilemma in the patent literature, where manufacturer M might be disincentivized from creating the supply chain at time 1 if they lose market power at time 2. Hence, preserving market power at time 2 becomes an important consideration for manufacturer M. This desire to preserve market power is arguably another important reason why we might see integration. By integrating suppliers at time 1, M obtains market power in the input market at time 2.

Movers and Shakers

Our model considers three tools that the manufacturer can use to achieve coordination: (i) insuring suppliers, (ii) paying markups, and (iii) integrating suppliers. Potentially, additional tools might be available to the manufacturer besides these three. Akerlof and Holden (2016) considers the role that "movers and shakers" play in coordinating parties. It considers, for instance, the role that financial institutions or real estate developers play in creating expectations that a project will succeed. Movers and shakers, according to this view, create a virtuous cycle where each party's confidence raises the confidence of the other parties.

In terms of the present model, we might view movers and shakers as changing the impulses of suppliers (γ_i). They might raise a supplier's γ_i by working to sell them on the project. Since raising γ_i reduces the overall markup that the manufacturer needs to pay, we might also expect movers and shakers to command large rents for the role that they play.

Keiretsu and Chaebol

Our model assumes that the manufacturer's relationship with suppliers is short-term. If instead the manufacturer engages in repeated transactions with suppliers, establishing long-term relationships, they can threaten to withhold future business if suppliers ever fail to deliver. Consequently, the manufacturer has greater ability to punish suppliers (i.e. the limited liability constraint is weaker). Long-term relationships thus serve as an additional tool to achieve coordination and may even substitute for other tools such as integration.

In Japan and Korea, we see an example of long-term relationships substituting for integration. In Japan, most companies are organized into *keiretsu*, and Korea has a similar type of arrangement called *chaebol*. Companies within a keiretsu or chaebol operate according to norms of loyalty to one another.[10] Many chaebol and many keiretsu (the so-called "vertical keiretsu") have a structure akin to our model where there is a central company and many smaller suppliers. Toyota is a leading example. Notably, the Japanese and Korean automobile industries are far less integrated than their American counterparts. For instance, by one estimate, 70 percent of all production is subcontracted in the Japanese auto industry, versus 30 percent in the US.[11]

International Trade

Suppose manufacturer M decides to obtain an input from overseas rather than domestically. How might this affect the integration decision? One possibility is that trust between the parties declines. We might think of this as a reduction in γ_i. At the same time, the congruence of preferences between M and the supplier might

[10] See Aoki (1990) and Berglöf and Perotti (1994) for a discussion of the relational contracts that exist within keiretsu.

[11] See Miyashita and Russell (1995) and Nagaoka et al. (2008) for a discussion of the Japanese auto industry and Kim et al. (2021) for a discussion of the Korean auto industry.

decline as well. We might think of this as an increase in β_i. According to Proposition 3, a decline in γ_i pushes the manufacturer toward integrating the supplier while an increases in β_i pushes the manufacturer against integrating the supplier. Hence, there are competing considerations. The model suggests that whether integration increases or decreases when production moves abroad depends upon which of these considerations dominates.

Industrial Policy

A number of rationales used to justify industrial policy in advanced economies in the 1960s and 1970s have been thoroughly debunked. For instance, industrial policy was sometimes justified on the grounds of job creation, exchange-rate or balance-of-payments management, or sectoral externalities. These arguments did not stand up to scrutiny (see, inter alia, Scitovsky (1954), Bhagwati et al. (1978)).

Our theory offers a different rationale for the potential efficacy of industrial policy. If there are projects with positive surplus that are not occurring, a government can play a coordinating role. This could happen through the moving-and-shaking channel we discussed above, or it could happen by the government providing capital. By providing a manufacturer with additional capital, a government could facilitate coordination where it otherwise would not occur, or help it be achieved at lower cost by reducing the need to pay markups.

Complex issues such as the green-energy transition exhibit obvious coordination problems, suggesting a potentially important role for "coordination-based" industrial policy.

5 Concluding Remarks

We have provided a model of supply chains that emphasizes the importance of coordination and speaks to relationships between a single manufacturer and multiple suppliers—or "spiders" as they have become known. Our model not only provides stark predictions that are consistent with existing empirical evidence;

it is also tractable enough to facilitate the analysis of a number of applications. These include important issues in international trade and industry policy.

At the heart of our approach is a novel theory of the firm based on the importance of coordination. Coase (1937) first raised the deep question of why firms exist if markets are an efficient means of allocating resources. The *Transaction Costs* theory of the firm that Coase suggested and Williamson developed (Williamson (1971), Williamson (1975)) was the first alternative to the rather unsatisfying *Neoclassical* theory of the firm. The *Principal-Agent Approach* (see for instance Alchian and Demsetz (1972) and Holmstrom (1982)), in contrast to transaction cost theory, emphasizes the importance of agency problems for hiring employees or outsourcing production to independent contractors.

Grossman and Hart (1986) and Hart and Moore (1990) pioneered the *Property Rights Theory* of the firm. PRT emphasizes the importance of assets ownership in alleviating hold-up problems and hence encouraging ex ante relationship-specific investments.

A fourth major approach to theory of the firm is the *Incentives Approach* of Holmstrom and Milgrom (1994). Those authors view make-or-buy decisions as two alternative systems for managing incentives. Gibbons (2005) provides an excellent overview of these theories.

We offer a distinct approach—a *Coordination-Based* theory of the firm.[12] This approach emphasizes the role that a firm (the manufacturer in our model) plays in coordinating the provision of inputs from a number of disparate suppliers. The firm can encourage provision of an input either through contract—by providing insurance to a supplier in the event that they supply the input but other suppliers do not—or via control of decision rights, whereby they can compel provision of the input from a given supplier. Our coordination-based theory provides sharp predictions about when exchange is mediated by contract, and when integration is preferable. We have shown that the coordination-based theory provides starkly

[12] Alonso et al. (2008) emphasize the role of coordination in a different way to us. They focus on the need to adapt decisions to local conditions but also coordinate those decisions in a multi-divisional firm. As such they focus on the internal structure of an organization whereas we focus on how the boundaries of firms are determined–i.e. the make-or-buy decision.

different predictions than PRT.

Our view is that all five of these rationales for the existence of firms are important; but to date, the coordinating role that firms play has received little formal attention. We hope that our contribution has partially remedied this oversight.

References

Acemoglu, Daron, Pol Antràs, and Elhanan Helpman, "Contracts and technology adoption," *American Economic Review*, 2007, *97* (3), 916–943.

Aghion, Philippe and Jean Tirole, "Formal and real authority in organizations," *Journal of political economy*, 1997, *105* (1), 1–29.

Akerlof, Robert and Richard Holden, "Movers and shakers," *Quarterly Journal of Economics*, 2016, *131* (4), 1849–1874.

_ **and** _ , "Capital assembly," *The Journal of Law, Economics, and Organization*, 2019, *35* (3), 489–512.

Alchian, Armen A and Harold Demsetz, "Production, information costs, and economic organization," *The American economic review*, 1972, *62* (5), 777–795.

Alfaro, Laura, Davin Chor, Pol Antras, and Paola Conconi, "Internalizing global value chains: A firm-level analysis," *Journal of Political Economy*, 2019, *127* (2), 508–559.

Alonso, Ricardo, Wouter Dessein, and Niko Matouschek, "When does coordination require centralization?," *American Economic Review*, 2008, *98* (1), 145–179.

Antràs, Pol, "Firms, contracts, and trade structure," *Quarterly Journal of Economics*, 2003, *118* (4), 1375–1418.

_ , "Incomplete contracts and the product cycle," *American Economic Review*, 2005, *95* (4), 1054–1073.

_ **and Davin Chor**, "Organizing the global value chain," *Econometrica*, 2013, *81* (6), 2127–2204.

— **and** — , "Global Value Chains," *Handbook of International Economics*, 2022, *5*, 297–376.

— **and Elhanan Helpman**, "Global sourcing," *Journal of Political Economy*, 2004, *112* (3), 552–580.

— **and** — , "Contractual Frictions and Global Sourcing," *The Organization of Firms in a Global Economy*, 2008.

Antras, Pol, Teresa C Fort, and Felix Tintelnot, "The margins of global sourcing: Theory and evidence from us firms," *American Economic Review*, 2017, *107* (9), 2514–2564.

Aoki, Masahiko, "Toward an economic model of the Japanese firm," *Journal of economic literature*, 1990, *28* (1), 1–27.

Baldwin, Richard and Anthony Venables, "Spiders and snakes: offshoring and agglomeration in the global economy," *Journal of International Economics*, 2013, *90* (2), 245–254.

Berglöf, Erik and Enrico Perotti, "The governance structure of the Japanese financial keiretsu," *Journal of financial Economics*, 1994, *36* (2), 259–284.

Bhagwati, Jagdish N et al., "Foreign trade regimes and economic development: Anatomy and consequences of exchange control regimes," *NBER Books*, 1978.

Chassang, Sylvain, "Building routines: Learning, cooperation, and the dynamics of incomplete relational contracts," *American Economic Review*, 2010, *100* (1), 448–465.

Chor, Davin and Lin Ma, "Contracting frictions in global sourcing: Implications for welfare," *mimeo*, 2021.

Coase, Ronald, "The Nature of the Firm," *Economica*, 1937, *4* (16), 386–405.

Corsetti, Giancarlo, Bernardo Guimaraes, and Nouriel Roubini, "International lending of last resort and moral hazard: A model of IMF's catalytic finance," *Journal of Monetary Economics*, 2006, *53* (3), 441–471.

Crawford, Vincent P, Miguel A Costa-Gomes, and Nagore Iriberri, "Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications," *Journal of Economic Literature*, 2013, *51* (1), 5–62.

Ellison, Glenn and Richard Holden, "A theory of rule development," *The Journal of Law, Economics, & Organization*, 2014, *30* (4), 649–682.

Gereffi, Gary, "International trade and industrial upgrading in the apparel commodity chain," *Journal of International Economics*, 1999, *48* (1), 37–70.

Gibbons, Robert, "Four formal (izable) theories of the firm?," *Journal of Economic Behavior & Organization*, 2005, *58* (2), 200–245.

_ **and Rebecca Henderson**, *What do managers do?: Exploring persistent performance differences among seemingly similar enterprises*, Citeseer, 2012.

Grossman, Sanford J and Oliver D Hart, "The costs and benefits of ownership: A theory of vertical and lateral integration," *Journal of Political Economy*, 1986, *94* (4), 691–719.

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

Holmstrom, Bengt, "Moral hazard in teams," *The Bell journal of economics*, 1982, pp. 324–340.

_ **and Paul Milgrom**, "The firm as an incentive system," *The American economic review*, 1994, pp. 972–991.

Kets, Willemien and Alvaro Sandroni, "A theory of strategic uncertainty and cultural diversity," *The Review of Economic Studies*, 2021, *88* (1), 287–333.

Kim, Chulsik, Jun Ho Jeong, and Hyung Je Jo, "Detecting dynamic changes in Hyundai Motor's parts supply system as an industry latecomer," *Journal of asian sociology*, 2021, *50* (1), 55–90.

Miller, Chris, *Chip War: The Fight for the World's Most Critical Technology*, Simon and Schuster, 2022.

Miyashita, Kenichi and David Russell, *Keiretsu: Inside the hidden Japanese conglomerates*, McGraw-Hill, 1995.

Morris, Stephen and Hyun Song Shin, "Catalytic finance: When does it work?," *Journal of International Economics*, 2006, *70* (1), 161–177.

Nagaoka, Sadao, Akira Takeishi, and Yoshihisa Noro, "Determinants of firm boundaries: Empirical analysis of the Japanese auto industry from 1984 to 2002," *Journal of the Japanese and International Economies*, 2008, *22* (2), 187–206.

Scitovsky, Tibor, "Two concepts of external economies," *Journal of political Economy*, 1954, *62* (2), 143–151.

Williamson, Oliver, *Markets and Hierarchies: Analysis and Antitrust Implications*, New York: Free Press, 1975.

Williamson, Oliver E., "The Vertical Integration of Production: Market Failure Considerations," *American Economic Review*, 1971, *61* (2), 112–123.

6 Appendix

Proof of Proposition 3. Suppose that a set of suppliers I are integrated. Let us consider how adding supplier i to set I changes the total markup. Let $\Delta(I)$ denote the change in the total markup from integrating i:

$$\Delta(I) \equiv \Phi^{I\cup\{i\}}(K') - \Phi^{I}(K),$$

where K is the optimal insurance choice when suppliers I are integrated and K' is the optimal choice when suppliers $I \cup \{i\}$ are integrated. To prove the result, it is sufficient to show that $\Delta(I)$ is weakly increasing in β_i and γ_i. There are four cases we need to consider:

1. I is nonempty and $i \notin K$.

2. I is nonempty and $i \in K$.

3. I is empty and $i \notin K$.

4. I is empty and $i \in K$.

In case 1, corollary 1 implies that $K' = K$. We find that:

$$\Delta(I) = c_i \cdot [\beta_i - \frac{1 - \Gamma_{-i-I} \cdot \gamma_M}{\Gamma_{-i-I} \cdot \gamma_M}] - \sum_{j \notin I \cup \{i\} \cup K} c_j \cdot [\frac{1 - \Gamma_{-j-I} \cdot \gamma_M}{\Gamma_{-j-I} \cdot \gamma_M} - \frac{1 - \Gamma_{-i-j-I} \cdot \gamma_M}{\Gamma_{-i-j-I} \cdot \gamma_M}]$$

It is clear that $\Delta(I)$ is increasing in β_i. γ_i appears in the second term of the above expression: a higher γ_i increases Γ_{-j-I}, which increases $\Delta(I)$.

In case 2, integrating i frees up some capital to insure other suppliers. We have $K' = [K - \{i\}] \cup A$, where A denotes the additional insured suppliers. We find that:

$$\Delta(I) = c_i \cdot \beta_i - \sum_{j \notin I \cup K} c_j \cdot [\frac{1 - \Gamma_{-j-I} \cdot \gamma_M}{\Gamma_{-j-I} \cdot \gamma_M} - \frac{1 - \Gamma_{-i-j-I} \cdot \gamma_M}{\Gamma_{-i-j-I} \cdot \gamma_M}] - \sum_{j \in A} c_j \cdot [\frac{1 - \Gamma_{-i-j-I} \cdot \gamma_M}{\Gamma_{-i-j-I} \cdot \gamma_M}]$$

It is clear that $\Delta(I)$ is increasing in β_i. γ_i appears in the second term of the above expression: a higher γ_i increases Γ_{-j-I}, which increases $\Delta(I)$.

In case 3, corollary 1 implies that $K' = K$. We find that:

$$\Delta(I) = c_i \cdot [\beta_i - \frac{1 - \Gamma_{-i}}{\Gamma_{-i}}] - \sum_{j \notin \{i\} \cup K} c_j \cdot [\frac{1 - \Gamma_{-j}}{\Gamma_{-j}} - \frac{1 - \Gamma_{-j-i} \cdot \gamma_M}{\Gamma_{-j-i} \cdot \gamma_M}]$$

It is clear that $\Delta(I)$ is increasing in β_i. γ_i appears in the second term of the above expression: a higher γ_i increases Γ_{-j}, which increases $\Delta(I)$.

In case 4, integrating i frees up some capital to insure other suppliers. We have $K' = [K - \{i\}] \cup A$, where A denotes the additional insured suppliers. We find that:

$$\Delta(I) = c_i \cdot \beta_i - \sum_{j \notin K} c_j \cdot [\frac{1 - \Gamma_{-j}}{\Gamma_{-j}} - \frac{1 - \Gamma_{-j-i} \cdot \gamma_M}{\Gamma_{-j-i} \cdot \gamma_M}] - \sum_{j \in A} c_j \cdot [\frac{1 - \Gamma_{-j-i} \cdot \gamma_M}{\Gamma_{-j-i} \cdot \gamma_M}]$$

It is clear that $\Delta(I)$ is increasing in β_i. γ_i appears in the second term of the above expression: a higher γ_i increases Γ_{-j}, which increases $\Delta(I)$.

This completes the proof. $\square$

Proof of Proposition 4.

The reasoning behind Points 1-3 is straightforward and given in the text. To establish Point 4, let us focus on the case where there is no supplier heterogeneity, $k = 0$, and $\gamma_M = \gamma$. In this case, the markup when g firms are integrated is:

$$\Phi(g) = g \cdot \beta + (n - g)c(\gamma^{g-n} - 1)$$

Differentiating with respect to g:

$$\Phi'(g) = \beta + c + c \cdot \gamma^{g-n}[(n - g)\log \gamma - 1]$$

Observe that $\Phi'(n) = \beta$ and $\Phi'(0) = \beta + c + c \cdot \gamma^{-n}[n \log \gamma - 1]$. If $\beta > 0$, $\gamma < 1$, $c > 0$, and n is large, $\Phi'(n) > 0$ and $\Phi'(0) < 0$. Hence, the value of g for which the markup is lowest is interior ($0 < g^* < n$). This establishes the result. $\square$