Firm Size, Strategic Communication, and Organization

Huanxing Yang*
Department of Economics, Ohio State University

December 2012

Abstract

This paper studies the relationship between firm size and the optimal organization structure by extending the two-division model of Alonso et. al (2008) to finite number of divisions. The firm must resolve the tradeoff between coordination and adaptation; relevant information for decision making is dispersed and communication is strategic. We compare the overall performance of centralization and decentralization as the size of the firm grows and show that the impact of firm size on the optimal organization structure depends on divisional managers' own-division bias and the incentive need of coordination. In an extension endogenize the number of divisions or the optimal size of the firm.

Keywords: coordination, firm size, centralization, decentralization, cheap talk
JEL classification: D23, D83, L23

1 Introduction

Multi-divisional organizations face an inevitable tradeoff between coordination and adaptation: if the activities across different divisions become more synchronized, they necessarily are less adapted to the local conditions of each individual division, and vice versa. Given that division managers are usually better informed about the local conditions of their own divisions, efficient communication among division managers and the decision makers is essential to achieve better coordination or adaptation. In such situations, it is natural to ask whether a centralized organization or a decentralized organization can achieve better overall performance? This important question was addressed in Alonso et. al (2008, ADM hereafter).

Specifically, in a two-division model ADM show that centralized organizations have an advantage in coordinating the activities among different divisions, while decentralized organizations have an advantage in adapting decisions to local conditions. Moreover, decentralization achieves a better overall performance if either the incentives of division managers are sufficiently aligned or the importance of coordination is sufficiently low. In particular, decentralization can be optimal even if coordination is very important.

*Email: yang.1041@osu.edu. I would like to thank Ching-Jen Sun for his input in the early stage of this paper.

This paper extends ADM's two-division model to finite number of divisions. We interpret the number of divisions as the size of a firm or an organization.[1] As the firm's size grows, coordination naturally becomes more important since the decision of each existing division now needs to be coordinated with those of the additional divisions. The following questions naturally arise. As the firm's size grows, under each organization structure will the quality of communication inside the firm improve or deteriorate? How does the firm's size affect the relative performance of centralization and decentralization? Will a bigger firm be more likely or less likely to adopt a decentralized organization structure? These questions are addressed in this paper.

Our model is a simple extension of ADM's two-division model. Specifically, the firm has N symmetric divisions and a decision needs to be made for each division. Each division's profit depends on how close its decision is to its local condition (adaptation) and how close its decision is to the decisions of other divisions (coordination). Information about the local condition of each division is only observed by the manager of that division, thus communication inside the firm is essential to achieve effective coordination and adaptation. Communication is strategic and is modeled as cheap talk, a la Crawford and Sobel (1982). Division managers' incentives are biased as they care more about the profits of their own divisions than the firm's overall profit. The firm can adopt either a centralized or a decentralized organization structure. Under centralization the headquarter retains all the decision rights and division managers communicate vertically with the headquarter before decisions are made, while under decentralization the decision rights of all divisions are delegated to the corresponding division managers and they communicate horizontally among themselves.

We distinguish *intensive* need of coordination from *extensive* need of coordination, with the former referring to the importance of coordination between the decisions of any two divisions and the latter referring to the number of other divisions' decisions that each division's decision needs to coordinate with. As the number of divisions or the size of the firm grows, the extensive need of coordination typically increases while the intensive need of coordination remains the same. As in ADM, when the intensive need of coordination increases the quality of communication deteriorates under centralization, while under decentralization it actually improves. When the size of the firm hence the extensive need of coordination increases, the quality of communication is worsened under both organization structures, and it deteriorates faster under centralization. Intuitively, as the number of divisions increases under both organization structures the decision of each division becomes less responsive to the report of the local information of any other division. As a result, each division manager has a stronger incentive to exaggerate its own information in order to pull the decisions of other divisions closer to the local condition of his own division, leading to more noisy communication.

[1] An alternative interpretation is that the number of divisions correspond to the degree of specialiation. See Section 2 for details.

Our main results are as follows. When the own-division bias of divisional managers is small, decentralization is always optimal regardless of the size of the firm. When the own-division bias is big and the intensive need of coordination is high, centralization is always optimal regardless of the size of the firm. When the own-division bias is relatively small but the intensive need of coordination is high, decentralization becomes more likely to be optimal as the number of divisions increases. Finally, when the own-division bias is big but the intensive need of coordination is low, centralization becomes more likely to be optimal as the size of the firm increases. The third and fourth results are counter-intuitive as simple intuition suggests the opposite. As the number of division increases, the increase in the overall importance of coordination is large (small) when the intensive need of coordination is high (low). Since centralization has an advantage in coordination, one would think that centralization should become more likely to be optimal when the intensive need of coordination is high and decentralization should become more likely to be optimal when the intensive need of coordination is low. This simple reasoning is flawed since it does take into account the endogenous adjustments in decision making and the quality of communication when the extensive need of coordination increases.

Finally we tried to endogenize the number of divisions or the optimal size of the firm. It turns out that firms with a big own-division bias and a high intensive need of coordination tend to be small and centralized, while firms with a small own-division bias and a low intensive need of coordination tend to be big and decentralized. Thus there is a positive correlation between firm size and decentralization.

The rest of the paper is organized as follows. In the next subsection we discuss related literature. Section 2 presents the model and the equilibria under both organization structures are derived in Section 3. In Section 4 we compare the overall performance of centralization and decentralization as the firm's size grows. The optimal number of divisions is endogenized in Section 5 and a conclusion is offered in Section 6. All the missing proofs in the text can be found in the appendix.

1.1 Related literature

As mentioned earlier, the most closely related to this paper is ADM. Rantakari (2008) develops a model similar to ADM's. In his model, division managers care exclusively about their own division and different managers might care about coordination to different degrees. The main difference between our model and the above ones is that they only consider two divisions while the current model studies more than two divisions, which enables us to shed light on the relationship between optimal organization form and firm size.[2]

This paper is also related to the literature on hierarchy (Williamson, 1967; Calvo and Wellisz, 1978, 1979; Keren and Levhari, 1979; Qian, 1994). This literature studies the optimal hierarchy

[2] For more related papers on coordination in organizations, see the references in ADM.

structure in organizations: the number of hierarchical tiers, span of control for each supervisor, and the wages for each hierarchical tier. Those papers focus on the tradeoff between hierarchical depth (the number of hierarchical tiers) and span of control: increasing the hierarchical depth can reduce the span of control for each manager thus monitoring workers becomes easier, but with the cost that there is a bigger cumulative losses across hierarchical tiers and more managers to pay. In some sense, the firm size in our model can be considered as the span of control for the headquarter. The main difference between this literature and our paper is that while they focus on incentive provision, we emphasize the tradeoff between coordination and adaptation with endogenous communication.

Another literature that is related to our paper is cheap talk with multiple senders. In Krishna and Morgan (2001) and Battaglini (2002) multiple senders observe the same piece of information. This is different from the current model as multiple senders have different information. McGee and Yang (2010) studies a model in which two senders have different and non-overlapping information. In all of those models the principal has a single decision to make thus there is no issue of coordination. Since decentralization has the feature of delegating decision rights to divisional managers, this paper is also related to the literature on delegation (Melumard and Shibano, 1991; Aghion and Tirole, 1997; Dessein, 2002; Alonso and Matouschek, 2008).

2 Model

A firm or an organization consists of $N \geq 2$ divisions and potentially one headquarter (HQ). Each division i has a local condition θ_i, which is uniformly distributed on $[-s, s]$, $s > 0$. All the θ_is are mutually independent. Regarding division i, a decision d_i needs to be made. Denote $d = (d_1, d_{2,}..., d_i, ..., d_N)$ as a profile of decisions. The profit generated by division i, π_i, depends on d and θ_i. In particular,

$$\pi_i = K - (d_i - \theta_i)^2 - \sum_{j \neq i} \delta(d_i - d_j)^2, \tag{1}$$

The constant K captures the base profit generated by a division. The second term measures the "adaptation loss", resulting from failing to adapt to local conditions. The last term is the "coordination loss" resulting from the miscoordination among the decisions of different divisions, where $\delta > 0$ represents the importance of coordination. The overall profit of the firm is $\pi = \sum_{i=1}^{N} \pi_i$.

There are two interpretations of the model. The first interpretation is in terms of firm size. Adding one more division or expanding to a new business adds a base profit of K to the firm. But to accommodate the new division efficiently, all the existing divisions have to coordinate their decisions to the new division's. Otherwise additional coordination losses will be incurred. The second interpretation is in terms of specialization. More specialization implies dividing the existing business of the firm into more specialized tasks, and setting up more divisions with each division specializing in a single task. Deepening specialization one step further (adding one more division) brings a base profit of K to the firm, but it requires more coordination as well. Throughout the paper, we will stick to the first interpretation.

Each division is run by a manager. The manager of division i, who we call manager i, observes only the realization of θ_i. The HQ does not observe the realizations of any θ_i. We assume that manager i's objective is to maximize $\lambda\pi_i + (1-\lambda)\sum_{j\neq i}\pi_j$, where $\lambda \in [1/2, 1]$. That is, manager i puts more weight on the profit of his own division than on those of other divisions. As λ increases, the own-division bias increases. The HQ's objective is to maximize the firm's overall profit $\sum_{i=1}^{N}\pi_i$.

Following the previous literature, we assume that only ex ante decision rights are contractible. As in ADM, we first consider two allocations of decision rights: *Centralization or Decentralization*. Under Centralization, the decision rights are retained in HQ, while under Decentralization the decision rights are within each decision manager. The sequence of events is as follows, given the organization form. Under Centralization, first each division manager i observes θ_i, then all divisional managers simultaneously communicate with the HQ by sending messages. After hearing all the messages, the HQ makes decisions $d = (d_1, d_{2,} ..., d_i, ..., d_N)$. Under decentralization, after observing their respective local conditions θ_i each divisional manager i simultaneously sends message m_i to all the other managers. The massage m_i sent by manager i to any other manager is the same.[3] After hearing messages from managers $-i$, each manager i simultaneously decides d_i. We model the message exchanges as a cheap talk game, a la Crawforad and Sobel (1982).

Observing (1), we see that an increase in δ and an increase in N both increase the need of coordination for division i. In the first case, the need of coordination between division i and any other existing division increases. On the other hand, an increase in N implies that division i needs to coordinate with more divisions. To distinguish these two needs of coordination, we call δ as the *intensive need of coordination,* and while when N increases, we say that *the extensive need of coordination* increases.

3 Equilibrium

In this section, first we characterize the decision making under each organization form, taking posterior beliefs about $\theta \equiv (\theta_1, \theta_2, ..., \theta_N)$ as given. Then, we characterize equilibrium information transmission in the strategic communication games. Finally, we derive the performance under each organization structure.

[3] Since all the divisions are symmetric, a manager has no incentive to send different messages to different managers.

3.1 Decision Making

Denote the messages sent by manager i as m_i, and $m \equiv (m_1, ..., m_N)$ as a profile of messages. Under centralization, given m the HQ chooses d to maximize $E[\sum_{i=1}^{N} \pi_i | m]$, which is equivalent to maximize

$$E[-\sum_{i=1}^{N}(d_i - \theta_i)^2 - \delta \sum_{i=1}^{N} \sum_{j \neq i}(d_i - d_j)^2 | m].$$

The first order condition with respect to d_i yields (the superscript C denotes centralization)

$$d_i^C = \frac{1}{1 + 2\delta(N-1)} E(\theta_i | m) + \frac{2\delta}{1 + 2\delta(N-1)} \sum_{j \neq i} d_j^C.$$

After manipulation, we get

$$d_i^C = \frac{1 + 2\delta}{1 + 2\delta N} E(\theta_i | m) + \frac{2\delta}{1 + 2\delta N} \sum_{j \neq i} E(\theta_j | m). \tag{2}$$

From (2), we observe that for decision d_i^C the HQ puts more weight on $E(\theta_i|m)$ than on $E(\theta_j|m)$. As δ increases, these two weights become closer. As N increases, both weights decrease since more terms of $E(\theta_j|m)$ are added.

Under decentralization, manager i chooses d_i to maximize $E[\lambda \pi_i + (1-\lambda) \sum_{j \neq i} \pi_j | \theta_i, m]$. The first order condition with respect to d_i yields (the superscript D denotes decentralization)

$$d_i^D = \frac{\lambda}{\lambda + \delta(N-1)} \theta_i + \frac{\delta}{\lambda + \delta(N-1)} \sum_{j \neq i} E(d_j^D | m).$$

After manipulation, we get

$$\begin{aligned} E[d_i^D | m] &= \frac{\lambda + \delta}{\lambda + \delta N} E(\theta_i | m) + \frac{\delta}{\lambda + \delta N} \sum_{j \neq i} E(\theta_j | m), \\ d_i^D &= \frac{\lambda}{\lambda + \delta(N-1)} \theta_i \\ &+ \frac{\delta}{\lambda + \delta(N-1)} [\frac{\delta(N-1)}{\lambda + \delta N} E(\theta_i | m) + \frac{\lambda + \delta(N-1)}{\lambda + \delta N} \sum_{j \neq i} E(\theta_j | m)]. \end{aligned} \tag{3}$$

From (3) we see that, fixing N, as δ increases manager i puts less weight on θ_i and more weight on the weighted average of the posterior beliefs. Comparing (2) and (3), it is evident that d_i^C and d_i^D converge in the limit as δ goes to infinity. Fixing δ, as N increases manager i put less weight on θ_i, and the weights on $E(\theta_i|m)$ and $E(\theta_j|m)$ decrease as well. As $N \to \infty$, all the weights under both d_i^C and d_i^D go to zero. However, d_i^C and d_i^D do not converge in the limit. To see this, lets compute the ratio of the weight of $E(\theta_i|m)$ to that of $E(\theta_j|m)$. For d_i^C this ratio is always $(1 + 2\delta)/2\delta > 1$, while under d_i^D this ratio converges to 1 as N goes to infinity.

3.2 Strategic Communication

We first identify individual managers' incentives to misrepresent information. Consider manager 1 (all the other managers' incentives are similar). Suppose manager 1 can credibly induce posterior belief $v_1 = E(\theta_1|m)$. Clearly, manager 1 would like to induce v_1 such that his expected payoff is maximized:

$$v_1^* = \arg\max_{v_1} E\{-\lambda[(d_1-\theta_1)^2 + \delta\sum_{j\neq 1}(d_1-d_j)^2] - (1-\lambda)[\sum_{j\neq 1}(d_j-\theta_j)^2 + \sum_{j\neq 1}\sum_{i\neq j}(d_i-d_j)^2]|\theta_1\}. \quad (4)$$

Under centralization, the d_is in (4) equal to d_i^C, defined in (2). For $j \neq 1$, assume that manager 1's posterior belief about θ_j, $E[E[\theta_j|m]] = 0$. This property will be shown later hold in equilibrium. After some calculation, we get

$$v_1^* - \theta_1 = b_C\theta_1, \text{ where } b_C = \frac{\delta(2\lambda-1)(N-1)(1+4\delta)}{\lambda(1+2\delta)^2 + \delta(N-1)[1+(1-\lambda)4\delta]}. \quad (5)$$

Under decentralization, the d_is in (4) equal to d_i^D, defined in (3). Again assume that $E[E[\theta_j|m]] = 0$. We derived the desired v_1 for manager 1:

$$v_1^* - \theta_1 = b_D\theta_1, \text{ where } b_D = \frac{(2\lambda-1)[\lambda+\delta(N-1)]}{\delta\lambda + (1-\lambda)[\lambda+\delta(N-1)]}. \quad (6)$$

Like the two-division model of ADM, manager 1 has incentive to exaggerate his information unless $\theta_1 = 0$. From (5) and (6), it can be shown that $db_C/d\lambda \geq 0$ and $db_D/d\lambda \geq 0$. That is, under either organization structure the incentive to misrepresent information increases in own-division bias. On the other hand, $db_C/d\delta \geq 0$ but $db_D/d\delta \leq 0$. As the intensive need of coordination increases, under centralization each individual manager has more incentive to misrepresent information, while under decentralization the incentive to misrepresent information is reduced. This is because when δ increases, under centralization d_j^C becomes less sensitive to $E(\theta_i|m)$, while under decentralization d_j^D becomes more sensitive to $E(\theta_i|m)$. These results are the same as those in ADM.

Regarding the changes in N, it can be shown that $db_C/dN \geq 0$ and $db_D/dN \geq 0$. That is, when the extensive need of coordination (the number of divisions) increases, under either organization structure the incentive to misrepresent information increases. This suggests that changes in N and changes in δ have different impacts. In other words, changes in the intensive need of coordination and changes in the extensive need of coordination might lead to different results. To understand why $db_C/dN \geq 0$ and $db_D/dN \geq 0$, consider (2) and (3), the decision making. Under centralization, as N increases d_j^C becomes less sensitive to $E(\theta_i|m)$. Under decentralization, as N increases d_j^D becomes less sensitive to $E(\theta_i|m)$ as well. This is because as N increases, manager j has to worry about the coordination with additional divisions. As a result, manger j put less weight on the posterior of existing divisions, and the weight he puts on his own information is reduced as well. Since each individual manager j puts more weight on its own adaptation loss, he tends to "exaggerate" his own information more in order to pull other divisions' decisions toward his own ideal decision.

Now we characterize communication equilibria under both organization structures. A communication equilibrium under an organization structure is characterized by: (1) communication rules for division managers, $\mu_i(m_i|\theta_i)$, (2) decision rules for the decision makers, either $d^C(m)$ under centralization or $d_i^D(\theta_i, m)$ under decentralization, and (3) belief functions for message receivers, $g_i(\theta_i|m_i)$. We adopt perfect Bayesian equilibrium as our solution concept, which requires: (1) communication rules are optimal given the decision rules, (2) decision rules are optimal given beliefs, and (3) the beliefs are consistent with the communication rules.

The communication equilibria are qualitatively the same as those in ADM. All the equilibria are interval equilibria. The state space $[-s, s]$ is partitioned into intervals, and manager i only reveals in which interval θ_i lies. Denote $\overline{m}_j = E(\theta_j|m_j)$. Given the independence of θ_is and the fact that $E[\theta_i] = 0$, for any $i \neq j$ we have $E[\overline{m}_j] = E[\theta_i \overline{m}_j] = E[\overline{m}_i \overline{m}_j] = 0$. Moreover, $E[\theta_j \overline{m}_j] = E[\overline{m}_j^2]$.

Proposition 1 *All the communication equilibria are interval equilibria. For any profile of positive integers $n \equiv (n_1, n_2, ..., n_N)$, there is one equilibrium with a profile of partition points $a \equiv (a_1, a_2, ..., a_N)$, $a_j = (a_{j,(-n_j)}, ..., a_{j,n_j})$, where a_j, under governance structure g, $g = C, D$, are characterized by the following difference equations:*

$$\begin{aligned} a_{j,i+1} - a_{j,i} &= a_{j,i} - a_{j,i-1} + 4b_g a_{j,i}; \\ a_{j,-(i+1)} - a_{j,-i} &= a_{j,-i} - a_{j,-(i-1)} + 4b_g a_{j,-i}. \end{aligned}$$

Proof. The proof is a straightforward extension of the proof of Proposition 1 in ADM. The only difference is that when we consider manager 1's problem, in ADM the expectation is taking over θ_2, in our model the expectation is taking over θ_{-1}. Given that all the θ_is are independent and $E[\theta_i] = 0$, with slight modifications the proof of ADM applies. ■

Like ADM, in our model there are multiple communication equilibria with different profiles of the numbers of partition elements n. Following ADM, which is also standard in the literature of cheap talk, we will focus on the most informative equilibrium. Similar to the results of Proposition 2 in ADM, we can show that in the most informative equilibrium $n_j \to \infty$ for all j.[4]

Let $\sigma^2 = s^2/3$ be the variance of θ_i. Similar to the results of Lemma 1 in ADM, in the most informative equilibrium

$$E(\overline{m}_j^2) = \frac{1 + b_g}{3 + 4b_g} s^2 = (1 - S_g)\sigma^2, \text{ where } S_g = \frac{b_g}{3 + 4b_g}.$$

[4] In the most informative equilibrium, for any j the partitions around 0 are infinitely fine but the partitions around the extreme states are coarse.

It follows that the residual variance $E[(\theta_j - E(\overline{m}_j))^2] = S_g\sigma^2$, which measure the (negative) quality of communication. A bigger S_g means less information is transmitted in equilibrium, or communication is noisier. Following (5) and (6), we get

$$S_C = \frac{\delta(2\lambda-1)(N-1)(1+4\delta)}{3\lambda(1+2\delta)^2 + \delta(N-1)[(8\lambda-1)+4\delta(5\lambda-1)]}; \tag{7}$$

$$S_D = \frac{(2\lambda-1)[\lambda+\delta(N-1)]}{3\delta\lambda + (5\lambda-1)[\lambda+\delta(N-1)]}. \tag{8}$$

Proposition 2 *(i)* $S_D > S_C$ *if* $\lambda > 1/2$*, and* $S_D = S_C = 0$ *if* $\lambda = 1/2$*. (ii)* $\partial S_D/\partial\lambda > \partial S_C/\partial\lambda > 0$*. (iii)* $\partial S_C/\partial\delta > 0 > \partial S_D/\partial\delta$ *and* $\lim_{\delta\to\infty} S_C = \lim_{\delta\to\infty} S_D$*. (iv)* $\partial S_C/\partial N > \partial S_D/\partial N > 0$*, and* $\lim_{N\to\infty} S_C < \lim_{N\to\infty} S_D$*.*

Part (i)-(iii) of Proposition 2 are the same as Proposition 3 in ADM. Part (i) says that centralization enjoys communication advantage. Part (ii) says that as λ increases, under both organization structure communication becomes noisier. Moreover, under decentralization the quality of communication deteriorates faster than under centralization. Part (iii) implies that as δ increases, the quality of communication improves under decentralization, but deteriorates under centralization.

Part (iv) of Proposition 2 implies that as N increases, communication under both organization structures becomes noisier. However, as N increases the communication advantage enjoyed by centralization decreases, but it does not converges to zero in the limit. This pattern is illustrated in figure 1.

Intuitively, as the number of divisions increases, under both organization structure the relevant decisions becomes less responsive to m_i (the weights of relevant decisions spread to more m_is). As a result, manager i has a stronger incentive to exaggerate his own information, leading to a deterioration in the quality of communication. The reason that the communication quality deteriorates faster under centralization is that, as N increases, under centralization d_i^C becomes less sensitive to $\overline{m}_i$ at a faster rate than d_j^D to $\overline{m}_i$ under decentralization.[5] Under decentralization, as N increases manager j tends to reduce the weight of d_j^D on his own information θ_j, which reduces the speed at which the weight of d_j^D on $\overline{m}_i$ decreases as N increases. This mitigating effect is absent under centralization. Therefore, the weight of d_j^D on $\overline{m}_i$ decreases at a lower speed under decentralization than the weight of d_i^C on $\overline{m}_i$ does under centralization as N increases. As a result, as N increases the quality of communication deteriorates at a slower rate under decentralization and the communication advantage enjoyed by centralization decreases. As N goes to infinity, centralization still enjoys communication advantage or S_C and S_D do not converge because in the limit the relative weights of relevant decisions on $\overline{m}_i$ and $\overline{m}_j$ are still different, as mentioned earlier.

[5] More formally,

$$\partial\left(\frac{1+2\delta}{1+2\delta N} - \frac{\delta}{\lambda+\delta N}\right)/\partial N = (1-2\lambda)(1+2\lambda+4\delta N) < 0.$$

.

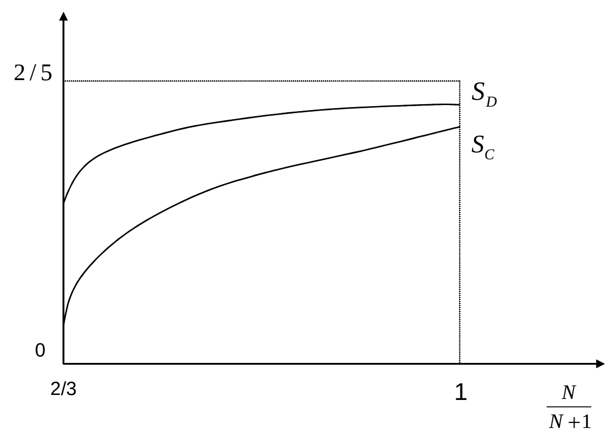

Figure 1: Communication Qualities as N Increases

3.3 Organization Performance

Now we compute the expected profits of the firm. Denote the expected profit under Centralization as $\Pi_C(N)$, and that under Decentralization as $\Pi_D(N)$. The profits are given in the following proposition.

Proposition 3 *The expected profits under centralization and under decentralization are given by*

$$\Pi_C(N) = KN - N\sigma^2[\frac{2\delta\,(N-1)}{1+2\delta N} + \frac{1+2\delta}{1+2\delta N}S_C] \tag{9}$$

$$\begin{aligned}\Pi_D(N) = {} & KN - N\sigma^2\{\frac{\delta(N-1)(2\lambda^2+\delta N)}{[\lambda+\delta N]^2} \\ & + \frac{\delta^2(N-1)[4\lambda^3+2\lambda^2\delta(2N-1)+\delta^2(N-1)N-\lambda^2]}{[\lambda+\delta(N-1)]^2[\lambda+\delta N]^2}S_D\}\end{aligned} \tag{10}$$

4 Firm Size and Optimal Organization Structure

As in ADM, coordination is achieved better under centralization while adaptation is achieved better under decentralization. The first result is due to the own-division bias of divisional managers, and the second result comes from the fact that some local information is lost in strategic communication.

Define AL_g^i and CL_g^i as the adaptation loss and coordination loss (for individual divisions) respectively under organization structure g. That is $AL_g^i \equiv E[(d_i^g - \theta_i)^2]$, and $CL_g^i \equiv \delta E[(d_i^g - d_j^g)^2]$. Since all divisions are symmetric, the total adaptation loss AL_g and total coordination loss CL_g can be expressed as $AL_g = N \times AL_g^i$ and $CL_g = N(N-1) \times CL_g^i$. It can be verified that $AL_C^i > AL_D^i$, and $CL_C^i < CL_D^i$. In other words, centralization enjoys coordination advantage and decentralization enjoys adaptation advantage.

As the number of divisions N increases, coordination becomes relatively more important. To see this, suppose N increases by 1. Now there are $N+1$ terms of individual adaptation losses and $(N+1)N$ terms of individual coordination losses. The ratio of the number of terms of individual coordination losses to that of individual adaptation losses increases from N to $N+1$. Combining with the fact that centralization enjoys coordination advantage, one might naively think that centralization becomes more likely to be optimal as N increases since coordination becomes relatively more important. But, as the analysis below shows, that is not always the case since it does not take into account endogenous decision making and endogenous quality of communication.

Now we formally compare the performance of two organization structures as N varies. From the expressions of (9) and (10), it is evident that two organization structures achieve the same outcome if $\lambda = 1/2$ or $\delta = 0$. For other cases, the following lemma compares the relative performance under two organization structures.

Lemma 1 *For $\lambda \in (1/2, 1]$ and $\delta > 0$, which organization structure is better depends on the sign of $f_N(\lambda, \delta)$. Specifically, $Sgn\{\Pi_C(N) - \Pi_D(N)\} = Sgn\{f_N(\lambda, \delta)\}$, where*

$$\begin{aligned} f_N(\lambda,\delta) = {} & \delta^3[12\lambda(2\lambda-1) + N(N-1)(100\lambda^2 - 90\lambda + 17)] \\ & +\delta^2[5N(N-1) + \lambda(-20 + 54N - 33N^2) + \lambda^2(66 - 144N + 40N^2) + \lambda^3(-40 + 100N)] \\ & +\delta\lambda[-5 + 6N + \lambda(25 - 38N) + \lambda^2(-30 + 40N)] - \lambda^2(5\lambda - 1). \end{aligned} \tag{11}$$

Define $\overline{\lambda}_N(\delta)$ as the value of λ as a function of δ such that $f_N(\lambda, \delta) = 0$, fixing N. The $\overline{\lambda}_N(\delta)$ curve demarcates the space of (λ, δ). Figure 2 plots the $\overline{\lambda}_N(\delta)$ curve for $N = 2, 4, 6$ (with $\delta/(1+\delta)$ as the vertical axis). In a three dimensional figure, Figure 3 shows how $\overline{\lambda}_N(\delta)$ shifts as N changes. In Figure 2, centralization is optimal for the area above (to the northeast) the $\overline{\lambda}_N(\delta)$ curve and decentralization is optimal for the area (to the southwest) under it. To see this, note that $\lim_{\lambda \to 1, \delta \to \infty} f_N(\lambda, \delta) > 0$ and $\lim_{\delta \to 0} f_N(\lambda, \delta) < 0$. Define $\overline{\delta}_N$ as the value of δ such that $\overline{\lambda}_N(\delta) = 1$.

We are interested in how changes in N affects the $\overline{\lambda}_N(\delta)$ curve. From the figures we see the following pattern. As N increases, the $\overline{\lambda}_N(\delta)$ curve rotates clockwisely. Specifically, the northwest part of the $\overline{\lambda}_N(\delta)$ shifts east, and the southeast part of the $\overline{\lambda}_N(\delta)$ shifts south. More formally, the following proposition shows how the relative performance of two organization structure changes as the number of divisions increases.

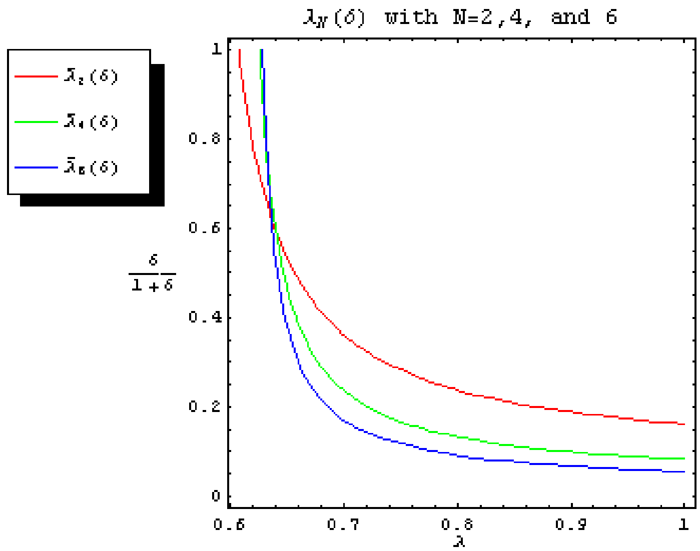

Figure 2: The $\overline{\lambda}_N(\delta)$ Curve as N Changes

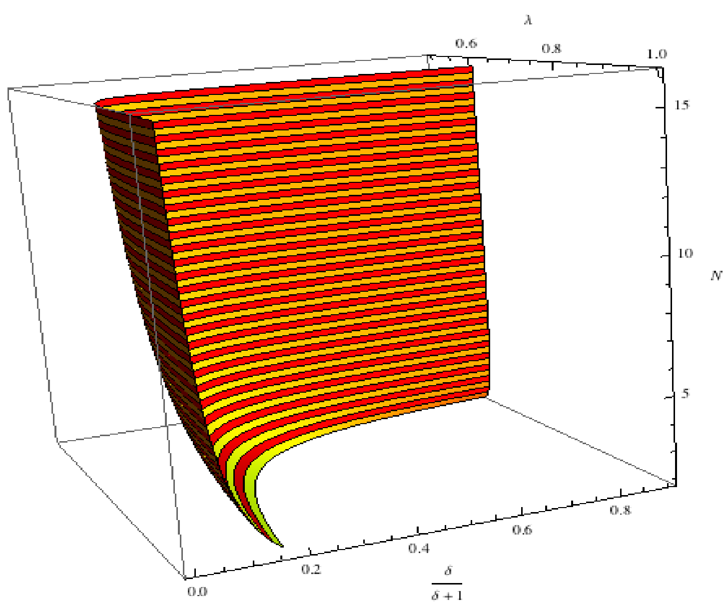

Figure 3: Three Dimensional Figure of the $\overline{\lambda}_N(\delta)$ Curves

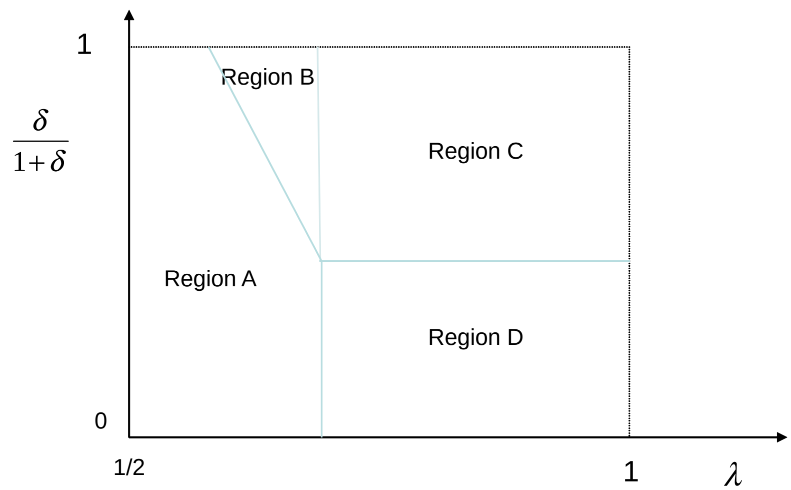

Figure 4: The Demarcation of Regions

Proposition 4 *(Centralization versus Decentralization as N increases). (i) If $\lambda \geq 0.75$, then as N increases the area (in the space of (λ, δ)) in which centralization is optimal expands (the $\overline{\lambda}_N(\delta)$ curve shifts downward); (ii) If $\lambda \leq 0.625$, then as N increases the area in which centralization is optimal shrinks (the $\overline{\lambda}_N(\delta)$ curve shifts to the right); (iii) for $\lambda \in (0.625, 0.75)$, the $\overline{\lambda}_N(\delta)$ curve and $\overline{\lambda}_{N+1}(\delta)$ curve intersect at least once; (iv) $\lim_{\delta\to\infty} \overline{\lambda}_N(\delta)$ is increasing in N, and $\lim_{N\to\infty,\delta\to\infty} \overline{\lambda}_N(\delta) = 0.63028$; $\overline{\delta}_N$ is decreasing in N and $\lim_{N\to\infty,\lambda\to 1} \overline{\delta}_N = 0$.*

Implications Proposition 4 has several implications. First, if the own-division bias is small enough ($\lambda < \overline{\lambda}_2 = 17/28$), then decentralization is always optimal regardless of N and δ, the need of extensive and intensive coordination. Second, as N increases the firm's optimal organization structure might change. More specifically, the parameter space of (λ, δ) can be roughly divided into four regions, as shown in the following figure.

Region A: The own-division bias (λ) is small. In this region, the optimal organization structure is always decentralization.

Region C: Both the own-division bias and the intensive need of coordination (both λ and δ) are large. In this region, the optimal organization structure is always centralization.

Region B: The own-division bias is relatively small but the intensive need of coordination is large (λ relatively small but δ large). In this region, the optimal organization structure depends on firm size. Specifically, as the number of divisions N increases decentralization becomes more likely to be optimal. This implies the following pattern on the firm's expansion path. If the firm starts with decentralization, then it will remains decentralized when the number divisions grows. If it starts with an centralized organization, then as the size of the firm grows, at some point it might switches to decentralization and then remains decentralized if the the number of divisions grows further.

Region D: The own-division bias is large but the intensive need of coordination is small (λ large but δ small). In this region, as the number of divisions N increases centralization becomes more likely to be optimal, which implies the following pattern of firm expansion. If the firm starts with centralization, then it will remains centralized when the number divisions grows. If it starts with an decentralized organization, then as the size of the firm grows, at some point it might switches to centralization and then remains centralized if the the number of divisions grows further.

The results regarding region B and region D are surprising. Simple reasoning would suggest the following: since centralization has coordination advantage and decentralization has adaptation advantage, with a high intensive need of coordination adding one more division will significantly increase the overall need of coordination and makes centralization more likely to be optimal, while the opposite is true when the intensive need of coordination is low. But our results indicate the opposite. As the number of divisions increases, it is the firms with a high intensive need of coordination (big δ, region B) that are become more likely to adopt decentralization, while firms with a low intensive need of coordination (small δ, region D) become more likely to adopt centralization.

Intuition To understand the results in Proposition 4, consider the impacts of increasing the number of divisions, N, by 1. To ease exposition, we introduce the following notation. Define $\Delta CL \equiv CL_D^i - CL_C^i$, the relative coordination loss, and $\Delta AL = AL_C^i - AL_D^i$, the relative adaptation loss. Note that $\Delta CL > 0$ and $\Delta AL > 0$, as decentralization has adaptation advantage and centralization has coordination disadvantage. For any point on the $\overline{\lambda}_N(\delta)$ curve,we have $\Delta AL = (N-1) \times \Delta CL$, or N terms of adaptation advantage of decentralization are balanced against $N(N-1)$ terms of coordination advantage of centralization. Note that as N increases by 1, under both organization structures divisions' decisions become closer, leading to an increase in individual adaptation loss and a decrease in individual coordination loss. Those endogenous adjustments in decisions will change the magnitudes of ΔAL and ΔCL, favoring either centralization or decentralization.

The magnitude of adjustments are different for different points on the $\overline{\lambda}_N(\delta)$ curve. To fix ideas, consider two points (λ', δ') and (λ'', δ'') on the $\overline{\lambda}_N(\delta)$ curve, with the first one being in region B (λ' small and δ' large) and the second one being in region D (λ'' big and δ'' small). First consider point (λ'', δ''). Since λ is large and δ is small, both the decision makings and the quality of communication under two organization structures are far apart. Specially, a big λ implies that the interests of division managers not aligned well, and a small δ makes division managers have little incentive to coordinate their decisions. As N increases by 1, The decisions and quality of communication under both organization structures will endogenously adjust. But since δ is small, adding one more division does not change the overall need of coordination much. Thus those adjustments will be small, which implies that ΔAL and ΔCL will not change much. As a result, with $N+1$ divisions the total adaptation advantage of decentralization will be smaller than the total coordination advantage of centralization, as the ratio of the terms of ΔCL to that of ΔCL increases from $N-1$ to N. Therefore, on point (λ'', δ'') centralization will dominate decentralization if N increases by 1.

Now consider point (λ', δ'). Since λ is small and δ is large, both the decision makings and the quality of communication under both organization structures are pretty close. In particular, a small λ implies that the interests of division managers are almost aligned, and a large δ makes division managers have strong incentive to coordinate their decisions. Now suppose N increases by 1. Since δ is large, adding one more division will have significant impacts on endogenous decisions and the quality of communication. Specifically, δ being large and λ being small implies that an increase in N will bring the decisions under both organizations significantly closer[6]. These endogenous adjustments in decisions tend to reduce ΔCL, the coordination advantage of centralization. Another effect of an increase in N is that it makes communication under both organization structures noisier, and it reduces the communication advantage of centralization. The decrease in communication quality under decentralization tends to increase ΔCL as it reduces divisions' ability to coordinate, while the decrease in communication quality under centralization tends to increase ΔAL as it reduces the HQ's ability to adapt. But since the reduction in communication quality is more significant under centralization, the increase in ΔAL tends to outweigh the increase in ΔCL. Combine all the effects mentioned above, as N increases by 1 the ratio of $\Delta AL/\Delta CL$ could adjust upward enough (from $N-1$) such that it exceeds N, making decentralization optimal.

Another way to understand the results is the following. When λ is small and δ is large, relative to centralization, decentralization can achieve coordination pretty well; while adaptation cannot be achieved well under centralization due to the noisiness of communication. As one more division is added, the concern for adaptation is going to outweigh the concern for coordination, which makes the region such that decentralization is optimal expands. On the other hand, when λ is big and δ

[6] It can be verified that the coeffecients in the expressions of d_i^C and d_i^D ((2) and (3)) are closer to each other as N increases.

is small, relative to centralization, decentralization cannot achieve coordination well, due to a big own division-bias and much more noisy communication under decentralization. Now adding one more division the concern for coordination is going to outweigh that for adaptation. This implies that the region in which centralization is optimal expands when N increases.

5 Endogenizing the Number of Divisions

In this section, we study the optimal number of divisions or the optimal size of the firm under two organization structures. Note that under either organization structure, the revenue function is KN, or the marginal revenue for each additional division is always K. We define the cost function under organization structure g as $L_g(N)$. In particular:

$$\begin{aligned} L_C(N,\lambda,\delta) &\equiv N\sigma^2[\frac{2\delta\,(N-1)}{1+2\delta N}+\frac{1+2\delta}{1+2\delta N}S_C]; \\ L_D(N,\lambda,\delta) &\equiv N\sigma^2\{\frac{\delta(N-1)(2\lambda^2+\delta N)}{[\lambda+\delta N]^2}+\frac{\delta^2(N-1)[4\lambda^3+2\lambda^2\delta(2N-1)+\delta^2(N-1)N-\lambda^2]}{[\lambda+\delta(N-1)]^2[\lambda+\delta N]^2}S_D\}. \end{aligned}$$

Moreover, we define the average cost under organization structure g as $AL_g \equiv L_g/N$. To ease analysis, although it is an integer we treat $N \geq 2$ as a continuous variable. This enables us to take the relevant derivatives and define marginal cost under organization structure g as $ML_g \equiv \partial L_g/\partial N$.

Lemma 2 *(i) Under centralization: the average cost is increasing in N, $\frac{\partial AL_C}{\partial N} > 0$; both the average cost and the marginal cost converge in the limit $\lim_{N\to\infty} AL_C = \lim_{N\to\infty} ML_C = \sigma^2$; the marginal cost curve shifts up as λ or δ increases, $\frac{\partial ML_C}{\partial \lambda} > 0$ and $\frac{\partial ML_C}{\partial \delta} > 0$; the marginal cost is increasing in N or $\frac{\partial ML_C}{\partial N} > 0$ if $-3\lambda+\delta(-1-4\lambda+\delta(8\lambda-4)) < 0$, otherwise it is decreasing in N. (ii) Under decentralization: the average cost is increasing in N, $\frac{\partial AL_D}{\partial N} > 0$; both the average cost and the marginal cost converge in the limit, $\lim_{N\to\infty} AL_D = \lim_{N\to\infty} ML_D = \sigma^2$; the marginal cost curve shifts up as λ increases, $\frac{\partial ML_D}{\partial \lambda} > 0$; the marginal cost curve shifts up as δ increases ($\frac{\partial ML_D}{\partial \delta} > 0$) if λ is relatively small or δ is relatively small; the marginal cost is increasing in N or $\frac{\partial ML_D}{\partial N} > 0$ if both λ and δ are relatively small.*

Under both organization structures, the average cost is always increasing in N. This is because an additional division increases the coordination loss of each existing division. The same pattern does not always hold for marginal costs. Under centralization, the marginal cost is either always increasing in N or always decreasing in N, depending on whether $-3\lambda+\delta(-1-4\lambda+\delta(8\lambda-4)) < 0$. Under decentralization, no clear pattern exists regarding whether the marginal is increasing or decreasing in N. Part (ii) of Lemma 2 just provides a sufficient condition under which the ML_D curve is upward sloping.

Denote the optimal number of divisions under organization structure g as N_g^*. To make sure that N_g^* exists, we assume that $K/\sigma^2 < 1$. This condition guarantees that the firm will not choose to have infinite number of divisions. If the marginal cost curve is upward sloping, then the optimal N_g^* is typically determined by the intersection of the (constant) marginal revenue curve and the marginal cost curve. Since the marginal cost curve is not always well behaved, to carry out analysis we have to put some restrictions on the parameter space. In particular, we define

$$\begin{aligned}
\Omega(\lambda,\delta) &\equiv \{(\lambda,\delta): -3\lambda+\delta(-1-4\lambda+\delta(8\lambda-4))<0\},\\
\Lambda(\lambda,\delta) &\equiv \{(\lambda,\delta): \frac{\partial ML_D}{\partial N}>0\},\\
\Lambda_\delta(\lambda,\delta) &\equiv \{(\lambda,\delta): \frac{\partial ML_D}{\partial \delta}>0\}.
\end{aligned}$$

Proposition 5 *(i) Under centralization: N_C^* is unique for any parameter values. Moreover, if $\Omega(\lambda,\delta)\in(\lambda,\delta)$, then N_C^* is decreasing in both λ and δ, and increasing in K/σ^2; if $(\lambda,\delta)\notin\Omega(\lambda,\delta)$, then $N_C^*=2$, independent of δ and λ. (ii) Under decentralization: if $(\lambda,\delta)\in\Lambda(\lambda,\delta)$, then N_D^* is unique, and N_D^* is decreasing in λ and increasing in K/σ^2. If $(\lambda,\delta)\in\Lambda(\lambda,\delta)\cap\Lambda_\delta(\lambda,\delta)$, then N_C^* is decreasing in δ.*

Note that $(\lambda,\delta)\in g(\lambda,\delta)$ implies that both λ and δ are relatively big. Thus Proposition 5 implies that for relatively big λ and δ the size of a centralized firm is very small (the low bound $N=2$) and is independent of λ and δ. For either small λ or small δ, the size of a centralized firm is decreasing both in λ and δ. Combine the results from the previous section that centralization is optimal only if both λ and δ are relatively big, we conclude that the size of a centralized firm tends to be small. Regarding decentralization, note that $(\lambda,\delta)\in\Lambda(\lambda,\delta)\cap\Lambda_\delta(\lambda,\delta)$ implies that both λ and δ are small. According to Proposition 5, in this parameter space the size of a decentralized firm is decreasing in both λ and δ. Combine the results from the previous section that decentralization is optimal only if either λ or δ is relatively small, we conclude that the size of a decentralized firm tends to be big. To summarize, in the parameter space of λ and δ, the optimal size and optimal organization of the firm is as follows. When both λ and δ are small (roughly region A), the firm is decentralized and has a relatively big size, while when both λ and δ are big (roughly region C), the firm is centralized and has a relatively small size.[7]

6 Conclusion

This paper studies the relationship between firm size and the optimal organization structure by extending ADM's two-division model to finite number of divisions. Organization structure not only affects the tradeoff between coordination and adaptation but also impacts on the quality of communication, which is strategic and endogenously determined. We show that under both

[7] No general conclusion can be drawn for Region B (small λ and big δ) and Region D (big λ and small δ) as we are not able to pin down the optimal size of the firm under decentralization in these regions.

centralization and decentralization communication becomes more noisy as the size of the firm grows (or more divisions), and the quality of communication under two organization structures getting closer.

Our central result is that the impact of firm size on the optimal organization structure depends on divisional managers' own-division bias and the intensive need of coordination. When the own-division bias of divisional managers is small, decentralization is always optimal regardless of the size of the firm. When the own-division bias is big and the intensive need of coordination is high, centralization is always optimal regardless of the size of the firm. When the own-division bias is relatively small but the intensive need of coordination is high, decentralization becomes more likely to be optimal as the number of divisions increases. Finally, when the own-division bias is big but the intensive need of coordination is low, centralization becomes more likely to be optimal as the size of the firm increases. In an extension we endogenize the number of divisions and study the optimal size of the firm. It turns out that firms with a big own-division bias and a high intensive need of coordination tend to be small and centralized, while firms with a small own-division bias and a low intensive need of coordination tend to be big and decentralized. Thus there is a positive correlation between firm size and decentralization.

References

[1] Alonso, R., W. Dessein, and N. Matouschek. "When Does Coordination Requires Centralization," *American Economic Review*, 2008, 98(1), 145-179.

[2] Alonso, R. and N. Matouschek. "Optimal Delegation," *Review of Economic Studies* 75(1), 2008, 259-293.

[3] Battaglini, M. "Multiple Referrals and Multidimensional Cheap Talk," *Econometrica* 70(4), 2002, 1379-1401.

[4] Calvo, G. and S. Wellisz. "Supervision, Loss of Control and the Optimal Size of the Firm," *Journal of Political Economy*, 1978, 86, 943-952.

[5] Calvo, G. and S. Wellisz. "Hierarchy, Ability and Income Distribution," *Journal of Political Economy,* 1979, 87, 991-1010.

[6] Crawford, V. and J. Sobel. "Strategic Information Transmission," *Econometrica* 50(6), 1982, 1431-1451.

[7] Dessein, W. "Authority and Communication in Organizations," *Review of Economic Studies* 69, 2002, 811-838.

[8] Keren, M. and D. Levhari. "The Optimal Span of Control in a Pure Hierarchy," *Management Science*, 25, 1162-1172.

[9] McGee, A. and H. Yang. "Cheap Talk with Two Senders and Complementary Information," 2010, working paper.

[10] Melumad, N. and T. Shibano. "Communication in Settings with No Transfers," *Rand Journal of Economics* 22(2), 1991, 173-198.

[11] Krishna, V. and J. Morgan. "A Model of Expertise," *Quarterly Journal of Economics* 116, 2001a, 747-775.

[12] Qian, Y. "Incentives and Loss of Control in an Optimal Hierarchy," *Review of Economic Studies*, 1994, 61(3), 527-544.

[13] Rantakari, H. "Governing Adaptation," *Review of Economic Studies*, 2008, 75(4), 1257-1285.

[14] Williamson, O. "Hierarchical Control and Optimal Firm Size," *Journal of Political Economy*, 1967, 75, 123-138.

7 Appendix

Proof of Proposition 2.
Proof. By (7) and (8),

$$S_D - S_C = \frac{\lambda(2\lambda-1)[3\lambda(1+2\delta)^2 + 3\delta(N-1)(1+\lambda+3\delta) + 3\delta^2(N-1)^2]}{\{3\delta\lambda + (5\lambda-1)[\lambda+\delta(N-1)]\}\{3\lambda(1+2\delta)^2 + \delta(N-1)[(8\lambda-1)+4\delta(5\lambda-1)]\}},$$

which is strictly great than 0 unless $\lambda = 1/2$. This proves part (i).

$$\begin{aligned}
\partial S_C/\partial\lambda &= \frac{3\delta(N-1)(1+4\delta)(1+2\delta)(1+2\delta N)}{\{3\lambda(1+2\delta)^2 + \delta(N-1)[(8\lambda-1)+4\delta(5\lambda-1)]\}^2} > 0, \\
\partial S_D/\partial\lambda &= \frac{3[\delta(N-1)(2\lambda+\delta N) + \lambda^2(1+2\delta)]}{\{3\delta\lambda + (5\lambda-1)[\lambda+\delta(N-1)]\}^2} > 0.
\end{aligned}$$

Taking the difference of the above two terms and simplifying, one can show that $\partial S_D/\partial\lambda - \partial S_C/\partial\lambda > 0$. This proves part (ii).

$$\begin{aligned}
\partial S_C/\partial\delta &= \frac{3\lambda(2\lambda-1)(N-1)[(1+2\delta)(1+6\delta) + 4\delta^2(N-1)]}{\{3\lambda(1+2\delta)^2 + \delta(N-1)[(8\lambda-1)+4\delta(5\lambda-1)]\}^2} > 0, \\
\partial S_D/\partial\delta &= \frac{-3\lambda^2(2\lambda-1)}{\{3\delta\lambda + (5\lambda-1)[\lambda+\delta(N-1)]\}^2} < 0.
\end{aligned}$$

In the limit,

$$\lim_{\delta\to\infty} S_C = \frac{(2\lambda-1)(N-1)}{3\lambda+(N-1)(5\lambda-1)} = \lim_{\delta\to\infty} S_D.$$

This proves part (iii).

$$\begin{aligned}
\partial S_C/\partial N &= \frac{3\lambda\delta(2\lambda-1)(1+4\delta)(1+2\delta)^2}{\{3\lambda(1+2\delta)^2+\delta(N-1)[(8\lambda-1)+4\delta(5\lambda-1)]\}^2} > 0, \\
\partial S_D/\partial N &= \frac{3\lambda\delta^2(2\lambda-1)}{\{3\delta\lambda+(5\lambda-1)[\lambda+\delta(N-1)]\}^2} > 0.
\end{aligned}$$

Calculating the difference,

$$sgn(\partial S_C/\partial N - \partial S_D/\partial N) = sgn\{(1+2\delta)^2\varpi_1 + (1+2\delta)^2(N-1)\varpi_2 + \delta^2(N-1)^2\varpi_3\},$$

where

$$\begin{aligned}
\varpi_1 &= (1+4\delta)\lambda^2(5\lambda-1)^2 + \delta\lambda^2[15(2\lambda-1)+\delta(120\lambda-51)] > 0, \\
\varpi_2 &= 2\lambda\delta(5\lambda-1)^2 + \lambda\delta^2[100\lambda(2\lambda-1)+2\lambda+8] > 0, \\
\varpi_3 &= (5\lambda-1)^2 + \delta[66\lambda(2\lambda-1)+7] + 12\delta^2(5\lambda-1)(3\lambda-1) > 0.
\end{aligned}$$

Therefore, $\partial S_C/\partial N - \partial S_D/\partial N > 0$. As $N \to \infty$,

$$\lim_{N\to\infty} S_C = \frac{(2\lambda-1)(1+4\delta)}{(8\lambda-1)+4\delta(5\lambda-1)} < \frac{2\lambda-1}{5\lambda-1} = \lim_{N\to\infty} S_D.$$

This proves part (iv). ■

Proof of Proposition 3.

Proof. By definition,

$$\Pi_C(N) = KN - E[\sum_{i=1}^{N}(d_i^C-\theta_i)^2 + \delta\sum_{i=1}^{N}\sum_{j\neq i}(d_i^C-d_j^C)^2].$$

From (2), we have

$$\begin{aligned}
E[(d_i^C-\theta_i)^2] &= \sigma^2 - \frac{1+2\delta}{1+2\delta N}(2-\frac{1+2\delta}{1+2\delta N})E(\overline{m}_i^2) + (\frac{1+2\delta}{1+2\delta N})^2\sum_{j\neq i}E(\overline{m}_j^2), \\
E[(d_i^C-d_j^C)^2] &= (\frac{1}{1+2\delta N})^2[E(\overline{m}_i^2)+E(\overline{m}_j^2)].
\end{aligned}$$

From the above three equations, we get

$$\Pi_C(N) = KN - N\sigma^2 + \frac{1+2\delta(N+1)+4\delta^2 N}{(1+2\delta N)^2}\sum_{i=1}^{N}E(\overline{m}_i^2),$$

and (9) can be readily derived.

Under Decentralization,

$$\Pi_D(N) = KN - E[\sum_{i=1}^{N}(d_i^D - \theta_i)^2 + \delta \sum_{i=1}^{N}\sum_{j\neq i}(d_i^D - d_j^D)^2].$$

From (3), we have

$$\begin{aligned} E[(d_i^D - \theta_i)^2] &= \frac{\delta^2(N-1)^2}{[\lambda+\delta(N-1)]^2}\sigma^2 - \frac{\delta^3(N-1)^2[2\lambda+\delta(2N-1)]}{[\lambda+\delta(N-1)]^2[\lambda+\delta N]^2}E(\overline{m}_i^2) + \frac{\delta^2}{[\lambda+\delta N]^2}\sum_{j\neq i}E(\overline{m}_j^2), \\ E[(d_i^D - d_j^D)^2] &= \frac{2\lambda^2}{[\lambda+\delta(N-1)]^2}\sigma^2 - \frac{\delta\lambda^2[2\lambda+\delta(2N-1)]}{[\lambda+\delta(N-1)]^2[\lambda+\delta N]^2}[E(\overline{m}_i^2) + E(\overline{m}_j^2)]. \end{aligned}$$

From the above three equations, we get

$$\Pi_D(N) = KN - N\sigma^2\{\frac{\delta(N-1)(2\lambda^2+\delta N)}{[\lambda+\delta N]^2} + \frac{\delta^2(N-1)[4\lambda^3 + 2\lambda^2\delta(2N-1) + \delta^2(N-1)N - \lambda^2]}{[\lambda+\delta(N-1)]^2[\lambda+\delta N]^2}S_D\},$$

from which (10) can be readily derived. ■

Proof of Lemma 1.

Proof. By (9) and (10), the difference between $\Pi_C(N)$ and $\Pi_D(N)$ can be calculated as:

$$\Pi_C(N) - \Pi_D(N) = \frac{\delta\lambda(2\lambda-1)N(N-1)\sigma^2}{NM_{CD}}f_N(\lambda,\delta),$$

where $NM_{CD} > 0$ is given by the following expression

$$\begin{aligned} &\{3\lambda(1+2\delta)^2 + \delta(N-1)[(8\lambda-1) + 4\delta(5\lambda-1)]\} \\ &\times\{3\delta\lambda + (5\lambda-1)[\lambda+\delta(N-1)]\}[\lambda+\delta(N-1)](\lambda+\delta N), \end{aligned}$$

and $f_N(\lambda,\delta)$ is given by (11). ■

Proof of Proposition 4.

Proof. Let $\Delta f_N(\lambda,\delta) = f_{N+1}(\lambda,\delta) - f_N(\lambda,\delta)$. By (11),

$$\begin{aligned} \Delta f_N(\lambda,\delta) &= \delta^3 2N[100\lambda^2 - 90\lambda + 17] + \delta^2[\lambda(21 - 104\lambda + 100\lambda^2) + 2N(5\lambda-1)(8\lambda-5)] \quad (12) \\ &+2\delta\lambda[(5\lambda-1)(4\lambda-3)]. \end{aligned}$$

Note that the first term in the bracket, $100\lambda^2 - 90\lambda + 17$, is increasing in λ for $\lambda \in [1/2, 1]$, and its value is zero when $\lambda = 0.63028$. For the term $21 - 104\lambda + 100\lambda^2$, its value is zero when $\lambda = 0.76576$, negative when $\lambda < 0.76576$, and positive when $\lambda > 0.76576$. For the term $(5\lambda-1)(8\lambda-5)$, it is increasing in λ for $\lambda \in [1/2, 1]$, and its value is zero when $\lambda = 0.625$. As N increases, the cutoff λ

such that the term in the second bracket is zero decreases. Observing all three terms in the bracket of (12), there are all negative when $\lambda \leq 0.625$, and there are all positive when $\lambda \geq 0.75$. Therefore, $\Delta f_N(\lambda, \delta) > 0$ if $\lambda \geq 0.75$ and $\Delta f_N(\lambda, \delta) < 0$ if $\lambda \leq 0.625$. This implies that for $\lambda \geq 0.75$, the area such that Centralization performs better expands as N increases, or the $\overline{\lambda}_{N+1}(\delta)$ curve lies below the $\overline{\lambda}_N(\delta)$ curve. Likewise, for $\lambda \leq 0.625$ the area such that Decentralization performs better shrinks as N increases, or the $\overline{\lambda}_{N+1}(\delta)$ curve lies to the right of the $\overline{\lambda}_N(\delta)$ curve. This proves part (i) and (ii).

To show part (iii), note that the $\overline{\lambda}_{N+1}(\delta)$ curve lies above the $\overline{\lambda}_N(\delta)$ curve for $\lambda \leq 0.625$ and it lies below the $\overline{\lambda}_N(\delta)$ curve for $\lambda \geq 0.75$. By the continuity of the $\overline{\lambda}_{N+1}(\delta)$ curve and the $\overline{\lambda}_N(\delta)$ curve, these two curves must intersect at least once for $\lambda \in (0.625, 0.75)$. Moreover, the number of intersections must be odd, since otherwise the relative position of the two curves must be the same for $\lambda \leq 0.625$ and $\lambda \geq 0.75$.

By (11), $\lim_{\delta \to \infty} \overline{\lambda}_N(\delta)$ is the solution to

$$12\lambda(2\lambda - 1) + N(N-1)(100\lambda^2 - 90\lambda + 17) = 0.$$

Since the first term is positive, the second term must be negative at the solution. As a result, $\lim_{\delta \to \infty} \overline{\lambda}_N(\delta)$ must increase with N as the weight of the second term increases, and it converges to 0.63028 from left as N goes to infinity. By (11), $\overline{\delta}_N$ is the solution to

$$[12\delta^3 + 6\delta^2 + (8N - 10)\delta - 4] + N\delta^2[27(N-1)\delta + (12N + 5)] = 0.$$

Inspecting the LHS of the above expression, we can see that all the terms involving δ are positive and it increases with both N and δ. Therefore, as N increases $\overline{\delta}_N$ must decrease in order to restore the equation; moreover, $\overline{\delta}_N$ converges to 0 as N goes to infinity. This proves part (iv). ■

Proof of Lemma 2.

Proof. Part (i) (centralization). It can be readily verified that $\lim_{N\to\infty} AL_C = \lim_{N\to\infty} ML_C = \sigma^2$ and

$$\frac{\partial AL_C}{\partial N} \sim 3\lambda\delta(1+2\delta)^2[(8\lambda - 1) + 4\delta(5\lambda - 1)] > 0.$$

Taking the relevant derivatives, we get:

$$\begin{aligned}
\frac{\partial ML_C}{\partial \lambda} &= \frac{3\delta(1+4\delta)(1+2\delta)^2[(N-1)\delta(1+4\delta) + \lambda(-3 - 4\delta + 8\delta^2 + 2N(3 + 2\delta(4+\delta)))]}{[-(N-1)\delta(1+4\delta) + \lambda(3 + 4\delta(1 - 2\delta + N(2+5\delta)))]^3}\sigma^2 > 0, \\
\frac{\partial ML_C}{\partial \delta} &= \frac{3\lambda(1+2\delta)[(N-1)\delta(1+4\delta) + \lambda(-3 - 4\delta + 8\delta^2 + 2N(3 + 2\delta(4+\delta)))](8\lambda - 1 + 6\delta(4\lambda - 1))}{[-(N-1)\delta(1+4\delta) + \lambda(3 + 4\delta(1 - 2\delta + N(2+5\delta)))]^3}\sigma^2 \\
&> 0.
\end{aligned}$$

To check whether the marginal cost is increasing in N, we compute

$$\frac{\partial ML_C}{\partial N} = -\frac{6(1+2\delta)^2\delta\lambda(8\lambda - 1 + 4\delta(5\lambda - 1))(-3\lambda + \delta(-1 - 4\lambda + \delta(8\lambda - 4)))}{[-(N-1)\delta(1+4\delta) + \lambda(3 + 4\delta(1 - 2\delta + N(2+5\delta)))]^4}\sigma^2.$$

From the above expression, it can be seen that $\frac{\partial ML_C}{\partial N} > 0$ if $-3\lambda + \delta(-1 - 4\lambda + \delta(8\lambda - 4)) < 0$ and $\frac{\partial ML_C}{\partial N} \leq 0$ otherwise.

Part (ii) (decentralization). Straightforward calculation shows that $\lim_{N\to\infty} AL_D = \sigma^2$ and $\lim_{N\to\infty} ML_D = \sigma^2$. The derivatives under decentralization are much more complicated. We did the calculations with the help of Maple. Here we just briefly describe what we did and report the results. First we take the relevant derivative. Then we arrange the terms according to the powers of δ. Denote the coefficient for δ^k as ϖ_k. Note that ϖ_k is a function of λ and N. Then we check the sign of each ϖ_k. As to $\frac{\partial AL_D}{\partial N}$, there are five coefficients, ϖ_k, $k = 0, .., 4$. All the $\varpi_k s$ are positive. Thus, $\frac{\partial AL_D}{\partial N} > 0$. As to $\frac{\partial ML_D}{\partial \lambda}$, there are eight coefficients, ϖ_k, $k = 0, ..., 7$. All the coefficients $\varpi_k s$ are positive as well. Therefore, $\frac{\partial ML_D}{\partial \lambda} > 0$. The case for $\frac{\partial ML_D}{\partial \delta}$ is more complicated. There are eight coefficients ϖ_k, $k = 0, ..., 7$. While ϖ_k is always positive for $k = 0, 1, 2, 3, 4$, and 7, it is not always the case for ϖ_5 and ϖ_6. Specifically, if λ is relatively small, then both ϖ_5 and ϖ_6 are positive, hence $\frac{\partial ML_D}{\partial \delta}$ is positive. If λ is relatively big, then both ϖ_5 and ϖ_6 are negative. However, if δ is relatively small (less than 1), then ϖ_k, $k = 0, 1, 2, 3, 4$ will dominate, and $\frac{\partial ML_D}{\partial \delta}$ again is positive. Finally, as to $\frac{\partial ML_D}{\partial N}$ there are nine coefficients, ϖ_k, $k = 0, ..., 8$. While ϖ_k is always positive for $k = 0, 1, 2, 3$, and 4, ϖ_8 is always negative. For ϖ_5, ϖ_6 and ϖ_6, they are all positive if λ is relatively small. Therefore, $\frac{\partial ML_D}{\partial \delta}$ is positive if both λ and δ are relatively small. ■

Proof of Proposition 5.

Proof. Part (i) (centralization). We first consider the case that $(\lambda, \delta) \in \Omega(\lambda, \delta)$. By Lemma 2, $\frac{\partial ML_C}{\partial N} > 0$. This means that the marginal cost curve is upward sloping. Combining with the fact that $\lim_{N\to\infty} ML_C = \sigma^2$, $ML_C < \sigma^2$ for any finite N. It follows that there is a unique N_C^*, which is determined by the intersection of the marginal revenue curve K and marginal cost curve ML_C. Since $\frac{\partial ML_C}{\partial \lambda} > 0$ and $\frac{\partial ML_C}{\partial \delta} > 0$, an increase in λ or δ will lead to an upward shift of the marginal cost curve ML_C, while the marginal revenue curve K is independent of either λ or δ. As a result, N_C^* is decreasing in either λ or δ. By similar logic, N_C^* is increasing in K/σ^2.

Next consider the case that $(\lambda, \delta) \notin \Omega(\lambda, \delta)$. By Lemma 2, $\frac{\partial ML_C}{\partial N} \leq 0$, or the marginal cost curve is downward sloping. Again by Lemma 2, we have $\frac{\partial AL_C}{\partial N} > 0$, $\lim_{N\to\infty} AL_C = \sigma^2$. This implies that the marginal cost curve ML_C is always above the average cost curve AL_C. Since $\lim_{N\to\infty} ML_C = \sigma^2$, the fact that ML_C is decreasing in N implies that $ML_C > \sigma^2$ for any finite N. By our assumption $K/\sigma^2 < 1$, the marginal revenue curve K is always below the marginal cost curve ML_C. Thus the optimal solution is the corner solution, that is, $N_C^* = 2$.

Part (ii) (decentralization). By Lemma 2, for $(\lambda, \delta) \in \Lambda(\lambda, \delta)$ we have $\frac{\partial ML_D}{\partial N} > 0$, or the marginal cost curve is upward sloping. Since $\lim_{N\to\infty} ML_D = \sigma^2$, for any finite N we have $ML_D < \sigma^2$. It follows that N_D^* is determined by the intersection of the marginal cost curve ML_D and the marginal revenue curve K, which is unique. Since by Lemma 2, $\frac{\partial ML_D}{\partial \lambda} > 0$, we have N_D^* decreasing in λ, and increasing in K/σ^2. If $(\lambda, \delta) \in \Lambda(\lambda, \delta) \cap \Lambda_\delta(\lambda, \delta)$, then $\frac{\partial ML_D}{\partial N} > 0$ and $\frac{\partial ML_D}{\partial \delta} > 0$. It follows that

N_C^* is decreasing in δ. ■