Organizing Modular Production*

Niko Matouschek
Northwestern University

Michael Powell
Northwestern University

Bryony Reich
Northwestern University

March 15, 2024

Abstract

Products are increasingly made by assembling separately produced modules. Motivated by the notion that a firm's production function drives its organization, we explore how modular production shapes a firm's communication structure. Decisions are partitioned into modules and require closer coordination within modules than across. Each agent knows the state his decision must be adapted to. The principal decides whom each agent tells about his state, given that each communication link comes at a cost. We show that optimal communication networks follow a simple threshold rule and exhibit the *threshold property.* We discuss comparative statics, applications, and empirical implications.

Keywords: organization, communication, network, modular

JEL classifications: D23, D85, L23

*We thank Heski Bar-Isaac, Hector Chade, Krishna Dasaratha, Wouter Dessein, Matt Elliott, Itay Fainmesser, Evgenii Fadeev, Ben Golub, Matt Jackson, Willemien Kets, Elliot Lipnowski, Luis Rayo, Jeroen Swinkels, Alireza Tahbaz-Salehi, Eduard Talamàs, Yiqing Xing and participants at various conferences and seminars for comments and suggestions. We are especially grateful to Andrea Prat for discussing this paper at the NBER Organizational Economics workshop and to Hossein Alidaee and Alexander Boche for excellent research assistance. Matouschek: n-matouschek@kellogg.northwestern.edu; Powell: mike-powell@kellogg.northwestern.edu; Reich: bryony.reich@kellogg.northwestern.edu.

1 Introduction

Modular production used to be the exception, now it is the norm. Ever since IBM introduced the first modular computer in 1964, firms in a wide range of industries have designed modular versions of their products (Baldwin and Clark 2000). Boeing's Dreamliner is one prominent example.[1] Other examples range from smartphones to residential homes to software programs.[2] Nowadays so many products are made by assembling separately produced modules that our times have been called the *Modular Age* (Garud et al. 2009).[3]

This paper is rooted in the notion that a firm's production function drives its organization—that the technological interdependencies among a firm's decisions shape the organization of those who make them. As such, we expect a firm with a modular product to have a distinct organization and a firm's organization to change as its product becomes more modular. Our goal is to understand the impact of modular production on the organization of firms.

To this end, we develop a model of a single firm with two key ingredients. The first is a modular production function, which we model as a network of decisions that is partitioned into *modules*, sets of decisions that require more coordination with each other than with decisions in the other modules. The second ingredient is a communication network. Each decision is made by a different agent who observes the relevant local conditions. The communication network specifies whom each agent tells about his local conditions, after which they all make their decisions. As in Arrow (1974), each communication link comes at a cost, capturing the time and energy it takes to communicate. The organizational problem is to design an optimal communication network, trading off the efficiency of decision making with the cost of communication.[4]

The challenge in designing an optimal network is the abundance of possibilities and absence of any apparent way to order them. Our main result shows that, despite the rich set of possibilities, optimal communication networks are characterized by a simple threshold rule. The key object is *module cohesion,* which captures how distinct a module is from the rest of the production network.

[1] See Section 7 and the references therein.

[2] See Baldwin and Clark (1997, 2000). See also the Wikipedia entries for Modular Design, Modular Programming, and Modular Building and the references therein.

[3] Herbert Simon anticipated the rise of modular product designs in an article in 1962, in which he observed that complex systems—large firms, mechanical watches, the human body—tend to be made up of modules, groups of elements with stronger within than across group interactions (Simon 1962). The prevalence of modular structures has since been corroborated by the literature on *community detection*, which has documented them in a wide variety of complex systems (Guimera et al. 2005, Meunier et al. 2009, and Fortunato 2010).

[4] As Kenneth Arrow put it: "*Since information is costly, it is clearly optimal, in general, to reduce the internal transmission...That is, it pays to have some loss in value for the choice of terminal act in order to economize on internal communication channels. The optimal choice of internal communication structures is a vastly difficult question*" Arrow (1974, p. 54).

In an optimal communication network, each agent tells his state to the other agents in his module, provided the need to coordinate with them is not too low, and he tells his state to all the agents in modules whose cohesion is above a threshold. The threshold differs across agents depending on the characteristics of their decisions and modules and the degree of uncertainty about their states.

This characterization has implications for what optimal communication networks look like, what structures they exhibit. Specifically, it implies that optimal communication networks have the *threshold property*. Loosely speaking, they exhibit a common receiver ranking that orders agents by how many others tell them about their local conditions. The threshold property implies that optimal communication networks cannot take the form of a *tree* or a *matrix*, which are structures that are commonly used to describe the allocation of authority in firms. Instead, we show that under a natural condition, optimal communication networks have a *core-periphery structure*, in which modules are partitioned into an intensely communicative core and a sparsely communicative periphery. Such structures are prevalent among social networks.[5] Even if the condition is not satisfied, optimal communication networks still resemble core-periphery structures. This is so because exhibiting the threshold property implies a *generalized core-periphery structure* that differs from the standard one because of the presence of a third group of modules whose members communicate too much to be in the periphery and too little to be in the core.

The characterization of optimal communication allows us to explore what happens when products become more modular—when the within-module needs for coordination become stronger relative to the across-module ones. If the production function is only somewhat modular, module cohesions are small, and, thus, similar to each other. This similarity favors all-or-nothing communication in which each agent either tells his state to all the other agents, or to none of them. If the production function is sufficiently modular, in contrast, module cohesions are not only large but also very different from each other. These large differences can make it optimal for agents to tell their states to some of the agents—those in the most cohesive modules—but not to others. Modularity, therefore, causes the fragmentation of communication networks and favors targeted communication to subsets of agents.

Having characterized optimal communication and explored comparative statics, we turn to applications. Even though modular production has so far received little attention in economics, it has long been the focus of a literature in management and software engineering. We revisit two tenets of this literature, the *Mirroring Hypothesis*, which conjectures that the optimal way to organize modular production is to mirror the production function, and *information hiding*, the

[5] See, for instance, Borgatti and Everett (2000) and Rombach et al. (2017) and the references therein.

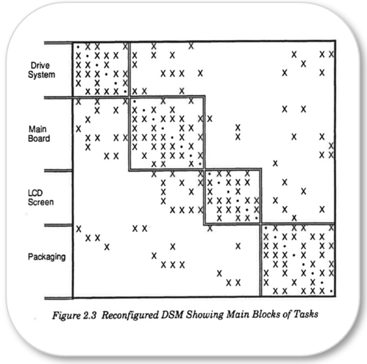

Figure 1: The *Design Structure Matrix* of a laptop in which each row and column corresponds to a task involved in designing the laptop and an "x" entry indicates a strong need for coordination between the corresponding tasks (replication of Figure 2.3 in McCord and Eppinger (1993)).

notion that modular production functions allow organizations to economize on communication costs by hiding information.[6] We conclude by discussing testable implications, where we focus, in particular, on the threshold nature of optimal communication.

The story of the first modular computer—IBM's *System/360*—illustrates the notion that modular production impacts organization.[7] Before the System/360, computers had been tightly integrated systems of their constituent parts. As a result, a change in a single critical component required the design of an entirely new computer. This feature made it difficult to adapt computers to changes in customer preferences and led IBM to seek a computer that could be made by assembling exchangeable modules. To design the new computer, IBM changed its organization. Engineers were divided into teams, each of which was put in charge of designing a different module. Across-team communication was limited, both because the teams were scattered across the globe and because the modular structure often made it unnecessary. This fragmented organization appears to have served IBM well. The System/360 became an enormous financial success and changed how computers have been designed ever since.[8]

Our approach to modeling modular production follows a path taken in systems engineering,

[6] See Thompson (1967), Conway (1968), Parnas (1972), Henderson and Clark (1990), Sanchez and Mahoney (1996), and, for a discussion of the literature, Colfer and Baldwin (2016).

[7] This account is based on Baldwin and Clark (1997, 2000).

[8] Baldwin and Clark (1997) argue that the organizational changes still reverberate today: "*But modularity also undermined IBM's dominance in the long run, as new companies produced their own so-called plug-compatible modules—printers, terminals, memory, software, and eventually even the central processing units themselves—that were compatible with, and could plug right into, the IBM machines. By following IBM's design rules but specializing in a particular area, an upstart company could often produce a module that was better than the ones IBM was making internally. Ultimately, the dynamic, innovative industry that has grown up around these modules developed entirely new kinds of computer systems that have taken away most of the mainframe's market share.*"

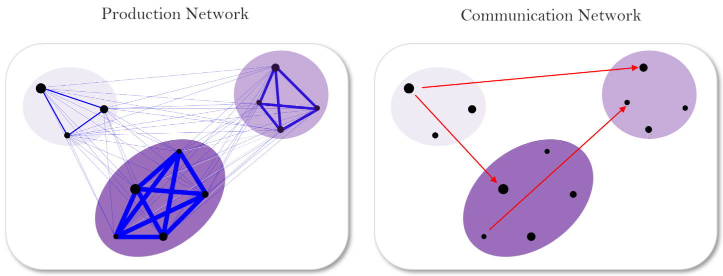

Figure 2: Left panel—The production function takes the form of a network with a *non-overlapping community structure*, where darker shading indicates more cohesive modules. Right panel—Given the production network, the principal designs optimal communication networks.

which characterizes products with *Design Structure Matrices* (Eppinger and Browning 2012). Figure 1 provides an example of such a matrix for a laptop. Each row and column refers to a task involved in designing the product, and the matrix entry indicates the need for coordination between the tasks. A product is *modular* if the Design Structure Matrix is a block matrix, as it is, at least approximately, in the case of the laptop. The blocks of the matrix form the modules, such as the LCD screen. The block structure implies a greater need for coordination between tasks involved in the LCD screen than between tasks in the LCD screen and, say, those in the Main Board.

Following this approach, we model the production function as a network of decisions with a *non-overlapping community structure* (Girvan and Newman 2002), as illustrated in the left panel in Figure 2. Every node represents a decision, an agent who makes the decision, and a state that captures the local conditions. The size of a node represents the importance of adapting the decision to its state, and the width of a link represents the importance of coordinating the two decisions it connects. Decisions are partitioned into groups whose members require more coordination with each other than with decisions in other groups. These *modules* are indicated by the shaded areas in the figure. This specification of the production network gives rise to an adjacency matrix with a block structure, similar to the Design Structure Matrix in Figure 1.

As we noted above, a key characteristic of the production network is the cohesions of its modules, which captures how distinct each is from the rest of the network. A module's cohesion is increasing in the number of decisions it contains and the need for coordination among them and is decreasing in the *degree of coupling*, the need for coordination between two decisions that belong to different modules. In Figure 2 a more cohesive module is indicated by darker shading.

This specification of the production network allows for a wide variety of differences in the technological structures of modular products. It allows for decisions to differ in their needs for adaptation and the degree of uncertainty about their states, for modules to differ in the number of decisions and the need for coordination between them, and for an arbitrary number of decisions and modules. The only substantive restriction is that coupling is homogeneous, that the need for coordination is the same across any two modules. We show in an extension that our results generalize naturally if we allow for heterogeneous coupling.

We embed this production network in a model of a firm's internal organization. To focus on communication costs, we follow the literature on team theory and abstract from incentive conflicts. We revisit this assumption in an extension where we show that our results continue to hold when we allow agents to have constant biases. Even though there are no incentive conflicts, agents may make poor decisions because they only observe their own states. What they learn about the other states is determined by a communication network, such as the one illustrated in the right panel in Figure 2. Each node is the same as in the production network, but the links are now directed and unweighted and indicate who tells whom his state. As we noted above, we follow Arrow (1974) in assuming that each directed link comes at a cost, which represents the time and effort it takes agents to understand each other, to learn the other's language or *code*.[9]

The principal's problem is to design the communication network. She takes the production network as given and builds the communication network, taking into account that each link comes at a cost. The contribution of this paper is to provide an analytical solution to this problem.

To get to the solution, suppose the principal adds a directed link to an arbitrary communication network. The marginal benefit of adding such a link is the additional expected revenue generated by the agents making better decisions. Learning the sender's state allows the receiver to coordinate his decision more closely with the sender's, which, in turn, allows the sender to adapt his decision more closely to his state. Communication facilitates coordination, which fosters adaptation.

The first key step to solving the principal's problem is finding that this marginal benefit is independent of what the receiver, or any other agent, knows about any other state. This feature implies that the principal's problem can be decomposed into independent subproblems. The principal can consider each agent separately and ask whom this agent should tell about his state.

The solution to each subproblem is driven by another property of the marginal benefit of adding a link to a communication network: the marginal benefit is larger, the more agents know the sender's

[9] As Kenneth Arrow put it: "*... I am thinking of the need for having made an adequate investment of time and effort to be able to distinguish one signal from another. Learning a foreign language is an obvious example of what I have in mind. The subsequent ability to receive signals in French requires this initial investment. There are in practice many other examples of codes that have to be learned in order to receive messages; the technical vocabulary of any science is a case in point. The issue here is that others have found it economical to use one of a large number of possible coding methods, and for any individual it is necessary to make an initial investment to acquire it*" Arrow (1974, pp. 39-42). See Cremer et al. (2007) for a formal investigation of this notion.

state. More generally, in equilibrium, expected revenue is supermodular in the set of agents who know any given state. This property implies that if the principal benefits from adding a link to a communication network she must also benefit from adding links from the sender to all the other agents in the receiver's module.

The second key step to solving the principal's problem is finding that the same logic applies across modules: if the principal benefits from adding a link across modules, she must also benefit from adding links from the sender to any agent in a module that is more cohesive than the receiver's.

Our main result then follows readily: in an optimal communication network, each agent tells his state to the other agents in his own module, provided the need to coordinate with them is not too low, and to any agent in another module whose cohesion is above a threshold. The threshold is lower, the more cohesive the sender's module is, the more important it is to adapt his decision to his state, and the more uncertainty there is about his state.

As we observed above, this result has implications for the structure of communication networks. We defer further discussion of these implications, as well as of comparative statics and applications, until after we have presented the model and formally derived the main result.

2 Related Literature

The existing literature on modular production is largely informal and lies mostly outside of economics. It goes back to Simon (1962), who observed that complex systems are often made up of modules and argued that this modular design facilitates adaptation. A similar point was made by Alexander (1964), who argued that a modular system design accelerates adaptation by allowing the system to adapt module by module.

We are not aware of papers that formalize these observations and do not attempt to do so in this paper. Instead, we follow the approach of a related literature in management that takes the modular design of products as given and explores its implications for the organization of firms. As we noted earlier, central arguments in this literature are the Mirroring Hypothesis, which posits that the internal organization of firms mirrors the modular design of the products they make, and the notion that information hiding can reduce communication costs in firms with modular products (see the references in Footnote 6).[10] Langlois and Robertson (1992) observed that modular production

[10] A related literature reverses the causality of the Mirroring Hypothesis and argues that the design of products mirrors the organization of the firms that designed them. In this view, a modular organization tends to design modular products. In software engineering, this view is known as *Conway's Law*, named after Melvin Conway, who observed that "*To the extent that an organization is not completely flexible in its communication structure, that organization will stamp out an image of itself in every design it produces*" Conway (1968, p. 30).

might not only affect the internal organization of firms but also their boundaries and, through this channel, the structure of industries. Baldwin and Clark (2000) document these dynamics in the context of IBM and the computer industry and provide an exhaustive discussion of modular production and its organization. We take a first step towards examining these issues through the lens of an economic model. In doing so, we take the boundaries of firms and the structure of their industries as given and examine the impact of modular production on their internal organization.

We focus, in particular, on the impact of modular production on communication structures. Such structures have long been recognized as an elemental feature of organizations (Arrow 1974). Moreover, recent empirical papers demonstrate that records of electronic communication make the patterns of such communication observable to outsiders (see, for instance, Yang et al. (2021) and Impink et al. (forthcoming)). This trend suggests that it may become possible to test predictions about communication structures and the information flows they generate, and to do so more readily than predictions about other aspects of internal organization (such as the allocation of decision rights, which have received much attention in the theoretical literature but have proven difficult to study empirically).

Communication structures, as well as information processing, are the focus of the large and long-running literature on team theory (for an early treatment see Marschak and Radner (1972) and for more recent surveys see Garicano and Prat (2013) and Garicano and Van Zandt (2013)). A central assumption in this literature is that agents share the same goal, but cognitive constraints make it difficult for them to communicate and process information. Our focus on communication structures and cognitive constraints places us firmly in this literature.

Communication structures depend on the technological interdependencies among the decisions agents make. In many settings, this interdependency arises because decisions must be both adapted to local conditions and coordinated with each other. March and Simon (1958) observed that this interdependency gives rise to a trade-off between adaptation and coordination that shapes the organization of firms. The contemporary organizational economics literature that studies how this trade-off affects organizations started with Dessein and Santos (2006), who explored implications for job design, and Alonso et al. (2008) and Rantakari (2008), who examined implications for the allocation of decision rights.

We relate to a set of papers that explore how the trade-off between adaptation and coordination shapes communication structures (Calvó-Armengol and de Martí Beltran 2009, Calvó-Armengol et al. 2015, Dessein et al. 2016, and Herskovic and Ramos 2020). These papers differ on various dimensions. Some focus on cognitive constraints while others allow for incentive conflicts. Some examine the centralized design of communication structures while others study their decentralized formation. And some assume that decisions need to be adapted to different states while others require them to be adapted to different signals about the same state. An assumption that is shared by all but one of these papers, though, is that the production network is complete, that the need for coordination between any two decisions is the same. This assumption rules out richer technologies, such as modular production.

The paper that allows for richer production networks, and is closest to ours, is Calvó-Armengol et al. (2015). They explore an organization in which each agent adapts his decision to the local conditions about which he is privately informed. In contrast to the above papers, but like us, they allow the needs for coordination to differ across decision pairs. They do not, however, assume that production has a non-overlapping community structure, and thus do not explore modular production. Their main result characterizes how much effort each agent puts into both explaining his state to others and understanding theirs.

To explore the impact of modular production on communication structures, we make use of the large toolbox of network economics. The payoff functions of our agents are quadratic, and their actions are continuous and exhibit strategic complementarities. This allows us to build on the literature on quadratic games on networks that started with Ballester et al. (2006). In recent contributions to this literature, Bergemann et al. (2017), Golub and Morris (2017), and Lambert et al. (2018) characterized optimal decision-making for general information and network structures. We draw on their results to determine the agents' decision-making for given communication networks. Our focus, though, is not on decision-making but on the prior stage in which the principal designs the communication network.

Finally, our paper contributes to a small but growing literature that studies centralized network design.[11] In an early paper in this literature, Baccara and Bar-Isaac (2008) explored the optimal design of a network among members of a criminal organization in which more links facilitate cooperation but also leave the organization more vulnerable to attack by law enforcement. The trade-off between the efficiency of interactions among members of a network and its increased vulnerability to attacks by outsiders is also at the center of Goyal and Vigier (2014), who were motivated by the optimal design and defense of computer networks. Even though we also explore centralized network design, we differ from these papers in both motivation and model.

[11] This literature is distinct from the large literature on endogenous network formation that started with Jackson and Wolinsky (1996) and Bala and Goyal (2000) and studies the emergence of networks from the decentralized decisions of agents. Some of the papers on communication structures we mentioned above, such as Herskovic and Ramos (2020), belong to this literature. In our model, instead, the network does not emerge endogenously from agents' communication decisions but is designed centrally by the principal.

3 Model

A firm consists of one principal and N agents. All parties are risk neutral and care only about the firm's profits. There are no incentive conflicts.

Production. Each agent $i \in \mathcal{N}$ makes a decision $d_i \in \left[-\overline{D}, \overline{D}\right]$ that is associated with a state $\theta_i \in \left[-\overline{\theta}, \overline{\theta}\right]$, where $\mathcal{N} = \{1, \ldots, N\}$ is the set of agents, and $\overline{D}$ and $\overline{\theta}$ are large but finite scalars. Revenue depends on how well each decision is adapted to its associated state and coordinated with the other decisions. Specifically, we follow Ballester et al. (2006) and assume that revenue is

$$r(d_1, \ldots, d_n) = \sum_{i=1}^{N} \left[-d_i^2 + 2a_i d_i \theta_i + \sum_{j=1}^{N} p_{ij} d_i d_j \right], \tag{1}$$

where $a_i > 0$ captures the importance of adapting decision d_i to its state θ_i, and the degree of strategic complementarity $p_{ij} \geq 0$ captures the need for coordination between decisions d_i and d_j.[12] The need for coordination is symmetric, that is, $p_{ij} = p_{ji}$, and p_{ii} is equal to zero. The interactions between decisions can, therefore, be represented by an undirected network, which we summarize in an $N \times N$ matrix $\boldsymbol{P}$ with entries p_{ij}. We assume that $\sum_{j=1}^{N} p_{ij} < 1$ for all $i \in \mathcal{N}$, which ensures that equilibrium decisions exist.

Modules. The decisions, and their associated states and agents, are partitioned into *modules*. There are M modules, and module $m \in \mathcal{M} = \{1, \ldots, M\}$ contains $n_m \geq 1$ decisions. Function $m(i)$ denotes the module decision d_i belongs to. For expositional convenience we adopt the convention that the first decision d_1, and its associated state and agent, belong to module 1, and assume that there are at least three modules, that is, $M \geq 3$.

The need for coordination between two decisions is stronger if they belong to the same module than if they belong to different ones. Specifically, the need for coordination between any two decisions d_i and d_j is given by $p_{ij} = p \geq 0$ if they belong to different modules and, abusing notation slightly, it is given by $p_{ij} = p_m \geq p$ if they belong to the same module m. The parameter p, therefore, captures the *degree of coupling*—the need for coordination across modules—while the parameter p_m captures the need for coordination within module m.

[12] A special case of this formulation is the widely-used payoff function in which payoffs are the weighted average of the quadratic difference between each decision and its state and between each pair of decisions (see, for instance, Alonso et al. (2008) and Calvó-Armengol and de Martí Beltran (2009)). Specifically, if $a_i = 1 - \sum_{j=1}^{N} p_{ij}$ for all $i \in \mathcal{N}$, we can re-write revenue as

$$r(d_1, \ldots, d_n) = \sum_{i=1}^{N} \left[-\left(1 - \sum_{j=1}^{N} p_{ij}\right)(d_i - \theta_i)^2 - \frac{1}{2}\sum_{j=1}^{N} p_{ij}(d_i - d_j)^2 \right] + \sum_{i=1}^{N} \left(1 - \sum_{j=1}^{N} p_{ij}\right)\theta_i^2,$$

where the last term is a constant.

Information. Each agent $i \in \mathcal{N}$ privately observes the realization of his state θ_i, which is independently drawn from a distribution with zero mean and variance σ_i^2.

Before the states are drawn, the principal can place communication links from any agent to any others. Each one of these links costs $\gamma > 0$, which captures the resources it takes for one agent to tell his state to another. If the principal places a communication link from agent i to agent j, agent i tells j the realization of his state θ_i. Communication, therefore, takes the form of a directed network, which we summarize in an $N \times N$ matrix $\boldsymbol{C}$. Entry c_{ij} is one if agent i tells agent j his state and it is zero if he does not. Moreover, since each agent i observes his own state, c_{ii} is always equal to one. Row $\boldsymbol{C}_i$ then summarizes the agents who learn θ_i and column $\boldsymbol{C}_{(j)}$ summarizes the states agent j learns about. The communication network, and all other information except for the agents' states, are common knowledge.

Organization. The principal's problem is to design an optimal communication network that maximizes expected revenue net of communication costs, that is, to solve

$$\max_{\boldsymbol{C}} \mathrm{E}[r(d_1, \ldots, d_N) \,|\, \boldsymbol{C}] - \gamma \sum_{i=1}^{N} \sum_{j \neq i} c_{ij}. \tag{2}$$

Timing. After the principal designs the communication network, agents learn their states and tell them to the other agents as specified by the network. Next, agents simultaneously make their decisions, after which the game ends.

We discuss key assumptions, such as homogeneous coupling and the absence of incentive conflicts in Section 8, after we solve the model in the next section, derive implications in Section 5, perform comparative statics in Section 6, and explore applications in Section 7.

4 Solving the Model

We start by determining Bayes-Nash equilibrium decisions for any given communication network. We then show that, given these equilibrium decision rules, we can simplify the principal's problem of designing an optimal communication network by decomposing it into independent subproblems. Finally, we characterize the solution to the principal's problem by solving these subproblems.

4.1 Decision-Making

After the agents have observed and communicated their states, they make the decisions that solve

$$\max_{d_i} \mathrm{E}\left[r(d_1, \ldots, d_N) \,\middle|\, \boldsymbol{C}_{(i)}\right] \text{ for all } i \in \mathcal{N}, \tag{3}$$

where $r(d_1, \dots, d_N)$ is revenue (1) and where $\boldsymbol{C}_{(i)}$ is the ith column of the communication matrix $\boldsymbol{C}$ that summarizes the states agent i knows. The best-response functions that follow from these optimization problems are given by

$$d_i = a_i \theta_i + \sum_{j=1}^{N} p_{ij} \mathrm{E}\left[d_j \,\middle|\, \boldsymbol{C}_{(i)}\right]. \tag{4}$$

Each agent's best response is the weighted sum of his state and the decisions he expects the other agents to make, where the weight on his state is a_i, and the weight on the decision he expects agent j to make is p_{ij}. To solve the system of best responses, note that $(\mathrm{diag}\boldsymbol{C}_j)\,\boldsymbol{P}\,(\mathrm{diag}\boldsymbol{C}_j)$ is the subgraph of the production network that consists of the nodes whose agents know state θ_j, and all the links between them. We can then state the following lemma.

LEMMA 1. *Equilibrium decisions are unique and given by*

$$d_i^* = \sum_{j=1}^{N} a_j \omega_{ij}(\boldsymbol{C}_j)\, \theta_j \text{ for all } i \in \mathcal{N}, \tag{5}$$

where $\omega_{ij}(\boldsymbol{C}_j)$ *denotes the* ij*th entry of* $(\boldsymbol{I} - (\mathrm{diag}\boldsymbol{C}_j)\,\boldsymbol{P}\,(\mathrm{diag}\boldsymbol{C}_j))^{-1}$.

The lemma shows that agent i's equilibrium decision d_i^* is the weighted sum of all states, where the weight on state θ_j is given by a_j, the importance of adapting decision d_j to θ_j, times $\omega_{ij}(\boldsymbol{C}_j)$, the ijth entry of $(\boldsymbol{I} - (\mathrm{diag}\boldsymbol{C}_j)\,\boldsymbol{P}\,(\mathrm{diag}\boldsymbol{C}_j))^{-1}$. This latter object has a natural interpretation in terms of walks on the production network: it is the value of all walks from d_i to d_j on the subgraph of the production network that consists only of decisions made by agents who know state θ_j.[13] If agent i does not know θ_s, for instance, d_i is not part of this subgraph, and so $\omega_{is}(\boldsymbol{C}_s) = 0$. Agent i puts no weight on θ_s, as one would expect. If, instead, θ_s is public, the subgraph encompasses the entire production network, and the weight agent i puts on θ_s is the value of all walks from d_i to d_s on the production network $\boldsymbol{P}$. Note that this is the case no matter what the agents know about the other states. This result reflects a general implication of the lemma that will be important for what follows: the weight agent i puts on state θ_s depends only on who knows θ_s and not on what agent i, or any other agent, knows about any other state.

The part of the equilibrium decision rules that will turn out to be important for the design of communication networks is the weight each agent's decision puts on his own state. To get an intuition for this weight, consider the simple example of a production network with just three modules, each of which consists of a single decision and associated state and agent. Suppose first

[13] A *walk* between d_i and d_j on the production network is a sequence of links that lead from d_i to d_j. Each link between two decisions in this sequence is associated with a discount factor, which is given by the need for coordination between them. The *value of a walk* is the product of these discount factors.

agent 1 does not tell his state to the other two agents. Agent 1 is then forced to adapt to his state autonomously, without the benefit of having the others coordinate their decisions with his. This limits the weight he puts on his own state to $a_1\omega_{11}((1,0,0)) = a_1$.

Suppose instead that agent 1 tells his state to agent 2 but not to agent 3. Since agent 2 cares about coordinating his decision with agent 1's, he puts some weight on θ_1. And since agent 1 also cares about coordinating his decision with agent 2's, this induces him to put more weight on his own state. Specifically, if agent 1 tells his state to agent 2, the weight he puts on θ_1 increases to

$$a_1\omega_{11}((1,1,0)) = a_1\left(1+\frac{p^2}{1-p^2}\right) > a_1.$$

Communication enables coordination, which, in turn, facilitates adaptation. The extent to which it does so is captured by the *coordination multiplier* $\omega_{11}(\boldsymbol{C}_1)$.[14]

A key property of the coordination multiplier is that it is supermodular. Suppose agent 1 tells his state to both agents 2 and 3. Since agents 2 and 3 care about coordinating with each other, and not just with agent 1, they put more weight on θ_1 than they would if agent 1 told his state to only one of them. This increase in the weights agents 2 and 3 put on θ_1, in turn, induces agent 1 to increase the weight he puts on his own state to

$$a_1\omega_{11}((1,1,1)) = a_1\left(1+\frac{p^2}{1-p^2}+\frac{p^2}{1-p^2}+\frac{2p^3}{(1-p^2)(1-2p)}\right),$$

where the last term in brackets captures the supermodularity.

These properties of the equilibrium decision rules hold in general, and we summarize them in the following corollary.

COROLLARY 1. *The weight $a_i\omega_{ii}(\boldsymbol{C}_i)$ that decision d_i^* puts on its state θ_i satisfies $\omega_{ii}(\boldsymbol{I}_i)\,a_i = a_i$, where $\boldsymbol{I}_i$ is the* ith *row of an $N \times N$ identity matrix, and it is increasing and supermodular in $\boldsymbol{C}_i$.*

Having characterized the agents' decision-making, we next turn to the principal's problem.

4.2 Simplifying the Principal's Problem

The principal's problem is to design a communication network that maximizes expected profits. It is useful to start by rewriting revenue (1) as

$$r(d_1,\ldots,d_N) = \sum_{i=1}^{N} a_i d_i \theta_i - \sum_{i=1}^{N} d_i\left(d_i - a_i\theta_i - \sum_{j=1}^{N} p_{ij} d_j\right).$$

[14] The coordination multiplier is related to the notion of *cycle centrality* in Talamàs and Tamuz (2017).

Substituting in the best-response functions (4), this expression simplifies to

$$r\left(d_1^*, \ldots, d_N^*\right) = \sum_{i=1}^{N} a_i d_i^* \theta_i + \sum_{i=1}^{N} \sum_{j=1}^{N} p_{ij} d_i^* \left(d_j^* - \mathrm{E}\left[d_j^* \,\middle|\, \boldsymbol{C}_{(i)}\right]\right). \tag{6}$$

Notice that by the law of iterated expectations the second term on the right-hand side is zero in expectation, which delivers the following result.

LEMMA 2. *Under equilibrium decision-making, expected revenue is given by*

$$R\left(\boldsymbol{C}\right) \equiv \mathrm{E}\left[r\left(d_1^*, \ldots, d_N^*\right)\right] = \sum_{i=1}^{N} a_i \mathrm{Cov}\left(d_i^*, \theta_i\right), \tag{7}$$

where $\mathrm{Cov}\left(d_i^*, \theta_i\right) = a_i \sigma_i^2 \omega_{ii}\left(\boldsymbol{C}_i\right)$.

The lemma shows that expected revenue boils down to how well each decision is adapted to its associated state. For expositional convenience, we interpret $a_i \mathrm{Cov}\left(d_i^*, \theta_i\right)$ as the expected revenue generated by agent $i \in \mathcal{N}$ and denote it by

$$R_i\left(\boldsymbol{C}_i\right) \equiv a_i \mathrm{Cov}\left(d_i^*, \theta_i\right) = a_i^2 \sigma_i^2 \omega_{ii}\left(\boldsymbol{C}_i\right).$$

The term $a_i^2 \sigma_i^2$ is the revenue agent i is expected to generate if he does not tell his state to any other agent and, thus, adapts his decision to his state autonomously. We refer to this term as the *value of autonomous adaptation* of decision d_i. The coordination multiplier captures how much more revenue agent i is expected to generate when he adapts his decision more closely to his state because other agents know his state.

The key property of agent i's expected revenue is that it depends on $\boldsymbol{C}_i$ but not on the rest of communication network $\boldsymbol{C}$. An additional agent learning θ_i increases agent i's coordination multiplier $\omega_{ii}\left(\boldsymbol{C}_i\right)$ and thus the weight $a_i \omega_{ii}\left(\boldsymbol{C}_i\right)$ he puts on his state. As a result, it also increases the expected revenue $a_i^2 \sigma_i^2 \omega_{ii}\left(\boldsymbol{C}_i\right)$ he generates. In contrast, agent i, or any other agent, learning any other state does not affect $\omega_{ii}\left(\boldsymbol{C}_i\right)$ and thus leaves the weight agent i puts on his own state, and the revenue he is expected to generate, unchanged.

This property of expected revenue is key because it implies that the principal's problem is separable. Instead of solving the overall problem (2) head on, the principal can consider each agent separately and ask whom this agent should tell about his state. The answer to whom agent $i \in \mathcal{N}$ should tell about θ_i is independent of whom any other agent should tell about his own state. We, therefore, have the following.

PROPOSITION 1. *An optimal communication network solves the principal's problem* (2) *if and only if it solves the N independent subproblems*

$$\max_{\boldsymbol{C}_i} R_i\left(\boldsymbol{C}_i\right) - \gamma \sum_{j \neq i} c_{ij}. \tag{8}$$

This separability result greatly facilitates the principal's quest for an optimal communication network. We can further simplify the problem by recalling that the coordination multiplier is supermodular. Together with the linearity of communication costs, this implies that the subproblems in the proposition are also supermodular. For any given parameter values, the principal's problem can, therefore, be solved using standard algorithms that maximize supermodular functions in polynomial time (see, for instance, chapter 10.2 in Murota (2003)). Our goal, though, is to solve the problem analytically, and we do so in the next section.

4.3 Optimal Communication Networks

The separability result in Proposition 1 allows us to solve the principal's problem by considering each agent separately and asking whom he should tell about his state. The answer depends critically on the *module cohesion* of each module, which we define as

$$x_m = \frac{1}{1 + p - \left(n_m - 1\right)\left(p_m - p\right)} \text{ for } m \in \mathcal{M}.$$

As we noted earlier, a module's cohesion captures how distinct it is from the rest of the production network. To see this interpretation, note that the last term in the denominator is the *excess need for coordination* of any decision d_i in the module, the difference between its total need for coordination $\sum_{j=1}^{N} p_{ij} = \left(n_m - 1\right) p_m + \left(N - n_m\right) p$ and $\left(N - 1\right) p$, the value $\sum_{j=1}^{N} p_{ij}$ would take if the need for coordination within module m were the same as that across the modules. A module is, therefore, more cohesive than another module if its decisions have a higher excess need for coordination.[15]

The first step in answering whom an agent should tell about his state is to express his expected revenue in terms of the model's primitives, which we do in the next lemma. To economize on notation, and without loss, the lemma and subsequent discussion focus on agent 1 who, recall, belongs to module 1.

LEMMA 3. *Agent 1's expected revenue is given by*

$$R_1\left(\boldsymbol{C}_1\right) = a_1^2 \sigma_1^2 \left(\frac{1 + \left(p_1 - p\right) x_1\left(\tilde{n}_1\right)}{1 + p_1} + \frac{p x_1\left(\tilde{n}_1\right)^2}{1 - p \sum_{m=1}^{M} \tilde{n}_m x_m\left(\tilde{n}_m\right)} \right), \tag{9}$$

[15] Our definition of module cohesion is close to the definition of cohesion in Morris (2000). Applied to our setting his, like ours, is increasing in n_m and p_m and decreasing in p.

where

$$x_m(\tilde{n}_m) = \frac{1}{1 + p - (\tilde{n}_m - 1)(p_m - p)} \text{ for } m \in \mathcal{M},$$

and $\tilde{n}_m$ is the number of agents in module m who know agent 1's state.

We already know that agent 1's expected revenue is the product of $a_1^2\sigma_1^2$—the value of autonomous adaptation of his decision—and the coordination multiplier $\omega_{11}(\boldsymbol{C}_1)$. The lemma shows that the coordination multiplier takes a simple form from which we can glean properties of optimal communication networks we show formally below.

Notice first that the coordination multiplier is convex in each $\tilde{n}_m$, the number of agents in module m who know agent 1's state. This convexity reflects the supermodularity of the coordination multiplier and implies that agent 1 either tells his state to all the agents in a module or to none of them. The remaining question then is what modules agent 1 should tell about his state.

Next notice that the coordination multiplier depends on the characteristics of modules $m > 1$ only through the sum $\tilde{n}_2 x_2(\tilde{n}_2) + \cdots + \tilde{n}_M x_M(\tilde{n}_M)$. If agent 1 does not tell his state to agents in module m, $\tilde{n}_m = 0$ and the characteristics of module m do not enter agent 1's expected revenue. If, instead, agent 1 does tell his state to agents in module m, $\tilde{n}_m = n_m$ and the characteristics of module m enter agent 1's expected revenue only through its scaled cohesion $n_m x_m$, where we are using the fact that $x_m = x_m(n_m)$.

Finally, notice that the coordination multiplier is convex in the sum of the scaled cohesions of the modules $m > 1$ that agent 1 tells about his state. This convexity once again reflects the supermodularity of the coordination multiplier. It is important here because it implies that if it is profitable to expand the set of informed modules by one additional module, it must also be profitable to expand it further by adding a second module, provided that the second is no less cohesive than the first. The answer to whom agent 1, or any other agent, should tell about his state follows directly from this claim.

PROPOSITION 2. *There exist thresholds $\lambda_i \geq 0$ and $\mu_i \geq 0$ such that it is optimal for agent $i \in \mathcal{N}$ to tell his state to a different agent $j \in \mathcal{N}$ if and only if:*

(*i.*) *agent j belongs to the same module $m(j) = m(i)$ with coordination need $p_{m(i)} \geq \mu_i$, or*

(*ii.*) *agent j belongs to a different module $m(j) \neq m(i)$ with cohesion $x_{m(j)} \geq \lambda_i$.*

Threshold λ_i is increasing in γ, decreasing in $a_i^2\sigma_i^2$, p, p_m, and n_m for any $m \in \mathcal{M}$, and independent of $a_k^2\sigma_k^2$ for any $k \in \mathcal{N} \setminus \{i\}$. The comparative statics of threshold μ_i are the same, except that it is independent of $p_{m(i)}$.

The key result in the proposition is that across-module communication is determined by a threshold rule on module cohesion. An agent tells his state to agents in another module if that module is sufficiently cohesive. To see the intuition for why cohesion matters, suppose agent 1 tells his state to the agents in one additional module, say module 2. Since the agents in module 2 care about coordinating their decisions with agent 1's, this induces them to put some weight on θ_1. And since agent 1 also cares about coordinating his decision with theirs, this, in turn, allows him to adapt his decision more aggressively to his state. Communication facilitates coordination, which fosters adaptation, as we noted before.

The agents in module 2, though, do not only care about coordinating their decisions with agent 1's, they also care about coordinating them with each other's. This is where the cohesion of module 2 comes into play. The more cohesive the module is—the higher the excess need for coordination of its decisions is—the *more* its agents care about coordinating their decisions with each other and, thus, the *more* weight their decisions put on θ_1. Cohesion, therefore, matters because it strengthens the link between communication and coordination which, ultimately, leads to more adaptation.

Cohesion also matters for within-module communication and does so for the same reason. Suppose we decompose module 1 into agent 1 and a *submodule* that consists of all the other agents and, thus, has cohesion $x_1(n_1 - 1)$. The benefit of agent 1 telling the agents in the submodule about his state is once again higher the more cohesive the submodule is. The difference between within and across-module communication is that agent 1's decision is more tightly coupled to the submodule than to any other module.[16] This is why the rule that determines whether agent 1 should tell his state to the other agents in his module is a threshold on p_1 rather than on $x_1(n_1 - 1)$. And it is why, other things equal, an agent is more likely to tell his state to the other agents in his own module than to those in another module, as shown in the next corollary.

COROLLARY 2. *Suppose agent i's module $m(i)$ is at least as cohesive as another module m, that is, $x_{m(i)} \geq x_m$. It cannot be optimal for agent i to tell his state to the agents in module m but not to the other agents in module $m(i)$.*

To get an intuition for the comparative statics in the proposition, consider Figure 3, where we again focus on agent 1. There are four modules labeled in decreasing order of their cohesion: $x_1 \geq x_2 \geq x_3 \geq x_4$. We denote the revenue agent 1 is expected to generate if he tells agents in modules 1 to $\ell \in \{1, ..., 4\}$ about his state by $R_1(\ell)$. The blue curve in the figure is the piecewise linear extension of expected revenue, which we denote by $\overline{R}_1(\ell)$, and the blue line is a continuous representation of communication costs $\gamma\left(\sum_{m=1}^{\ell} n_m - 1\right)$. The changing curvature of expected revenue $\overline{R}_1(\ell)$ reflects the economic forces at work in our model. Supermodularity pushes towards

[16] More precisely, the need for coordination between agent 1's decision and the submodule, given by p_1, is higher than that between agent 1's decision and those in any other module, which is given by $p \leq p_1$.

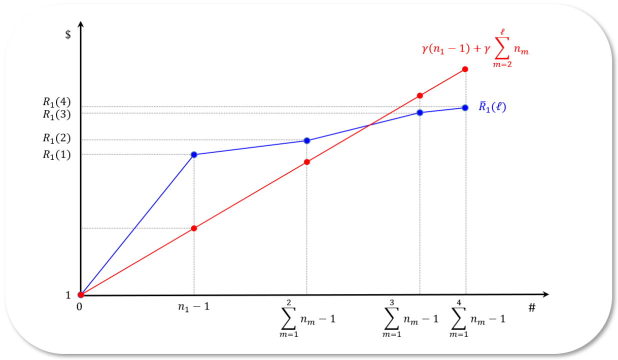

Figure 3: Determining optimal communication networks for agent 1 (drawn for parameter values $n_1 = n_2 = n_3 = 5$, $n_4 = 2$, $p_1 = p_2 = p_3 = 0.2$, $p_4 =$, $p = 0.01$, and $a_1\sigma_1 = 1$).

convexity while the modular structure of the production function pushes towards concavity.

Now consider the effect of changes in the parameters on the cost and benefit curves. The slope of each line segment in the benefit curve is the marginal expected revenue generated by telling the additional agents in the corresponding module about θ_1, divided by the number of additional agents. This per node marginal benefit is larger, the higher the value of autonomous adaptation for agent 1's decision is. And it is larger, the more agents know θ_1 and the higher the need for coordination among them is. An increase in $a_1^2\sigma_1^2$, n_m, p_m, or p, therefore, steepens the benefit curve, which favors telling agents in more modules about θ_1. In contrast, an increase in the communication costs γ steepens the cost curve, which favors telling agents in fewer modules about θ_1.

The characterization of the optimal communication networks in Proposition 2 is our main result. To derive it, we first showed that the principal's problem can be decomposed into independent subproblems. While the subproblems are independent, though, their solutions are related in a way that has implications for what types of structures optimal communication networks exhibit. We explore these implications next.

5 Implications for Network Structures

We now turn to the implications of the characterization result in Proposition 2 for the structure of optimal communication networks. Our starting point is that optimal communication networks have the *threshold property*. Once we have established this result, we show that the class of networks that have this property excludes a number of well-known structures and includes others.

Intuitively, a communication network has the threshold property if it exhibits a *common* receiver

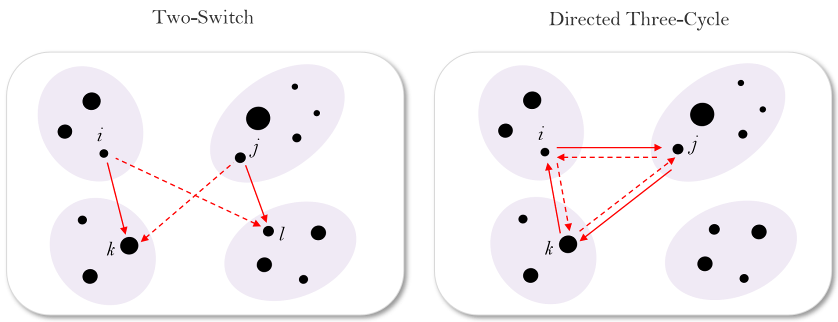

Figure 4: Communication networks exhibiting two-switches and directed three-cycles, where shaded areas indicate modules and a dashed arrow indicates the absence of a communication link.

ranking—if agents can be ranked by whether they are told about a state in a different module, and the *same* ranking determines the order in which they are told about any state that belongs to a different module. Formally, we have the following.

DEFINITION. *A communication network has the "threshold property" if there exist sender thresholds* $\{s_1, \ldots, s_N\}$ *and receiver scores* $\{r_1, \ldots, r_N\}$ *such that for any two agents* $i, j \in \mathcal{N}$ *who belong to different modules, agent* i *tells agent* j *his state if and only if* $r_j \geq s_i$.

In line with the above intuition, the definition requires a common receiver ranking. Suppose agents i and j belong to different modules, and agent i's receiver score is larger than agent j's, $r_i > r_j$. Agent i then has a higher receiver rank than agent j does, he is told about any state in any module $m \in \mathcal{M} \setminus \{m(i), m(j)\}$ that agent j is told about, and possibly others.

Optimal communication networks exhibit a common receiver ranking. Suppose agent i's module is more cohesive than agent j's, $x_{m(i)} > x_{m(j)}$. The characterization result in Proposition 2 then implies that, in an optimal communication network, agent i is told about any state in any module $m \in \mathcal{M} \setminus \{m(i), m(j)\}$ that agent j is told about, and possibly others. We can, therefore, obtain the receiver scores the definition calls for by setting $r_i = x_{m(i)}$, and the sender thresholds by setting $s_i = \lambda_i$, for all $i \in \mathcal{N}$. Doing so delivers the result.

COROLLARY 3. *Optimal communication networks have the threshold property.*

The class of networks that exhibit the threshold property excludes a number of well-known structures and includes others. To describe what it excludes, we adapt Cloteaux et al. (2014) and show that the threshold property rules out certain forbidden subgraphs. This then allows us to rule out any structure that contains one or more of these subgraphs. The relevant subgraphs are *two-switches* and *directed three-cycles*, which we illustrate in Figure 4 and define as follows.

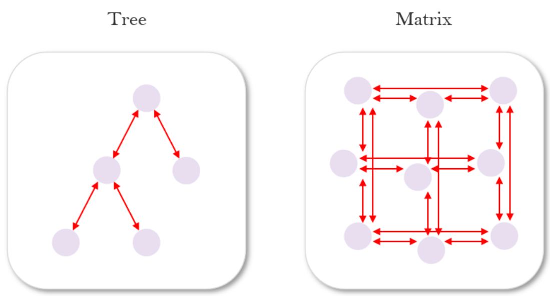

Figure 5: Illustration of trees and matrices, where shaded areas indicate modules and an arrow indicates communication links from all agents in one module to all the agents in the other.

DEFINITION. *A communication network contains a "two-switch" if there are four agents* i, j, k, $\ell \in \mathcal{N}$ *who belong to different modules such that agent* i *tells his state to* k *but not to* ℓ *and agent* j *tells his state to* ℓ *but not to* k.

A communication network contains a "directed three-cycle" if there are three agents $i, j, k \in \mathcal{N}$ *who belong to different modules such that agent* i *tells his state to* j*, but not the reverse, agent* j *tells his state to* k*, but not the reverse, and agent* k *tells his state to* i*, but not the reverse.*

The next lemma shows that the threshold property rules out these subgraphs.

LEMMA 4. *A communication network with the threshold property contains no two-switches or directed three-cycles.*

To see why two-switches cannot be part of an optimal communication network, consider agents i, j, k, and ℓ from the definition of two-switches and their illustration in Figure 4. If it is optimal for agent i to tell his state to agent k but not to agent ℓ, it must be that module $m(k)$ is more cohesive than module $m(\ell)$. If module $m(k)$ is more cohesive than module $m(\ell)$, though, it cannot be optimal for agent j to tell his state to agent ℓ but not to agent k. The intuition for why three-cycles cannot be part of an optimal communication network is similar.

The lemma is useful because it rules out the many well-known network structures that do contain two-switches or three-cycles (or both). Two structures are of particular relevance to us: trees and matrices. These structures—which we illustrate in Figure 5 and define below—are commonly used to describe the allocation of decision rights in firms. As the examples in the figure suggest, though, they all contain at least one two-switch. It then follows from the lemma, together with Corollary 3, that any optimal communication structure must be distinct from these well-known ones.

To state the result, we first define the two structures as follows.

DEFINITION. *A communication network is a "tree" if there is a partition of the set of modules into hierarchical levels* $1, \ldots, H$ *with* $H \geq 3$ *such that there is one module in level* 1 *and at least two modules in level* $h \in \{2, \ldots, H\}$. *Each module in level* $h \in \{2, \ldots, H\}$ *is associated with a unique "predecessor module" in level* $h-1$. *Any two agents* $i, j \in \mathcal{N}$ *who belong to different modules tell each other their states if and only if* $m(i)$ *is a predecessor of* $m(j)$ *or vice versa.*

A communication network is a "matrix" if there are two partitions of the set of modules into horizontal and vertical teams, with at least two horizontal teams containing at least two modules, and such that any two modules in the same horizontal team are in different vertical teams. Any two agents $i, j \in \mathcal{N}$ *who belong to different modules tell each other their states if and only if their modules belong to the same horizontal or vertical team.*

We then have the following.

COROLLARY 4. *A communication network is not optimal if it is a tree or matrix.*

The fact that the threshold property rules out well-known network structures raises the question of what structures it is consistent with. We show next that under a natural condition, the threshold property gives rise to a *core-periphery structure* with intense communication in the core, no across-module communication within the periphery, and sparse communication between the core and the periphery. As we noted in the introduction, such structures are prevalent among social networks.

DEFINITION. *A communication network has a "core-periphery structure" if the set of agents can be partitioned into a core and a periphery such that* (*i.*) *an agent in the core tells his state to all the agents in other modules in the core, and* (*ii.*) *an agent in the periphery does not tell his state to other agents in the periphery who do not belong to his module.*

The next proposition shows that optimal communication networks have a core-periphery structure if the agents who have a low sender threshold also have a high receiver rank. Loosely speaking, agents who talk a lot then also hear a lot, and thus form the core, while those who hear little also talk little and find themselves in the periphery.

PROPOSITION 3. *Suppose that for any agents* $i, j \in \mathcal{N}$, $\lambda_i \leq \lambda_j$ *if and only if* $x_{m(i)} \geq x_{m(j)}$. *Any optimal communication network then has a core-periphery structure in which the agents who belong to the most cohesive modules form the core.*

The condition in the proposition is satisfied if decisions that belong to more cohesive modules have sufficiently higher values of autonomous adaptation $a_i^2 \sigma_i^2$ than those that belong to less cohesive ones. This follows from the results in Proposition 2 that an increase in $a_i^2 \sigma_i^2$ reduces λ_i but leaves λ_j, and cohesions $x_{m(i)}$ and $x_{m(j)}$, unchanged. The condition is satisfied, for instance, if decisions differ only in the size of the modules they belong to, as in the example illustrated in Figure 6.

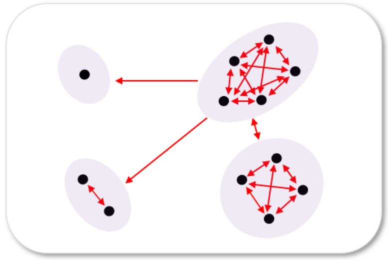

Figure 6: An optimal communication network with a core-periphery structure when decisions differ only in the size of the modules they belong to ($\gamma = 0.00015$, $n_1 = 5$, $n_2 = 4$, $n_3 = 2$, $n_4 = 1$, $p = 0.01$, $a_i\sigma_i = 1$, and $p_{m(i)} = 0.05$ for all $i \in \mathcal{N}$). An arrow from one module to another indicates communication links from all agents in the former to all agents in the latter.

Even if the condition in the proposition does not hold, the structures of optimal communication networks are akin to core-periphery structures. In particular, we show in Lemma 5 in the appendix that exhibiting the threshold property implies a *generalized core-periphery structure*: modules can be partitioned into a core, a periphery, and what we refer to as a *suburban periphery.* Suburban periphery modules are involved in too much communication to the periphery and too little communication from the core to belong to either.

6 The Effect of an Increase in Modularity

Having characterized optimal communication networks for modular production functions, we now explore how these networks change as production becomes more modular—as the within-module needs for coordination become stronger relative to the across-module ones. We show that such a change leads to the fragmentation of communication and explain what form it takes.

Our starting point is a *weakly modular* production function in which the difference in the within- and across-module needs for coordination is small.

DEFINITION. *Production network $\boldsymbol{P}$ is "weakly modular" if for all $m \in \mathcal{M}$, $n_m \geq 2$ and*

$$p_m - p < \frac{p^3\,(1+p)}{(n_m - 1)\,(1 + p\,(1 + n_m))}.$$

In the previous sections we saw that optimal communication is shaped by two economic forces. On the one hand, supermodularity of expected revenue pushes towards corner solutions in which each agent either tells his state to all the other agents or to none of them. On the other hand, the modular structure of the production network pushes towards interior solutions in which agents tell their states to some of the other agents but not to all of them. Our first result here is that if a production network is only weakly modular, supermodularity dominates. Optimal communication is then *all or nothing*, which we define as follows.

DEFINITION. *For a given production network $\boldsymbol{P}$, optimal communication is "all or nothing" if there exist thresholds γ_i for all $i \in \mathcal{N}$ such that it is optimal for agent i to tell his state to all agents if $\gamma \in [0, \gamma_i]$ and to no agents if $\gamma \in [\gamma_i, \infty)$.*

We then have the following.

PROPOSITION 4. *If production network $\boldsymbol{P}$ is weakly modular, optimal communication is all or nothing.*

Intuitively, if the production network is only weakly modular, its modules cannot be very pronounced. Their cohesions must be small and, thus, similar to each other. Supermodularity then implies that it cannot be optimal for an agent to tell his state to agents in some modules but not to those in others.

Next we turn to how the communication network changes if the production network becomes more modular. Recall that the total need for coordination of decision d_i is given by

$$\sum_{j=1}^{N} p_{ij} = \left(n_{m(i)} - 1\right) p_{m(i)} + \left(N - n_{m(i)}\right) p \text{ for any } i \in \mathcal{N}. \tag{10}$$

To avoid conflating the effect of an increase in modularity with those due to changes in the total needs for coordination, we focus on increases that are *weight-neutral*, ones that do not change (10).

DEFINITION. *A "weight-neutral increase in modularity of size μ" of production network $\boldsymbol{P}$ reduces the degree of coupling from p to $p(\mu) \equiv (1-\mu)\, p$ and increases the within-module needs for coordination from p_m to*

$$p_m(\mu) \equiv p_m + \mu p \frac{N - n_m}{n_m - 1} \text{ for all } m \in \mathcal{M},$$

where $\mu \in (0,1)$. We denote the resulting production network by $\boldsymbol{P}(\mu)$.

Suppose now we start with a weakly modular production function and increase its modularity in a weight-neutral manner. Our next result shows that if the increase is sufficiently large, optimal communication changes from all or nothing to being *fully fragmented*: in addition to the corner solutions, each interior solution is optimal for some communication costs.

DEFINITION. *For a given production network $\boldsymbol{P}$, optimal communication is "fully fragmented" if for each $i \in \mathcal{N}$, there exist thresholds $\gamma_i^M < \gamma_i^{M-1} < \cdots < \gamma_i^1$ such that it is optimal for agent i to tell his state to all agents if $\gamma \in \left[0, \gamma_i^M\right]$, to all agents in some number of modules n, $1 \leq n \leq M-1$, if $\gamma \in \left[\gamma_i^{n+1}, \gamma_i^n\right]$, and to no one if $\gamma \in [\gamma_i^1, \infty)$.*

With this definition at hand, we can state the result.

PROPOSITION 5. *If production network $\boldsymbol{P}$ is weakly modular, and the modules differ in their cohesions, there exists a $\overline{\mu} \in (0,1)$ such that for any $\boldsymbol{P}(\mu)$ with $\mu \in [\overline{\mu}, 1)$:* (i.) *optimal communication is fully fragmented and* (ii.) *if it is optimal for agent $i \in \mathcal{N}$ to tell his state to some module $m \in \mathcal{M}$, it is optimal for him to tell his state to the agents in his own module $m(i)$.*

The intuition for the first part of the proposition follows from the effect of an increase in modularity on module cohesion. As the production network becomes more modular, its modules become more cohesive. Because the increase in cohesion is larger the more cohesive a module is, this causes the module cohesions to diverge. At some point, the modules have such different cohesions that the modular structure of the production network dominates the supermodularity of expected revenue. There then exist communication costs for which it is optimal for agents to tell their states to the agents in some modules but not to those in others. This conclusion does not hold, though, if the modules all have the same cohesion before the increase in modularity. If they did have the same cohesion before the increase, they would still have the same afterwards. Communication would then be all or nothing, no matter how much modularity increased.

The second part of the proposition speaks to whom agents tell about their states. We already know from the characterization result that agents tell their states to agents in other, more cohesive modules before they tell them to those in other, less cohesive ones. The second part of the proposition adds to this result by showing that, after a sufficient increase in modularity, agents tell their states to the agents in their own modules before they tell them to those in others.

Together Propositions 4 and 5 show that an increase in modularity is associated with the fragmentation of communication networks and describe what form it takes. An increase in modularity is associated with fragmentation because it makes modules more distinct from each other and does so on the dimension that matters—their cohesions. This change can lead to the removal of communication links from agents to others in the least cohesive modules, and it can lead to the addition of links from agents to others in their own modules and in the most cohesive, other modules.

7 Applications

We now turn to broader applications of the model. We do so in the context of the Mirroring Hypothesis and information hiding, the two tenets of the management and software engineering literatures on modular production. After discussing each, we return to the theoretical predictions of the model and provide guidance for empirical work.

7.1 Mirroring Hypothesis

The *Mirroring Hypothesis* conjectures that the optimal way to organize modular production is to mirror the production function, to enable intense communication within modules and accept sparse communication across (for references, see Footnote 6). In our setting, an organization mirrors its production function if the principal places communication links within modules but not across.

DEFINITION. *An organization "mirrors" the production function if agent $i \in \mathcal{N}$ tells agent $j \in \mathcal{N}$ about his state if and only if they belong to the same module.*

The Boeing Company's experience with the 787 Dreamliner illustrates the Mirroring Hypothesis and why it may fail.[17] The Dreamliner was designed to be modular precisely because it allowed Boeing to outsource the development and production of most modules to independent suppliers. Suppliers delivered the finished modules to Boeing's factory in Everett, where its workers put them together with the tail fin, the only major module still made by Boeing itself. To the extent that firm boundaries hamper communication, this way of organizing the production of the Dreamliner is broadly in line with the Mirroring Hypothesis.[18]

The intention of Boeing's organizational strategy was to speed up the development of the Dreamliner and save production costs. Instead, persistent coordination problems among the suppliers, and between them and Boeing, caused long delays and large cost overruns. These problems were so severe that Boeing was forced to modify its organizational strategy. One change it made was to bring the production of some major modules, specifically those forming the fuselage, back in-house by acquiring the relevant factories from its suppliers. The other major change was to establish an organizational unit—the *Production and Integration Center*—which was tasked with facilitating communication and collaboration between Boeing and its remaining, independent suppliers.[19] Eventually, these efforts succeeded in overcoming Boeing's development and production problems.[20]

Boeing's experience illustrates that mirroring might fail because the need to coordinate across modules can necessitate intense across-module communication, even when products are highly

[17] This account is based on McDonald and Kotha (2015) and Brown and Garthwaite (2016). See also Tadelis and Williamson (2013).

[18] For evidence that vertical integration fosters the flow of intangibles see, for example, Atalay et al. (2014).

[19] The center provided an around-the-clock communication channel between Boeing and its suppliers: "*An industry observer explained: 'Suppliers as far afield as Australia, Italy, Japan and Russia could call in through translators and show Boeing engineers in the center close-up images of their components using high-definition handheld video cameras. ... Immediate multimedia communications have eliminated the problem of unclear email exchanges between distant engineers who work on opposite ends of the clock.'*" (McDonald and Kotha 2015, p. 10)

[20] For instance, McDonald and Kotha (2015, p. 11): "*Industry experts agree that the PIC* [Production Integration Center] *was pivotal in stabilizing the 787's supply chain, as measured by fewer delays stemming from design changes due to flight tests and less traveled work. (Traveled work is work assigned to a supplier but later sent to Everett, for scheduling reasons, for Boeing workers to complete.) Thanks to improved communication and collaboration, the time devoted to problem resolution between partner engineers and Boeing was significantly shortened.*"

modular. This is in line with the supermodularity of expected revenue that is at the heart of the model. Across-module communication always improves decision-making but it does so especially if agents tell their states to others in their own modules. For across-module communication to be unprofitable nevertheless, the degree of coupling must be sufficiently low. Just how low it needs to be depends on the characteristics of the production network, as described in the next proposition.

PROPOSITION 6. *There exists some $\gamma > 0$ such that mirroring is optimal if and only if $p \leq \overline{p}$, where $\overline{p}$ is decreasing in the module characteristics n_m and p_m for all $m \in \mathcal{M}$.*

For mirroring to be optimal, then, there cannot be any modules that consist of too many decisions or require too much coordination. If there are, across-module communication remains important. Arguably, this is why mirroring failed at Boeing. Its production function was modular but still complex, comprising modules that involved many decisions which required a high degree of coordination. The model suggests that, for such a production function, across-module communication can be essential, even when the degree of coupling is low.

This brings us back to Boeing's decision to respond to the failure of its initial organizational strategy by insourcing the production of the fuselage and establishing the Production and Integration Center. This response created a core-periphery structure in which the in-house modules formed the core and the outsourced ones the periphery. To the extent that the tail fin and the fuselage were the most cohesive ones, this response is consistent with the optimal design of communication networks in our model. The formation of an organizational unit tasked with improving communication between Boeing and its suppliers is also in line with the formation of a core-periphery structure, which emphasizes communication between the core and the periphery rather than within the periphery itself. Overall, Boeing's experience with the development and production of the Dreamliner is broadly consistent with the model.

7.2 Information Hiding

The development of OS/360—the operating system for the System/360—illustrates the danger of failing to economize on communication costs in large-scale development projects.[21] Frederick Brooks, the manager in charge of the project, insisted on communication of all project-relevant information among its hundreds of programmers. He did so with the help of a continuously updated *workbook* that documented all aspects of the project and to which all programmers had access at all times.[22] The problems with this approach soon became apparent (Brooks 1995, p. 77):

[21] This account is based on Brooks (1995) and Langlois (2002).
[22] See chapter seven in Brooks (1995).

"*Our project had not been under way six months before we hit another problem. The workbook was about five feet thick! If we stacked up the 100 copies serving programmers in our offices in Manhattan's Time-Life Building, they would have towered above the building itself. Furthermore, the daily change distribution averaged two inches, some 150 pages to be interfiled in the whole. Maintenance of the workbook began to take a significant time from each workday.*"

To mitigate these problems, Brooks invested in reducing communication costs by switching to microfiche (Brooks 1995, p. 77). In terms of the model, he reduced per-link communication cost γ. The underlying problem, though, remained. In fact, the observation Brooks is best known for today is that, because of the increasing costs of communication, adding programmers to a delayed software project will only delay it further.[23]

What Brooks did not do was to engage in *information hiding*, to economize on communication costs by reducing the number of communication links, rather than the cost per link. This approach, which has since become a basic principle of computer programming, was first advocated by David Parnas, who also coined the term (Parnas 1972).[24] Specifically, Parnas suggested partitioning programmers into organizational units in a way that minimizes technological interdependencies and thus reduces, or even eliminates, the need for communication between them. Brooks was aware of Parnas' notion of information hiding but dismissed it promptly out of concerns for coordination problems between organizational units (Brooks 1995, p. 78):

"*D. L. Parnas of Carnegie Mellon University has proposed a still more radical solution. His thesis is that the programmer is most effective if shielded from, rather than exposed to the details of construction of system parts other than his own. ... While that is definitely sound design, dependence upon its perfect accomplishment is a recipe for disaster.*"

The question Parnas (1972) raised is how to carve an organization into informationally isolated units without causing the coordination problems Brooks worried about. Our model provides a lens through which to explore this question. To avoid confusion between properties of the production function and communication network, we refer to Parnas' organizational units as *teams* and say that a communication network that is partitioned into teams has a *team structure*.

[23] See the Wikipedia entries for *The Mythical Man-Month* (the title of Brooks' book) and for *Brooks' Law*. See also Brooks (1995, pp. 17-19): "*In tasks that can be partitioned but which require communication among the subtasks, the effort of communication must be added to the amount of work to be done. ... If each part of the task must be separately coordinated with each other part, the effort increases as $n(n-1)/2$. ... The added effort of communicating may fully counteract the division of the original tasks and bring us to the situation of Fig. 2.4* [which shows a U-shaped relationship between project completion time and number of workers]. *... Adding more men then lengthens, not shortens, the schedule.*"

[24] See, for instance, the Wikipedia entry for Information Hiding.

DEFINITION. *A communication network has a "team structure" if it partitions agents into two or more subsets, or "teams," such that two agents tell each other their states if and only if they belong to the same team.*

There are many possible team structures, including the type of mirroring we discussed in the previous section, in which each module forms a separate team. The next corollary draws on different aspects of optimal communication we derived above to characterize the type of team structures that can be optimal.

COROLLARY 5. *If an optimal communication network has a team structure, it has the following properties:* (*i.*) *a team with two or more agents either includes all agents of a module or none of them,* (*ii.*) *there is at most one team whose agents belong to two or more modules, and* (*iii.*) *a team whose agents belong to two or more modules includes the agents from the most cohesive modules.*

A first implication of the model is that if one carves up a communication network, one may cut through a module and put its agents into different teams. In this case, though, one must go all the way and carve the module into as many teams as there are agents in the module. This implication follows readily from supermodularity and its push towards all-or-nothing communication.

A second implication is that there can be at most one team consisting of multiple modules. The presence of two or more such teams violates the fact that optimal communication precludes two-switches, which we established in Lemma 4. There may be many single-agent or single-module teams, but there can be at most one large, multi-module one.

A third and final implication, which follows from the characterization result, is that if there is a multi-module team, it consists of the most cohesive modules. The team structure then resembles a core-periphery structure of the type we discussed above, albeit a stark one with no communication between the core and the periphery.

This last point brings up a broader lesson the model offers about information hiding. Even though team structures can be optimal in our setting, they are special. There is no inherent reason why such structures ought to be more pervasive than others that do not divide agents neatly into informationally isolated units. The model suggests that instead of investing in microfiche, Brooks could have cut communication costs by reducing across-module communication to programmers in the least cohesive modules (and, possibly, even within-module communication in such modules). Done optimally, the resulting communication network would have taken the form of a threshold graph, or even a core-periphery structure with intense communication between programmers in the most cohesive modules but little communication between those in the less cohesive ones. This suggests David Parnas was right to propose cutting communication costs by removing communication links. The quest of how to hide information optimally, though, should go beyond team structures and encompass core-periphery ones.

7.3 A Network Measure of Threshold Violations

Our model generates a number of predictions about communication within firms with modular production functions. Our goal here is to focus on the model's core prediction, that optimal communication networks exhibit the threshold property. Since one would not expect this property to hold exactly, we propose a network measure of the extent to which it may be violated. An advantage of the measure is that it does not depend on the specifics of the production network beyond its modular structure, such as the needs for adaptation and coordination and the degrees of uncertainty. Instead, the measure only needs information about what modules decisions belong to, such as that used to construct Design Structure Matrices (see, for example, Eppinger and Browning (2012)), and about directed communication patterns, such as that used in the systems engineering literature (see, for example Eppinger (2015) and the references therein) and more recently within economics (see, for instance, Yang et al. (2021) and Impink et al. (forthcoming)).

The measure is inspired by the Fulkerson-Chen-Anstee Theorem (Fulkerson 1960, Chen 1980), which provides inequalities involving in-degrees and out-degrees of nodes that hold with equality if and only if the network exhibits the threshold property. We show that one can interpret slack in these inequalities as a measure of the extent to which the threshold property is violated.

The threshold property is a property of across-module communication. We say that the $M \times M$ submatrix $\boldsymbol{\Psi}$ is *induced* by the communication matrix $\boldsymbol{C}$ if it consists of one agent from each module and the directed communication links between them. We next describe the *threshold gap* of $\boldsymbol{\Psi}$, which is our measure.

To this end, denote agent i's out-degree by δ_i^+ and agent i's in-degree by δ_i^-. Order the agents by out-degree, so that agent 1 has the highest out-degree and agent M the lowest. If multiple agents have the same out-degree, order them by their in-degree. Next, define $\overline{\delta}_m = \sum_{i<m} 1_{\delta_i^- \geq m-1} + \sum_{i>m} 1_{\delta_i^- \geq m}$, which counts how many agents are told the states of at least m others (with an adjustment for placement in the ordering). The Fulkerson-Chen-Anstee Theorem shows that the inequalities $\sum_{m=1}^{k} \delta_m^+ \leq \sum_{m=1}^{k} \overline{\delta}_m$ hold for each $1 \leq k \leq M-1$, and that they hold with equality for $k = M$. Cloteaux et al. (2014), in turn, shows that a directed network exhibits the threshold property if and only if each of these inequalities holds with equality.

We now define the *threshold gap* of $\boldsymbol{\Psi}$ as $\Delta(\boldsymbol{\Psi}) \equiv \sum_{m=1}^{M} \max\left\{\overline{\delta}_m - \delta_m^+, 0\right\}$. Moreover, we define a *swap* as removing a link from agent i to agent j and adding a link from agent k to an agent

ℓ to whom he does not yet have a link. We can then state the result.

PROPOSITION 7. *If $\boldsymbol{C}^*$ is an optimal communication network, then the threshold gap $\Delta(\boldsymbol{\Psi})$ is zero for any network $\boldsymbol{\Psi}$ induced by $\boldsymbol{C}^*$. If network $\boldsymbol{\Psi}$ has $\Delta(\boldsymbol{\Psi}) > 0$, then $\lceil \Delta(\boldsymbol{\Psi})/2 \rceil$ is the minimum number of swaps required for the resulting communication network to have a threshold gap of zero and thus exhibit the threshold property.*

The proposition shows that the threshold gap is tied to the number of swaps needed to restore the threshold property. If $\Delta(\boldsymbol{\Psi}) = 0$, the network satisfies the property, and no swaps are needed. If $\Delta(\boldsymbol{\Psi}) > 0$, the network does not satisfy the threshold property and the minimum number of swaps needed to restore it is given by the smallest integer weakly larger than $\Delta(\boldsymbol{\Psi})/2$. As such, the threshold gap provides a measure of the extent to which the threshold property is violated.

8 Relaxing Assumptions

Having solved the model and explored its implications, we return to the model assumptions in Section 3 to discuss how our results change as we relax them.

Noisy communication. Our model assumes that communication is perfectly informative. Suppose, instead, that the principal chooses the precision with which agents tell others their states, with more precise communication being more costly. Specifically, suppose states are normally distributed and that each agent $i \in \mathcal{N}$ receives a noisy and conditionally independent signal $s_{ij} = \theta_j + \eta_{ij}$ about state $\theta_j \sim N\left(0, \sigma_j^2\right)$, where $\eta_{ij} \sim N\left(0, 1/\tau_{ij}\right)$ and where τ_{ij} is the precision of the signal. The costs of agent $j \in \mathcal{N}$ telling others about his state are given by $\sum_{i \neq j} k\left(\varphi_{ij}\right)$, where $\varphi_{ij} = \tau_{ij}/\left(\tau_{ij} + 1/\sigma_j^2\right)$ is the signal-to-noise ratio for the signal agent i receives about agent j's state. The rest of the model is as in Section 3.

In Appendix B, we show that the separability result continues to hold in this setting. We further show that if within-module communication is free, the precision of the signals an agent sends others about his state is the same for any two receiving agents who belong to the same module, that is, $\varphi_{ij} = \varphi_{kj}$ if $m(i) = m(k)$. If $k\left(\varphi_{ij}\right)$ is linear, then optimal signal-to-noise ratios are never interior and optimal communication networks coincide with those in our main model. If, instead, $k\left(\varphi_{ij}\right)$ is sufficiently convex so that optimal signal-to-noise ratios are interior, agents send more precise signals to agents who belong to more cohesive modules.

Correlated states. An assumption we share with Calvó-Armengol, de Martí, and Prat (2015) is that states are independent. If, instead, states were correlated, an agent who is told about one state would also learn some information about the other states, reducing the benefit of telling him about them. As a result, the problems of whom each agent should tell about his state would be interdependent. In such a setting, expected revenue $R(\boldsymbol{C})$ could still be written as the sum of the expected revenue generated by each agent, $R_i(\cdot)$, that is, Lemma 2 would still hold. The expected revenue generated by each agent, though, would depend on the entire communication network $\boldsymbol{C}$, causing the principal's subproblems in Proposition 1 to become interdependent.

Re-transmission of information. Next, we explore the assumption that agents do not re-transmit information, which is an assumption we share with Calvó-Armengol and de Martí Beltran (2009), Calvó-Armengol et al. (2015), and Herskovic and Ramos (2020).[25] This assumption captures the idea that, even though we model each state as simply a number, it refers to a complex set of conditions and circumstances that only the associated agent can describe appropriately.

One way to relax this assumption is to allow agents to *synthesize* information, to combine their knowledge of their own state with information they receive about other states. Suppose, for instance, that after learning state θ_i, agent j can communicate a summary statistic of (θ_i, θ_j) to agent k and do so at cost γ. A communication link from agent i to j then does not only affect what agent j knows about θ_i but also what he can tell others. As a result, the principal's subproblems in Proposition 1 become interdependent, and the separability result no longer holds.

Another way to allow for re-transmission is to suppose that after being told state θ_i, agent j must incur cost γ to tell θ_i to agent k. Agents can tell others the states they have been told about and they can do so at the same cost at which they can tell them about their own states. In this case, any equilibrium with re-transmission is payoff equivalent to one without. The separability result in Proposition 1 continues to hold, as does the characterization result in Proposition 2.

Heterogeneous coupling. The assumption that coupling is homogenous applies to settings, such as the laptop in Figure 1, in which the needs for coordination across any two modules are (roughly) the same. Even though such settings are common, there are others in which some modules need to be coordinated more closely with each other than with other modules. The law and business schools of a university, for instance, may require more coordination with each other than with the schools of engineering and natural sciences, and vice versa.

To explore such heterogeneous coupling, suppose we partition the set of modules into clusters that differ in their degrees of coupling, such as in the example in the left panel in Figure 7. Suppose, in particular, that each node i belongs to a module $m(i)$, and each module m belongs to a cluster $k(m) \in \mathcal{K}$, where $\mathcal{K} = \{1, \ldots, K\}$. As in the main model, the need for coordination between any two decisions d_i and d_j is given by $p_{ij} = p_m \geq 0$ if they belong to the same module m. In contrast

[25] For an exploration of hierarchical communication with re-transmission see Migrow (2021).

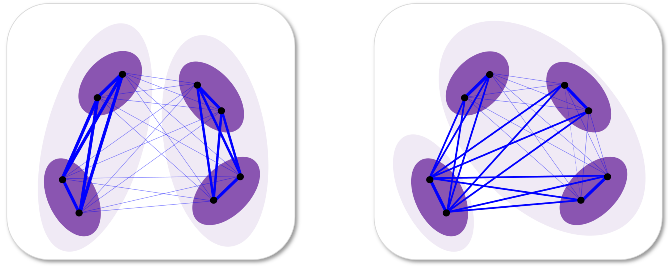

Figure 7: Left panel—Production network with *heterogeneous coupling*, where the dark shaded areas indicate modules and the lighter ones indicate clusters of modules. Right panel—Production network with an *interface module*.

to the main model, however, the need for coordination between the two decisions need no longer be given by p if they belong to different modules. Instead, it is given by $p_{ij} = p^k \geq 0$ if their modules belong to the same cluster k, and it is given by p only if their modules belong to different clusters. The new parameters $p^1, \ldots, p^K$, therefore, capture the degrees of coupling in the different clusters. The rest of the model is as in Section 3.

In Appendix C we show that, even though the computations get considerably more involved, we can still derive a closed-form expression for the agents' expected revenue in terms of the model primitives, such as the one in Lemma 3. This result, in turn, allows us to show that the characterization result in Proposition 2 generalizes as follows.

PROPOSITION 8. *There exist thresholds $\lambda_i^k \geq 0$ such that it is optimal for agent $i \in \mathcal{N}$ to tell his state to agent $j \in \mathcal{N}$ with $m(j) \neq m(i)$* and *$m(j)$* in cluster *$k \in \mathcal{K}$ if and only if $x_{m(j)} \geq \lambda_i^k$.*

Optimal communication is, therefore, still determined by threshold rules on module cohesions: each agent tells his state to the agents in modules whose cohesion is sufficiently high. The only difference is that how high cohesion needs to be depends on the characteristics of the cluster that the receiving agent belongs to as well as whether agent i's module is in that cluster.

Interfaces. A feature of modular production that is important in some applications is the presence of an *interface module*—a module that all other modules have to be tightly coordinated with. To return to the university example, the various schools of a university may all need to be coordinated more closely with the office of the provost than with each other. The extension to heterogeneous coupling allows us to incorporate this feature.

Notice, in particular, that the extension allows for the across-cluster degree of coupling p to be higher than the within-cluster degrees of coupling $p^1, \ldots, p^K$. We can, therefore, specify a production network, such as the one in the right panel in Figure 7, in which the set of modules is divided into one cluster with a single module and another cluster with all the other ones. The across-cluster degree of coupling then captures how tightly the modules in the multi-module cluster must be coordinated with the single module, and the within-cluster degree of coupling captures how well they must be coordinated with each other. Setting the across-cluster degree of coupling higher than the within-cluster one turns the single module into an interface module.

Because this specification is a special case of the extension with heterogeneous coupling, Proposition 8 still applies. Even with an interface module, optimal communication networks are determined by threshold rules on module cohesion.

General production networks. Even though we focus on modular production, it is instructive to examine how our results change if we allow for general production networks. To this end, suppose the production network $\boldsymbol{P}$ can take any form, provided it still satisfies $p_{ii} = 0$, $p_{ij} = p_{ji}$, and $\sum_{j=1}^{N} p_{ij} < 1$ for all $i, j \in \mathcal{N}$. The proofs of Lemma 1, Lemma 2, and Proposition 1 allow for such production networks, so the separability result continues to hold. As such, the principal can still determine an optimal communication network by considering each agent separately. Moreover, we show in Proposition 9 in Appendix C that the principal's objective is still supermodular and can, therefore, be maximized using standard algorithms.

What can no longer hold is the characterization of optimal communication networks in Proposition 2, which uses the presence of modules. Optimal communication networks can now take many forms and need not exhibit the threshold property. The supermodularity of the principal's objective, though, ensures that comparative statics are still monotone. As in the main model, the principal will only ever respond to an increase in the value of adaptation or the need for coordination, or a decrease in the cost of communication, by adding communication links.

Incentive conflicts. Lastly, we explore how the results change if we depart from our team theoretic approach and allow for incentive conflicts. To this end we assume that agents have a constant bias, as in Hagenbach and Koessler (2010), Galeotti et al. (2013), and much of the literature on cheap talk that builds on Crawford and Sobel (1982). Specifically, suppose each agent $i \in \mathcal{N}$ cares about $r(d_1, \ldots, d_N) + 2a_i d_i b_i$, where $r(d_1, \ldots, d_N)$ is the firm's revenue (1) and b_i is agent i's bias, which is common knowledge. The rest of the model is as in Section 3.

We study this extension in Appendix D, where we show that the characterization of optimal communication networks in Proposition 2 continues to hold. The reason is that while the agents' biases distort their decisions, they do not distort how their decisions vary with their states. As a result, the agents' biases do not affect the benefits of communication links and, thus, do not affect the design of optimal communication networks.

The extension focuses on how incentive conflicts affect decision-making and, through this channel, impact optimal communication networks. In contrast to Hagenbach and Koessler (2010) and Galeotti et al. (2013), it does not allow for cheap talk communication, and thus does not explore how incentive conflicts affect the agents' incentives to engage in strategic communication.

9 Conclusions

The structure of technology drives the organization of firms. Based on this premise, this paper explored how the rise of modular production shapes the pattern of communication and the flow of information inside of firms. We conclude by suggesting several avenues for future research.

One avenue is to explore the organization of firms with non-modular production functions, especially that of multidivisional firms. Alfred Chandler documented the central role of multidivisional firms in the development of the US economy and spurred a large literature examining their organization (Chandler 1962). A goal of this literature is to understand the firms' choice between M, U, and matrix forms, between organizing by product, function, or a combination of both (see, for instance, Maskin et al. (2000)). Our paper suggests that this choice is shaped by the firms' production functions. These production functions often have an *overlapping community structure* rather than a non-overlapping one. The R&D decisions for one product, say, must be closely coordinated with both the manufacturing and marketing decisions for the same product and the R&D decisions for the firm's other products. Even though such production functions are not modular, they fall within the class of general production functions we examined in the extensions. The fact that the separability result continues to hold serves as a useful starting point for an exploration of when M, U, and matrix forms are optimal and what determines the choice among them.

A second avenue for future research is to explore the broader impact of modular production on the organization of firms. As we noted in the introduction, Baldwin and Clark (1997) observe that, while the introduction of the System/360 did lead to immediate changes in IBM's internal organization, its more enduring impact was to cause entry into the computer industry in the following decades. The entrants were often small, entrepreneurial firms that focused on the development and production of individual modules and whose innovative products allowed them to compete successfully with IBM's own, in-house module makers. In this telling, the introduction of the System/360 in the 1960s sowed the seeds for the subsequent disintegration of IBM and the other large mainframe manufacturers and gave rise to the competitive and innovative computer industry of today (see Footnote 8). There are many reasons why modular production may affect the boundaries of firms and the structure of industries. We leave their exploration for future research.

A final avenue for future research goes beyond the impact of modular production on organization and asks what explains its rise in the first place. Herbert Simon argued that modularity facilitates adaptation by confining adaptive changes to individual modules within a system (Simon 1962). In line with this intuition, firms such as IBM explain their development of modular products with the need to adapt quickly to the changing capabilities of their suppliers and needs of their customers. Yet, a full explanation for the rise of modular production also needs to account for its costs. It may be easier to adapt a modular product to its environment but, for a given environment, one would expect limitations in across-module interactions to affect its performance. After all, products have not always been modular, and even today many are not, suggesting that such designs also have significant downsides. Answering the questions of when and why firms develop modular products, and what trade-offs they face when they are doing so, would require moving beyond one of the foundational economic modeling assumptions, that production functions are given by nature and not designed by firms. As such, it is the most challenging question this paper highlights and, like the other open questions we sketched above, we leave it for future research.

References

[1] Alexander, Christopher. 1964. *Notes on the Synthesis of Form.* Harvard University Press.

[2] Alonso, Ricardo, Wouter Dessein, and Niko Matouschek. 2008. When Does Coordination Require Centralization? *American Economic Review.* 98(1): 145–79.

[3] Arrow, Kenneth. 1974. *The Limits of Organization.* W.W. Norton & Company.

[4] Atalay, Enghin, Ali Hortaçsu, and Chad Syverson. 2014. Vertical Integration and Input Flows. *American Economic Review.* 104(4): 1120–48.

[5] Baccara, Mariagiovanna and Heski Bar-Isaac. 2008. How to Organize Crime. *The Review of Economic Studies.* 75(4): 1039–67.

[6] Bala, Venkatesh and Sanjeev Goyal. 2000. A Noncooperative Model of Network Formation. *Econometrica.* 68(5): 1181–229.

[7] Baldwin, Carliss and Kim Clark. 1997. Managing in an Age of Modularity. *Harvard Business Review.* 75(5): 84–93.

[8] ——— and ——— . 2000. *Design Rules: The Power of Modularity* (Vol.1). MIT Press.

[9] Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou. 2006. Who's Who in Networks. Wanted: The Key Player. *Econometrica.* 74(5): 1403–17.

[10] Bergemann, Dirk, Tibor Heumann, and Stephen Morris. 2017. Information and Interaction. Cowles Foundation Discussion Paper No. 2088.

[11] Borgatti, Stephen and Martin Everett. 2000. Models of Core/Periphery Structures. *Social Networks.* 21(4): 375–95.

[12] Brooks, Frederick. 1995. The Mythical Man-Month (Anniversary Edition). Addison-Wesley.

[13] Brown, Jennifer and Craig Garthwaite. 2016. Global Aircraft Manufacturing, 2002–2011. *Kellogg School of Management Case Study.* KEL 938.

[14] Calvó-Armengol, Antoni and Joan de Martí Beltran. 2009. Information Gathering in Organizations: Equilibrium, Welfare, and Optimal Network Structure. *Journal of the European Economic Association.* 7(1): 116–61.

[15] ———, Joan de Martí, and Andrea Prat. 2015. Communication and Influence. *Theoretical Economics*. 10(2): 649–90.

[16] Chandler, Alfred. 1962. Strategy and Structure: Chapters in the History of the Industrial Empire. Cambridge Mass.

[17] Chen, Wai-Kai. 1980. Subgraphs and Their Degree Sequences of a Digraph. *Journal of the Franklin Institute*. 310(6): 349–63.

[18] Cloteaux, Brian, Michael Drew LaMar, Elizabeth Moseman, and James Shook. 2014. Threshold Digraphs. *National Institute of Standards and Technology.* 119. 227-34.

[19] Colfer, Lyra and Carliss Baldwin. 2016. The Mirroring Hypothesis: Theory, Evidence, and Exceptions. *Industrial and Corporate Change*. 25(5): 709–38.

[20] Conway, Melvin. 1968. How Do Committees Invent? *Datamation*. 14(4): 84–93.

[21] Crawford, Vincent and Joel Sobel. 1982. Strategic Information Transmission. *Econometrica.* 50(6): 1431–51.

[22] Cremer, Jacques, Luis Garicano, and Andrea Prat. 2007. Language and the Theory of the Firm. *Quarterly Journal of Economics*. 122(1): 373–407.

[23] Dessein, Wouter and Tano Santos. 2006. Adaptive Organizations. *Journal of Political Economy*. 114(5): 956–95.

[24] ———, Andrea Galeotti, and Tano Santos. 2016. Rational Inattention and Organizational Focus. *American Economic Review*. 106(6): 1522–36.

[25] Eppinger, Steven and Tyson Browning. 2012. Design Structure Matrix Methods and Applications. MIT Press.

[26] ———. 2015. Design Structure Matrix Extensions and innovations: A Survey and New Opportunities. *IEEE Transactions on Engineering Management*. 63(1): 27–52.

[27] Fulkerson, Delbert. 1960. Zero-one Matrices with Zero Trace. *Pacific Journal of Mathematics*. 10.3: 831–36.

[28] Fortunato, Santo. 2010. Community Detection in Graphs. *Physics Reports*. 486(3–5): 75–174.

[29] Galeotti, Andrea, Christian Ghiglino, and Francesco Squintani. 2013. Strategic Information Transmission on Networks. *Journal of Economic Theory*. 148: 1751–69.

[30] Garicano, Luis and Andrea Prat. 2013. Organizational Economics with Cognitive Costs. *Advances in Economics and Econometrics*. 1: 342–88.

[31] ——— and Timothy Van Zandt. 2013. Hierarchies and the Division of Labor. In *The Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts. Princeton University Press: 604–54.

[32] Garud, Raghu, Arun Kumaraswamy, and Richard Langlois, eds. 2009. *Managing in the Modular Age: Architectures, Networks, and Organizations*. John Wiley & Sons.

[33] Girvan, Michelle and Mark Newman. 2002. Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences*. 99(12): 7821–26.

[34] Golub, Benjamin and Stephen Morris. 2017. Expectations, Networks, and Conventions. Mimeo.

[35] Goyal, Sanjeev and Adrien Vigier. 2014. Attack, Defence, and Contagion in Networks. *The Review of Economic Studies*. 81(4): 1518–42.

[36] Guimera, Roger, Stefano Mossa, Adrian Turtschi, and Luis Nunes Amaral. 2005. The Worldwide Air Transportation Network: Anomalous Centrality, Community Structure, and Cities' Global Roles. *Proceedings of the National Academy of Sciences*. 102(22): 7794–99.

[37] Hagenbach, Jeanne and Frėdėric Koessler. 2010. Strategic Communication Networks. *The Review of Economic Studies*. 77(3): 1072–99.

[38] Henderson, Rebecca and Kim Clark. 1990. Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly*. 35(1): 9–30.

[39] Herskovic, Bernard and Joao Ramos. 2020. Acquiring Information through Peers. *American Economic Review*. 110(7): 2128–52.

[40] Impink, Stephen, Andrea Prat, and Raffaella Sadun. Forthcoming. Communication within Firms: Evidence from CEO Turnovers. *Management Science*.

[41] Jackson, Matthew and Asher Wolinsky. 1996. A Strategic Model of Social and Economic Networks. *Journal of Economic Theory*. 71(1): 44–74.

[42] Lambert, Nicolas, Giorgio Martini, and Michael ostrovsky. 2018. Quadratic Games. NBER Working Paper Series, No. 24914.

[43] Langlois, Richard and Paul Robertson. 1992. Networks and Innovation in a Modular System: Lessons from the Microcomputer and Stereo Component Industries. *Research Policy*. 21(4): 297–313.

[44] ———. 2002. Modularity in Technology and Organization. *Journal of Economic Behavior & Organization*. 49(1): 19–37.

[45] March, James and Herbert Simon. 1958. Organizations. John Wiley & Sons.

[46] Marschak, Jacob and Roy Radner. 1972. Economic Theory of Teams. Yale University Press.

[47] Maskin, Eric, Yingyi Qian, and Chenggang Xu. 2000. Incentives, Information, and Organizational Form. *Review of Economic Studies*. 67(2): 359–78.

[48] McCord, Kent and Steven Eppinger. 1993. Managing the Integration Problem in Concurrent Engineering. MIT Sloan School of Management, Cambridge, MA, Working Paper No. 3594.

[49] McDonald, Rory and Suresh Kotha. 2015. Boeing 787: Manufacturing a Dream. *Harvard Business School Case Study*. 9-615-048.

[50] Meunier, David, Renaud Lambiotte, Alex Fornito, Karen Ersche, and Edward Bullmore. 2009. Hierarchical Modularity in Human Brain Functional Networks. *Frontiers in Neuroinformatics*. 3: 37.

[51] Migrow, Dimitri. 2021. Designing Communication Hierarchies. *Journal of Economic Theory*. 198: 105349.

[52] Morris, Stephen. 2000. Contagion. *The Review of Economic Studies*. 67(1): 57–78.

[53] Murota, Kazuo. 2003. Discrete Convex Analysis. Society for Industrial and Applied Mathematics.

[54] Parnas, David. 1972. On the Criteria to be Used in Decomposing Systems in Modules. In *Pioneers and Their Contributions to Software Engineering*, eds. Manfred Broy and Ernst Denert. Springer: 479–98.

[55] Rantakari, Heikki. 2008. Governing Adaptation. *Review of Economic Studies*. 75(4): 1257–85.

[56] Rombach, Puck, Mason Porter, James Fowler, and Peter Mucha. 2017. Core-Periphery Structure in Networks (Revisited). *SIAM Journal on Applied Mathematics*. 59(3): 619–46.

[57] Sanchez, Ron and Joseph Mahoney. 1996. Modularity, Flexibility, and Knowledge Management in Product and Organization Design. *Strategic Management Journal*. 17(2): 63–76.

[58] Simon, Herbert. 1962. The Architecture of Complexity. *Proceedings of the American Philosophical Society*. 106(6): 467–82.

[59] Tadelis, Steven and Oliver Williamson. 2013. Transaction Cost Economics. In *The Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts. Princeton University Press: 159–90.

[60] Talamàs, Eduard and Omer Tamuz. 2017. Network Cycles and Welfare. Mimeo.

[61] Thompson, James. 1967. Organizations in Action: Social Science Bases of Administrative Theory. McGraw-Hill.

[62] Yang, Longqi, David Holtz, Sonia Jaffe, Siddharth Suri, Shilpi Sinha, Jeffrey Weston, Connor Joyce, Neha Shah, Kevin Sherman, Brent Hecht, and Jaime Teevan. 2021. The Effects of Remote Work on Collaboration Among Information Workers. *Nature Human Behaviour*: 1–12.

Appendix A: Proofs

We first introduce some notation. Throughout the appendix, for notational compactness, we will denote by $\mathrm{E}_i[\cdot]$ the expectation over $\theta = (\theta_1, \ldots, \theta_N)$ given the information agent i has under communication network $\boldsymbol{C}$. That is, for a random variable Z, $\mathrm{E}_i[Z] \equiv \mathrm{E}\left[Z|\, \boldsymbol{C}_{(i)}\right]$. Next, a strategy for agent i is a mapping $\tilde{d}_i : \left[-\overline{\theta}, \overline{\theta}\right]^N \to \left[-\overline{D}, \overline{D}\right]$, where $\tilde{d}_i(\theta)$ denotes the decision that agent i makes in state θ. We denote a strategy profile by $\tilde{d} = \times_{i=1}^N \tilde{d}_i$. To ensure equilibrium strategies involve interior decisions, define $\overline{p} = \max_i \sum_j p_{ij}$, and assume that $\overline{D} \geq \frac{\overline{\theta}}{1-\overline{p}}$.

LEMMA 1. *Equilibrium decisions are unique and given by*

$$d_i^* = \sum_{j=1}^N a_j \omega_{ij}(\boldsymbol{C}_j)\, \theta_j \textit{ for all } i \in \mathcal{N},$$

where $\omega_{ij}(\boldsymbol{C}_j)$ *denotes the* ijth *entry of* $(\boldsymbol{I} - (\operatorname{diag} \boldsymbol{C}_j)\, \boldsymbol{P}\, (\operatorname{diag} \boldsymbol{C}_j))^{-1}$.

Proof of Lemma 1. This proof parallels the approach of Golub and Morris (2017), Appendix A1: We take the communication network $\boldsymbol{C}$ as given and show that $d^* = \times_{i=1}^N d_i^*$ is the unique strategy profile that survives iterated elimination of strictly dominated strategies and is therefore the unique Bayesian-Nash equilibrium.

Step 1: *Show that there is a unique Bayesian-Nash equilibrium by showing that there is a unique strategy profile that survives iterated elimination of strictly dominated strategies.*

Given any $\boldsymbol{C}$, the game played by the agents is a game of strategic complements: if we denote

$$\hat{d}_i\left(\theta, \tilde{d}_{-i}\right) = a_i\theta_i + \sum_{j=1}^N p_{ij} \mathrm{E}_i\left[\tilde{d}_j(\theta)\right],$$

agent i's best response to the strategy profile $\tilde{d}_{-i}$ in state θ, then $\hat{d}_i$ is increasing in each $\tilde{d}_j$ under the partial order given by $\tilde{d}_j \succeq \tilde{d}'_j$ if and only if $\tilde{d}_j(\theta) \geq \tilde{d}'_j(\theta)$ for all θ.

Define the set $S_i(k)$ as the set of i's pure strategies surviving k rounds of iterated elimination of strictly dominated strategies. Since $d_i(\theta) \in \left[-\overline{D}, \overline{D}\right]$, the first set in the sequence is

$$S_i(0) = \left\{ \tilde{d}_i \middle| -\overline{D} \leq \tilde{d}_i(\theta) \leq \overline{D} \text{ for all } \theta \right\}.$$

Next, as this is a game of strategic complements, an upper bound on $S_i(1)$ is i's best response to the maximal strategy profile $\tilde{d}_{-i} \in S_{-i}(0)$, where $\tilde{d}_{-i} = \times_{j \neq i} \tilde{d}_j$ and $S_{-i}(k) = \prod_{j \neq i} S_j(k)$, and a lower bound on $S_i(1)$ is i's best response to the minimal strategy profile $\tilde{d}_{-i} \in S_{-i}(0)$. That is,

$$S_i(1) = \left\{ \tilde{d}_i \middle| a_i\theta_i - \sum_{j=1}^N p_{ij}\overline{D} \leq \tilde{d}_i(\theta) \leq a_i\theta_i + \sum_{j=1}^N p_{ij}\overline{D} \right\}.$$

Next, suppose that for $k > 1$, the set $S_i(k)$ takes the form

$$S_i(k) = \left\{ \tilde{d}_i \middle| \underline{d}_i^k(\theta) \leq \tilde{d}_i(\theta) \leq \overline{d}_i^k(\theta) \text{ for all } \theta \right\},$$

where

$$\begin{aligned}
\overline{d}_i^k(\theta) &= a_i\theta_i + \sum_{m=1}^{k-1} \beta_{im} + \sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\cdots\sum_{j_k=1}^{N} p_{ij_1}p_{j_1j_2}\cdots p_{j_{k-1}j_k}\overline{D} \\
\underline{d}_i^k(\theta) &= a_i\theta_i + \sum_{m=1}^{k-1} \beta_{im} - \sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\cdots\sum_{j_k=1}^{N} p_{ij_1}p_{j_1j_2}\cdots p_{j_{k-1}j_k}\overline{D},
\end{aligned}$$

and

$$\beta_{im} = \sum_{j_1=1}^{N}\cdots\sum_{j_m=1}^{N} p_{ij_1}p_{j_1j_2}\cdots p_{j_{m-1}j_m} a_{j_m} E_i E_{j_1}\cdots E_{j_{m-1}}[\theta_{j_m}].$$

Then an upper bound on $S_i(k+1)$ is agent i's best response to the maximal strategy profile $\tilde{d}_{-i} \in S_{-i}(k)$, and a lower bound on $S_i(k+1)$ is agent i's best response to the minimal strategy profile $\tilde{d}_{-i} \in S_{-i}(k)$. That is,

$$S_i(k+1) = \left\{ \tilde{d}_i \middle| \underline{d}_i^{k+1}(\theta) \leq \tilde{d}_i(\theta) \leq \overline{d}_i^{k+1}(\theta) \text{ for all } \theta \right\}.$$

To show that the upper and lower bounds of $S_i(k)$ converge to the same value, we show that

$$\lim_{k\to\infty} \sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\cdots\sum_{j_k=1}^{N} p_{ij_1}p_{j_1j_2}\cdots p_{j_{k-1}j_k}\overline{D} = 0.$$

This term converges to zero as long as the row sum of the production matrix to the kth power, $\boldsymbol{P}^k$, converges to zero as $k \to \infty$. This result follows since $\sum_{j=1}^{N} p_{ij} < 1$ for all i, and therefore the spectral radius of $\boldsymbol{P}$ is strictly less than one. By the sandwich theorem, we therefore have

$$\lim_{k\to\infty} \underline{d}_i^k(\theta) = \lim_{k\to\infty} \overline{d}_i^k(\theta) = a_i\theta_i + \sum_{m=1}^{\infty} \beta_{im}.$$

This result implies that $\lim_{k\to\infty} S_i(k)$ is a singleton for all i. As this is a supermodular game, the resulting strategy profile is the unique Bayesian-Nash equilibrium of the game.

Step 2: *Show that the unique Bayesian-Nash equilibrium strategy profile is a linear combination of $\theta_1, \ldots, \theta_N$, that is, $d_i^*(\theta) = \sum_{j=1}^{N} \alpha_{ij}\theta_j$ for some scalars $\{\alpha_{ij}\}_{j=1}^{N}$.*

First, observe that $\mathrm{E}_i\mathrm{E}_{j_1}\cdots\mathrm{E}_{j_{m-1}}[\theta_{j_m}]$ is zero if some $j \in \{i, j_1, \ldots, j_{m-1}\}$ does not know θ_{j_m} under $\boldsymbol{C}$, and $\mathrm{E}_i\mathrm{E}_{j_1}\cdots\mathrm{E}_{j_{m-1}}[\theta_{j_m}] = \theta_{j_m}$ if all $j \in \{i, j_1, \ldots, j_{m-1}\}$ know θ_{j_m} under $\boldsymbol{C}$. This result follows by an induction argument and the law of iterated expectations. For the $m = 1$

case, $\mathrm{E}_i[\theta_{j_1}] = 0$ if i does not know θ_{j_1}, and $\mathrm{E}_i[\theta_{j_1}] = \theta_{j_1}$ if i does know θ_{j_1}. Next, suppose all $j \in \{i, j_1, \ldots, j_{m-1}\}$ know $\theta_{j_{m+1}}$. Then $\mathrm{E}_{j_m}\mathrm{E}_i\mathrm{E}_{j_1}\cdots\mathrm{E}_{j_{m-1}}\left[\theta_{j_{m+1}}\right] = \mathrm{E}_{j_m}\left[\theta_{j_{m+1}}\right]$, which is 0 if j_m does not know $\theta_{j_{m+1}}$ and is $\theta_{j_{m+1}}$ if j_m does know $\theta_{j_{m+1}}$. Finally, suppose there is some $j \in \{i, j_1, \ldots, j_{m-1}\}$ who does not know $\theta_{j_{m+1}}$. Then $\mathrm{E}_{j_m}\mathrm{E}_i\mathrm{E}_{j_1}\cdots\mathrm{E}_{j_{m-1}}\left[\theta_{j_{m+1}}\right] = \mathrm{E}_{j_m}[0] = 0$.

The result in the previous paragraph ensures that each β_{im} from Step 1 is a linear combination of $\theta_1, \ldots, \theta_N$, and therefore $d_i^*(\theta) = a_i\theta_i + \sum_{m=1}^{\infty}\beta_{im} = \sum_{j=1}^{N}\alpha_{ij}\theta_j$ for some scalars $\{\alpha_{ij}\}_{j=1}^{N}$.

Step 3: *Show that* $\alpha_{ij} = a_j\omega_{ij}(\boldsymbol{C}_j)\theta_j$, where $\omega_{ij}(\boldsymbol{C}_j)$ *denotes the* ijth *entry of the matrix* $(\boldsymbol{I} - (\operatorname{diag}\boldsymbol{C}_j)\boldsymbol{P}(\operatorname{diag}\boldsymbol{C}_j))^{-1}$.

The network associated with $(\operatorname{diag}\boldsymbol{C}_j)\boldsymbol{P}(\operatorname{diag}\boldsymbol{C}_j)$ is the subgraph of the production network induced by nodes that know θ_j, and therefore $\omega_{ij}(\boldsymbol{C}_j)$ is the sum of the values of all walks from node i to node j on the production network that pass only through nodes that know θ_j.

Note that $p_{ij_1}p_{j_1j_2}\cdots p_{j_{m-1}j_m}$ describes the value of a walk of length m from node i to node j_m on the production network. If any node in a walk $ij_1, j_1j_2, \cdots, j_{m-1}j_m$ does not know θ_{j_m}, then from the argument in step 2, $\mathrm{E}_i\mathrm{E}_{j_1}\cdots\mathrm{E}_{j_{m-1}}[\theta_{j_m}] = 0$. Otherwise, $\mathrm{E}_i\mathrm{E}_{j_1}\cdots\mathrm{E}_{j_{m-1}}[\theta_{j_m}] = \theta_{j_m}$. Thus, β_{im} is the sum of the values of all walks of length m from node i to node j_m on the production network that pass only through nodes that know θ_{j_m}. The result then follows. ■

COROLLARY 1. *The weight* $a_i\omega_{ii}(\boldsymbol{C}_i)$ *that decision* d_i^* *puts on its state* θ_i *satisfies* $\omega_{ii}(\boldsymbol{I}_i)a_i = a_i$, *where* $\boldsymbol{I}_i$ *is the* i*th row of an* $N \times N$ *identity matrix, and it is increasing and supermodular in* $\boldsymbol{C}_i$.

Proof of Corollary 1. This result follows from the proofs of Lemma 1 and Proposition 9 in Appendix C. ■

LEMMA 2. *Under equilibrium decision-making, expected revenue is given by*

$$R(\boldsymbol{C}) \equiv \mathrm{E}\left[r\left(d_1^*, \ldots, d_N^*\right)\right] = \sum_{i=1}^{N} a_i \operatorname{Cov}\left(d_i^*, \theta_i\right),$$

where $\operatorname{Cov}(d_i^*, \theta_i) = a_i\sigma_i^2\omega_{ii}(\boldsymbol{C}_i)$.

Proof of Lemma 2. Given equilibrium decision-making, revenue in state θ can be written as

$$\sum_{i=1}^{N} a_i d_i^* \theta_i - \sum_{i=1}^{N} d_i^* \left[d_i^* - a_i\theta_i - \sum_{j=1}^{N} p_{ij} d_j^*\right].$$

Next, substitute in the best responses $d_i^* = a_i\theta_i + \sum_{j=1}^{N} p_{ij}\mathrm{E}_i\left[d_j^*\right]$. The term in square brackets is therefore equal to $\sum_{j=1}^{N} p_{ij}\left[\mathrm{E}_i\left[d_j^*\right] - d_j^*\right]$, and expected revenue can be written as

$$\sum_{i=1}^{N} a_i \mathrm{E}\left[d_i^*\theta_i\right] - \sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij}\mathrm{E}\left[d_i^*\left[\mathrm{E}_i\left[d_j^*\right] - d_j^*\right]\right] = \sum_{i=1}^{N} a_i \mathrm{E}\left[d_i^*\theta_i\right] - \sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij}\mathrm{E}\left[\mathrm{E}_i\left[d_i^*\right]\underbrace{\left[\mathrm{E}_i\left[d_j^*\right] - \mathrm{E}_i\left[d_j^*\right]\right]}_{=0}\right],$$

where the first equality holds by the law of iterated expectations. The result follows because $\operatorname{Cov}(d_i^*, \theta_i) = \mathrm{E}[d_i^* \theta_i] - \mathrm{E}[d_i^*]\,\mathrm{E}[\theta_i]$, and $\mathrm{E}[\theta_i] = 0$ for all i. ■

PROPOSITION 1. *An optimal communication network solves the principal's problem* (2) *if and only if it solves the N independent subproblems*

$$\max_{\boldsymbol{C}_i} R_i(\boldsymbol{C}_i) - \gamma \sum_{j \neq i} c_{ij}.$$

Proof of Proposition 1. Optimal communication networks maximize expected revenues minus communication costs. Using the expected revenue expression derived in Lemma 2, optimal communication networks solve

$$\max_{\boldsymbol{C}} \left[\sum_{i=1}^{N} a_i^2 \sigma_i^2 \omega_{ii}(\boldsymbol{C}_i) - \gamma \sum_{i=1}^{N} \sum_{j \neq i} c_{ij} \right].$$

Since $\omega_{ii}(\boldsymbol{C}_i)$ depends only on $\boldsymbol{C}_i$ and not the rest of the communication network $\boldsymbol{C}$, and the objective is additively separable in i, a communication network $\boldsymbol{C}^*$ solves this problem if and only if $\boldsymbol{C}_i^*$ solves

$$\max_{\boldsymbol{C}_i} a_i^2 \sigma_i^2 \omega_{ii}(\boldsymbol{C}_i) - \gamma \sum_{j \neq i} c_{ij}$$

for all $i \in \mathcal{N}$. ■

LEMMA 3. *Agent* 1*'s expected revenue is given by*

$$R_1(\boldsymbol{C}_1) = a_1^2 \sigma_1^2 \left(\frac{1 + (p_1 - p)\, x_1(\tilde{n}_1)}{1 + p_1} + \frac{p x_1(\tilde{n}_1)^2}{1 - p \sum_{m=1}^{M} \tilde{n}_m x_m(\tilde{n}_m)} \right), \tag{11}$$

where

$$x_m(\tilde{n}_m) = \frac{1}{1 + p - (\tilde{n}_m - 1)(p_m - p)} \text{ for } m \in \mathcal{M},$$

and $\tilde{n}_m$ is the number of agents in module m who know agent 1*'s state.*

Proof of Lemma 3. Suppose $\tilde{n}_1 \geq 1$ nodes in module 1 know θ_1 and $\tilde{n}_m \geq 0$ nodes in module m know θ_1 for $m > 1$. The restriction that $\tilde{n}_1 \geq 1$ reflects the fact that node 1 knows θ_1. The labeling of the other modules is immaterial, so we will denote the modules $m > 1$ for which $\tilde{n}_m \geq 1$ as modules $2, \ldots, \ell$. From Lemma 2, the expected revenue generated by agent 1 is $R_1(\mathbf{C}_1) = a_1^2 \sigma_1^2 \omega_{11}(\mathbf{C}_1)$. We will derive $\omega_{11}(\mathbf{C}_1)$ in four steps.

Step 1: *Derive a representation of $\omega_{11}(\mathbf{C}_1)$ as the value of walks on a modified module-level production network, and show that $\omega_{11}(\mathbf{C}_1)$ is the $(1,1)$ entry of a matrix $\mathbf{Q}^{-1}$.*

First, let us suppose $\tilde{n}_1 \geq 2$. The value $\omega_{11}(\mathbf{C}_1)$ is the sum of the value of all walks from node 1 back to itself on the subgraph of the production network consisting of nodes who know state θ_1. Denote this value by v_0. Next, let v_k be the sum of the values of all walks from an informed node in module k to node 1 on this same subgraph. These values can be written recursively as a system of equations.

$$\begin{aligned}
v_0 &= 1 + p_1(\tilde{n}_1 - 1)\,v_1 + p\tilde{n}_2 v_2 + \cdots + p\tilde{n}_\ell v_\ell \\
v_1 &= p_1 v_0 + p_1(\tilde{n}_1 - 2)\,v_1 + p\tilde{n}_2 v_2 + \cdots + p\tilde{n}_\ell v_\ell \\
&\vdots \\
v_\ell &= p v_0 + p(\tilde{n}_1 - 1)\,v_1 + p\tilde{n}_2 v_2 + \cdots + p_\ell(\tilde{n}_\ell - 1)\,v_\ell.
\end{aligned}$$

The right-hand side of the first equation describes the value of all walks from node 1 back to node 1 in the following way: the first term, 1, is the value of walks that pass only through node 1; the second term, $p_1(\tilde{n}_1 - 1)\,v_1$, is the value of all walks that initially pass to one of the $\tilde{n}_1 - 1$ other informed nodes in module 1; the $k+1$ term, $p\tilde{n}_k v_k$, is the value of all walks that initially pass to one of the $\tilde{n}_k$ informed nodes in module k. The right-hand side of the second equation captures the value of all walks from an informed node $j \neq 1$ in module 1 back to node 1 in the following way: the first term is the value of all walks that initially pass back to node 1; the second term is the value of all walks that initially pass to one of the other $\tilde{n}_1 - 2$ informed nodes in module 1; the $k+1$ term is the value of all walks that initially pass to one of the $\tilde{n}_k$ informed nodes in module k. The remaining equations are interpreted analogously.

This system of $\ell + 1$ equations can be written in matrix form:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & -p_1(\tilde{n}_1 - 1) & -p\tilde{n}_2 & \cdots & -p\tilde{n}_\ell \\ -p_1 & 1 - p_1(\tilde{n}_1 - 2) & -p\tilde{n}_2 & \cdots & -p\tilde{n}_\ell \\ -p & -p(\tilde{n}_1 - 1) & 1 - p_2(\tilde{n}_2 - 1) & \cdots & -p\tilde{n}_\ell \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -p & -p(\tilde{n}_1 - 1) & -p\tilde{n}_2 & \cdots & 1 - p_\ell(\tilde{n}_\ell - 1) \end{bmatrix} \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ \vdots \\ v_\ell \end{bmatrix}.$$

Denote this $(\ell + 1) \times (\ell + 1)$ matrix by $\boldsymbol{Q}$. Then v_0 is the $(1,1)$ entry of the inverse matrix $\boldsymbol{Q}^{-1}$, and by the definition of a matrix inverse, $v_0 = \det \tilde{\mathbf{Q}} / \det \mathbf{Q}$, where $\tilde{\mathbf{Q}}$ is the matrix obtained by removing the first row and column of $\boldsymbol{Q}$.

Next, to account for the possibility of $\tilde{n}_1 = 1$ (i.e., agent 1 does not tell his state to others in his own module), we can define the $\ell \times \ell$ matrix $\boldsymbol{R}$ and the $(\ell - 1) \times (\ell - 1)$ matrix $\tilde{\boldsymbol{R}}$, where $\boldsymbol{R}$ is the matrix obtained by removing the second row and column of $\boldsymbol{Q}$, and $\tilde{\boldsymbol{R}}$ is the matrix obtained by removing the first and second rows and columns of $\boldsymbol{Q}$. In this case,

$$v_0 = \frac{\det \tilde{\mathbf{R}}}{\det \mathbf{R}} = \frac{(1+p_1)\det \tilde{\mathbf{R}}}{(1+p_1)\det \mathbf{R}} = \frac{\det \tilde{\mathbf{Q}}}{\det \mathbf{Q}},$$

where the final equality follows by the Laplace expansion. In either case, it therefore suffices to solve for $\det \tilde{\mathbf{Q}}/\det \mathbf{Q}$.

Step 2: *Show that* $\det \mathbf{Q} = \frac{(1+p_1)(1-p_1(\tilde{n}_1-1))}{x_2(\tilde{n}_2)\cdots x_\ell(\tilde{n}_\ell)}\left(1-\lambda\sum_{m=2}^{\ell}\tilde{n}_m x_m(\tilde{n}_m)\right)$, *where* $\lambda = \frac{p}{1-p\tilde{n}_1 x_1(\tilde{n}_1)}$.

We can write $\mathbf{Q}$ in block-matrix form $\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$, where

$$\begin{aligned}
\mathbf{A} &= \begin{bmatrix} 1 & -p_1(\tilde{n}_1-1) \\ -p_1 & 1-p_1(\tilde{n}_1-2) \end{bmatrix},\ \mathbf{B} = \begin{bmatrix} -p\tilde{n}_2 & \cdots & -p\tilde{n}_\ell \\ -p\tilde{n}_2 & \cdots & -p\tilde{n}_\ell \end{bmatrix} \\
\mathbf{C} &= \begin{bmatrix} -p & -p(\tilde{n}_1-1) \\ \vdots & \vdots \\ -p & -p(\tilde{n}_1-1) \end{bmatrix},\ \mathbf{D} = \begin{bmatrix} 1-p_2(\tilde{n}_2-1) & \cdots & -p\tilde{n}_\ell \\ \vdots & \ddots & \vdots \\ -p\tilde{n}_2 & \cdots & 1-p_\ell(\tilde{n}_\ell-1) \end{bmatrix}.
\end{aligned}$$

By the block matrix determinant formula, $\det \mathbf{Q} = \det(\mathbf{A})\det\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)$. We first calculate $\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$. The expression for $\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ can be written as the sum of a diagonal matrix and a rank-one matrix:

$$\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B} = \mathbf{X}^{-1} - \lambda \mathbf{u}\mathbf{v}^T,$$

where $\lambda = \frac{p}{1-p\tilde{n}_1 x_1(\tilde{n}_1)}$, and

$$\mathbf{X}^{-1} = \begin{bmatrix} x_2(\tilde{n}_2)^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & x_\ell(\tilde{n}_\ell)^{-1} \end{bmatrix}, \mathbf{u} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} \tilde{n}_2 \\ \vdots \\ \tilde{n}_\ell \end{bmatrix}.$$

By the matrix determinant lemma,

$$\det\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right) = \left(1-\lambda\mathbf{v}^T\mathbf{X}\mathbf{u}\right)\det \mathbf{X}^{-1} = \frac{1-\lambda\sum_{m=2}^{\ell}\tilde{n}_m x_m(\tilde{n}_m)}{x_2(\tilde{n}_2)\cdots x_\ell(\tilde{n}_\ell)}.$$

We therefore have that

$$\det \mathbf{Q} = \frac{(1+p_1)(1-p_1(\tilde{n}_1-1))}{x_2(\tilde{n}_2)\cdots x_\ell(\tilde{n}_\ell)}\left(1-\lambda\sum_{m=2}^{\ell}\tilde{n}_m x_m(\tilde{n}_m)\right).$$

Step 3: *Show that* $\det \tilde{\mathbf{Q}} = \frac{1-p_1(\tilde{n}_1-2)}{x_2(\tilde{n}_2)\cdots x_\ell(\tilde{n}_\ell)}\left(1-\tilde{\lambda}\sum_{m=2}^{\ell}\tilde{n}_m x_m(\tilde{n}_m)\right)$, *where* $\tilde{\lambda} = \frac{p}{1-p(\tilde{n}_1-1)y_1(\tilde{n}_1)}$ *and* $y_1(\tilde{n}_1) = \frac{1}{1-p_1(\tilde{n}_1-2)+p(\tilde{n}_1-1)}$.

This step proceeds similarly to step 2. Recall that $\tilde{\mathbf{Q}}$ is the matrix derived by eliminating the first row and column from the matrix $\mathbf{Q}$. Partition $\tilde{\mathbf{Q}}$ into the block matrix $\begin{bmatrix} \tilde{\mathbf{A}} & \tilde{\mathbf{B}} \\ \tilde{\mathbf{C}} & \tilde{\mathbf{D}} \end{bmatrix}$ by letting $\tilde{\mathbf{A}} = 1 - p_1 (\tilde{n}_1 - 2)$ and setting $\tilde{\mathbf{B}}$, $\tilde{\mathbf{C}}$, and $\tilde{\mathbf{D}}$ accordingly. Then $\det \tilde{\mathbf{Q}} = \det \left(\tilde{\mathbf{A}}\right) \det \left(\tilde{\mathbf{D}} - \tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\right)$. Again, we can write

$$\tilde{\mathbf{D}} - \tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}} = \mathbf{X}^{-1} - \tilde{\lambda}\mathbf{u}\mathbf{v}^T,$$

where $\mathbf{X}^{-1}$, $\mathbf{u}$, and $\mathbf{v}$ are the same as in step 2, $\tilde{\lambda} = \frac{p}{1-p(\tilde{n}_1-1)y_1(\tilde{n}_1)}$, and $y_1 (\tilde{n}_1) = \frac{1}{1-p_1(\tilde{n}_1-2)+p(\tilde{n}_1-1)}$. As above, the matrix determinant lemma,

$$\det \left(\tilde{\mathbf{D}} - \tilde{\mathbf{C}}\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\right) = \left(1 - \tilde{\lambda}\mathbf{v}^T\mathbf{X}\mathbf{u}\right) \det \mathbf{X}^{-1} = \frac{1 - \tilde{\lambda} \sum_{m=2}^{\ell} \tilde{n}_m x_m (\tilde{n}_m)}{x_2 (\tilde{n}_2) \cdots x_\ell (\tilde{n}_\ell)},$$

then gives us the result.

Step 4. *Show that* $\omega_{11} (\mathbf{C}_1 (\ell)) = \frac{1+(p_1-p)x_1(\tilde{n}_1)}{1+p_1} + \frac{px_1(\tilde{n}_1)^2}{1-p\sum_{m=1}^{\ell} \tilde{n}_m x_m(\tilde{n}_m)}$.

By the preceding three steps,

$$\begin{aligned} \omega_{11} (\mathbf{C}_1 (\ell)) &= v_0 = \frac{\det \tilde{\mathbf{Q}}}{\det \mathbf{Q}} = \frac{(1 - p_1 (\tilde{n}_1 - 2)) \left(1 - \tilde{\lambda} \sum_{j=2}^{\ell} \tilde{n}_j x_j (\tilde{n}_j)\right)}{(1 + p_1) (1 - p_1 (\tilde{n}_1 - 1)) \left(1 - \lambda \sum_{j=2}^{\ell} \tilde{n}_j x_j (\tilde{n}_j)\right)} \\ &= \frac{1 + (p_1 - p) x_1 (\tilde{n}_1)}{1 + p_1} + \frac{p x_1 (\tilde{n}_1)^2}{1 - p \sum_{m=1}^{\ell} \tilde{n}_m x_m (\tilde{n}_m)} \\ &= \frac{1 + (p_1 - p) x_1 (\tilde{n}_1)}{1 + p_1} + \frac{p x_1 (\tilde{n}_1)^2}{1 - p \sum_{m=1}^{M} \tilde{n}_m x_m (\tilde{n}_m)}, \end{aligned}$$

where the last equality holds because $\tilde{n}_m = 0$ for all $m > \ell$. The lemma then follows because $R_1 (\boldsymbol{C}_1 (\ell)) = a_1^2 \sigma_1^2 \omega_{11} (\boldsymbol{C}_1)$. ■

PROPOSITION 2. *There exist thresholds $\lambda_i \geq 0$ and $\mu_i \geq 0$ such that it is optimal for agent $i \in \mathcal{N}$ to tell his state to a different agent $j \in \mathcal{N}$ if and only if:*

(i.) agent j belongs to the same module $m(j) = m(i)$ with coordination need $p_{m(i)} \geq \mu_i$, or

(ii.) agent j belongs to a different module $m(j) \neq m(i)$ with cohesion $x_{m(j)} \geq \lambda_i$.

Threshold λ_i is increasing in γ, decreasing in $a_i^2 \sigma_i^2$, p, p_m, and n_m for any $m \in \mathcal{M}$, and independent of $a_k^2 \sigma_k^2$ for any $k \in \mathcal{N} \setminus \{i\}$. The comparative statics of threshold μ_i are the same, except that it is independent of $p_{m(i)}$.

Proof of Proposition 2. Fix a node i. First note that agent i's expected revenue expression is given by Lemma 3, mutatis mutandis:

$$R_i (\boldsymbol{C}_i) = a_i^2 \sigma_i^2 \left(\frac{1 + \left(p_{m(i)} - p\right) x_{m(i)} \left(\tilde{n}_{m(i)}\right)}{1 + p_{m(i)}} + \frac{p x_{m(i)} \left(\tilde{n}_{m(i)}\right)^2}{1 - p \sum_{m=1}^{M} \tilde{n}_m x_m (\tilde{n}_m)} \right).$$

This expression is convex in $\tilde{n}_m$ for each m, which implies that in any optimal communication network, agent i tells his state to all the agents in a module or to none of them.

Next, let $\mathcal{K}$ be an arbitrary set of modules, define $S(\mathcal{K}) = \sum_{m \in \mathcal{K} \setminus \{m(i)\}} n_m x_m$, and denote by $\mathbf{C}_i(\mathcal{K})$ the row of the communication matrix in which agent i tells θ_i to agent j if and only if $m(j) \in \mathcal{K}$. Then $R_i(\mathbf{C}_i(\mathcal{K}))$ can be written as an increasing and convex function of $S(\mathcal{K})$, $h(\mathcal{S}(\mathcal{K}))$.

Now suppose that it is optimal to inform all modules in $\mathcal{K} \setminus \{m(i)\}$. Then it must be the case that for all $m \in \mathcal{K} \setminus \{m(i)\}$

$$\gamma \leq \frac{h(S(\mathcal{K})) - h(S(\mathcal{K} \setminus m))}{n_m} = x_m \frac{h(S(\mathcal{K})) - h(S(\mathcal{K} \setminus m))}{S(\mathcal{K}) - S(\mathcal{K} \setminus m)} < x_m h'(S(\mathcal{K})),$$

where the last inequality holds because h is convex. Suppose further that it is not optimal to also inform some module $m' \notin \mathcal{K} \setminus \{m(i)\}$. Then it must be the case that

$$\gamma > \frac{h(S(\mathcal{K} \cup \{m'\})) - h(S(\mathcal{K}))}{n_{m'}} = x_{m'} \frac{h(S(\mathcal{K} \cup \{m'\})) - h(S(\mathcal{K}))}{S(\mathcal{K} \cup \{m'\}) - S(\mathcal{K})} > x_{m'} h'(S(\mathcal{K})).$$

These two inequalities imply that $x_m > x_{m'}$ for all modules m that are optimally told about θ_i and modules m' that are optimally not told about θ_i. In other words, there is some threshold λ_i such that agent i tells θ_i to agent j in module $m(j) \neq m(i)$ if and only if $x_{m(j)} \geq \lambda_i$. This establishes part $(ii.)$

Next, suppose it is optimal for agent i to tell θ_i to agent j if and only if $m(j) \in \mathcal{K}$. Then $m(i) \in \mathcal{K}$ if and only if

$$\frac{R_i(\mathbf{C}_i(\mathcal{K} \cup \{m(i)\})) - R_i(\mathbf{C}_i(\mathcal{K} \setminus \{m(i)\}))}{n_{m(i)} - 1} \geq \gamma,$$

Since the expression on the left-hand side of the inequality is increasing in $p_{m(i)}$, this establishes part $(i.)$.

The comparative statics follow from Proposition 9 in Appendix C, which shows that for general production networks $\mathbf{P}$ satisfying $p_{ii} = 0$, $p_{ij} = p_{ji}$, and $\sum_{j=1}^N p_{ij} < 1$, the principal's objective for the subproblem involving who agent i should tell about θ_i exhibits increasing differences in $\left(\{c_{ij}\}_j, a_i^2\sigma_i^2, \{p_{ij}\}_{ij}, -\gamma\right)$. By Topkis's theorem, then, $\mathbf{C}_i^*$ is increasing in $a_i^2\sigma_i^2$ and each p_{ij}, $j \neq i$, and it is decreasing in γ. It is therefore increasing in $a_i^2\sigma_i^2$, p, p_m, and $-\gamma$, and therefore the thresholds are decreasing in these parameters. To establish the comparative static with respect to n_m, consider an expanded production network that is the same as $\mathbf{P}$ except that it has an additional "ghost" node ℓ in module m but with $p_{\ell i} = 0$ for all $i \in \mathcal{N}$. The optimal $\mathbf{C}_i^*$ does not change with the inclusion of this ghost node. By Topkis's theorem, $\mathbf{C}_i^*$ is increasing in $p_{\ell i}$ for all $i \in \mathcal{N}$ and is therefore higher when $p_{\ell i} = p_m$ for all i such that $m(i) = m$ and $p_{\ell i} = p$ for all i such that $m(i) \neq m$ than it is when $p_{\ell i} = 0$ for all i. Adding a node to module m, therefore, increases $\mathbf{C}_i^*$. ■

COROLLARY 2. *Suppose agent i's module $m(i)$ is at least as cohesive as another module m, that is, $x_{m(i)} \geq x_m$. It cannot be optimal for agent i to tell his state to the agents in module m but not to the other agents in module $m(i)$.*

Proof of Corollary 2. We will show that if it is optimal for agent i to tell his state to agents in module m with $x_{m(i)} \geq x_m$, it must also be optimal for agent i to tell his state to the other agents in $m(i)$.

Suppose it is optimal for agent i to tell his state to all agents in the set of modules $\mathcal{K}^*$. Suppose there exists m such that $x_{m(i)} \geq x_m$, and $m \in \mathcal{K}^*$ but $m(i) \notin \mathcal{K}^*$. Define $\mathcal{K} = \mathcal{K}^* \setminus \{m\}$. Since $m \in \mathcal{K}^*$, we must have that

$$\gamma \leq \frac{R_i\left(\mathbf{C}_i\left(\mathcal{K} \cup \{m\}\right)\right) - R_i\left(\mathbf{C}_i\left(\mathcal{K}\right)\right)}{n_m}.$$

It remains to show that if $x_{m(i)} \geq x_m$, then

$$\frac{R_i\left(\mathbf{C}_i\left(\mathcal{K}^* \cup \{m(i)\}\right)\right) - R_i\left(\mathbf{C}_i\left(\mathcal{K}^*\right)\right)}{n_{m(i)} - 1} \geq \frac{R_i\left(\mathbf{C}_i\left(\mathcal{K} \cup \{m\}\right)\right) - R_i\left(\mathbf{C}_i\left(\mathcal{K}\right)\right)}{n_m},$$

which implies that agent i would want to tell his state to agents in his own module, contradicting the claim that $m(i) \notin \mathcal{K}^*$. Suppose $x_{m(i)} \geq x_m$. Then

$$\begin{aligned}
& \frac{R_i\left(\mathbf{C}_i\left(\mathcal{K}^* \cup \{m(i)\}\right)\right) - R_i\left(\mathbf{C}_i\left(\mathcal{K}^*\right)\right)}{n_{m(i)} - 1} \\
\geq \quad & a_i^2 \sigma_i^2 \frac{p\left(\frac{1}{1+p}\right)^2}{1 - p n_{m(i)} x_{m(i)} - p n_m x_m - p \sum_{\ell \in \mathcal{K}} n_\ell x_\ell} \frac{p}{1 - p\frac{1}{1+p} - p n_m x_m - p \sum_{\ell \in \mathcal{K}} n_\ell x_\ell} \frac{n_{m(i)} x_{m(i)} - \frac{1}{1+p}}{n_{m(i)} - 1} \\
\geq \quad & a_i^2 \sigma_i^2 \frac{p\left(\frac{1}{1+p}\right)^2}{1 - p n_{m(i)} x_{m(i)} - p n_m x_m - p \sum_{\ell \in \mathcal{K}} n_\ell x_\ell} \frac{p x_{m(i)}}{1 - p\frac{1}{1+p} - p n_m x_m - p \sum_{\ell \in \mathcal{K}} n_\ell x_\ell} \\
\geq \quad & a_i^2 \sigma_i^2 \frac{p\left(\frac{1}{1+p}\right)^2}{1 - p\frac{1}{1+p} - p n_m x_m - p \sum_{\ell \in \mathcal{K}} n_\ell x_\ell} \frac{p x_m}{1 - p\frac{1}{1+p} - p \sum_{\ell \in \mathcal{K}} n_\ell x_\ell} = \frac{R_i\left(\mathbf{C}_i\left(\mathcal{K} \cup \{m\}\right)\right) - R_i\left(\mathbf{C}_i\left(\mathcal{K}\right)\right)}{n_m}
\end{aligned}$$

where the first inequality follows because $x_{m(i)} > \frac{1}{1+p}$, and the second inequality follows because $n_{m(i)} x_{m(i)} - \frac{1}{1+p} \geq \left(n_{m(i)} - 1\right) x_{m(i)}$. ■

COROLLARY 3. *Optimal communication networks have the threshold property.*

Proof of Corollary 3. This result follows from Proposition 2 with $s_i = \lambda_i$ and $r_j = x_{m(j)}$ for all i, j. ■

LEMMA 4. *A communication network with the threshold property contains no two-switches or directed three-cycles.*

Proof of Lemma 4. To establish this result, we will use Theorem 1 of Cloteaux et al. (2014), which establishes a forbidden subgraph characterization of threshold directed graphs. To do so, we first introduce two definitions.

Let $\mathbf{C}$ be a communication network, and say that $\mathbf{\Psi}$ is *induced by* $\mathbf{C}$ if $\mathbf{\Psi}$ is a subgraph of $\mathbf{C}$ consisting of $\mathbf{M}$ nodes $\{i_1, \ldots, i_M\}$ with $m(i_\ell) = \ell$ and m, m' entry $\psi_{mm'} = c_{i_m i_{m'}}$ if $m \neq m'$ and $\psi_{mm} = 0$. The network $\mathbf{\Psi}$ is a *threshold directed graph* if for all m, m', m'' distinct, if $\sum_\ell \psi_{m\ell} \geq \sum_\ell \psi_{m'\ell}$ (and $\sum_\ell \psi_{\ell m} \geq \sum_\ell \psi_{\ell m'}$ if $\sum_\ell \psi_{m\ell} = \sum_\ell \psi_{m'\ell}$), then $\psi_{m'm''} = 1$ implies $\psi_{mm''} = 1$.

We begin with an alternate characterization of threshold directed graphs that parallels our threshold condition.

Step 1: *Show $\mathbf{\Psi}$ is a threshold directed graph if and only if there exists two sequences of nonnegative real numbers $\{\tilde{s}_1, \ldots, \tilde{s}_M\}$ and $\{\tilde{r}_1, \ldots, \tilde{r}_M\}$ such that $\psi_{mm'} = 1$ if and only if $\tilde{s}_m \leq \tilde{r}_{m'}$.*

Suppose there exists $\{\tilde{s}_1, \ldots, \tilde{s}_M\}$ and $\{\tilde{r}_1, \ldots, \tilde{r}_M\}$ such that $\psi_{mm'} = 1$ if and only if $\tilde{s}_m \leq \tilde{r}_{m'}$. Arrange the nodes so that $\tilde{s}_1 \leq \tilde{s}_2 \leq \cdots \leq \tilde{s}_M$, and $\tilde{s}_m = \tilde{s}_{m+1}$ implies $\tilde{r}_m \geq \tilde{r}_{m+1}$, and define $\overline{\beta}_m = \max_{\ell \neq m} \{\ell | \, \tilde{s}_\ell \leq \tilde{r}_m\}$. Let

$$\beta_m = \sum_{\ell=1}^{M} \psi_{\ell m} = \begin{cases} \overline{\beta}_m & \text{if } \overline{\beta}_m \leq m \\ \overline{\beta}_m - 1 & \text{if } \overline{\beta}_m > m. \end{cases}$$

Then $\psi_{mm'} = 1$ if and only if $\tilde{s}_m \leq \tilde{r}_{m'}$ if and only if $m \leq \beta_{m'}$ if $m < m'$ and $m \leq \beta_{m'} + 1$ if $m > m'$. We therefore have that

$$\psi_{mm'} = \begin{cases} 1 & \text{if } m < m' \text{ and } m \leq \beta_{m'} \\ 1 & \text{if } m > m' \text{ and } m \leq \beta_{m'} + 1 \\ 0 & \text{otherwise,} \end{cases}$$

and therefore by Corollary 3 of Cloteaux et al. (2014), Ψ is a threshold directed graph.

Conversely, suppose $\mathbf{\Psi}$ is a threshold directed graph. If we arrange the nodes so that $\sum_\ell \psi_{m\ell} \geq \sum_\ell \psi_{m+1,\ell}$ (and $\sum_\ell \psi_{\ell m} \geq \sum_\ell \psi_{\ell,m+1}$ if $\sum_\ell \psi_{m\ell} = \sum_\ell \psi_{m+1,\ell}$) and let $\beta_m = \sum_\ell \psi_{\ell m}$, then by Corollary 3 of Cloteaux et al. (2014),

$$\psi_{mm'} = \begin{cases} 1 & \text{if } m < m' \text{ and } m \leq \beta_{m'} \\ 1 & \text{if } m > m' \text{ and } m \leq \beta_{m'} + 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $\tilde{s}_m = m$ and $\tilde{r}_m = \min_\ell \{\ell | \, \psi_{\ell m} = 1\}$. Then $\psi_{mm'} = 1$ if $\tilde{s}_m \leq \tilde{r}_{m'}$ and $\psi_{mm'} = 0$ otherwise, establishing the claim.

Step 2: *Show that if* $\mathbf{C}$ *has the threshold property, for all* $\mathbf{\Psi}$ *induced by* $\mathbf{C}$, $\mathbf{\Psi}$ *is a threshold directed graph.*

This step follows from step 1 by setting $\tilde{s}_m = s_{i_m}$ and $\tilde{r}_m = r_{i_m}$ for all m.

Step 3: *Argue that if* $\mathbf{C}$ *has the threshold property, it has no two-switches or directed three-cycles.*

By Theorem 1 of Cloteaux et al. (2014), $\mathbf{\Psi}$ is a threshold directed graph if and only if it has no two-switches or directed three-cycles. Suppose $\mathbf{C}$ has a two-switch or a directed three-cycle. Then consider $\mathbf{\Psi}$ induced by $\mathbf{C}$ containing the two-switch or the directed three-cycle. Then $\mathbf{\Psi}$ is not a threshold directed graph, so $\mathbf{C}$ does not have the threshold property. ■

COROLLARY 4. *A communication network is not optimal if it is a tree or matrix.*

Proof of Corollary 4. Suppose $\mathbf{C}$ is a tree. Let i denote an agent in the module in level 1, j and k be agents in two different modules in level 2, and ℓ be an agent in level 3 whose unique predecessor module is $m(j)$. This is without loss of generality, since levels 2 and 3 have at least two modules with a unique predecessor module. Then agent i tells his state to agent k but not to agent ℓ, and agent j tells his state to agent ℓ but not to agent k. The communication network $\mathbf{C}$ therefore contains a two-switch and by Lemma 4 is not optimal.

Next, suppose $\mathbf{C}$ is a matrix**.** Consider two agents i and k in different modules but in the same horizontal team. Consider a different horizontal team containing two agents j and ℓ in two different modules. Since any two modules in the same horizontal team are in different vertical teams, then of the four modules containing i, j, k, and ℓ, either these modules are partitioned into two vertical teams, or they are partitioned into more than two vertical teams.

Suppose they are partitioned into two vertical teams. Suppose agent i's module and agent j's module are in one vertical team, and agent k's module and agent ℓ's module are in another vertical team. Then i tells his state to k but not ℓ, and j tells his state to ℓ but not k, so $\mathbf{C}$ contains a two-switch and is therefore not optimal. Suppose instead i, j, k, and ℓ are partitioned into at least three vertical teams. Then at least one module from each horizontal team, the modules containing i and j, say, must be in a vertical team without any of the other four modules containing i, j, k, or ℓ. Then i tells his state to k but not ℓ, and j tells his state to ℓ but not k, so $\mathbf{C}$ contains a two-switch and is therefore not optimal. ■

PROPOSITION 3. *Suppose that for any agents* $i, j \in \mathcal{N}$, $\lambda_i \leq \lambda_j$ *if and only if* $x_{m(i)} \geq x_{m(j)}$. *Any optimal communication network then has a core-periphery structure in which the agents who belong to the most cohesive modules form the core.*

Proof of Proposition 3. We construct a partition of the set of agents that has a core-periphery structure as described. Label the modules by their cohesion, with the most cohesive labeled 1 and the least cohesive labeled M, that is $x_1 \geq x_2 \geq \cdots \geq x_M$. Find the highest $k \in \{2, \ldots, M\}$ such that some agent j in module k tells his state to all agents in module $k-1$. The ordering of modules by cohesion is unique up to modules with the same cohesion. We deal with the case where no such k exists below.

We first show that agents in modules $1, \ldots, k-1$ are in the core. By Proposition 2, agent j tells his state to all agents in modules $1, \ldots, k-1$ and, since $\lambda_j \geq \lambda_i$ for all agents i in modules weakly more cohesive than j's module,then all agents in modules $1, \ldots, k-1$ tell their state to all agents in other modules $1, \ldots, k-1$. Thus we say modules $1, \ldots, k-1$ are in the core.

Next we examine whether agents in module k are in the core. If all agents in module $k-1$ tell their state to agents in module k, since $\lambda_i \leq \lambda_j$ if and only if $x_{m(i)} \geq x_{m(j)}$, then all agents in modules $1, \ldots, k-1$ tell their state to all agents in module k. Then any agents in module k that tell their state to agents in modules $1, \ldots, k-1$ are in the core. By Proposition 2, any agents in module k that do not tell their state to agents in module $k-1$ also do not inform any agents in modules $k+1, \ldots, M$, and we say they are in the periphery. If instead some agents in module $k-1$ do not tell their state to agents in module k then, by Proposition 2, those agents in module $k-1$ do not tell their state to any agents in modules $k, \ldots, M$. Since $\lambda_i \leq \lambda_j$ if and only if $x_{m(i)} \geq x_{m(j)}$, then agents in module k also do not tell their state to any agents in modules $k+1, \ldots, M$ and we say they are in the periphery.

We next show that agents in modules $k+1, \ldots, M$ do not tell their states to agents in modules $k, \ldots, M$ outside their own module, and we say they are in the periphery. By definition of module k, there is no agent in module $k+1$ who tells his state to agents in module k. Then by Proposition 2 and since $\lambda_i \leq \lambda_j$ if and only if $x_{m(i)} \geq x_{m(j)}$, no agent in module $k+1, \ldots, M$ tells his state to any agents in modules $k, \ldots, M$, aside from possibly those agents in his own module.

If there is no $k \in \{2, \ldots, M\}$ such that some agent j in module k tells his state to all agents in module $k-1$, then no agent in module 2 tells his state to any agent in module 1. By Proposition 2 and since $\lambda_i \leq \lambda_j$ if and only if $x_{m(i)} \geq x_{m(j)}$, then no agent in modules $2, \ldots, M$ tells their state to any agent outside their own module. We say agents in module 1 are in the core, and all other agents are in the periphery. ■

For the next lemma, we introduce a definition.

DEFINITION. *Let* $\boldsymbol{\Psi}$ *be a directed graph consisting of* M *nodes. We say that* $\boldsymbol{\Psi}$ *has a "generalized core-periphery structure" if its nodes can be partitioned into a core (denoted* $\mathcal{M}^C$*), a periphery (denoted* $\mathcal{M}^P$*), and a suburban periphery (denoted* $\mathcal{M}^{SP}$*), at least two of which are non-empty, satisfying the following properties:*

(*i.*) *the core is a clique:* $\psi_{mm'} = 1$ *for all* $m, m' \in \mathcal{M}^C$,

(*ii.*) *the periphery is an independent set:* $\psi_{mm'} = 0$ *for all* $m, m' \in \mathcal{M}^P$, *and*

(*iii.*) *for each* $m \in \mathcal{M}^{SP}$, *there exists some* $m' \in \mathcal{M}^C$ *(if non-empty) and* $m'' \in \mathcal{M}^P$ *(if non-empty) such that either* (*a*) $\psi_{m'm} = 0$ *and* $\psi_{m''m} = 1$ *or* (*b*) $\psi_{mm'} = 0$ *and* $\psi_{mm''} = 1$.

LEMMA 5. *Let* $\mathbf{C}$ *be an optimal communication network, and suppose* $\mathbf{\Psi}$ *is induced by* $\mathbf{C}$. *Then* $\mathbf{\Psi}$ *has a generalized core-periphery structure.*

Proof of Lemma 5. For the definition of what it means for $\mathbf{\Psi}$ to be induced by $\mathbf{C}$, see the proof of Lemma 4. Arrange the nodes $\{i_1, \ldots, i_M\}$ in Ψ so that $\lambda_1 \leq \cdots \leq \lambda_M$, and consider a trivial partition of $\{i_1, \ldots, i_M\}$ in which $m(i_\ell) \in \mathcal{M}^{SP}$ for all ℓ. There are two cases two consider. In the first case, suppose for all m, $\psi_{i_M i_m} = 0$. Consider an alternative partition in which $m(i_\ell) \in \mathcal{M}^{SP}$ for all $\ell < M$, and $m(i_M) \in \mathcal{M}^P$. In the second case, suppose $\psi_{i_M i_m} = 1$ for some $m < M$. Then take $\overline{m} = \text{argmax}_{m<M}\, x_m$. By construction, $\psi_{i_M i_{\overline{m}}} = 1$ and therefore $\psi_{i_m i_{\overline{m}}} = 1$ for all $m \neq \overline{m}$. Consider an alternative partition in which $m(i_\ell) \in \mathcal{M}^{SP}$ for all $\ell \neq \overline{m}$, and $\overline{m} \in \mathcal{M}^C$. In both cases, therefore, $\mathbf{\Psi}$ has a partition into a core, periphery, and suburban periphery, at least two of which are non-empty, and therefore $\mathbf{\Psi}$ has a generalized core-periphery structure. ■

PROPOSITION 4. *If production network* $\boldsymbol{P}$ *is weakly modular, optimal communication is all or nothing.*

Proof of Proposition 4. We will show that if $\mathbf{P}$ is weakly modular, then the per-node return to informing another module about θ_i is always increasing in the number of modules whose agents know θ_i. This result implies that it is either optimal for i to tell his state to all agents or none of them.

To make this argument, consider agent 1, and order the remaining modules in decreasing order of their cohesion: $x_2 \geq x_3 \geq \cdots \geq x_M$. The proof proceeds in four steps. First, we provide conditions under which if it is profitable to inform any module $m \in \{2, \ldots, M\}$, it is profitable to inform all modules $m \in \{2, \ldots, M\}$. Second, we provide conditions under which if it is profitable to inform only module 1, then it is also profitable to inform module $m = 2$ and therefore by the first step, all modules. Third, we provide conditions under which if it is profitable to inform only module $m = 2$, then it is also profitable to inform module 1. Finally, we show that if $\mathbf{P}$ is weakly modular, all these conditions are satisfied, and therefore optimal communication is all or nothing.

Step 1: *If* $p_m - p \leq \frac{2p(1+p)}{(1+3p)(n_m-1)}$ *for all* m, *then if it is profitable to inform modules* $2, \ldots, \ell - 1$, *it is also profitable to inform module* ℓ.

Suppose modules $\{1, \ldots, \ell - 1\}$ are told about θ_1, and consider the per-node returns to informing module ℓ. If we define $S_\ell = \sum_{m=1}^{\ell} n_m x_m$, then we can write expected revenues as

$a_1^2\sigma_1^2\left(\frac{1+(p_1-p)x_1}{1+p_1}+\frac{px_1^2}{1-pS_\ell}\right)$, so the per-node returns to informing module ℓ are

$$x_\ell\frac{a_1^2\sigma_1^2p^2x_1^2}{(1-pS_\ell)(1-pS_{\ell-1})}\equiv x_\ell b_\ell.$$

Next, note that $x_\ell b_\ell < x_{\ell+1}b_{\ell+1}$ if and only if

$$x_\ell - x_{\ell+1} < \frac{n_\ell x_\ell + n_{\ell+1}x_{\ell+1}}{1-pS_{\ell+1}}px_{\ell+1}.$$

Since $x_m \geq \frac{1}{1+p}$ for all m, the left-hand side of this inequality is less than $x_\ell - \frac{1}{1+p}$. And since $n_m \geq 1$ for all m, the right-hand side is greater than $\frac{2}{1+p}\frac{p}{1+p}$. We therefore have that a sufficient condition for $x_\ell b_\ell < x_{\ell+1}b_{\ell+1}$ for all ℓ is that

$$x_\ell - \frac{1}{1+p} < \frac{2}{1+p}\frac{p}{1+p}.$$

This inequality is satisfied for all ℓ if

$$p_\ell - p < \frac{2p(1+p)}{(1+3p)(n_\ell-1)}.$$

Similarly, suppose modules $\{2,\ldots,\ell-1\}$ are told about θ_1, and consider the per-node returns to informing module ℓ. If we define $\tilde{S}_\ell = \frac{1}{1+p}+\sum_{m=2}^{\ell} n_m x_m$, then we can write expected revenues as $a_1^2\sigma_1^2\left(\frac{1+(p_1-p)\frac{1}{1+p}}{1+p_1}+\frac{p\left(\frac{1}{1+p}\right)^2}{1-p\tilde{S}_\ell}\right)$, so the per-node returns to informing module ℓ are

$$x_\ell\frac{a_1^2\sigma_1^2p^2\left(\frac{1}{1+p}\right)^2}{\left(1-p\tilde{S}_\ell\right)\left(1-p\tilde{S}_{\ell-1}\right)}\equiv x_\ell\tilde{b}_\ell.$$

Following the same argument as above, a sufficient condition for $x_\ell\tilde{b}_\ell < x_{\ell+1}\tilde{b}_{\ell+1}$ for all ℓ is

$$p_\ell - p < \frac{2p(1+p)}{(1+3p)(n_\ell-1)}$$

for all ℓ.

Step 2: *If* $p_1 - p \leq p^2\frac{p(n_1-1)}{1-p(n_1-1)}$, *then if it is profitable to inform module* 1, *it is also profitable to inform module* 2.

The per-node returns to informing module 1, given no other modules are informed, are

$$a_1^2\sigma_1^2\frac{p_1}{1+p_1}\frac{p_1}{1-(n_1-1)p_1}.$$

The per-node returns to informing module 2, given that module 1 is informed, are

$$x_2\frac{a_1^2\sigma_1^2p^2x_1^2}{(1-pn_1x_1-pn_2x_2)(1-pn_1x_1)} = x_2\frac{a_1^2\sigma_1^2p^2x_1}{(1-pn_1x_1-pn_2x_2)(1-(n_1-1)p_1)},$$

which are higher than the returns to informing only module 1 if

$$\frac{p_1^2}{1+p_1} \leq x_2 \frac{p^2 x_1}{1 - pn_1 x_1 - pn_2 x_2}$$

The left-hand side is less than $\frac{p_1^2}{1+p}$ since $p_1 \geq p$. And the right-hand side is greater than $\frac{p^2}{1+p}\frac{1}{1-p(n_1-1)}$ because $x_2 \geq \frac{1}{1+p}$ and $pn_2x_2 \geq 0$. A sufficient condition for this inequality is therefore

$$p_1^2 - p^2 \leq p^2 \frac{p\,(n_1-1)}{1-p\,(n_1-1)},$$

and since $p_1^2 - p^2 = (p_1+p)\,(p_1-p) \leq p_1 - p$, a sufficient condition for this inequality is that

$$p_1 - p \leq p^2 \frac{p\,(n_1-1)}{1-p\,(n_1-1)},$$

which establishes the claim.

Step 3: *If $p_2 - p \leq p\,(1+p)$, then if it is profitable to inform module 2, it is also profitable to inform module 1.*

The per-node returns to informing module 2, given that no other modules are informed, are

$$a_1^2\sigma_1^2 \frac{px_2}{1-p\left(\frac{1}{1+p}+n_2x_2\right)} \frac{p}{1+p}.$$

The per-node returns to informing module 1, given that module 2 is informed, are

$$x_1 \frac{a_1^2\sigma_1^2}{1+p}\left(\left(\frac{p_1-p}{1+p_1} + p\frac{x_1+\frac{1}{1+p}}{1-p\,(n_1x_1+n_2x_2)}\right)(p_1-p) + x_1 \frac{p\frac{1}{1+p}}{1-p\,(n_1x_1+n_2x_2)} \frac{p}{1-p\left(\frac{1}{1+p}+n_2x_2\right)} \frac{1+p_1}{1+p}\right),$$

which are increasing in p_1 and are therefore greater than

$$\frac{a_1^2\sigma_1^2}{1+p} \frac{p\left(\frac{1}{1+p}\right)}{1-p\left(n_1\frac{1}{1+p}+n_2x_2\right)} \frac{1}{1-p\left(\frac{1}{1+p}+n_2x_2\right)} \frac{p}{1+p},$$

which is the same expression but with $p_1 = p$.

A sufficient condition for the claim is that

$$x_2 \leq \left(\frac{1}{1+p}\right)^2 \frac{1}{1-p\left(n_1\frac{1}{1+p}+n_2x_2\right)}$$

The right-hand side is greater than $\frac{1}{1+p}\frac{1}{1+p}\frac{1}{1-pn_2\frac{1}{1+p}}$, so a sufficient condition for this claim is

$$x_2 \leq \frac{1}{1+p}\frac{1}{1-p\,(n_2-1)},$$

which is equivalent to

$$p_2 - p \leq p\,(1+p)\,.$$

Step 4: *Suppose* **P** *is weakly modular. Then* p_m *satisfies the conditions in step 1 for all* m, p_1 *satisfies the condition in step 2, and* p_2 *satisfies the condition in step 3. Optimal communication is therefore all or nothing.*

Since **P** is weakly modular, $n_m \geq 2$ and $p_m - p < \frac{p^3(1+p)}{(n_m-1)(1+p(1+n_m))}$ for all m. We therefore have that

$$\frac{p^3\,(1+p)}{(n_m-1)\,(1+p\,(1+n_m))} \leq \frac{2p\,(1+p)}{(1+3p)\,(n_m-1)}$$

for all m,

$$\frac{p^3\,(1+p)}{(n_1-1)\,(1+p\,(1+n_1))} \leq p^2\frac{p\,(n_1-1)}{1-p\,(n_1-1)},$$

and

$$\frac{p^3\,(1+p)}{(n_2-1)\,(1+p\,(1+n_2))} \leq p\,(1+p)\,,$$

and so $p_m - p \leq \min\left\{\frac{2p(1+p)}{(1+3p)(n_m-1)}, p^2\frac{p(n_1-1)}{1-p(n_1-1)}, p\,(1+p)\right\}$. We therefore have that if it is optimal for agent 1 to tell his state to any other agent, it is optimal to tell his state to all agents. An analogous argument establishes that the same is true for any arbitrary agent i, which establishes the proposition. ■

PROPOSITION 5. *If production network* $\boldsymbol{P}$ *is weakly modular, and the modules differ in their cohesions, there exists a* $\overline{\mu} \in (0,1)$ *such that for any* $\boldsymbol{P}(\mu)$ *with* $\mu \in [\overline{\mu}, 1)$: (i.) *optimal communication is fully fragmented and* (ii.) *if it is optimal for agent* $i \in \mathcal{N}$ *to tell his state to some module* $m \in \mathcal{M}$*, it is optimal for him to tell his state to the agents in his module* $m(i)$*.*

Proof of Proposition 5. Without loss we consider agent 1 in module 1, and we order the remaining modules by cohesion with the most cohesive remaining module labelled 2.

Step 1: *There exists some* $\tilde{\mu} < 1$ *such that for all* $\mu \geq \tilde{\mu}$*, if agent* 1 *tells his state to agents in module* 2*, then he tells his state to agents in his own module* 1.

The per-node benefit of informing agents in module 1 if no one else is informed is

$$a_1^2\sigma_1^2\frac{p_1^2 x_1}{(1-pn_1x_1)\,(1+p_1)},$$

and the per-node benefit of informing the next most cohesive module if module 1 is informed is

$$a_1^2\sigma_1^2\frac{p^2x_1^2x_2}{(1-pn_1x_1)\left(1-p\sum_{m=1}^{2}n_mx_m\right)}.$$

The per-node benefit of telling agents in module 1 if no one else is told minus the per-node benefit of telling agents in module 2 if module 1 is told is

$$\frac{a_1^2\sigma_1^2x_1}{1-pn_1x_1}\left(\frac{p_1^2}{1+p_1}-\frac{p^2x_1x_2}{1-p\sum_{m=1}^{2}n_mx_m}\right).$$

Denote the term in brackets by the function $H(\mu)$. Then $H(1)>0$, since the second term in brackets is zero when $p=0$, and the first term in brackets is strictly positive. Since $H(\mu)$ is continuous in μ, there exists some $\tilde{\mu}<1$ such that for all $\mu\geq\tilde{\mu}$, the per-node benefit of informing module 1 is higher than the per-node benefit of informing module 2 given module 1 is informed. Then, by supermodularity, the per-node benefit of agent 1 telling agents in module 1 given agents in module 2 are told is higher than the per-node benefit of telling agents in module 2 alone, which establishes the claim.

Moreover, if agent 1 tells his state to agents in any other module, he tells his state to agents in module 2. And for $\mu\geq\tilde{\mu}$, if he tells his state to agents in module 2, he also tells his state to agents in module 1. This argument establishes the second part of the proposition.

Step 2: *There exists some $\hat{\mu}<1$ such that for all $\mu\geq\hat{\mu}$, if agent 1 tells his state to agents in his own module and tells his state to agents in other modules in order of their cohesion, then the per-node benefit of agent 1 telling another module his state is always decreasing in the number of modules whose agents know his state.*

Recall from the proof of Proposition 4 that if we order the modules in decreasing order of their cohesion, we can write the per-node returns to informing module ℓ as

$$x_\ell\frac{a_1^2\sigma_1^2p^2x_1^2}{(1-pS_\ell)(1-pS_{\ell-1})}\equiv x_\ell b_\ell.$$

Take a production network $\mathbf{P}$, and denote by $x_\ell(\mu)$, $b_\ell(\mu)$, and $S_\ell(\mu)$ the objects defined above that correspond to the weight-neutral increase in modularity of size μ of $\mathbf{P}$. We will show that there is a $\hat{\mu}<1$ such that if $\mu\geq\hat{\mu}$, $x_\ell(\mu)b_\ell(\mu)>x_{\ell+1}(\mu)b_{\ell+1}(\mu)$, that is, the per-node returns to informing another module about a state θ_i is always decreasing in the number of modules whose agents know θ_i. To this end, note that $x_{\ell+1}(\mu)b_{\ell+1}(\mu)-x_\ell(\mu)b_\ell(\mu)$ is equal to

$$\frac{a_1^2\sigma_1^2p(\mu)^2x_1(\mu)^2}{(1-p(\mu)S_{\ell+1}(\mu))(1-p(\mu)S_\ell(\mu))}\underbrace{\left[\frac{n_\ell x_\ell(\mu)+n_{\ell+1}x_{\ell+1}(\mu)}{1-p(\mu)S_{\ell-1}(\mu)}p(\mu)x_\ell(\mu)-(x_\ell(\mu)-x_{\ell+1}(\mu))\right]}_{\equiv H_\ell(\mu)}.$$

Moreover, $H_\ell(1)\leq 0$ for all ℓ. To see why, note that $x_\ell>x_{\ell+1}$ and $\frac{d}{d\mu}(x_\ell(\mu)-x_{\ell+1}(\mu))=pN\left(x_\ell(\mu)^2-x_{\ell+1}(\mu)^2\right)>0$ implies $x_\ell(1)>x_{\ell+1}(1)$. We therefore have that $H_\ell(1)=x_{\ell+1}(1)-$

$x_\ell(1) < 0$. Let $\overline{\mu} = \sup_{\mu\in(0,1),\ell}\{H_\ell(\mu)\}$. Since $H_\ell(\mu)$ is continuous in μ, we have that $\hat{\mu} < 1$, and for all $\mu > \hat{\mu}$, $x_\ell(\mu)\, b_\ell(\mu) > x_{\ell+1}(\mu)\, b_{\ell+1}(\mu)$ for all ℓ, which establishes the result.

Step 3: *There exists some $\overline{\mu} < 1$ such that for all $\mu \geq \overline{\mu}$, optimal communication is fully fragmented.*

Let $\overline{\mu} = \max\{\tilde{\mu}, \hat{\mu}\}$. By steps 1 and 2, for all $\mu \geq \overline{\mu}$, if it is optimal for agent 1 to tell his state to agents in some module, then it is optimal for him to tell his state to agents in his own module. Let γ_1^ℓ denote the per-node benefit of telling agents in module ℓ given those in modules $1,\ldots,\ell-1$ know his state. Then by step 2 (and since if agent 1 tells any module, he tells his own module) for $\mu \geq \overline{\mu}$ and $\ell \in \{2,\ldots,M-1\}$, if $\gamma \in \left[\gamma_1^{\ell+1}, \gamma_1^\ell\right]$, it is optimal for agent 1 to inform agents in modules $1,\ldots,\ell$, and no agents in modules $\ell+1,\ldots,M$. And if $\gamma \geq \gamma_1^M$, it is optimal to tell agents in all modules. Step 1 gives the value of the per-node benefit of informing agents in module 1 only, which gives the value of γ_1^1. ■

PROPOSITION 6. *There exists some $\gamma > 0$ such that mirroring is optimal if and only if $p \leq \overline{p}$, where $\overline{p}$ is decreasing in the module characteristics n_m and p_m for all $m \in \mathcal{M}$.*

Proof of Proposition 6. The proof proceeds in three steps. For the first step, we fix a node i and derive thresholds $\overline{\gamma}^i$ and $\overline{p}^i(\gamma)$ such that it is optimal for i to tell his state to only his own module if and only if $\gamma \leq \overline{\gamma}^i$ and $p \leq \overline{p}^i(\gamma)$. For the second step, we show that mirroring is optimal if and only if $\gamma \leq \overline{\gamma} \equiv \min_{i\in\mathcal{N}} \overline{\gamma}^i$ and $p \leq \overline{p}(\gamma) \equiv \min_{i\in\mathcal{N}} \overline{p}^i(\gamma)$. The first part of the proposition then follows from choosing $\gamma < \overline{\gamma}$ arbitrary and setting $\overline{p} = \overline{p}(\gamma)$. The third step establishes the comparative statics, which completes the proof.

Step 1: *For each i, there exist $\overline{\gamma}^i$ and $\overline{p}^i(\gamma)$ such that it is optimal for i to tell his state to only his own module if and only if $\gamma \leq \overline{\gamma}^i$ and $p \leq \overline{p}^i(\gamma)$.*

Denote agent i's revenues if he tells his state to all agents in modules $m \in \mathcal{K}_i \subset \mathcal{M}$ and to none of the agents in modules $m \notin \mathcal{K}_i$ by $R_i(\mathbf{C}_i(\mathcal{K}_i))$. Agent i's per-node returns to informing only module $m(i)$ are therefore

$$\frac{R_i(\mathbf{C}_i(\{m(i)\})) - R_i(\mathbf{C}_i(\emptyset))}{n_{m(i)} - 1} = a_i^2 \sigma_i^2 \frac{p_{m(i)}}{1 + p_{m(i)}} \frac{p_{m(i)}}{1 - \left(n_{m(i)} - 1\right) p_{m(i)}}.$$

Define this value to be $\overline{\gamma}^i$. Denote by $\mathcal{K}_i^*$ the set of modules agent i optimally tells his state to.

First, note that if $\gamma > \overline{\gamma}^i$, then $m(i) \in \mathcal{K}_i^*$ only if there is some $m \neq m(i)$ such that $m \in \mathcal{K}_i^*$. To see why, suppose $p = 0$. Then the per-node returns to informing any modules other than $m(i)$ are zero, and since $\gamma > \overline{\gamma}^i$, the per-node returns to informing agent i's own module are less than the communication costs γ, so $\mathcal{K}_i^* = \emptyset$. By definition, agent i's per-node returns to informing only his own module is independent of p, so an increase in p can expand $\mathcal{K}_i^*$, but if it does so, it necessarily adds modules $m \neq m(i)$.

Next, for γ and $\mathcal{K} \subset \mathcal{M} \backslash \{m(i)\}$ arbitrary, define $\overline{p}^i(\mathcal{K}, \gamma)$ to be such that

$$\frac{R_i\left(\mathbf{C}_i\left(\mathcal{K} \cup \{m(i)\}\right)\right) - R_i\left(\mathbf{C}_i\left(\{m(i)\}\right)\right)}{\sum_{m \in \mathcal{K}} n_m} = \gamma.$$

This value $\overline{p}^i(\mathcal{K}, \gamma)$ is the degree of coupling at which the per-node returns to informing all the agents in modules $m \in \mathcal{K} \cup \{m(i)\}$ relative to informing only those in $m(i)$ are exactly equal to the costs of doing so. Define $\overline{p}^i(\gamma) \equiv \min_{\mathcal{K} \subset \mathcal{M}} \overline{p}^i(\mathcal{K}, \gamma)$. Then $\mathcal{K}_i^* = \{m(i)\}$ if and only if $\gamma \leq \overline{\gamma}^i$ and $p \leq \overline{p}^i(\gamma)$.

Step 2: *Mirroring is optimal if and only if* $\gamma \leq \overline{\gamma} \equiv \min_{i \in \mathcal{N}} \overline{\gamma}^i$ *and* $p \leq \overline{p}(\gamma) \equiv \min_{i \in \mathcal{N}} \overline{p}^i(\gamma)$.

Define $\overline{\gamma} \equiv \min_{i \in \mathcal{N}} \overline{\gamma}^i$ and $\overline{p}(\gamma) \equiv \min_{i \in \mathcal{N}} \overline{p}^i(\gamma)$. If $\gamma \leq \overline{\gamma}$, then $m(i) \in \mathcal{K}_i^*$ for all i. If $p \leq \overline{p}(\gamma)$, then for all i and $m \neq m(i)$, $m \notin \mathcal{K}_i^*$, so mirroring is optimal. This establishes the "if" part of the claim.

To establish the "only if" part of the claim, suppose $\gamma > \overline{\gamma}$. Then for some i, by the argument in step 1, if $m(i) \in \mathcal{K}_i^*$, there exists some $m \neq m(i)$ such that $m \in \mathcal{K}_i^*$, so mirroring is not optimal. And if $\gamma \leq \overline{\gamma}$ but $p > \overline{p}(\gamma)$, then there exists some i such that $p > \overline{p}^i(\gamma)$. This implies that there exists some $m \neq m(i)$ such that $m \in \mathcal{K}_i^*$, so again, mirroring is not optimal.

Step 3: *For each i, the threshold $\overline{p}^i(\gamma)$ is decreasing in n_m and p_m for all m. Therefore $\overline{p}(\gamma)$ is decreasing, in n_m and p_m for all m.*

Take i and $\mathcal{K} \subset \mathcal{M} \backslash \{m(i)\}$ arbitrary, and note that by Proposition 9, the expression

$$\frac{R_i\left(\mathbf{C}_i\left(\mathcal{K} \cup \{m(i)\}\right)\right) - R_i\left(\mathbf{C}_i\left(\{m(i)\}\right)\right)}{\sum_{m \in \mathcal{K}} n_m} - \gamma$$

is increasing in n_m and p_m for all m, and it is increasing in p. This implies that $\overline{p}^i(\mathcal{K}, \gamma)$, the value of p at which this expression is zero, is decreasing in n_m and p_m for all m. Since i and $\mathcal{K}$ were arbitrary, it follows that $\overline{p}(\gamma)$ is decreasing in n_m and p_m for all m. ■

COROLLARY 5. *If an optimal communication network has a team structure, it has the following properties: (i.) a team with two or more agents either includes all agents of a module or none of them, (ii.) there is at most one team whose agents belong to two or more modules, and (iii.) a team whose agents belong to two or more modules includes the agents from the most cohesive modules.*

Proof of Corollary 5. Part (*i.*) follows from the first part of the proof of Proposition 2: if agent i optimally tells his state to agent j, then he also tells his state to all other agents in module $m(j)$. And so if i and j are in a team, then so too must be all k with $m(k) = m(j)$.

For part (*ii.*), suppose there are at least two teams in the optimal communication network, each of which contains agents who belong to at least two different modules. Without loss, suppose the first team contains agents i and j with $m(i) = 1$ and $m(j) = 2$, and the second team contains agents k and ℓ with $m(k) = 3$ and $m(\ell) = 4$. Then the communication network contains a two-switch, and by Lemma 4, it is not optimal.

For part (*iii.*), suppose there is a team whose agents belong to two or more modules. Suppose module m is part of this team. By Proposition 2, an agent i in module m tells his state to an agent j outside his module if and only if $x_{m(j)} \geq \lambda_i$. The symmetric argument holds for all individuals within all modules in this team. Therefore this team contains the modules m with the highest values of x_m. ■

For the next proposition, let $\mathbf{C}$ be a communication network, and say that $\mathbf{\Psi}$ is *induced by* $\mathbf{C}$ if $\mathbf{\Psi}$ is a subgraph of $\mathbf{C}$ consisting of M nodes $\{i_1, \ldots, i_M\}$ with $m(i_\ell) = \ell$ and m, m' entry $\psi_{mm'} = c_{i_m i_{m'}}$ if $m \neq m'$ and $\psi_{mm} = 0$. Given $\mathbf{\Psi}$, define $\delta^+ = \left(\delta_1^+, \ldots, \delta_M^+\right)^T$, $\delta^- = \left(\delta_1^-, \ldots, \delta_M^-\right)^T$, and $\overline{\delta} = \left(\overline{\delta}_1, \ldots, \overline{\delta}_M\right)^T$, where $\delta_m^+ = \sum_{j \neq m} \psi_{mj}$ is node m's *out-degree*, $\delta_m^- = \sum_{j \neq m} \psi_{jm}$ is node m's *in-degree*, and $\overline{\delta}_m = \sum_{j<m} 1_{\delta_j^- \geq m-1} + \sum_{j>m} 1_{\delta_j^- \geq m}$ is node m's *conjugate out-degree*. Define $\Delta(\mathbf{\Psi}) = \sum_{m=1}^M \max\left\{\overline{\delta}_m - \delta_m^+, 0\right\}$ to be the *threshold gap* of $\mathbf{\Psi}$. We say that $\mathbf{\Psi}$ is *properly ordered* if $m < m'$ if and only if $\delta_m^+ > \delta_{m'}^+$ or $\delta_m^+ = \delta_{m'}^+$ and $\delta_m^- \geq \delta_{m'}^-$. Given $\mathbf{\Psi}$, we say $s = ij \to k\ell$ is a *valid swap* if $\psi_{ij} = 1$ and $\psi_{k\ell} = 0$, and we define $\tilde{\mathbf{\Psi}} = \mathbf{\Psi} + s$ to be an $M \times M$ matrix with $(m, m') = (i, j)$ entry equal to 0, $(m, m') = (k, \ell)$ entry equal to one, and all other (m, m') entries equal to $\psi_{mm'}$.

PROPOSITION 7. *If $\boldsymbol{C}^*$ is an optimal communication network, then the threshold gap $\Delta(\mathbf{\Psi})$ is zero for any network $\mathbf{\Psi}$ induced by $\boldsymbol{C}^*$. If network $\mathbf{\Psi}$ has $\Delta(\mathbf{\Psi}) > 0$, then $\lceil \Delta(\mathbf{\Psi})/2 \rceil$ is the minimum number of swaps required for the resulting communication network to have a threshold gap of zero and thus exhibit the threshold property.*

Proof of Proposition 7. For the first part of the proposition, recall from step 2 of Lemma 4 that if $\mathbf{C}$ has the threshold property, so does any $\mathbf{\Psi}$ induced by $\mathbf{C}$. By Cloteaux et al. (2014), Theorem 1, if $\mathbf{\Psi}$ is properly ordered, then the M Fulkerson-Chen inequalities hold with equality. Thus, for all m, $\delta_m^+ = \overline{\delta}_m$, and hence $\Delta(\mathbf{\Psi}) = 0$. This part of the proposition then follows because if $\mathbf{C}^*$ is an optimal communication network, then by Corollary 3, it has the threshold property.

For the second part of the proposition, we will establish that if Ψ is properly ordered, then $\lceil \Delta(\mathbf{\Psi})/2 \rceil$ is a lower bound on the number of sequential valid swaps on Ψ such that the resulting $\tilde{\Psi}$, properly ordered, has $\Delta\left(\tilde{\mathbf{\Psi}}\right) = 0$. To do so, we will show that if $\mathbf{\Psi}$ is properly ordered, and $\Delta(\mathbf{\Psi}) > 0$, then if $\tilde{\mathbf{\Psi}} = \mathbf{\Psi} + s$ for some valid swap s, and $\tilde{\mathbf{\Psi}}'$ is a proper reordering of $\tilde{\mathbf{\Psi}}$,

$\Delta\left(\tilde{\mathbf{\Psi}}'\right) \geq \Delta(\mathbf{\Psi}) - 2$. One challenge the proof has to overcome is that $\tilde{\mathbf{\Psi}} = \mathbf{\Psi} + s$ may not be properly ordered. The proof proceeds in three steps.

Step 1: *Suppose $\mathbf{\Psi}$ is not properly ordered, and let $\mathbf{\Psi}'$ be a proper reordering of $\mathbf{\Psi}$. Then $\Delta(\mathbf{\Psi}') \leq \Delta(\mathbf{\Psi})$.*

If $\mathbf{\Psi}$ is not properly ordered, then by the Bubble-Sorting Algorithm, there exists a finite sequence $\left\{\tilde{\mathbf{\Psi}}^k\right\}_{k=0}^K$ such that $\mathbf{\Psi} = \tilde{\mathbf{\Psi}}^0$, $\mathbf{\Psi}' = \tilde{\mathbf{\Psi}}^K$, and each $\tilde{\mathbf{\Psi}}^k$ is an adjacent reordering of $\tilde{\mathbf{\Psi}}^{k-1}$, swapping the ℓ_k and ℓ_k+1 rows and columns of $\tilde{\mathbf{\Psi}}^{k-1}$, where in $\tilde{\mathbf{\Psi}}^{k-1}$, either $\delta^+_{\ell_k} < \delta^+_{\ell_k+1}$ or $\delta^+_{\ell_k} = \delta^+_{\ell_k+1}$ and $\delta^-_{\ell_k} < \delta^-_{\ell_k+1}$. We will show that each adjacent reordering weakly reduces Δ. Specifically, suppose node ℓ should out-rank node $\ell+1$. That is, either $\delta^+_\ell < \delta^+_{\ell+1}$ or $\delta^+_\ell = \delta^+_{\ell+1}$ and $\delta^-_\ell < \delta^-_{\ell+1}$. Construct $\tilde{\mathbf{\Psi}}$ from $\mathbf{\Psi}$ by swapping the ℓ and $\ell+1$ rows and columns.

Let $\zeta = \overline{\delta} - \delta^+ = \sum_{m=1}^M \overline{\Xi}_m - \delta^+$, where

$$\overline{\Xi}_m = \begin{bmatrix} 1_{\delta^-_m \geq 1} \\ \vdots \\ 1_{\delta^-_m \geq m-1} \\ 0 \\ 1_{\delta^-_m \geq m} \\ \vdots \\ 1_{\delta^-_m \geq N-1} \end{bmatrix}.$$

Then if we let $x = 1_{\delta^-_{\ell+1} \geq \ell} - 1_{\delta^-_\ell \geq \ell} + \delta^+_{\ell+1} - \delta^+_\ell$, we can write

$$\tilde{\zeta} = \zeta - x \cdot e_\ell + x \cdot e_{\ell+1},$$

where e_m is an $M \times 1$ vector with m^{th} entry 1 and other entries 0. Note that $x \geq 0$ because either $\delta^+_\ell < \delta^+_{\ell+1}$ or $\delta^+_\ell = \delta^+_{\ell+1}$ and $\delta^-_\ell < \delta^-_{\ell+1}$.

Next, notice that

$$\zeta_\ell = \zeta_{\ell+1} + \sum_{m<\ell} 1_{\delta^-_m = \ell-1} + \sum_{m>\ell} 1_{\delta^-_m = \ell} + 1_{\delta^-_{\ell+1} \geq \ell} - 1_{\delta^-_\ell \geq \ell} + \delta^+_{\ell+1} - \delta^+_\ell = \zeta_{\ell+1} + z + x,$$

where

$$z = \sum_{m<\ell} 1_{\delta^-_m = \ell-1} + \sum_{m>\ell} 1_{\delta^-_m = \ell}.$$

Let $h(w) = \max\{w, 0\}$. We then have

$$\begin{aligned}
\Delta\left(\tilde{\boldsymbol{\Psi}}\right) &= \Delta(\boldsymbol{\Psi}) + h\left(\tilde{\zeta}_\ell\right) + h\left(\tilde{\zeta}_{\ell+1}\right) - h(\zeta_\ell) - h\left(\zeta_{\ell+1}\right) \\
&= \Delta(\boldsymbol{\Psi}) + \left[h(\zeta_\ell - x) - h(\zeta_\ell)\right] + \left[h\left(\zeta_{\ell+1} + x\right) - h\left(\zeta_{\ell+1}\right)\right] \\
&= \Delta(\boldsymbol{\Psi}) + \left[h\left(\zeta_{\ell+1} + x\right) - h\left(\zeta_{\ell+1}\right)\right] - \left[h\left(\zeta_{\ell+1} + x + z\right) - h\left(\zeta_{\ell+1} + z\right)\right] \\
&\leq \Delta(\boldsymbol{\Psi}),
\end{aligned}$$

since $h(\cdot)$ is a convex function.

We therefore have

$$\Delta(\boldsymbol{\Psi}) = \Delta\left(\tilde{\boldsymbol{\Psi}}^0\right) \leq \Delta\left(\tilde{\boldsymbol{\Psi}}^1\right) \leq \cdots \leq \Delta\left(\tilde{\boldsymbol{\Psi}}^K\right) = \Delta\left(\boldsymbol{\Psi}'\right),$$

which establishes the claim.

Step 2: *Suppose $\boldsymbol{\Psi}$ is properly ordered, and $s = ij \to k\ell$ is a valid swap. Let $\tilde{\boldsymbol{\Psi}} = \boldsymbol{\Psi} + s$, not necessarily properly ordered. Then $\Delta(\boldsymbol{\Psi}) - 2 \leq \Delta\left(\tilde{\boldsymbol{\Psi}}\right) \leq \Delta(\boldsymbol{\Psi}) + 2$.*

Since s is a valid swap, we have $\tilde{\delta}^+ = \delta^+ + e_k - e_i$ and $\tilde{\delta}^- = \delta^- + e_\ell - e_j$. We can write $\tilde{\overline{\delta}} = \overline{\delta} + \left(\tilde{\overline{\Xi}}_\ell - \overline{\Xi}_\ell\right) + \left(\tilde{\overline{\Xi}}_j - \overline{\Xi}_j\right)$, where

$$\tilde{\overline{\Xi}}_\ell - \overline{\Xi}_\ell = \begin{bmatrix} 1_{\delta_\ell^- \geq 0} \\ \vdots \\ 1_{\delta_\ell^- \geq \ell-2} \\ 0 \\ 1_{\delta_\ell^- \geq \ell-1} \\ \vdots \\ 1_{\delta_\ell^- \geq N-2} \end{bmatrix} = e_m \text{ and } \tilde{\overline{\Xi}}_j - \overline{\Xi}_j = \begin{bmatrix} 1_{\delta_j^- \geq 1} \\ \vdots \\ 1_{\delta_j^- \geq j-1} \\ 0 \\ 1_{\delta_j^- \geq j} \\ \vdots \\ 1_{\delta_j^- \geq N-1} \end{bmatrix} = e_{m'}$$

for some m, m', since s is a valid swap. We therefore have that $\tilde{\zeta} = \zeta + e_m + e_i - e_{m'} - e_k$, and so if we let $\Delta\left(\tilde{\boldsymbol{\Psi}}\right) = \sum_{t=1}^M \max\left\{\tilde{\zeta}_t, 0\right\}$, we have $\Delta(\boldsymbol{\Psi}) - 2 \leq \Delta\left(\tilde{\boldsymbol{\Psi}}\right) \leq \Delta(\boldsymbol{\Psi}) + 2$.

Step 3. *Suppose $\boldsymbol{\Psi}$ is properly ordered, and $s = ij \to k\ell$ is a valid swap. Let $\tilde{\boldsymbol{\Psi}} = \boldsymbol{\Psi} + s$, not necessarily properly ordered, and let $\tilde{\boldsymbol{\Psi}}'$ be a proper reordering of $\tilde{\boldsymbol{\Psi}}$. Then $\Delta\left(\tilde{\boldsymbol{\Psi}}'\right) \geq \Delta(\boldsymbol{\Psi}) - 2$.*

Given s, suppose $\tilde{\boldsymbol{\Psi}} = \boldsymbol{\Psi} + s$ is properly ordered. Then the result follows immediately from Step 2, since $\tilde{\boldsymbol{\Psi}} = \tilde{\boldsymbol{\Psi}}'$.

Suppose instead that $\tilde{\boldsymbol{\Psi}} = \boldsymbol{\Psi} + s$ is not properly ordered. Let $\hat{\boldsymbol{\Psi}}$ be a reordering of $\boldsymbol{\Psi}$ such that nodes are ordered as in $\tilde{\boldsymbol{\Psi}}'$. Then by Step 1, we have $\Delta(\boldsymbol{\Psi}) \leq \Delta\left(\hat{\boldsymbol{\Psi}}\right)$. By Step 2, $\Delta\left(\tilde{\boldsymbol{\Psi}}'\right) \geq \Delta\left(\hat{\boldsymbol{\Psi}}\right) - 2$. Putting these inequalities together, we have $\Delta\left(\tilde{\boldsymbol{\Psi}}'\right) \geq \Delta\left(\hat{\boldsymbol{\Psi}}\right) - 2 \geq \Delta(\boldsymbol{\Psi}) - 2$, which establishes the claim. ■

Appendix B: Noisy Communication

This appendix derives the principal's problem when communication is imperfect, but its precision is a choice. Specifically, suppose each $\theta_j \sim N\left(0, \sigma_j^2\right)$, and each agent i receives a noisy and conditionally independent signal of θ_j, denoted by $s_{ij} = \theta_j + \eta_{ij}$, where $\eta_{ij} \sim N\left(0, 1/\tau_{ij}\right)$, and η_{ij} is independent of $\eta_{k\ell}$ for all $(i,j) \neq (k,\ell)$. Let $\varphi_{ij} = \tau_{ij} / \left(\tau_{ij} + \sigma_j^{-2}\right)$ be the signal-to-noise ratio for i's signal of θ_j, and denote by $k\left(\varphi_{ij}\right)$ the cost of agent j ensuring that agent i's signal of θ_j has signal-to-noise ratio φ_{ij}. Assume, as in the main model, that the total communication costs are additively separable across i and j. Further, assume that $p > 0$. Then the principal's problem is to choose a matrix $\mathbf{\Phi}$ with ij^{th} element φ_{ij} to solve

$$\max_{\mathbf{\Phi}} E\left[r\left(d_1, \ldots, d_N\right) \middle| \mathbf{\Phi}\right] - \sum_{i=1}^{N} \sum_{j=1}^{N} k\left(\varphi_{ij}\right).$$

As in the main model, the timing of the game is as follows. First, the principal chooses $\mathbf{\Phi}$. Then agents learn their states and send each other signals with signal-to-noise ratios as specified in $\mathbf{\Phi}$. Next, agents simultaneously make their decisions, and the game ends.

The following proposition establishes existence of equilibrium decision rules that are linear in parties' signals, derives expressions for the coefficients on these decision rules, and shows that the principal's problem can be decomposed into N subproblems, as in Proposition 1.

PROPOSITION B1. *Given signals $\{s_{ij}\}_{ij}$, there are unique linear equilibrium decisions given by*

$$d_i^* = \sum_{j=1}^{N} a_j \omega_{ij}\left(\mathbf{\Phi}_j\right) s_{ij} \text{ for all } i \in \mathcal{N},$$

where $\omega_{ij}\left(\mathbf{\Phi}_j\right)$ denotes the ij^{th} entry of $\left(I - \left(\operatorname{diag} \mathbf{\Phi}_j\right) \mathbf{P}\right)^{-1}$.

Proof of Proposition B1. By Lambert, Martini, and Ostrovsky (2018), Theorem 2, given a communication structure there is a unique linear-in-signals equilibrium. These equilibrium strategies take the form

$$d_j^*\left(s_{j1}, s_{j2}, \ldots, s_{jN}\right) = \beta_j + \sum_{k=1}^{N} \alpha_{jk} s_{jk}$$

for each $j \in \mathcal{N}$, where β_j and α_{jk} are scalars.

We next establish the equilibrium values of β_j and α_{jk} for all $j, k \in \mathcal{N}$. The best response functions for each agent are as in the main model

$$d_i = a_i \theta_i + \sum_{j=1}^{N} p_{ij} E\left[d_j \middle| \mathbf{\Phi}_{(i)}\right].$$

Using the linear equilibrium strategies above, agent i's expectation of j's equilibrium strategy can be written

$$E\left[d_j^* \middle| \Phi_{(i)}\right] = \beta_j + \sum_{k=1}^{N} \alpha_{jk} E\left[s_{jk} \middle| \Phi_{(i)}\right] = \beta_j + \sum_{k=1}^{N} \alpha_{jk} E\left[\theta_k + \eta_{jk} \middle| \Phi_{(i)}\right] = \beta_j + \sum_{k=1}^{N} \alpha_{jk} \varphi_{ik} s_{ik}.$$

Plugging equilibrium decisions into the best response function, we have

$$d_i^* = a_i \theta_i + \sum_{j=1}^{N} p_{ij} E\left[d_j^* \middle| \Phi_{(i)}\right] = a_i s_{ii} + \sum_{j=1}^{N} p_{ij} \left[\beta_j + \sum_{k=1}^{N} \alpha_{jk} \varphi_{ik} s_{ik}\right].$$

We therefore have a system of equations

$$a_i s_{ii} + \sum_{j=1}^{N} p_{ij} \left[\beta_j + \sum_{k=1}^{N} \alpha_{jk} \varphi_{ik} s_{ik}\right] = \beta_i + \sum_{k=1}^{N} \alpha_{ik} s_{ik}$$

for each $i \in \mathcal{N}$. So for each $i \in \mathcal{N}$, the following equilibrium conditions hold:

$$\begin{aligned} \sum_{j=1}^{N} p_{ij} \beta_j &= \beta_i \\ \alpha_{ik} &= \varphi_{ik} \sum_{j=1}^{N} p_{ij} \alpha_{jk} \text{ for all } k \in \mathcal{N},\ k \neq i, \text{ and} \\ \alpha_{ii} &= a_i + \varphi_{ii} \sum_{j=1}^{N} p_{ij} \alpha_{ji}. \end{aligned}$$

We can write the equilibrium conditions for a given $k \in \mathcal{N}$ and each $i \in \mathcal{N}$ as a system of N equations

$$\alpha_k = \vec{a}_k + (\operatorname{diag} \mathbf{\Phi}_k)\, \mathbf{P} \alpha_k,$$

where $\vec{a}_k$ is a $1 \times N$ vector of zeros and value a_k in the kth position, and α_k is a $1 \times N$ vector with ith entry α_{ik}. Rearranging this system, we have

$$\alpha_k = \left[I - (\operatorname{diag} \mathbf{\Phi}_k)\, \mathbf{P}\right]^{-1} \vec{a}_k,$$

where $I - (\operatorname{diag} \mathbf{\Phi}_k)\, \mathbf{P}$ is invertible since the spectral radius of $\mathbf{P}$ is strictly less than 1. Therefore α_{ij} is equal to the ijth entry of $a_j \left[I - (\operatorname{diag} \mathbf{\Phi}_k)\, \mathbf{P}\right]^{-1}$. Finally, the values of β_j satisfy $\mathbf{P}\beta = \beta$, where β is a vector with ith entry β_i. Since $\mathbf{P}$ is full rank and has spectral radius strictly less than 1, we must have $\beta = 0$. ■

PROPOSITION B2. *An optimal communication pattern $\mathbf{\Phi}^*$ solves the principal's problem if and only if $\varphi_{ij}^* = \varphi_{m(i)j}^*$ for all i, j, and $\mathbf{\Phi}_j^*$ solves*

$$\max_{\mathbf{\Phi}_j} R_j\left(\mathbf{\Phi}_j\right) - \sum_{m=1}^{M} n_m k\left(\varphi_{mj}\right).$$

Moreover, if within-module communication is free, then

$$R_j\left(\mathbf{\Phi}_j\right)=a_j^2\sigma_j^2\left(\frac{1+\left(p_{m(j)}-p\right)x_{m(j)}}{1+p_{m(j)}}+\frac{px_{m(j)}^2}{1-p\left(\sum_{m=1}^{M}\varphi_{mj}n_m x_m\left(\varphi_{mj}\right)\right)}\right),$$

where $\varphi_{m(j)j}=1$ and $x_m\left(\varphi_{mj}\right)=\left(1+\varphi_{mj}\left(p-\left(n_m-1\right)\left(p_m-p\right)\right)\right)^{-1}$.

Proof of Proposition B2. This proof proceeds in three steps. First, we establish the separability result. Second, we establish that in any optimal communication pattern, $\varphi_{ij}^*=\varphi_{m(i)j}^*$ for all i,j. Third, we compute the expression for agent j's expected revenues given $\mathbf{\Phi}_j$.

Step 1. *Show that an optimal communication pattern solves the principal's problem if and only if it solves i's sub-problem for all i.*

Using the same argument as in Lemma 1, we have

$$E\left[r\left(d_1,\ldots,d_N\right)\middle|\,\mathbf{\Phi}\right]=\sum_{i=1}^{N}R_i\left(\mathbf{\Phi}\right)=\sum_{i=1}^{N}a_i\operatorname{Cov}\left(d_i^*,\theta_i\right).$$

Since θ_i and s_{ij} are independent, then

$$\operatorname{Cov}\left(d_i^*,\theta_i\right)=a_i\omega_{ii}\left(\mathbf{\Phi}_i\right)\sigma_i^2,$$

where by Proposition B1, $\omega_{ii}\left(\mathbf{\Phi}_i\right)$ is the value of all directed walks from i to i on $\left(\operatorname{diag}\mathbf{\Phi}_i\right)\mathbf{P}$. Since $R_i\left(\mathbf{\Phi}\right)$ depends only on $\mathbf{\Phi}_i$, this establishes the first step.

Step 2. *Show that* $\varphi_{ij}^*=\varphi_{m(i)j}^*$ *for all* j.

Consider informing agents about state θ_1. Let $\varphi_1^*=\left(\varphi_{21}^*,\ldots,\varphi_{N1}^*\right)$ denote the optimal signal-to-noise ratios for agent 1's communication. Suppose that for agent i and agent j in the same module, we have $\varphi_{i1}^*>\varphi_{j1}^*$. Denote by $\omega_{11}\left(\varphi_{21},\ldots,\varphi_{N1}\right)$ the value of all closed walks from agent 1 to itself on $\left(\operatorname{diag}\mathbf{\Phi}_1\right)\mathbf{P}$ for $\mathbf{\Phi}_1$ arbitrary.. Then $\omega_{11}\left(\varphi_{i1},\varphi_{-i1}^*\right)-\omega_{11}\left(0,\varphi_{-i1}^*\right)$ is the value of all closed walks from agent 1 to itself when agent i's signal s_{i1} has signal-to-noise ratio φ_{i1} and agent $j\neq i$'s signal s_{j1} has signal-to-noise ratio φ_{j1}^*. This difference is a measure of the additional value generated by informing agent i with signal-to-noise ratio φ_{i1} versus not informing him at all. Since φ_{i1}^* is optimal, it follows that $\omega_{11}\left(\varphi_{i1},\varphi_{-i1}^*\right)-\omega_{11}\left(0,\varphi_{-i1}^*\right)-k\left(\varphi_{i1}\right)$ is maximized at φ_{i1}^*.

Similarly, let $\omega_{11}\left(\varphi_{j1},\varphi_{-j1}^*\right)-\omega_{11}\left(0,\varphi_{-j1}^*\right)$ be the value of all closed walks from agent 1 to itself when agent j's signal s_{j1} has signal-to-noise ratio φ_{j1} and agent $i\neq j$'s signal s_{i1} has signal-to-noise ratio φ_{i1}^*. We will show that $\omega_{11}\left(\varphi_{j1},\varphi_{-j1}^*\right)-\omega_{11}\left(0,\varphi_{-j1}^*\right)-k\left(\varphi_{j1}\right)$ is not maximized at $\varphi_{j1}^*<\varphi_{i1}^*$, and so φ_{j1}^* cannot be optimal because then we could increase total profits by changing φ_{j1} and holding all other φ_{-j1} fixed.

Write $\omega_{11}\left(\varphi, \varphi^*_{-j1}\right) - \omega_{11}\left(0, \varphi^*_{-j1}\right) - k(\varphi)$ as

$$\begin{aligned}
&\left(\omega_{11}\left(\varphi, \varphi^*_{-i1}\right) - \omega_{11}\left(0, \varphi^*_{-i1}\right) - k(\varphi)\right) \\
&+ \left[\left(\omega_{11}\left(\varphi, \varphi^*_{-j1}\right) - \omega_{11}\left(0, \varphi^*_{-j1}\right)\right) - \left(\omega_{11}\left(\varphi, \varphi^*_{-i1}\right) - \omega_{11}\left(0, \varphi^*_{-i1}\right)\right)\right].
\end{aligned}$$

From above, the term $\omega_{11}\left(\varphi, \varphi^*_{-i1}\right) - \omega_{11}\left(0, \varphi^*_{-i1}\right) - k(\varphi)$ is higher at $\varphi = \varphi^*_{i1}$ that at the lower value $\varphi = \varphi^*_{j1}$. It remains to show that the remaining term is positive and increasing in φ.

The term $\omega_{11}\left(\varphi, \varphi^*_{-i1}\right) - \omega_{11}\left(0, \varphi^*_{-i1}\right)$ gives the value of all walks in the network $(\operatorname{diag} \mathbf{\Phi}_1)\mathbf{P}$ with $\varphi_{i1} = \varphi$ and $\varphi_{-i1} = \varphi^*_{-i1}$ that pass through agent i. We can partition this set of walks into walks that pass through i but not j and walks that pass through i and j. Similarly, the term $\omega_{11}\left(\varphi, \varphi^*_{-j1}\right) - \omega_{11}\left(0, \varphi^*_{-j1}\right)$ gives the value of all walks in the network $(\operatorname{diag} \mathbf{\Phi}_1)\mathbf{P}$ with $\varphi_{j1} = \varphi$ and $\varphi_{-j1} = \varphi^*_{-j1}$ that pass through agent j. We can partition this set of walks into walks that pass through j but not i and walks that pass through i and j. The only difference between the networks is that in the first one, we have $\varphi_{i1} = \varphi$ and $\varphi_{j1} = \varphi^*_{j1}$ and in the second network, we have $\varphi_{j1} = \varphi$ and $\varphi_{i1} = \varphi^*_{i1}$. Since i and j are in the same module and so otherwise symmetric, the value of all walks that pass through i but not j in the first network is the same as the value of all walks that pass through j but not i in the second network.

Now consider the walks that pass through both i and j. Write the value of such a walk as

$$p_{1j_1}p_{j_1j_2}\cdots p_{j_{\ell-1}1}\varphi_{j_11}\cdots\varphi_{j_{\ell-1}1}.$$

Each pair of links in this walk that go to and then from node i contributes to the value of this walk in two ways: as part of a $p_{ki}p_{i\ell}$ term and a φ_{i1} term. Similarly, each link in this walk that goes into j contributes to a $p_{hj}p_{jm}$ term and a φ_{j1} term. For each such walk we can find an equivalent walk where any pair of links to and from i are replaced by a pair of links to and from j, and any pair of links to and from j are replaced by a pair of links to and from i. Because of the symmetry of the production network and because i and j are in the same module, then $p_{ki}p_{i\ell} = p_{kj}p_{j\ell}$ and so there is no difference in the value of the walk coming from the production network. The difference in the value of the walks comes from the φ_{i1} and φ_{j1} terms. In the network $(\operatorname{diag} \mathbf{\Phi}_1)\mathbf{P}$ with $\varphi_{i1} = \varphi$ and $\varphi_{-i1} = \varphi^*_{-i1}$ consider a walk that passes out of node i m times and passes out of node j n times and so its value is multiplied by $\varphi^m\left(\varphi^*_{j1}\right)^n$. We can find an equivalent walk in network $(\operatorname{diag} \mathbf{\Phi}_1)\mathbf{P}$ with $\varphi_{j1} = \varphi$ and $\varphi_{-j1} = \varphi^*_{-j1}$, where we replace j with i, and so it passes out of node j m times and out of node i n times, and so its value is multiplied by $\varphi^m\left(\varphi^*_{i1}\right)^n$. Since $\varphi^m\left(\varphi^*_{i1}\right)^n > \varphi^m\left(\varphi^*_{j1}\right)^n$, and the difference in value between these walks is increasing in φ, this establishes the claim.

Step 3. *Assume within-module communication is free. Show that*

$$R_j\left(\mathbf{\Phi}_j\right)=a_j^2\sigma_j^2\left(\frac{1+\left(p_{m(j)}-p\right)x_{m(j)}}{1+p_{m(j)}}+\frac{px_{m(j)}^2}{1-p\left(\sum_{m=1}^{M}\varphi_{mj}n_m x_m\left(\varphi_{mj}\right)\right)}\right),$$

where $\varphi_{m(j)j}=1$ and $x_m\left(\varphi_{mj}\right)=\left(1+\varphi_{mj}\left(p-\left(n_m-1\right)\left(p_m-p\right)\right)\right)^{-1}$.

This argument parallels the argument in the proof of Lemma 3. Without loss of generality, we compute $R_1\left(\mathbf{\Phi}_1\right)$. By the same argument as in step 1 of the proof of Lemma 3, if we let v_0 represent the sum of the value of all walks from node 1 back to itself on the reweighted network $\left(\operatorname{diag}\mathbf{\Phi}_j\right)\mathbf{P}$, and v_k be the sum of the value of all walks from a node in module k to node 1 on $\left(\operatorname{diag}\mathbf{\Phi}_j\right)\mathbf{P}$, then

$$\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}=\mathbf{Q}\begin{bmatrix}v_0\\v_1\\\vdots\\v_M\end{bmatrix},$$

where

$$\mathbf{Q}=\begin{bmatrix}1 & -p_1\left(n_1-1\right) & -pn_2 & \cdots & -pn_M\\ -p_1 & 1-p_1\left(n_1-2\right) & -pn_2 & \cdots & -pn_M\\ -\varphi_{21}p & -\varphi_{21}p\left(n_1-1\right) & 1-\varphi_{21}p_2\left(n_2-1\right) & \cdots & -\varphi_{21}pn_M\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ -\varphi_{M1}p & -\varphi_{M1}p\left(n_1-1\right) & -\varphi_{M1}pn_2 & \cdots & 1-\varphi_{M1}p_M\left(n_M-1\right)\end{bmatrix}.$$

If we let $\tilde{\mathbf{Q}}$ be the matrix obtained by removing the first row and column of $\mathbf{Q}$, then $v_0=\det\tilde{\mathbf{Q}}/\det\mathbf{Q}$. Carrying out the same calculations as in the proof of Lemma 3, we

$$\det\mathbf{Q}=\frac{\left(1+p_1\right)\left(1-p_1\left(n_1-1\right)\right)}{x_2\left(\varphi_{21}\right)\cdots x_M\left(\varphi_{M1}\right)}\left(1-\lambda\sum_{m=2}^{M}\varphi_{m1}n_m x_m\left(\varphi_{m1}\right)\right),$$

where $\lambda=p\left(1+\frac{pn_1}{1-p_1\left(n_1-1\right)}\right)$ and $x_m\left(\varphi_{mj}\right)=\left(1+\varphi_{mj}\left(p-\left(n_m-1\right)\left(p_m-p\right)\right)\right)$. Similarly,

$$\det\tilde{\mathbf{Q}}=\frac{1-p_1\left(n_1-2\right)}{x_2\left(\varphi_{21}\right)\cdots x_M\left(\varphi_{M1}\right)}\left(1-\tilde{\lambda}\sum_{m=2}^{M}\varphi_{m1}n_m x_m\left(\varphi_{m1}\right)\right),$$

where $\tilde{\lambda}=p\left(1+\frac{p\left(n_1-1\right)}{1-p_1\left(n_1-2\right)}\right)$. Putting these results together, we have

$$\begin{aligned}R_1\left(\Phi_1\right) &= a_1^2\sigma_1^2 v_0=a_1^2\sigma_1^2\frac{\det\tilde{\mathbf{Q}}}{\det\mathbf{Q}}\\ &= a_1^2\sigma_1^2\left(\frac{1+\left(p_1-p\right)x_1}{1+p_1}+\frac{px_1^2}{1-p\left(n_1x_1+\sum_{m=2}^{M}\varphi_{m1}n_m x_m\left(\varphi_{m1}\right)\right)}\right),\end{aligned}$$

which establishes the result. ■

PROPOSITION B3. *Suppose within-module communication is free. Then if $k(\varphi_{ij})$ is linear in φ_{ij}, then optimal communication patterns satisfy $\varphi^*_{mj} \in \{0,1\}$ with $\varphi^*_{mj} \geq \varphi^*_{m'j}$ if and only if $x_m \geq x_{m'}$. If k is convex, and optimal communication patterns have $\varphi^*_{mj} \in (0,1)$ for all m and j, then $\varphi^*_{mj} \geq \varphi^*_{m'j}$ if and only if $x_m \geq x_{m'}$.*

Proof of Proposition B3. By Proposition B2, agent j's sub-problem is

$$\max_{\{\varphi_{mj}\}_{m\neq m(j)}} a_j^2\sigma_j^2\left(\frac{1+\left(p_{m(j)}-p\right)x_{m(j)}}{1+p_{m(j)}}+\frac{px^2_{m(j)}}{1-p\left(\sum_{m=1}^M \varphi_{mj}n_m x_m\left(\varphi_{mj}\right)\right)}\right)-\sum_{m\neq m(j)} n_m k\left(\varphi_{mj}\right),$$

where $\varphi_{m(j)j}=1$. The marginal benefit of increasing $\varphi_{\ell j}$ for $m\neq m(j)$ is therefore

$$\begin{aligned}
& a_j^2\sigma_j^2\frac{px^2_{m(j)}p}{\left(1-p\left(\sum_{m=1}^M \varphi_{mj}n_m x_m\left(\varphi_{mj}\right)\right)\right)^2}\left(\varphi_{\ell j}x'_\ell\left(\varphi_{\ell j}\right)+x_\ell\left(\varphi_{\ell j}\right)\right) \\
= \quad & a_j^2\sigma_j^2\frac{p^2x^2_{m(j)}x_\ell\left(\varphi_{\ell j}\right)^2}{\left(1-p\left(\sum_{m=1}^M \varphi_{mj}n_m x_m\left(\varphi_{mj}\right)\right)\right)^2}.
\end{aligned}$$

This marginal benefit is increasing in $\varphi_{\ell j}$, and so if $k(\varphi_{\ell j})$ is linear in $\varphi_{\ell j}$, then any optimal solution will be a corner solution with

$$a_j^2\sigma_j^2\frac{p^2x^2_{m(j)}x^2_\ell}{\left(1-p\left(\sum_{m=1}^M \varphi_{mj}n_m x_m\left(\varphi^*_{mj}\right)\right)\right)^2}\geq k(1)\geq a_j^2\sigma_j^2\frac{p^2x^2_{m(j)}x^2_m}{\left(1-p\left(\sum_{m=1}^M \varphi_{mj}n_m x_m\left(\varphi^*_{mj}\right)\right)\right)^2}$$

for all ℓ with $\varphi^*_{\ell j}=1$ and m with $\varphi^*_{mj}=0$.

If $k(\varphi_{\ell j})$ is convex, and optimal communication patterns are interior, then they satisfy the first-order conditions

$$k'\left(\varphi^*_{\ell j}\right)=a_j^2\sigma_j^2\frac{p^2x^2_{m(j)}x_\ell\left(\varphi^*_{\ell j}\right)^2}{\left(1-p\left(\sum_{m=1}^M \varphi_{mj}n_m x_m\left(\varphi^*_{mj}\right)\right)\right)^2}$$

and so

$$\frac{k'\left(\varphi^*_{\ell j}\right)}{k'\left(\varphi^*_{mj}\right)}=\left(\frac{x_\ell\left(\varphi^*_{\ell j}\right)}{x_m\left(\varphi^*_{mj}\right)}\right)^2.$$

Since $x_\ell\geq x_m$ if and only if $x_\ell(\varphi)\geq x_m(\varphi)$ for all φ, we have that $\varphi^*_{\ell j}\geq\varphi^*_{mj}$ if and only if $x_\ell\geq x_m$. ■

Appendix C: Heterogeneous Coupling

Each node i belongs to a module $m(i) \in \mathcal{M}$, and each module m belongs to a cluster $k(m) \in \mathcal{K}$, where $\mathcal{K} = \{1, \ldots, K\}$. As in the main model, the need for coordination between two decisions depends on whether they are in the same module. In contrast to the main model, the need for coordination between two decisions in different modules depends on whether they are in the same cluster. That is, $p_{ij} = p_m$ if $m(i) = m(j)$, $p_{ij} = p^k$ if $m(i) \neq m(j)$ but $k(m(i)) = k(m(j))$, and $p_{ij} = p$ otherwise. Note that throughout this extension, we will use subscripts to denote module-level characteristics and superscripts to denote cluster-level characteristics.

The proof of this proposition proceeds by establishing two results, which parallel Lemma 3 and Proposition 2. First, Lemma C1 derives a closed-form expression for the expected revenues that result when node 1 informs an arbitrary set of modules in an arbitrary set of clusters. Then, Proposition 8 uses the convexity of the resulting function in each of its arguments to show that it implies a cluster-level threshold property for optimal communication networks.

LEMMA C1. *Suppose agent 1 tells his state to all agents in an arbitrary set of modules $\mathcal{M}^*$ that includes module $m(1)$. Agent 1's expected revenue is then given by*

$$R_1(\mathbf{C}_1) = a_1^2\sigma_1^2 \left(\frac{1+\left(p_1 - p^1\right)x_1}{1+p_1} + \frac{\left(p^1 - p\right)x_1^2}{1-\left(p^1-p\right)S^1} + \left(\frac{1}{1-\left(p^1-p\right)S^1}\right)^2 \frac{px_1^2}{1-p\sum_{k=1}^{K}\frac{S^k}{1-\left(p^k-p\right)S^k}} \right),$$

where

$$S^k = \sum_{m \in \mathcal{M}^*: k(m)=k} n_m x_m \text{ and } x_m = \frac{1}{1+p^{k(m)} - (n_m - 1)\left(p_m - p^{k(m)}\right)}.$$

If $m(1) \notin \mathcal{M}^*$, then replace n_1 with 1 *in these expressions.*

Proof of Lemma C1. Suppose agent 1 is in module $m = 1$ in cluster $k = 1$, and suppose $\mathcal{M}^*$ contains k_1 modules in cluster $k = 1$, $k_2 - k_1$ modules in cluster $k = 2$, and $k_K - k_{K-1}$ modules in cluster $k = K$. Number these modules so that modules $m = 1$ to $m = k_1$ are in cluster 1, modules $m = k_1 + 1$ to $m = k_2$ are in cluster 2, and modules $m = k_{K-1} + 1$ to $m = k_K$ are in cluster K.

By the same argument as in step 1 of the proof of Lemma 3, if we let v_0 represent the sum of the value of all walks from node 1 back to itself on the subgraph of the production network consisting of nodes in modules whose agents know state θ_1 and v_ℓ the sum of the values of all walks from a node in module ℓ to node 1 on this same subgraph, then

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{Q} \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{n_{k_K}} \end{bmatrix},$$

where

$$\boldsymbol{Q} = \begin{array}{c} \begin{array}{cccccccccccc} \text{module 1} & & \text{cluster 1} & & & & \text{cluster 2} & & & & \text{cluster } K & \end{array} \\ \left[\begin{array}{cccccccccccc}
1 & -p_1(n_1-1) & -p^1 n_2 & \cdots & -p^1 n_{k_1} & -pn_{k_1+1} & \cdots & -pn_{k_2} & \cdots & -pn_{k_{K-1}+1} & \cdots & -pn_{k_K} \\
-p_1 & 1-p_1(n_1-2) & -p^1 n_2 & \cdots & -p^1 n_{k_1} & -pn_{k_1+1} & \cdots & -pn_{k_2} & \cdots & -pn_{k_{K-1}+1} & \cdots & -pn_{k_K} \\
-p^1 & -p^1(n_1-1) & x_2^{-1}-p^1 n_2 & \cdots & -p^1 n_{k_1} & -pn_{k_1+1} & \cdots & -pn_{k_2} & \cdots & -pn_{k_{K-1}+1} & \cdots & -pn_{k_K} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
-p^1 & -p^1(n_1-1) & -p^1 n_2 & \cdots & x_{k_1}^{-1}-p^1 n_{k_1} & -pn_{k_1+1} & \cdots & -pn_{k_2} & \cdots & -pn_{k_{K-1}+1} & \cdots & -pn_{k_K} \\
-p & -p(n_1-1) & -pn_2 & \cdots & -pn_{k_1} & x_{k_1+1}^{-1}-p^2 n_{k_1+1} & \cdots & -p^2 n_{k_2} & \cdots & -pn_{k_{K-1}+1} & \cdots & -pn_{k_K} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
-p & -p(n_1-1) & -pn_2 & \cdots & -pn_{k_1} & -p^2 n_{k_1+1} & \cdots & x_{k_2}^{-1}-p^2 n_{k_2} & \cdots & -pn_{k_{K-1}+1} & \cdots & -pn_{k_K} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
-p & -p(n_1-1) & -pn_2 & \cdots & -pn_{k_1} & -pn_{k_1+1} & \cdots & -pn_{k_2} & \cdots & x_{k_{K-1}+1}^{-1}-p^K n_{k_{K-1}+1} & \cdots & -p^K n_{k_K} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
-p & -p(n_1-1) & -pn_2 & \cdots & -pn_{k_1} & -pn_{k_1+1} & \cdots & -pn_{k_2} & \cdots & -p^K n_{k_{K-1}+1} & \cdots & x_{k_K}^{-1}-p^K n_{k_K}
\end{array}\right] \end{array}$$

If we let $\widetilde{\mathbf{Q}}$ be the matrix obtained by removing the first row and column of $\mathbf{Q}$, then $v_0 = \det \widetilde{\mathbf{Q}} / \det \mathbf{Q}$. This proof calculates this value by carrying out several decompositions of $\mathbf{Q}$ and $\widetilde{\mathbf{Q}}$. First, we will show step-by-step how $\det \mathbf{Q}$ is calculated. The same steps are then used to calculate $\det \widetilde{\mathbf{Q}}$ and therefore v_0.

Step 1: *Factor out terms involving links within module* 1.

First, write $\mathbf{Q}$ in block-matrix form $\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$, where $\mathbf{A} = \begin{bmatrix} 1 & -p_1(n_1-1) \\ -p_1 & 1-p_1(n_1-2) \end{bmatrix}$ captures the terms describing links within module 1. The submatrices $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are defined accordingly. Then by the block-matrix determinant formula, $\det \mathbf{Q} = \det(\mathbf{A}) \det\left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)$.

Step 2: *Factor out terms involving the remaining modules in cluster 1, that is, modules* 2 *to* k_1.

We can write $\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ in block matrix form as $\begin{bmatrix} \mathbf{M} & \mathbf{N} \\ \mathbf{O} & \mathbf{P} \end{bmatrix}$, where $\mathbf{M}$ captures the terms describing links within cluster 1, and each of these submatrices is defined as follows.

$$\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B} = \left[\begin{array}{ccc|ccccccc}
x_2^{-1}-\lambda^A n_2 & \cdots & -\lambda^A n_{k_1} & -\lambda^B n_{k_1+1} & \cdots & -\lambda^B n_{k_2} & \cdots & -\lambda^B n_{k_{K-1}+1} & \cdots & -\lambda^B n_{k_K} \\
\vdots & \boldsymbol{M} & \vdots & \vdots & \ddots & \vdots & \ddots\ \boldsymbol{N} & \vdots & \ddots & \vdots \\
-\lambda^A n_2 & \cdots & x_{k_1}^{-1}-\lambda^A n_{k_1} & -\lambda^B n_{k_1+1} & \cdots & -\lambda^B n_{k_2} & \cdots & -\lambda^B n_{k_{K-1}+1} & \cdots & -\lambda^B n_{k_K} \\
\hline
-\lambda^B n_2 & \cdots & -\lambda^B n_{k_1} & x_{k_1+1}^{-1}-\lambda^2 n_{k_1+1} & \cdots & -\lambda^2 n_{k_2} & \cdots & -\lambda n_{k_{K-1}+1} & \cdots & -\lambda n_{k_K} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
-\lambda^B n_2 & \cdots & -\lambda^B n_{k_1} & -\lambda^2 n_{k_1+1} & \cdots & x_{k_2}^{-1}-\lambda^2 n_{k_2} & \cdots & -\lambda n_{k_{K-1}+1} & \cdots & -\lambda n_{k_K} \\
\vdots & \boldsymbol{O} & \vdots & \vdots & \ddots & \vdots & \ddots\ \boldsymbol{P} & \vdots & \ddots & \vdots \\
-\lambda^B n_2 & \cdots & -\lambda^B n_{k_1} & -\lambda n_{k_1+1} & \cdots & -\lambda n_{k_2} & \cdots & x_{k_{K+1}+1}^{-1}-\lambda^K n_{k_{K-1}+1} & \cdots & -\lambda^K n_{k_K} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
-\lambda^B n_2 & \cdots & -\lambda^B n_{k_1} & -\lambda n_{k_1+1} & \cdots & -\lambda n_{k_2} & \cdots & -\lambda^K n_{k_{K-1}+1} & \cdots & x_{k_K}^{-1}-\lambda^K n_{k_K}
\end{array}\right]$$

where $\lambda^A = p^1\left(1+\frac{p^1 n_1}{1-p_1(n_1-1)}\right)$, $\lambda^B = p\left(1+\frac{p^1 n_1}{1-p_1(n_1-1)}\right)$, $\lambda = p\left(1+\frac{p n_1}{1-p_1(n_1-1)}\right)$, and $\lambda^k = \left(p^k - p\right) + \lambda$.

Again, by the block-matrix determinant formula, $\det\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right) = \det\left(\mathbf{M}-\mathbf{N}\mathbf{P}^{-1}\mathbf{O}\right)\det\left(\mathbf{P}\right)$.

Step 3: *Write the terms involving clusters* 2 to K *as the sum of a diagonal matrix and a low-rank matrix.*

Next, note that we can write $\mathbf{P} = \mathbf{X}^{-1} + \mathbf{U}\mathbf{V}^T$, where

$$\mathbf{X} = \begin{bmatrix} x_{k_1+1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{k_K} \end{bmatrix}, \mathbf{U} = \begin{bmatrix} -\lambda^2 & \cdots & -\lambda \\ \vdots & \ddots & \vdots \\ -\lambda^2 & \cdots & -\lambda \\ \vdots & \ddots & \vdots \\ -\lambda & \cdots & -\lambda^K \\ \vdots & \ddots & \vdots \\ -\lambda & \cdots & -\lambda^K \end{bmatrix}, \mathbf{V} = \begin{bmatrix} n_{k_1+1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ n_{k_2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & n_{k_{K-1}+1} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & n_{k_K} \end{bmatrix}.$$

Using the Weinstein-Aronszajn identity, $\det\mathbf{P} = \det\left(\mathbf{X}^{-1}\right)\det\left(\mathbf{I}+\mathbf{V}^T\mathbf{X}\mathbf{U}\right)$. The second term in this identity can, in turn, be written as the sum of a diagonal matrix and a rank-one matrix: $\mathbf{I}+\mathbf{V}^T\mathbf{X}\mathbf{U} = \mathbf{E}+\mathbf{u}\mathbf{v}^T$, where

$$\mathbf{E} = \begin{bmatrix} 1-\left(\lambda^2-\lambda\right)S^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1-\left(\lambda^K-\lambda\right)S^K \end{bmatrix}, \mathbf{u} = \begin{bmatrix} S^2 \\ \vdots \\ S^K \end{bmatrix}, \mathbf{v} = \begin{bmatrix} -\lambda \\ \vdots \\ -\lambda \end{bmatrix},$$

where recall that $S^k = \sum_{m\in\mathcal{M}^*, k=k(m)} n_m x_m$. By the matrix determinant lemma, $\det\left(\mathbf{I}+\mathbf{V}^T\mathbf{X}\mathbf{U}\right) = \left(1+\mathbf{v}^T\mathbf{E}^{-1}\mathbf{u}\right)\det\mathbf{E}$.

Step 4: *Rewrite the terms linking cluster* 1 *to clusters* 2 *to* K *in terms of calculable matrices.*

Recall from step 2 that calculating $\det\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)$ requires calculating $\det\left(\mathbf{M}-\mathbf{N}\mathbf{P}^{-1}\mathbf{O}\right)$ and $\det\left(\mathbf{P}\right)$. Step 3 carried out the latter. This step shows how to calculate $\det\left(\mathbf{M}-\mathbf{N}\mathbf{P}^{-1}\mathbf{O}\right)$. Since $\mathbf{P} = \mathbf{X}^{-1}+\mathbf{U}\mathbf{V}^T$ is the sum of a diagonal matrix and a low-rank matrix, the Woodbury matrix identity allows us to write its inverse as follows

$$\mathbf{P}^{-1} = \mathbf{X} - \mathbf{X}\mathbf{U}\left(\mathbf{I}+\mathbf{V}^T\mathbf{X}\mathbf{U}\right)^{-1}\mathbf{V}^T\mathbf{X}.$$

Moreover, since $\mathbf{I}+\mathbf{V}^T\mathbf{X}\mathbf{U} = \mathbf{E}+\mathbf{u}\mathbf{v}^T$ is the sum of a diagonal matrix and a rank-one matrix, the Sherman-Morrison identity allows us to write its inverse as

$$\left(\mathbf{I}+\mathbf{V}^T\mathbf{X}\mathbf{U}\right)^{-1} = \mathbf{E}^{-1} - \frac{\mathbf{E}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{E}^{-1}}{1+\mathbf{v}^T\mathbf{E}^{-1}\mathbf{u}}.$$

Step 5: *Substitute in the expressions from steps 1 to 4 to give an expression for* $\det \mathbf{Q}$.

Putting together each of the preceding steps,we have the following:

$$\begin{aligned}
\det \mathbf{Q} &= \det(\mathbf{A}) \det\left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right) \\
&= \det(\mathbf{A}) \det\left(\mathbf{M} - \mathbf{N}\mathbf{P}^{-1}\mathbf{O}\right) \det \mathbf{P} \\
&= \det(\mathbf{A}) \det\left(\mathbf{M} - \mathbf{N}\mathbf{P}^{-1}\mathbf{O}\right) \det\left(\mathbf{X}^{-1}\right) \det\left(\mathbf{I} + \mathbf{V}^T\mathbf{X}\mathbf{U}\right) \\
&= \det(\mathbf{A}) \det\left(\mathbf{M} - \mathbf{N}\mathbf{P}^{-1}\mathbf{O}\right) \det\left(\mathbf{X}^{-1}\right) \left(1 + \mathbf{v}^T\mathbf{E}^{-1}\mathbf{u}\right) \det \mathbf{E} \\
&= \det(\mathbf{A}) \det\left(\mathbf{M} - \mathbf{N}\left(\mathbf{X} - \mathbf{X}\mathbf{U}\left(\mathbf{E}^{-1} - \frac{\mathbf{E}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{E}^{-1}}{1 + \mathbf{v}^T\mathbf{E}^{-1}\mathbf{u}}\right)\mathbf{V}^T\mathbf{X}\right)\mathbf{O}\right) \cdot \\
&\quad \det\left(\mathbf{X}^{-1}\right) \left(1 + \mathbf{v}^T\mathbf{E}^{-1}\mathbf{u}\right) \det \mathbf{E}.
\end{aligned}$$

Step 6: *Carry out the preceding steps for the matrix* $\widetilde{\mathbf{Q}}$ *to obtain an expression for* $\det \widetilde{\mathbf{Q}}$.

Recall that $\widetilde{\mathbf{Q}}$ is the matrix obtained by removing the first row and column of $\mathbf{Q}$. Carrying out steps 1 to 5 on this matrix, denote the corresponding matrices with tildes. For the analog of step 1, let $\widetilde{\mathbf{A}} = 1 - p_1(n_1 - 2)$ and define $\widetilde{\mathbf{B}}$, $\widetilde{\mathbf{C}}$, and $\widetilde{\mathbf{D}}$ correspondingly. For the analog of step 2, write $\widetilde{\mathbf{D}} - \widetilde{\mathbf{C}}\widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{B}}$ in block matrix form as $\begin{bmatrix} \widetilde{\mathbf{M}} & \widetilde{\mathbf{N}} \\ \widetilde{\mathbf{O}} & \widetilde{\mathbf{P}} \end{bmatrix}$, where each of the submatrices $\widetilde{\mathbf{M}}$, $\widetilde{\mathbf{N}}$, $\widetilde{\mathbf{O}}$, and $\widetilde{\mathbf{P}}$ are the same as $\mathbf{M}$, $\mathbf{N}$, $\mathbf{O}$, and $\mathbf{P}$, except that each of the λ^A, λ^B, λ, and λ^ℓ terms are replaced with $\tilde{\lambda}^A = p^1\left(1 + \frac{p^1(n_1-1)}{1-p_1(n_1-2)}\right)$, $\tilde{\lambda}^B = p\left(1 + \frac{p^1(n_1-1)}{1-p_1(n_1-2)}\right)$, $\tilde{\lambda} = p\left(1 + \frac{p(n_1-1)}{1-p_1(n_1-2)}\right)$, and $\tilde{\lambda}^k = \left(p^k - p\right) + \tilde{\lambda}$, respectively. For the analog of step 3, write $\widetilde{\mathbf{P}} = \mathbf{X}^{-1} + \widetilde{\mathbf{U}}\mathbf{V}^T$, where $\widetilde{\mathbf{U}}$ is the same as $\mathbf{U}$ but with λ and λ^ℓ replaced by $\tilde{\lambda}$ and $\tilde{\lambda}^\ell$. Additionally, $\mathbf{I} + \mathbf{V}^T\mathbf{X}\widetilde{\mathbf{U}} = \mathbf{E} + \mathbf{u}\tilde{\mathbf{v}}^T$, where $\tilde{\mathbf{v}}$ is the same as $\mathbf{v}$ but with λ replaced by $\tilde{\lambda}$.

Putting together each of these steps, we have

$$\begin{aligned}
\det \widetilde{\mathbf{Q}} &= \det\left(\widetilde{\mathbf{A}}\right) \det\left(\widetilde{\mathbf{M}} - \widetilde{\mathbf{N}}\left(\mathbf{X} - \mathbf{X}\widetilde{\mathbf{U}}\left(\mathbf{E}^{-1} - \frac{\mathbf{E}^{-1}\mathbf{u}\tilde{\mathbf{v}}^T\mathbf{E}^{-1}}{1 + \tilde{\mathbf{v}}^T\mathbf{E}^{-1}\mathbf{u}}\right)\mathbf{V}^T\mathbf{X}\right)\widetilde{\mathbf{O}}\right) \cdot \\
&\quad \det\left(\mathbf{X}^{-1}\right) \left(1 + \tilde{\mathbf{v}}^T\mathbf{E}^{-1}\mathbf{u}\right) \det \mathbf{E}
\end{aligned}$$

and therefore

$$\omega_{11}(\mathbf{C}_1) = v_0 = \frac{\det \widetilde{\mathbf{Q}}}{\det \mathbf{Q}} = \frac{\det\left(\widetilde{\mathbf{A}}\right)}{\det(\mathbf{A})} \frac{\det\left(\widetilde{\mathbf{M}} - \widetilde{\mathbf{N}}\left(\mathbf{X} - \mathbf{X}\widetilde{\mathbf{U}}\left(\mathbf{E}^{-1} - \frac{\mathbf{E}^{-1}\mathbf{u}\tilde{\mathbf{v}}^T\mathbf{E}^{-1}}{1+\tilde{\mathbf{v}}^T\mathbf{E}^{-1}\mathbf{u}}\right)\mathbf{V}^T\mathbf{X}\right)\widetilde{\mathbf{O}}\right)}{\det\left(\mathbf{M} - \mathbf{N}\left(\mathbf{X} - \mathbf{X}\mathbf{U}\left(\mathbf{E}^{-1} - \frac{\mathbf{E}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{E}^{-1}}{1+\mathbf{v}^T\mathbf{E}^{-1}\mathbf{u}}\right)\mathbf{V}^T\mathbf{X}\right)\mathbf{O}\right)} \frac{1 + \tilde{\mathbf{v}}^T\mathbf{E}^{-1}\mathbf{u}}{1 + \mathbf{v}^T\mathbf{E}^{-1}\mathbf{u}}.$$

Making the appropriate substitutions, we have

$$\omega_{11}(\mathbf{C}_1) = \frac{1 + \left(p_1 - p^1\right)x_1}{1 + p_1} + \frac{\left(p^1 - p\right)x_1^2}{1 - \left(p^1 - p\right)S^1} + \left(\frac{1}{1 - \left(p^1 - p\right)S^1}\right)^2 \frac{p x_1^2}{1 - p\sum_{k=1}^K \frac{S^k}{1-\left(p^k-p\right)S^k}},$$

which completes the proof of the part of the proposition with $m(1) \in \mathcal{M}^*$.

Now suppose $m(1) \notin \mathcal{M}^*$ (i.e., agent 1 does not tell his state to others in his own module). Then we can define $\mathbf{Q}$ and $\tilde{\mathbf{Q}}$ as above, plugging in $n_1 = 1$. And we can define the matrices $\mathbf{R}$ and $\tilde{\mathbf{R}}$, where $\mathbf{R}$ is the matrix obtained by removing the second row and column of $\mathbf{Q}$, and $\tilde{\mathbf{R}}$ is the matrix obtained by removing the first and second rows and columns of $\mathbf{Q}$. In this case,

$$v_0 = \frac{\det \tilde{\mathbf{R}}}{\det \mathbf{R}} = \frac{(1+p_1)\det \tilde{\mathbf{R}}}{(1+p_1)\det \mathbf{R}} = \frac{\det \tilde{\mathbf{Q}}}{\det \mathbf{Q}},$$

with $n_1 = 1$ substituted into the definition of $\mathbf{Q}$. The last equality holds by the Laplace expansion of the determinant. This completes the proof. ■

PROPOSITION 8. *There exist thresholds $\lambda_i^k \geq 0$ such that it is optimal for agent $i \in \mathcal{N}$ to tell his state to agent $j \in \mathcal{N}$ with $m(j) \neq m(i)$* and $m(j)$ in cluster $k \in \mathcal{K}$ *if and only if $x_{m(j)} \geq \lambda_i^k$.*

Proof of Proposition 8. First, suppose agent i tells his state to all the agents in his own module. Let $\tilde{S}^1 = S^1 - n_1 x_1$ and define the function $h\left(\tilde{S}^1, S^2, S^3, \ldots, S^K\right)$ to be

$$\frac{\left(p^1 - p\right) x_1^2}{1 - \left(p^1 - p\right)\left(n_1 x_1 + \tilde{S}^1\right)} + \frac{1}{\left(1 - \left(p^1 - p\right)\left(n_1 x_1 + \tilde{S}^1\right)\right)^2} \frac{p x_1^2}{1 - p\left(\frac{\left(n_1 x_1 + \tilde{S}^1\right)}{1 - \left(p^1 - p\right)\left(n_1 x_1 + \tilde{S}^1\right)} + \sum_{k=2}^K \frac{S^k}{1 - \left(p^k - p\right) S^k}\right)}.$$

We will establish that this function is convex in each of its arguments and use this fact to argue that optimal communication networks have the multi-threshold property described in the lemma.

First, we show that h is convex in $\tilde{S}^1$. Let $W = \sum_{k=2}^K \frac{S^k}{1-\left(p^k - p\right)S^k}$. Then h is convex in $\tilde{S}^1$ if and only if

$$\tilde{h}\left(\tilde{S}^1\right) = \frac{\left(p^1 - p\right) x_1^2}{1 - \left(p^1 - p\right)\left(n_1 x_1 + \tilde{S}^1\right)} + \frac{1}{\left(1 - \left(p^1 - p\right)\left(n_1 x_1 + \tilde{S}^1\right)\right)^2} \frac{p x_1^2}{1 - p\left(\frac{\left(n_1 x_1 + \tilde{S}^1\right)}{1 - \left(p^1 - p\right)\left(n_1 x_1 + \tilde{S}^1\right)} + W\right)}$$

is convex in $\tilde{S}^1$. This function is convex because it is twice differentiable and

$$\tilde{h}''\left(\tilde{S}^1\right) = 2x_1^2 \left(\frac{p^1}{1 - \left(p^1 - p\right) S^1} + \frac{p}{1 - \left(p^1 - p\right) S^1} \frac{p^1 \frac{S^1}{1 - \left(p^1 - p\right) S^1} + pW}{1 - p^1 \frac{S^1}{1 - \left(p^1 - p\right) S^1} - pW}\right)^3 > 0,$$

since $1 - \left(p^1 - p\right) S^1 > 0$ and $1 - p^1 \frac{S^1}{1-\left(p^1-p\right)S^1} - pW > 0$.

Next, we show that h is convex in S^ℓ for $\ell > 1$. Let $W = \frac{\left(n_1 x_1 + \tilde{S}^1\right)}{1 - \left(p^1 - p\right)\left(n_1 x_1 + \tilde{S}^1\right)} + \sum_{k \neq 1, k \neq \ell} \frac{S^k}{1 - \left(p^k - p\right) S^k}$. Then h is convex in S^ℓ if and only

$$\tilde{h}\left(S^\ell\right) = \frac{1}{1 - p\left(W + \frac{S^\ell}{1 - \left(p^\ell - p\right) S^\ell}\right)}$$

is convex in S^{ℓ}. This function is convex because it is twice differentiable and

$$\tilde{h}''\left(S^{\ell}\right) = 2\frac{\tilde{h}\left(S^{\ell}\right)^3}{\left(1-\left(p^{\ell}-p\right)S^{\ell}\right)^3}p\left(p^{\ell}\left(1-pW\right)+\left(p\right)^2 W\right) > 0,$$

since $1-pW>0$ and $1-\left(p^{\ell}-p\right)S^{\ell}>0$.

Now suppose agent 1 does not tell his state to the other agents in his own module. We can define $\tilde{S}^1 = S^1 - \frac{1}{1+p^1}$ and $\hat{h}\left(\tilde{S}^1, S^2, S^3, \ldots, S^K\right)$ to be the same as above, except with $n_1 = 1$. Then the function $\hat{h}$ has the same properties as h derived above.

To establish the multi-threshold property, let $\mathcal{K}$ be an arbitrary set of modules with $m\left(i\right) \in \mathcal{K}$. Define $S^{\ell}\left(\mathcal{K}\right) = \sum_{m:m\in\mathcal{K},k(m)=\ell} n_m x_m$, and denote by $\mathbf{C}_i\left(\mathcal{K}\right)$ the row of the communication matrix in which agent i tells θ_i to agent j if and only if $m\left(j\right)\in\mathcal{K}$. By Lemma C1, $R_i\left(\mathbf{C}_i\left(\mathcal{K}\right)\right)$ is a linear function of $h\left(\tilde{S}^1\left(\mathcal{K}\right), S^2\left(\mathcal{K}\right), \ldots, S^K\left(\mathcal{K}\right)\right)$.

Now suppose that it is optimal to inform all modules in $\mathcal{K}$. Then it must be the case that for all $m\in\mathcal{K}$ and $k\left(m\right)=\ell$,

$$\begin{aligned}\gamma &\leq \frac{h\left(\tilde{S}^1\left(\mathcal{K}\right), S^2\left(\mathcal{K}\right), \ldots, S^K\left(\mathcal{K}\right)\right) - h\left(\tilde{S}^1\left(\mathcal{K}\right), S^2\left(\mathcal{K}\right), \ldots, S^{\ell}\left(\mathcal{K}\backslash\left\{m\right\}\right), \ldots, S^K\left(\mathcal{K}\right)\right)}{n_m} \\ &< x_m h_{\ell}\left(\tilde{S}^1\left(\mathcal{K}\right), S^2\left(\mathcal{K}\right), \ldots, S^K\left(\mathcal{K}\right)\right),\end{aligned}$$

where the second inequality holds because h is convex in $S^{\ell}\left(\mathcal{K}\right)$ and where h_{ℓ} denotes the partial derivative of h with respect to S^{ℓ}. Suppose further that it is not optimal to also inform some module $m'\notin\mathcal{K}$ with $k\left(m'\right)=\ell$. Then it must be the case that

$$\begin{aligned}\gamma &> \frac{h\left(\tilde{S}^1\left(\mathcal{K}\right), S^2\left(\mathcal{K}\right), \ldots, S^{\ell}\left(\mathcal{K}\cup\left\{m'\right\}\right), \ldots, S^K\left(\mathcal{K}\right)\right) - h\left(\tilde{S}^1\left(\mathcal{K}\right), S^2\left(\mathcal{K}\right), \ldots, S^K\left(\mathcal{K}\right)\right)}{n_{m'}} \\ &> x_{m'} h_{\ell}\left(\tilde{S}^1\left(\mathcal{K}\right), S^2\left(\mathcal{K}\right), \ldots, S^K\left(\mathcal{K}\right)\right).\end{aligned}$$

These two inequalities imply that $x_m > x_{m'}$ for all modules m in cluster ℓ that are are optimally told about θ_i and modules m' in cluster ℓ that are optimally not told about θ_i. In other words, there is a threshold λ_i^{ℓ} such that agent i tells θ_i to agent j in module $m\left(j\right)\neq m\left(i\right)$ and cluster $k\left(m\left(j\right)\right)=\ell$ if and only if $x_{m(j)}\geq\lambda_i^{\ell}$. ■

We conclude this appendix by establishing that the principal's problem is supermodular for arbitrary production networks satisfying $p_{ii}=0$, $p_{ij}=p_{ji}$, *and* $\sum_{j=1}^{N}p_{ij}<1$.

PROPOSITION 9. *As long as the production network* $\boldsymbol{P}$ *satisfies* $p_{ii}=0$, $p_{ij}=p_{ji}$, *and* $\sum_{j=1}^{N}p_{ij}<1$, *optimal communication networks* $\boldsymbol{C}^*$ *are increasing in the value of autonomous adaptation* $a_i^2\sigma_i^2$ *and the needs for coordination* p_{ij} *for all* $i,j\in\mathcal{N}$, *and decreasing in communication costs* γ.

Proof of Proposition 9. For general production networks $\boldsymbol{P}$ satisfying $p_{ii} = 0$, $p_{ij} = p_{ji}$, and $\sum_{j=1}^{N} p_{ij} < 1$, Lemmas 1 and 2 and Proposition 1 continue to hold. As long as $\omega_{ii}(\boldsymbol{C}_i)$ is supermodular in $\boldsymbol{C}_i$, then the principal's objective for the subproblem involving who should agent i inform about θ_i is supermodular in $\boldsymbol{C}_i$ and exhibits increasing differences in $\left(a_i^2\sigma_i^2, \{p_{ij}\}_{ij}, \{c_{ij}\}_{ij}, -\gamma\right)$, so the comparative statics results follow from Topkis's theorem. It remains, therefore, to show that $\omega_{ii}(\boldsymbol{C}_i)$ is supermodular in $\boldsymbol{C}_i$.

To show that $\omega_{ii}(\cdot)$ is supermodular, let $\mathcal{J} \subset \mathcal{N}$ denote a subset of agents, and denote by $\mathbf{c}(\mathcal{J})$ the $1 \times N$ vector with jth element equal to one if $j \in \mathcal{J}$ and equal to zero otherwise. We will show that the incremental value of informing agent 1 about θ_i is higher when agent 2 knows θ_i than when she does not. Take $\mathcal{J}$ to be a set of nodes that are informed throughout the exercise.

Denote by $\boldsymbol{P}(\mathcal{J}) = (\operatorname{diag}\mathbf{c}(\mathcal{J}))\,\boldsymbol{P}\,(\operatorname{diag}\mathbf{c}(\mathcal{J}))$ the subset of the production network consisting of the nodes j for which the jth element of $\mathbf{c}(\mathcal{J})$ is equal to one. Then

$$\Delta^k \equiv \boldsymbol{P}(\mathcal{J}\cup\{1,2\})^k - \boldsymbol{P}(\mathcal{J}\cup\{2\})^k - \left(\boldsymbol{P}(\mathcal{J}\cup\{1\})^k - \boldsymbol{P}(\mathcal{J})^k\right)$$

is the matrix whose ijth element is the value of the additional walks of length k from informing agent 1 when agents $\mathcal{J}\cup\{2\}$ are informed relative to when only agents $\mathcal{J}$ are informed. Since informing agent 1 adds more walks of all lengths to $\boldsymbol{P}(\mathcal{J}\cup\{2\})$ than it does to $\boldsymbol{P}(\mathcal{J})$, it follows that every element of Δ^k is nonnegative. Since this argument holds for all k, we have that the iith element of

$$\sum_{k=1}^{\infty}\Delta^k = \sum_{k=1}^{\infty}\boldsymbol{P}(\mathcal{J}\cup\{1,2\})^k - \sum_{k=1}^{\infty}\boldsymbol{P}(\mathcal{J}\cup\{2\})^k - \left(\sum_{k=1}^{\infty}\boldsymbol{P}(\mathcal{J}\cup\{1\})^k - \sum_{k=1}^{\infty}\boldsymbol{P}(\mathcal{J})^k\right)$$

is nonnegative. Recall that $\omega_{ii}(\mathbf{c}(\mathcal{J}))$ is the iith element of $(\boldsymbol{I} - \boldsymbol{P}(\mathcal{J}))^{-1} = \boldsymbol{I} + \sum_{k=1}^{\infty}\boldsymbol{P}(\mathcal{J})^k$. We therefore have that

$$\omega_{ii}(\mathbf{c}(\mathcal{J}\cup\{1,2\})) - \omega_{ii}(\mathbf{c}(\mathcal{J}\cup\{2\})) \geq \omega_{ii}(\mathbf{c}(\mathcal{J}\cup\{1\})) - \omega_{ii}(\mathbf{c}(\mathcal{J})),$$

so $\omega_{ii}(\cdot)$ has increasing differences in c_{i1} and c_{i2}. The choice of agents 1 and 2 was immaterial in this argument, and so $\omega_{ii}(\cdot)$ has increasing differences in c_{ij} and c_{ik} for all $j, k \neq i$ and is therefore supermodular. ■

Appendix D: Incentive Conflicts

Suppose the principal cares about revenues $r(d_1, \ldots, d_N)$, and each agent i cares instead about

$$u_i(d_1, \ldots, d_N) = r(d_1, \ldots, d_N) + 2a_i d_i b_i,$$

where b_i reflects agent i's *decision-making bias.*

PROPOSITION D1. *Given a vector $b = [b_1, \ldots, b_N]$ of decision-making biases, an optimal communication network solves*

$$\max_{\mathbf{C}} \sum_{i=1}^{N} a_i \operatorname{Cov}(d_i^*, \theta_i) - \begin{bmatrix} a_1 b_1 & \cdots & a_N b_N \end{bmatrix} (\mathbf{I} - \mathbf{P})^{-1} \begin{bmatrix} a_1 b_1 \\ \vdots \\ a_N b_N \end{bmatrix},$$

where

$$\operatorname{Cov}(d_i^*, \theta_i) = a_i \sigma_i^2 \omega_{ii}(\mathbf{C}_i).$$

Proof of Proposition D1. Agent i's best response function is

$$d_i^* = a_i(\theta_i + b_i) + \sum_{j=1}^{N} p_{ij} E_i\left[d_j^*\right].$$

Given communication network $\mathbf{C}$, the argument in the proof of Lemma 1 establishes that there are unique Bayes-Nash equilibrium decisions. These decisions satisfy, for all j,

$$d_i^* = \sum_{j=1}^{N} a_j \omega_{ij}(\mathbf{C}_j)\theta_j + \sum_{j=1}^{N} a_j w_{ij} b_j,$$

where $\omega_{ij}(\mathbf{C}_j)$ denotes the ijth entry of $(\mathbf{I} - (\operatorname{diag}\mathbf{C}_j)\mathbf{P}(\operatorname{diag}\mathbf{C}_j))^{-1}$ and w_{ij} denotes the ijth entry of $(\mathbf{I} - \mathbf{P})^{-1}$.

Given equilibrium decision-making, revenue in state θ can be written as

$$\sum_{i=1}^{N} a_i d_i^* \theta_i - \sum_{i=1}^{N} d_i^* \left[d_i^* - a_i\theta_i - \sum_{j=1}^{N} p_{ij} d_j^*\right].$$

Substituting in the best responses $d_i^* = a_i(\theta_i + b_i) + \sum_{j=1}^{N} p_{ij} E_i\left[d_j^*\right]$, the term in square brackets is therefore equal to $\sum_{j=1}^{N} p_{ij}\left[a_i b_i + E_i\left[d_j^*\right] - d_j^*\right]$, and expected revenue can be written as

$$\begin{aligned} & \sum_{i=1}^{N} a_i E\left[d_i^*\theta_i\right] - \sum_{i=1}^{N} a_i b_i d_i^* - \sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij} E\left[d_i^*\left[E_i\left[d_j^*\right] - d_j^*\right]\right] \\ = & \sum_{i=1}^{N} a_i \operatorname{Cov}(d_i^*, \theta_i) - \sum_{i=1}^{N} a_i b_i E\left[d_i^*\right], \end{aligned}$$

where the last term is zero by the law of iterated expectations, and $E\left[d_i^*\theta_i\right] = \operatorname{Cov}(d_i^*, \theta_i)$ because $E[\theta_i] = 0$.

Finally, note that

$$\begin{aligned} E\left[d_i^*\right] &= E\left[a_i\left(\theta_i + b_i\right) + \sum_{j=1}^{N} p_{ij} E_i\left[d_j^*\right]\right] \\ &= a_i b_i + \sum_{j=1}^{N} p_{ij} E\left[d_j^*\right], \end{aligned}$$

so that we can write

$$\begin{bmatrix} E\left[d_1^*\right] \\ \vdots \\ E\left[d_N^*\right] \end{bmatrix} = (\mathbf{I} - \mathbf{P})^{-1} \begin{bmatrix} a_1 b_1 \\ \vdots \\ a_N b_N \end{bmatrix},$$

and

$$\sum_{i=1}^{N} a_i b_i E\left[d_i^*\right] = \begin{bmatrix} a_1 b_1 & \cdots & a_N b_N \end{bmatrix} (\mathbf{I} - \mathbf{P})^{-1} \begin{bmatrix} a_1 b_1 \\ \vdots \\ a_N b_N \end{bmatrix},$$

which establishes the result. ∎