# Rational Inattention and Organizational Focus

preliminary draft – do not post online.

Wouter Dessein
Columbia University

Andrea Galeotti
University of Essex

Tano Santos
Columbia University

February 29, 2012

**Abstract**

Organizations exist to coordinate specialized agents in charge of complementary tasks. In order to do so, they must build up collective knowledge and allocate scarce organizational attention. We study when organizational knowledge is optimally specialized - and mainly pertains to one or a few tasks - and when it is balanced across all tasks. Focused organizations exhibit strong leadership and prioritize a select number of tasks or 'core competences' at the expense of others. Balanced organizations, in contrast, have shared leadership and do not prioritize tasks. We show how organizational focus, leadership and core competences naturally arise as the result of a fundamental complementarity between (i) the attention devoted by the organization to a particular task and (ii) the initiative taken by the agent in charge of that task. Shared leadership and balanced organizations, on the other hand, become more attractive as communication technology improves. These results shed light on new trends in organization design.

# 1 Introduction

At least since Adam Smith's "Wealth of Nations," ([1776] 1981), the importance of specialization has been one of the central ideas in economics. As Smith argued, the division of labor allows workers to develop specialized skills and knowledge, and therefore expands the production frontier. The role of organizations, then, is to coordinate specialized workers (Bolton and Dewatripont 1994, Garicano 2000, Dessein and Santos 2006). The economics literature is largely silent, however, on whether organizations themselves benefit from developping specialized knowledge, or how to define such organizational knowledge. The management literature, in contrast, has long argued that in order to be successful, firms should focus and nurture a limited set of "core competences" (Prahald and Hamel 1990). Such core competences represent the "collective learning in the organization" and the "capacity to coordinate diverse production skills". Similarly, Porter (1985, 1996) has argued that firms that aim to be "all things to all people," will be "caught in the middle" and fail.[1]

In this paper, we provide formal foundations for returns to specialization for organizations, as opposed to individuals. To do so, we propose a model of team production in which a number of complementary tasks, such as product development, engineering, manufacturing, marketing and selling must be implemented in a coordinated fashion. Coordination requires communication, but agents have limited attention or information-processing capabilities. *Organizational focus* takes the form of allocating more scarce organizational attention to one task – or one agent – than another task. In our model, an agent which receives a lot of attention from other members in the organization can respond more effectively to task-specific information, as other agents are then able to take the appropriate coordinating actions. In contrast, an agent who is ignored by others is forced to also largely ignore his own task-specific information. Responding to his own information, would then result in subtantial coordination failures with other tasks. It follows that a focussed organization excells in certain tasks, but underperform in others. It has specialized, task-specific organizational knowledge that allows

[1]The importance of organizational focus is a central idea in the Management literature. Both Prahalad and Hamel (1990) and Porter (1996), for example, where selected among Harvard Business Review's 10 Must Read articles (HBR's 10 Must Reads: The Essentials).

it to coordinate very well those tasks, but this comes at the expense of all others.

Why can organizational focus be optimal? Why should an organization prioritize certain tasks over others? The underlying mechanism we highlight is a fundamental complementarity in organizations between the communication devoted to a task and the adaptiveness of that task. As argued above, an agent who is not central in the communication network will not adapt his task much to task-specific information for fear of coordination failures. Conversely, it is a waste of resources to allocate scarce attention to an agent who ignores his local information and implements his task in a predictable way. Following the same logic, the more central an agent is in the communication network, the more this agent will take initiative and adapt his task. In turn, the more an agent takes initiative, the more valuable and important it becomes to devote scarce attention to this agent in order to ensure effective coordination. It follows that if there are constant marginal returns to communication, it is optimal center all communication around one or a few tasks. An optimal communication network then gives outsized influence to one or a few agents, which we call *leaders.* This result holds even though all tasks are equally important for production, and all agents are equally capable.

A necessary ingredient for organizational focus to be optimal is that the information-processing capacity of agents is constrained. In the absence of informational constraints, it is optimal to treat all tasks symmetrically in our model. Indeed, in the absence of any information frictions, task are perfectly coordinated with each other and perfectly adapted to their task-specific shocks.[2]

The specific way in which we model limits to communication and information-processing borrows from a recent literature on rational inattention (Sims 2003), which in turn is based on information theory (Cover and Thomas, 2006). By virtue of carrying out a task, we assume that each agent privately observes a local shock pertaining to his own task. In order to learn about the local shocks affecting other tasks, however, agents need to communicate with each other. Following information theory, the uncertainty regarding other tasks is expressed in terms of the *entropy* of the distribution of the local shocks. For normally distributed random

[2]Similarly, if communication networks were to be exogenously and symmetrically designed, all tasks would be equally responsive to their task-specific shocks, and equally well (or poorly) coordinated.

variables, our leading case, the entropy of a variable is proportional to the log of its variance. The *mutual information* agents have regarding a particular local shock is equivalent to the reduction in this entropy following communication.[3]

Rather than explicitly modeling the details of how information is processed, we follow information theory in positing that the mutual information regarding a task-specific shock is proportional to the communication devoted to that task. Consistent with information theory, organizations have a fixed communication capacity implying that also the *total amount of mutual information* than can be achieved is subject to this attention constraint. This attention constraint can be interpreted, for example, as the total time agents spend in meetings as opposed to production.[4] As in Hamel and Prahalad's definition of 'core competences', the mutual information in our model thus represents the collective learning of an organization which is useful for coordinating production processes. It can be interpreted as a form of *organizational knowledge,* distinct from the individual knowledge of the agents in the organization. While the total amount of organizational knowledge is fixed, the organization designer can decide to *focus attention* on one or a few tasks, resulting in more mutual information for those tasks at the expense of others. Put differently, organizational knowledge can be specialized or not.

An important and intuitive feature of the above communication technology is that it implies *decreasing marginal returns to communicating* about a particular task-specific shock. While it is easy to reduce uncertainty when the variance of a posterior is large, it is increasingly difficult to further reduce the residual variance when this posterior becomes increasingly precise.[5] In the absence of any trade-offs between adaptation and coordination, this provides a powerful force *against* focus. To see this, let all tasks be carried out by the same agent, so

[3]The concepts of entropy, communication capacity and mutual information have strong theoretical foundations in coding technology (Clover and Thomas 1990), and have proven extremely useful in a wide variety of settings in both applied and exact sciences.

[4]In most of the paper, we treat this attention constraint as an exogenous variable. In Section 5.2, however, we endogenize this constraint by allowing the organization to expand attention at some cost.

[5]For most common distributions (e.g. normal, uniform) the residual variance decreases at a logarithmic rate: If one unit of attention reduces the residual variance to 1/2 of the initial variance, two units are required to reduce it to 1/4, three units to reduce it to 1/8, and so on. In order to completely eliminate any uncertainty, unlimited attention is required.

that task coordination is perfect and the only remaining goal is to adapt tasks to local shocks. Moreover assume this agent does not observe any shocks himself, but must communicate with other agents who do. Decreasing marginal returns to communication and a quadratic pay-off structure then imply that a single agent in charge of all tasks spreads his attention evenly over all tasks. Thus a second necessary ingredient for organizational focus to be optimal is that tasks are carried out by a team of agents, who must *coordinate* task implementation.

The need for coordination and the scarcity of attention together then drive our results. On the one hand, if tasks need to be coordinated, there is a complementarity between how

communication is very effective, a true "team of leaders" may be optimal.

Over the last decades, there has been enormous technological innovations in communication and coordination technology (e-mail, wireless communication and computing, intra networks). Our model predicts that as communication technology improves, more decentralized communication networks become optimal in which there are more, but less influential, leaders. The resulting organization is less well coordinated, less cohesive, but has a broader focus – it pays attention to the task-specific information of a larger number of tasks. Our model therefore sheds light on new trends in organizational design away from hierarchies towards more network-like organizations where communication flows are lateral rather than vertical, and decision-making and influence is broadly shared in the organization.[6]

Our results further provide foundations as well as caveats for a dominant view in the strategic management literature that successful firms are those that "pick their strengths," that is those that are designed to excell in a relatively small number of dimensions. Michael Porter's management classic "Competitive Advantage" (Michael Porter 1985), for example, argues that firms should either minimize cost or produce differentiated, high-value products. Firms that pursue both goals simultaneously will be "caught in the middle". Put differently, the profits of an organization are maximized at some corner solution of the organizational design problem, where performance in one or a few tasks is maximized at the expense of others. Choosing a firm's strategy is then equivalent with choosing what dimensions or tasks to focus on. Our paper can be interpreted as giving credence to the importance of choosing a strategy.[7] In

[6]The above results stand in contrast with those obtained in recent team-theory models that model organizations as information-processing (Bolton and Dewatripont (1994)) or problem-solving institutions (Garicano (2000). While they also characterize optimal information flows in organizations, decentralization is seen as a way to save on communication costs. Hence improvements in communication technology result in more centralization, not less. The above approaches also do not allow for an analysis of network-like organizations where communication flows are lateral rather than vertical. They are therefore are less suited to shed light on recent trends in organization design away from hierarchies.

[7]Business strategy, in our paper, is fundamentally about focussing the organization's attention on a select number of functions or tasks. There is a lot of casual evidence that firms often have organizational cultures that favor certain functional areas over others. Google, for example, is well known to have an engineering-dominated culture. IBM has often been described as having a sales-centric or sales-oriented business culture.

fact, in our symmetric model, it is not important what the firm's strategy is, as long as a strategy is chosen. By the same token, we provide insights as to how focussed firms should be. Taken literally, our results suggest that choosing a narrow strategy may become less important as information technology relaxes the information and attention constraints of organizations. Aiming to be all things to all people may be more attractive now than it used to be.

Our paper is organized as follows. Section 2 describes our model. Most of the insights and intuition of our paper can be derived and illustrated in a simple model with two agents and two tasks, which is analyzed in Section 3. In Section 4, we generalize the model to $n$ agents and $n$ tasks. Section 5 develops some extensions and we conclude in Section 6.

## 2 The Model

We posit a team-theoretic model, based on Dessein and Santos (2006), in which production requires the combination of $n$ tasks, each carried out by a different agent. The implementation of a task is informed by the realization of a task-specific shock, only observed by the agent in charge of that task. Communication flows within the team allow for this private information to be partially shared with other members of the organization. Organizational trade-offs arise because agents need to adapt to the privately observed shock while maintaining coordination across different tasks. The model is symmetric in that, ex-ante, there are no differences across agents and across tasks. The paper studies the optimal communication network and, hence, the allocation of scarce organizational attention.[8]

Most of the intuitions regarding the trade-offs in our setup can be grasped in the two-agent case, n=2, which we cover in depth in this section. We leave for Section 4 the analysis of the case $n > 2$.

---

And according to industry observers, Proctor and Gamble used to be a "manufacturing company producing consumer products" in the 1970's and 1980's, but was subsequently turned into a consumer-driven organization under former CEO Lafley.

[8]In Dessein and Santos (2006), the communication network among agents is assumed to be exogenous. Instead, their focus is on the specialization of agents, that is the number of tasks per agents.

## 2.1 Production

Production involves the implementation of two tasks, each performed by one agent $i \in \{1, 2\}$. The profits of the organization depend on (i) how well each task is *adapted* to its organizational environment and (ii) how well each task is *coordinated* with the other task. For this purpose, agent $i$ must choose a primary action, $q_{ii}$, and a complementary action, $q_{ij}$, with $i \neq j$.

In particular, Agent $i$ observes a piece of information $\theta_i$, a shock with variance $\sigma_\theta^2$ and mean 0, which is relevant for the proper implementation of the assigned task. We refer to $\theta_i$ as the local information of agent $i$. The realization of this local information is independent across agents. In order to achieve perfect adaptation, agent $i$ should set his primary action $q_{ii}$ equal to $\theta_i$. In order to achieve perfect coordination with task $j$, agent $i$ should set his complementary action $q_{ij}$ equal to $q_{jj}$, the primary action of agent $j$. If tasks are imperfectly adapted or coordinated, the organization suffers adaptation and/or coordination losses. Formally, let $q_i = [q_{i1}, q_{i2}]$ with $i \in \{1, 2\}$. Given a particular realization of the string of local information, $\boldsymbol{\theta} = [\theta_1, \theta_2]$, and a choice of actions, $\mathsf{q} = [q_1, q_2]$, the realized profit of the organization is:

$$\pi\left(\mathsf{q}|\boldsymbol{\theta}\right) = -\left(q_{11} - \theta_1\right)^2 - \left(q_{22} - \theta_2\right)^2 - \beta\left[\left(q_{21} - q_{11}\right)^2 + \left(q_{12} - q_{22}\right)^2\right]. \tag{1}$$

In expression (1), the parameter $\beta > 0$ measures the importance of coordination relative to adaptation. The larger $\beta$, the more important it is to maintain coordination between tasks. The smaller $\beta$, the more important it is to adapt tasks to local information, relatively speaking. For simplicity, we normalize the importance of adaptation to 1.

## 2.2 The communication network

A communication network $\mathsf{t} = [t_1, t_2]$ represents the time or *attention* that the organization devotes to communication about task 1 and task 2. Communication about task $j$ yields a message $m_j$ to agent $i \neq j$ regarding the local information of agent $j$. Naturally, the precision of the message $m_j$ depends on the time or attention $t_j$ agents devote to communicate about local information $\theta_j$. We assume that the organization cannot devote an infinite amount of resources to communicate:

$$t_1 + t_2 \leq \tau, \tag{2}$$

where $\tau < \infty$. Limited resources to communication or attention capacity plays a critical role in our analysis in that focus, naturally, can only arise in the presence of such attention capacity constraints. We assume first that $\tau$ is an exogenous parameter, but we will later consider the case where $\tau$ is also an organizational design variable. For example, $\tau$ can be the length of a meeting, and $t_1$ and $t_2$ the time that agent 1 and 2 are allowed to speak. We say that an organization is focused on task 1 whenever it devotes more attention to that task, $t_1 > t_2$ and conversely for task 2. We refer to the agent in charge of the task that is the focus of the organization as the organization's *leader.* We say that an organization is balanced if it is not focused, that is $t_1 = t_2 = \tau/2$.

### 2.3 The communication technology

We now describe in more details the communication technology. A particular communication network $\mathsf{t} = [t_1, t_2]$ yields information sets for agents 1 and 2, $\mathcal{I}_1$ and $\mathcal{I}_2$. Information set $\mathcal{I}_i$ contains the agent $i$'s local shock, $\theta_i$, as well as the message received from the other agent $j$, $m_j$. The degree of precision of message $m_j$ depends on the time or attention $t_j$ agents devote to communicate about local information $\theta_j$. In particular, we assume that the agent $i$ receives a noisy message $m_j$, which is a random variable with mean zero, variance $\sigma_m^2$ and correlation

$$\rho(t_j) = \frac{cov(\theta_j, m_j)}{\sigma_\theta \sigma_m}.$$

**Assumption A.** The random variables $(\theta_j, m_j)$ are such that the conditional expectations are linear in the conditioning information, i.e., $E[\theta_j|m_j]$ is linear in $m_j$, and $E[m_j|\theta_j]$ is linear in $\theta_j$, for every $j \in \{1, 2\}$.

Assumption A is satisfied, for example, if messages and information are normally distributed or uniformly distributed (see example 1 and 2 below).[9] Assumption A implies that

$$E[\theta_j|m_j] = \frac{cov(\theta_j, m_j)^2}{\sigma_m^2} m_j,$$

where we are using that both $\theta_j$ and $m_j$ have zero mean. Using the law of total variance, we can then write the expected conditional variance of local shock $\theta_j$, referred to as residual

[9]As we shall show in section 3, Assumption A assures that, for every communication network, there is an equilibrium where actions are linear in the information possed by agents

variance, as follows:

$$\mathsf{RV}(t_j) = E[\mathrm{Var}\,(\theta_j|m_j)] = \sigma_\theta^2\left[1-\rho^2(t_j)\right]. \tag{3}$$

Let $\hat{\tau}$ be such that $\mathsf{RV}(\hat{\tau}) = 0$; if $\mathsf{RV}(t) > 0$ for every finite $t$, set $\hat{\tau} = \infty$.

**Assumption B.** We make the following assumptions; for every $j = 1, 2$:

1B. The role of communication among agents is to reduce the conditional variance of the local shock, i.e., $\mathsf{RV}(t_j)$ is a decreasing function of $t_j$.

2B. Agent $i$ cannot "pick up" any information on $\theta_j$ if the organization devotes no attention to task $j$, i.e., $\mathsf{RV}\,(t_j = 0) = \sigma_\theta^2$.

3B. There are limited resources for communication in that, for every communication network $\mathbf{t}$, total residual variance is strictly positive, i.e., $\tau < 2\hat{\tau}$.[10]

The following two examples of communication technologies, widely used in the literature, satisfy our formulation.

**Example 1. Normally distributed messages and information.** Assume first that $\theta_j \sim \mathcal{N}\left(0, \sigma_\theta^2\right)$, and that agent $i$ receives a noisy message

$$m_j = \theta_j + \varepsilon_j \quad \text{with} \quad \varepsilon_j \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\left(t_j\right)\right). \tag{4}$$

The fact that $\theta_j$ and $\epsilon_j$ are dr99Td[(a1TJ83310.909Tf10.95(at392(99Td[(aed)-392(33(f10.95(stricd[(su/F6c

with probability $p(t_j)$ in which case agent $i$ receives a message $m_j = \theta_j$. With the remaining probability $1 - p(t_j)$, $m_j$ is uniformly distibuted on $[-1, 1]$. Then $E[\theta_j | m_j] = p(t_j)m_j$ and $E[m_j|\theta_j] = p(t_j)\theta_j$, and hence Assumption A holds. The residual variance is

$$\mathsf{RV}(t_j) = \sigma_\theta^2 \left[1 - p(t_j)\right].$$

By assuming that $p'(\cdot) > 0$, $p(0) = 0$ and $p(\tau/2) < 1$, we obtain that $\mathsf{RV}(\cdot)$ satisfies Assumption B. ■

In order to characterize optimal communication networks, additional assumptions are required on the functional form of $\mathsf{RV}(t)$. In order to do so, we will borrow from the literature on rational inattention (Sims, 2003), which in turn builds and information theory (Cover and Thomas, 1991). This theory, which relies on the concept of entropy, has strong theoretical foundations in coding theory and has proven to be useful in wide variety of settings. For Normally distributed information (example 1), it has the intuitive feature that there are decreasing marginal returns to communication, that is $\mathsf{RV}'(\cdot) < 0$ but $\mathsf{RV}''(\cdot) > 0$. To highlight the intuition behind our results, however, it will be useful to first focus on a benchmark case where there are constant marginal returns to communication: $\mathsf{RV}''(\cdot) = 0$. The case where communication displays decreasing marginal returns to communication will be addressed in Section 3.3.

### 2.4 Timing and the organizational design program

The timing of our model goes as follows:

1. *Organizational design*: Optimal communication network is designed, that is, $\mathbf{t}$ is chosen.
2. The local information $\{\theta_i\}_{i=1,2}$ is realized and observed by the agent in charge of task $i$.
3. *Adaptation:* Primary actions $q_{11}$ and $q_{22}$ are chosen by each of the agents.
4. *Communication:* Agents allocate attention $t_i$, $i = 1, 2$, to task $i$.
5. *Coordination:* Agents choose complementary actions, $q_{12}$ and $q_{21}$.

Armed with this we can write the organizational design problem as

$$\max_{\mathbf{t}} E\pi\left(\mathbf{q}|\boldsymbol{\theta}\right) \qquad \text{subject to} \quad (2)\,. \tag{6}$$

We are interested in understanding whether and under what circumstances attention is focused, say, in a single task, or group of tasks, or whether attention is evenly divided across all the tasks needed for production.

# 3 Organizational focus in the two agent case

In this section we show how the combination of adaptation-coordination trade-offs and limited attention capacity lead to organizational focus. We also emphasize that, in the absence of any of these two ingredients, attention is evenly split among tasks. We first describe in Section 3.1 the equilibrium actions and expected profit of the organization for a given communication network. This will highlight the role of communication networks in improving coordination and allowing for enhanced adaptation. We then solve for the optimal communication network under different assumptions of the communication technology.

## 3.1 Actions and the expected profits of the organization

For a given communication network $\mathbf{t}$, the best response of agent 1 features

$$q_{11} = \frac{1}{1+\beta}\left[\theta_1 + \beta E\left[q_{21}|\mathcal{I}_1\right]\right] \qquad \text{and} \qquad q_{12} = E\left[q_{22}|\mathcal{I}_1\right], \tag{7}$$

and similarly for agent 2. We can go no further without making some assumptions about the structure of the conditional expectations. We therefore focus on characterizing equilibria in linear strategies. This is without loss of generality for the two leading examples of communication technologies (example 1 and example 2). We can write (7) as

$$q_{11} = a_{11}\left(t_1\right)\theta_1 \qquad \text{and} \qquad q_{12} = a_{12}\left(t_2\right)E\left[\theta_2|\mathcal{I}_1\right]. \tag{8}$$

Substituting the guess (8) into (7), and using Assumption A, we find that the equilibrium actions for agent 1 are

$$q_{11} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta \mathsf{RV}(t_1)}\theta_1 \qquad \text{and} \qquad q_{12} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta \mathsf{RV}(t_2)}E\left[\theta_2|\mathcal{I}_1\right], \tag{9}$$

and similarly for agent 2.

Note first that the larger the residual uncertainty $\mathsf{RV}(t_i)$ about taks $i$, the less adaptive is task $i$ to its environment. The impact of attention on adaptation, however, depends on the importance of coordination, $\beta$. As $\beta$ goes to 0, tasks become perfectly adaptive for any level of attention $t_i$. In contrast, as $\beta$ goes to infinity, task $i$ becomes unresponsive to its information unless attention is perfect ($t_i \geq \hat{\tau}$) and $\mathsf{RV}(t_i) = 0$. Second, note that when the organization devotes the same amount of communication to each of the two tasks, the tasks are equally responsive to their local information. However, if the organization focuses on, say, task 1, the residual variance of task 1 is lower relative to the one of task 2, and, consequently, the primary action of task 1 is more adaptive to the shock $\theta_1$.

Substituting (9) into (1) and taking unconditional expectations we find that

$$E\left[\pi\left(\mathsf{q}|\boldsymbol{\theta}\right)\right] = \left(\Omega\left(t_1\right) - 1\right)\sigma_\theta^2 + \left(\Omega\left(t_2\right) - 1\right)\sigma_\theta^2, \tag{10}$$

where

$$\Omega\left(t_i\right) = \frac{\operatorname{cov}\left(q_{ii}\left(t_i\right), \theta_i\right)}{\sigma_\theta^2} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta \mathsf{RV}(t_i)} \in [0,1] \tag{11}$$

neatly captures the *adaptiveness of task $i$ to its task-specific information.* When the organization is fully adaptive, that is $\operatorname{cov}\left(q_{ii}, \theta_i\right) = \sigma_\theta^2$, the expected profits are maximized and $E\left[\pi\left(\mathsf{q}|\boldsymbol{\theta}\right)\right] = 0$. From (9), however, a limited attention capacity $\tau < 2\hat{\tau}$ imposes limits to adaptation such that $\operatorname{cov}(q_{ii}, \theta_i) < \sigma_\theta^2$ and $E\left[\pi\left(\mathsf{q}|\boldsymbol{\theta}\right)\right] < 0$.

An alternative representation of the expected profit function that highlights the role of communication networks better than (10) is[11]

$$E\left[\pi\left(\mathsf{q}|\boldsymbol{\theta}\right)\right] = -\beta\Omega\left(t_1\right)\mathsf{RV}\left(t_1\right) - \beta\Omega\left(t_2\right)\mathsf{RV}\left(t_2\right). \tag{12}$$

The expected residual uncertainty regarding the local information of task $i$, as reprented by $\mathsf{RV}\left(t_i\right)$, is costly to the organization only to the extent task $i$ is adaptive, as represented by its adaptiveness (11). Expression (12) separates the two aspects of organizational design

[11] Expression (10) is a generalization of the expected profit function in Dessein and Santos (2006), Proposition 2. The key difference is that now the covariances of primary actions with the corresponding local information are allowed to be different across tasks. These differences result from possible asymmetries in the communication network which are ruled out in Dessein and Santos.

that are the focus of this paper: Coordination-adaptation tradeoffs as captured by $\Omega(t_i)$, and the residual uncertainty that results from limited attention devoted to communication about task $i$, $\mathsf{RV}(t_i)$. It is immediate, then, that there is a complementarity between adaptation and coordination on a given task and a lower residual expected variance on that same task: One wants to reduce the residual variance on those tasks on which the organization is most adaptive. In turn, from expression (9), those tasks that receive most attention and have the lowest residual variance, are also most adaptive.

The problem of organizational design is to maximize (10) with respect to $t_1$ subject to $t \in [0, \tau]$ and $t_2 = \tau - t_1$. Substituting $t_2 = \tau - t_1$, the derivative of the profit function with respect to $t_1$ is

$$\begin{aligned} \frac{\partial E\left[\pi\left(\mathsf{q}|\boldsymbol{\theta}\right)\right]}{\partial t_1} &= \frac{\partial \Omega(t_1)}{\partial t_1}\sigma_\theta^2 + \frac{\partial \Omega(\tau - t_1)}{\partial t_1}\sigma_\theta^2 && (13) \\ &= \beta\Omega^2(t_1)\left|\mathsf{RV}'(t_1)\right| - \beta\Omega^2(t_2)\left|\mathsf{RV}'(t_2)\right| && (14) \end{aligned}$$

where $\left|\mathsf{RV}'(t_i)\right|$ are the marginal returns to communicate about $\theta_i$ given $t = t_i$.

### 3.2 Organizational focus under constant marginal returns to communication

We first consider the case of communication technologies that exhibit constant marginal returns to communication, i.e., $\mathsf{RV}''(\cdot) = 0$. For example, with uniformly distributed information and messages (recall example 2), the assumption of constant marginal returns to communication signifies that the probability that communication is successful is linear in the time devoted to communication, i.e., $p(t) = \alpha t$, for some positive $\alpha$. Using (13), we obtain that if there are constant returns to communication, then

$$\frac{\partial E\left[\pi\left(\mathsf{q}|\boldsymbol{\theta}\right)\right]}{\partial t_1} > 0 \qquad \Longleftrightarrow \qquad \Omega(t_1) > \Omega(t_2) \qquad \Longleftrightarrow \qquad t_1 > t_2. \tag{15}$$

It follows that the expected profits reach a global minimum for $t_1 = t_2 = \tau/2$, that is when attention is equally divided between the two tasks. We have thus proved the following proposition:

**Proposition 1** *If there are constant marginal returns to communication, then the organization will focus on one task. In particular, if $\tau < \hat{\tau}$ then the organization will focus entirely on one*

*task and ignore the other, i.e.,* $t^* \in \{0, \tau\}$*. If* $\tau > \hat{\tau}$ *the organization will devote enough resources on one task so that the organization perfectly learns the local shock of that task, and will devote the remaining communication resources to learn about the other task, i.e.,* $t^* \in \{\tau - \hat{\tau}, \hat{\tau}\}$*.*

As an illustration of this result, consider the communication technology in example 2, and let messages be informative with probability $p(t) = \alpha t$. Proposition 1 implies that if communication resources $\tau$ are lower than $\hat{\tau} = 1/\alpha$, then the optimal organization will focus

as shown in the next section, the unfocussed outcome $(t_1, t_2) = (\tau/2, \tau/2)$ is the unique osptimum for $\beta = 0$ when there are decreasing marginal returns to communication. Similarly, as shown in Section 3.3, focus in never optimal when both tasks are carried out by one and the same agent who can ensure perfect coordination.

Our result of organizational focus has been derived under the assumption of constant marginal returns to communication, $\mathsf{RV}'' = 0$. Obviously the result in Proposition 1 holds if the communication technology displays increasing marginal returns to communication, i.e., $\mathsf{RV}''(\cdot) < 0$. In what follows we study the possibility of organizational focus in those contexts where communication technologies display decreasing marginal returns.

### 3.3 Decreasing marginal returns to communication

Next we put more structure on our framework and the function $\mathsf{RV}(t)$ by drawing on insights from the literature on rational inattention (Sims, 2003) and information theory (Cover and Thomas, 1991). Following this literature, we assume that the time or attention needed to communicate or process a signal about a random variable $\theta_i$ depends on the extent to which this signal reduces the differential entropy of $\theta_i$, where this time or attention plays the role of the finite (Shannon) capacity of a noisy communication channel. Formally, the communication capacity (time, attention) needed to communicate a message $m_i$ about $\theta_i$ is given by

$$I(\theta_i, m_i) = h(\theta_i) - h(\theta_i \| m_i)$$

where $h(\theta_i)$ is the differential entropy of $\theta_i$ and $h(\theta_i | m_i)$ is the differential entropy of $\theta_i$ conditional on observing $m_i$. We then posit that[12]

$$I(\theta_1, m_1) + I(\theta_2, m_2) \leq \tau \tag{16}$$

where $\tau$ is the (Shannon) capacity of communication channel between agent 1 and 2. We refer to Cover and Thomas (1991) for a thorough treatment of the foundations of Information Theory. Rather than for its axiomatic appeal, however, Shannon capacity is widely used because it has proven to be appropriate concept for studying information flows in a variety of disciplines:

[12] We assume here that $m_i$ is uncorrelated with $\theta_j$ whenever $i \neq j$.

probability theory, communication theory, computer science, mathematics, statics, as well as in both portfolio theory and macroeconomics. While there are arguably an unlimited number of ways to model communication and information-processing constraints, it is intuitively appealing – and limits the degrees of freedom of the modeler – to assume that those limits behave like finite Shannon capacity.[13]

In information theory, $I(\theta_i, m_i)$ is referred to as the *mutual information* between $\theta_i$ and $m_i$. While each agent is privately informed about the task-specific shock affecting his own task, the mutual information about $\theta_1$ and $\theta_2$ represents the collective knowledge of the organization. While the *total amount of organizational knowledge* is fixed, the organization, can decide to allocate a larger fraction of his communication capacity to, say, task 1. The question of this paper, then, is whether organizations optimally develop *specialized organizational knowledge* or not. A specialized organization has $I(\theta_i, m_i) > I(\theta_j, m_j)$ with an extreme form of organizational specialization being the case where $I(\theta_i, m_i) = \tau$ and $I(\theta_j, m_j) = 0$.

We will make the assumption, common in the literature on rational inattention, that $m_i = \theta_i + \varepsilon_i$ where both $\theta_i$ and $\varepsilon_i$ are independently normally distributed. This assumption is justified by its tractability and a well-known result in information theory, which states that the normal distribution minimizes the variance for a given level of entropy (see Sims 2006 for a discussion).[14] Since the entropy of a normal variable with variance $\sigma^2$ is given by $\frac{1}{2}\ln(2\pi e\sigma^2)$, then

$$I(\theta_i, m_i) = \frac{1}{2}\left(\ln \sigma_\theta^2 - \ln \operatorname{Var}(\theta_i|m_i)\right) \tag{17}$$

It follows that the time or attention needed to communicate a message $m_i$ about $\theta_i$ is linear in the reduction in the log residual variance of $\theta_i$ following communication. While it takes no time to communicate a message which is pure noise, it would be infinitely costly to communicate $\theta_i$ perfectly. Let $t_i$ be the communication capacity allocated to communicate $m_i$, with $t_1 + t_2 = \tau$,

[13]Sims (1998, 2003) uses exactly the same justification to advocate the use of finite Shannon capacity in modelling the limits of attention by economic agents to publicly available information and the resulting inertia in observed economic behavior.

[14]If, as assumed in information theory, agents optimally design the distribution of $F(\theta_i|m_i)$ subject a capacity constraint for the communication channel, they would choose $F(\theta_i|m_i)$ to be normally distributed in or to maximize our quadratic objective function. Cover and Thomas (1991) devote a whole chapter to Guassian Channels as they are the most commonly used to model information flows in a variety of settings.

we have that

$$\ln \mathsf{RV}\left(t_i\right) = \ln \sigma_\theta^2 - 2t_i \tag{18}$$

where, recall, $\mathsf{RV}(t_i) = Var(\theta_i|m_i)$.

An important and intuitive feature of the above communication technology is that it implies *decreasing marginal returns to communicating* about a particular task-specific shock. While initially it is easy to reduce the residual variance by devoting a small amount of attention, it is increasingly difficult to further reduce the residual variance as more attention has already been allocated. If it takes $\Delta t$ to reduce the expected residual variance from $\sigma_\theta^2$ to $\sigma_\theta^2/2$, it will take an additional $\Delta t$ to reduce the expected residual variance from $\sigma_\theta^2/2$ to $\sigma_\theta^2/4$, and so on. Only in the limit where $t_i$ goes to infinity will the residual variance go to zero. Formally, the marginal returns to attention/communication equal

$$\left|\mathsf{RV}'\left(t_i\right)\right| = 2\mathsf{RV}\left(t_i\right), \tag{19}$$

hence the lower the residual variance, the lower the marginal returns to further reduce this variance. While we have derived equation (18) using foundations in information theory, we believe it provides an intuitive, tractable and parsimonious way to model decreasing marginal returns to communication.[15] [16]

**Focussed versus balanced organizations** In Section 3.2, we highlighted a complementarity between attention and the adaptiveness of a task. It is optimal to allocate more attention to a task which is more adaptive. In turn, a task which receives more attention is more adaptive, making organizational focus optimal. The more interdependent are tasks, that is the

[15] Expression (18) also has a natural interpretation if $\theta_i$ and $m_i$ are uniformly distributed, as in Example 2. It is then equivalent with communication following a poisson process with a constant hazard rate of correctly learning the local shock. In this case, the function $p\left(t_i\right) = 1 - e^{-2t_i}$ is the probability that success has occurred prior to time $t_i$.

[16] We note that, within the framework of information theory, example 2 represents a case of constant marginal returns to communication. To see this note that, since $\theta_i$ is distributed uniformly in $[-1, 1]$, it follows that the differential entropy of $\theta_i$ is $\ln(2)$ and the conditional differential entropy of $\theta_i$ given the message $m_i$ is $[1-p(t_i)]\ln(2)$. So, the mutual information is $p(t_i)\ln(2)$. Hence, imposing $t_1+t_2 = \tau$ and $I(\theta_1, t_1)+I(\theta_2, t_2) = \tau$, we obtain that $p(t)$ must be linear in $t$, an example of constant marginal returns to communication.

larger is $\beta$, the stronger is this complementarity. Decreasing marginal returns to communication, however, provides a powerful force *against* focus. Indeed, now the more attention a task receives, the lower the marginal return to further increase attention, at least in terms of reducing residual uncertainty. There is then a race between increasing returns to coordination and decreasing returns to communication. Formally, it follows from the profit derivate (13) that a focussed organization with $(t_1, t_2) = (\tau, 0)$ is a *local maximum* if and only if

$$\underbrace{\Omega^2(\tau)}_{\text{Adaptiveness}} \quad * \quad \underbrace{\left|\mathsf{RV}'(\tau)\right|}_{\text{Marg. returns to comm.}} \quad > \quad \Omega^2(0) \quad * \quad \mathsf{RV}'(0). \tag{20}$$

As shown in the previous section, this condition is always satisfied and organizational focus is optimal if there are constant marginal returns to communication. A balanced organization $(t_1, t_2) = (\tau/2, \tau/2)$ may be optimal, however, when there are decreasing marginal returns to communication. Indeed, if the organization focusses on, say, task 1, then task 1 is more adaptive, that is $\Omega(\tau) > \Omega(0)$, but the marginal returns to communication are larger for task 2, that is $\left|\mathsf{RV}'(0)\right| > \left|\mathsf{RV}'(\tau)\right|$. As we show next, if either coordination is not very important ($\beta$ small) or attention is not very constrained ($\tau$ large), a focussed organization with $(t_1, t_2) = (\tau, 0)$ is then suboptimal.

Consider first the case where coordination is not very important. For $\beta$ small, both tasks are almost equally adaptive, that is $\Omega(\tau) \approx \Omega(0)$. At the same time, the marginal returns to communication are distinctly lower on task 1 than on task 2. Regardless of $\tau$, for $\beta$ sufficiently small, inequality (20) is then violated and $(\tau, 0)$ is not a local maximum. Intuitively, the complementarity between adaptiveness and the allocation of attention relies on the importance of coordination. In the limit, as $\beta$ goes to zero, this complementarity and the associated increasing returns to coordination disappear.

Next, consider the case where attention is not very constraint, that is $\tau$ is large. Since the marginal returns to communication on task 1, $\left|\mathsf{RV}'(\tau)\right|$, go to zero as $\tau$ goes to infinity, whereas $\Omega(0)$ is strictly positive, it follows that $(\tau, 0)$ is not a local maximum for $\tau$ sufficiently large.

In line with the above intuitions, the following proposition shows that a fully focussed organization is optimal if coordination is sufficiently important and attention is constraint:

**Proposition 2** *There exists a* $\widehat{\beta}$ *and* $\top(\beta)$ *such that*

1. *If* $\beta \leq \widehat{\beta}$ *then organizational balance is optimal:* $(t_1^*, t_2^*) = (\frac{\tau}{2}, \frac{\tau}{2})$.

2. *If* $\beta > \widehat{\beta}$ *then*

   (a) *Organizational focus is optimal,* $t_1^* \in \{0, \tau\}$, *if and only if* $\tau \leq \top(\beta)$

   (b) *Organizational balance is optimal,* $(t_1^*, t_2^*) = (\frac{\tau}{2}, \frac{\tau}{2})$, *if* $\tau > \top(\beta)$

   (c) $\top(\cdot)$ *is increasing in the importance of coordination,* $\beta$.

Proposition 2 shows that organizations which are 'somewhat' focussed are never optimal. Indeed, if full focus is not optimal, the organization divides its attention equally among both tasks. Intuitively, given the complementarities between the adaptiveness of a task and the attention devoted to a task, the organization either completely ignores a task, or it devotes a substantial amount of attention to it. At the threshold $\top(\beta)$, the organization makes this shift from no attention to one task, to a substantial amount of attention to both tasks.

Proposition 2 further yields an interesting comparative static results with respect to exogenous changes in the communication capacity $\tau$. Improvements in the communication technology (email, wireless communication devices, intranets, ...) can be interpreted as an exogenous increase in $\tau$. An implication of Proposition 2, therefore, is that such technological improvements result in a shift from focussed organizations which are centered around one task and excell on that task at the expense of others, towards more balanced organizations which aim to perform equally well on all tasks, but excell in none.

Finally, Proposition 2 has implications for the importance of leadership in teams. At the threshold $\top(\beta)$ the organization changes from having a single leader who monopolizes all information flows to a structure with shared leadership. Hence, an increased communication capacity may come at the expense of the original leader in an organization, who may face a discrete loss of power and influence in the organization. As a result, his task is less adapted to its environment and, typically, other tasks are less well coordinated with it. From having a complete monopoly on attention in the organization, this leader now must equally shares all attention

Underlying the above results is the notion, proposed in Dessein and Santos (2006), that there are *two ways to maintain coordination* in an organization. One way is for the organization to devote substantial attention to a task. The agent in charge of this task can then be very responsive to his local information as the other agents in the organization will likely be aware of his actions and take the appropriate coordinating actions. In Dessein and Santos (2006), this was referred to as *ex-post coordination,* by means of communication. An alternative way is for the agent to simply ignore his private information and always implement his task in the same manner. Other agents can then maintain coordination with this task without having to devote any attention to it. This is referred to as *ex-ante coordination.*

While in Dessein and Santos (2006) all tasks were treated symmetrically by assumption, the insight of Proposition 2 is that when attention is constraint, it is often optimal to coordinate ex-ante on one of the tasks and coordinate ex post on the other task. The first task is then very rigid and insensitive to its local information, so that the organization can afford to ignore this task and fully allocate its attention to the second task, allowing this task to be very flexible and adaptive. Despite a limited attention capacity, both tasks are then well coordinated, but only one task is very sensitive to its environment. In contrast, when attention is not very constrained, its optimal for both tasks to be very adaptive, as both tasks can then be coordinated ex post through communication.

**Team production versus individual/centralized production** In the next section, we generalize our results to production by an arbitrary number of players. Before doing so, however, we want to highlight that organizational focus only arises if there is production by a team of players. In particular, in a one-man organization, where all tasks are carried out by the same agent, organizational focus never arises. Indeed, it is team production which makes

respectively $\theta_1$ and $\theta_2$. We next show that under such individual or centralized production, there is no organizational focus. Note that if the single agent were to observe one or both of the random variables $\theta_1$ and $\theta_2$, the allocation of attention would be trivial or irrelevant.

Let $t_1$ be the amount attention a single agent allocates to learn about $\theta_1$, and $t_2 = \tau - t_1$, the time she allocates to task 2. The organizational design problem then consists of how much limited attention to devote to task 1 and task 2, given an attention constraint $\tau$. We maintain the assumption that information processing is subject to decreasing marginal returns, that is $\mathsf{RV}(t_i)$ is given by (18).[17] Let $\mathcal{I}_S$ be the information of the single agent, then she will choose

$$q_{21} = q_{11} = E\left[\theta_1 | \mathcal{I}_{HQ}\right] \qquad \text{and} \qquad q_{12} = q_{22} = E\left[\theta_2 | \mathcal{I}_{HQ}\right]. \tag{21}$$

Note that since $q_{12} = q_{11}$ and $q_{21} = q_{22}$, it is *as if* $\beta = 0$. Substituting the optimal actions into (1) and taking unconditional expectations we find that

$$E[\pi(\mathsf{q})|\theta] = -\left[\mathsf{RV}(t_1) + \mathsf{RV}(\tau - t_1)\right]. \tag{22}$$

Intuitively, if there is no division of labor, the importance of coordination is irrelevant for expected profits or the allocation of attention. The optimal allocation of attention is then the one that minimizes total residual variances. Given that there are decreasing marginal returns to attention, the following result is then direct:

**Proposition 3** *If all tasks are implemented by a single agent, then this agent optimally splits attention evenly among both tasks $t^* = \frac{\tau}{2}$.*

While optimal level of task specialization is not the focus of this paper (see Dessein and Santos 2006, which endogenizes the number of tasks per agent and, hence, the optimal team size), it is natural to wonder when team production is optimal. We address this question in Section 5. For now, we take team production as a necessary and exogenous feature of production.

---

[17] The specific functional form is irrelevant for Proposition 3 to hold. All what is required is that there are decreasing marginal returns to attention.

# 4 Organizational focus with many agents

## 4.1 The model with $n > 2$

We now extend our analysis to allow for an arbitrary number agents in the team. Once a team is composed of more than two agents, the way in which communication occurs– bilateral meetings **vs** public meetings – matters for how the attention constraint is defined. We therefore first remark on how the model introduced in section 2 extends to the case of $n > 2$. The characterization of the optimal communication network is left for the next section. To simplify the analysis, we immediately adopt the decreasing marginal returns to communication technology introduced in Section 2.

**Production.** Consider therefore a production process which involves the implementation of $n > 2$ tasks. As before, each task $i$ must be performed by a specialized agent $i \in \mathcal{N} \equiv \{1, ..., n\}$ who observes some task-specific information $\theta_i$ with mean 0 and variance $\sigma_\theta^2$. In order to implement task $i$, agent $i$ chooses a primary action $q_{ii}$, who must be adapted to the task-specific shock $\theta_i$, as well as $(n-1)$ coordinating actions $q$, who must be adapted to the primary actions chosing by the other agents $j \in \mathcal{N} \setminus \{i\}$. We denote by

$$q_i = [q_{i1}, q_{i2} ..., q_{ii}, ..., q_{in}], \tag{23}$$

the string of actions choosen by agent $i$. Denote by $\boldsymbol{\theta} = [\theta_1, ..., \theta_n]$ the vector of realized shocks and by $\mathbf{q} = [q_1, q_2, ..., q_n]$ the vector of actions, respectively; the realized profit of the organization is:

$$\pi(\mathbf{q}|\boldsymbol{\theta}) = -\sum_{i \in \mathcal{N}} \left[ (q_{ii} - \theta_i)^2 + \frac{\beta}{n-1} \sum_{j \in \mathcal{N} \setminus \{i\}} (q_{ii} - q_{ji})^2 \right]. \tag{24}$$

**Communication.** Following communication, each agent $i$ observes a string of messages

$$m_i = [m_{i1}, m_{i2}, ..., m_{ii}, ..., m_{in}],$$

where $m_{ii} = \theta_i$ and $m_{ij} = \theta_i + \varepsilon_{ij}$ with $\varepsilon_{ij}$ a random noise term. As in the two-agents case, we draw upon information theory and posit that communication constraints stem from a finite (Shannon) communication capacity $\tau$. Let $\theta_j$ and $m_{ij}$, for all $i, j \in \mathcal{N}$, be normally distributed,

and let $t_{ij}$ be the communication capacity (or attention) agent $i$ and $j$ devote to communication about $\theta_j$, then

$$\ln Var(\theta_j|m_{ij}) = \ln \sigma_\theta^2 - 2t_{ij}, \tag{25}$$

where it must be that

$$\sum_{j\in\mathcal{N}} \sum_{i\in N\setminus\{i\}} t_{ij} \leq \tau, \tag{26}$$

The above communication network $\mathsf{t} = \{t_{ij}\}_{i\neq j}$ is one where communication among agents is assumed to be bilateral and allows for a rich variety of asymmetries. On the one hand, agent $j$ may devote more attention to agent $i$ than another agent $k$, that is, $t_{ji} > t_{ki}$. On the other hand, agent $i$ may receive more attention from the organization than another agent $k$, that is, $\sum_j t_{ji} > \sum_j t_{jk}$. Next, we discuss some alternative communication technologies, and discuss how they result in equivalent information structures.

**Alternative models of communication.** In what follows we shall develop the analysis assuming bilateral communication, as this puts the least constraints on nature of communication flows. We will however show that in the optimal bilateral communication network, every agent $j$ pays the same attention to agent $i$, i.e.,

$$t_{ji} = t_{ki}, \quad \text{for all } j,k \in \mathcal{N}\setminus\{i\} \text{ and for all } i. \tag{27}$$

As we argue next, the insights we shall present about the properties of optimal communication networks therefore also hold when communication is *public* rather than bilateral and, hence, property (27) holds exogenously. We also discuss how our analysis can be applied when the communication constraint $\tau$ holds at the individual level, where each agent has its own communication channel with limited capacity, rather than at the organizational level.

*(i) Public communication.* An alternative model of communication is one in which communication occurs in public meetings, where only one agent can speak at a given time and all others listen. The organizational design variable is then the "air-time" or "attention" any agent $j$ receives. The communication network is given by $\mathsf{t} = \{t_1, ..., t_n\}$, where $t_j$ is the communication capacity devoted to task $j$, and the communication constraint is given by

$$\sum_j t_j \leq \tau.$$

Formally, one can think of a communication channel which can have only one input or sender, but has no limit to the number of receivers. The conditional variances are then defined by

$$\ln Var(\theta_j|m_{ij}) = \ln \sigma_\theta^2 - t_j$$

Under public communication, two agents $j, k \in \mathcal{N} \backslash \{i\}$ are constrained to pay the same amount of attention to agent $i$, a property that can be endogenously derived for the optimal bilateral communication networks. The following equivalence result, proven in appendix, therefore follows directly:

**Result 1:** *An optimal communication network $t = \{t_1, .., t_n\}$ given public communication and constraint $\tau$ satisfies*

$$t_j = t_{ij}^b \text{ for all } j, i \in \mathcal{N}$$

*where $t^b = \left\{t_{ij}^b\right\}_{i \neq j}$ is an optimal communication network under bilateral communication and constraint $\tau^b = (n-1)\tau$.*

In the above result, we use the terminology of 'an' optimal communication network as there are typically several optimal communication networks, where the organization focusses on the same number, but potentially different, tasks. Note that with 2 agents, even $\tau = \tau^b$, and there is no difference between public and bilateral communication networks.

*(ii) Individual communication channels and constraints.* Sofar, we have assumed that the communication constraint is determined at the organizational level. Alternatively, each agent may have a limited communication capacity $\tau$.[18] Formally, let each agent have access to an individual communication channel, whose finite capacity $\tau$ can be used to broadcast information to all other agents and/or to process information broadcasted by others. Each agent $i$ then optimally decides on a vector $t_i = [t_{i1}, t_{i2}, ..., t_{ii}, ..., t_{in}]$, where

$$\sum_{j \in N} t_{ij} \leq \tau \qquad \forall i \in \mathcal{N}, \tag{28}$$

and where $t_{ii}$ is the capacity devoted to broadcast information about $\theta_i$, and $t_{ij}$ is the capacity devoted to listen to the information broadcasted by agent $j \neq i$. The effective communication

[18] Note that this distinction again does not matter when $n = 2$, as both agents are then always involved at the same time.

flow between agents $j$ and $i$ regarding $\theta_j$ then equals $\min\{t_{ij}, t_{jj}\}$ such that[19]

$$\ln Var(\theta_j | m_{ij}) = \ln \sigma_\theta^2 - \min\{t_{jj}, t_{ij}\}.$$

In appendix we prove the following equivalence result which again relies on the property (27) of any optimal bilateral communication network:

**Result 2:** *An optimal communication network $t = \{t_{ij}\}_{i,j}$ with individual communication constraints $\tau$ satisfies*

$$t_{jj} = t_{ij} = t_{ij}^b \text{ for all } j, i \in \mathcal{N}$$

*where $t^b = \left\{t_{ij}^b\right\}_{i \neq j}$ is an optimal bilateral communication network with organization-wide constraints $\tau^b = (n-1)\tau$.*

Note again that with $n = 2$, there is no distinction between the model with bilateral meetings, the model with public meetings, and the model with individual communication constraints.

## 4.2 Organizational actions and performance

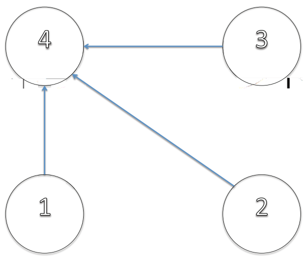
Having characterized how a particular communcation network maps into an information structure, we now characterize the organizational actions and performance for a given communication structure. For a given communication network $\mathbf{t}$ and a string of observed messages $m_i$, agent $i$ chooses the string of actions $q_i$, given in (23), in order to maximize

$$E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)|\mathcal{I}_i\right],$$

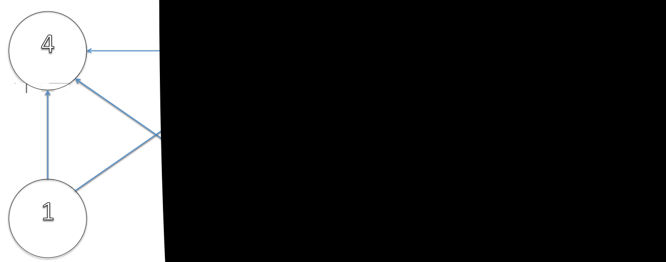
where the function $\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)$ is given by expression (24) and $\mathcal{I}_i$ is the information set of agent $i$ after communication with the rest of the other agents as prescribed by communication network $\mathbf{t}$. Primary and complementary actions are thus

$$q_{ii} = \frac{1}{1+\beta}\left[\theta_i + \frac{\beta}{n-1}\sum_{j \neq i} E\left[q_{ji}|\mathcal{I}_i\right]\right] \qquad \text{and} \qquad q_{ij} = E\left[q_{jj}|\mathcal{I}_i\right].$$

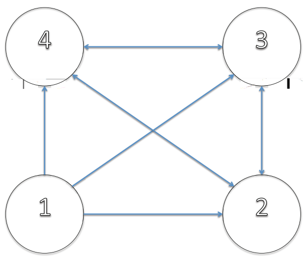
[19]For example, if agent $j$ communicates for 1 hour, but agent $i$ only listens for 1/2 hour, then the effective communication time is only 1/2 hour. The same holds if agent $i$ listens for 1 hour, but agent $j$ only communicates for a 1/2 hour.



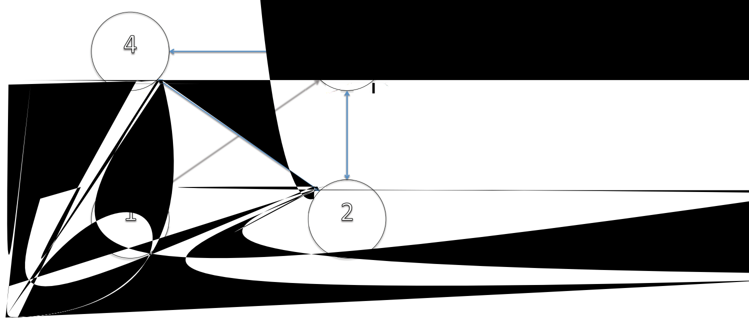

(a) 1-leader, $\tau_{i4} = \tau/3$, $i \neq 4$.




(b) 2-leaders, $\tau_{i4} = \tau_{j3}$




(c) 3-leaders, $\tau_{i4} = \tau_{j3} = \tau_{s2} = \tau/9$, $i \neq 4$, $j \neq 3$, $s \neq 2$




(d) 4-leaders, $\tau_{ij} = \tau/12$, $i \neq j$.

Figure 1: $\ell-$leader organization, $n = 4$

agents (including other leaders) pay *equal attention*, and a second class of agents to whom no other agent in the organization pays attention. Our main result is the following proposition.

**Proposition 4** *The optimal communication network is an $\ell-$leader organization with $\ell \in \{1, 2, \cdots, n\}$.*

The proof of Proposition 4 follows from the next two lemmas.

**Lemma 5** *In an optimal communication network all agents devote the same attention to a particular agent, that is, $t_{ji} = t_{ki}$ for all $j, k \neq i$*

The intuition behind lemma 5 is the following. Suppose it is optimal for the organization to devote a total amount of attention $t_i = \sum_{j \neq i} t_{ji}$ to task $i$. Then, the optimal way to distribute $t_i$ across communication links $\{t_{1i}, .., t_{i-1i}, t_{i+1i}, ..., t_{ni}\}$ is such that it minimizes the total residual variance about $\theta_i$ of the organization, i.e., it minimizes $\sum_j \mathsf{RV}(t_{ji})$. Since there are decreasing marginal returns to communication, it is optimal to split total attention devoted to $i$, $t_i$, equally across communication links $\{t_{1i}, .., t_{i-1i}, t_{i+1i}, ..., t_{ni}\}$.

**Lemma 6** *In an optimal communication network all agents who receive some positive attention from all other agents in the organization, receive the same attention, i.e., if $t_i = \sum_s t_{si} > 0$ and $t_j = \sum_s t_{sj} > 0$ then $t_i = t_j$, for all $i, j$.*

To see the intuition behind lemma 6, let $i$ and $j$ be two tasks with $\hat{t}_i = \sum_s t_{si}$ be the total attention devoted to task $i$ and $\hat{t}_j = \sum_s t_{sj}$ the total attention devoted to task $j$. Moreover, assume $\hat{t}_i > \hat{t}_j > 0$, in violation of lemma 6. In the case of two tasks, Proposition 2 has shown that either $t_1^* \in \{0, \tau\}$ , or $t_1^* = t_2^* = \tau/2$. Following the same logic, one can show that, keeping the attention allocated to all other tasks $k \notin \{i, j\}$ fixed, profits can always be strictly increased by either setting $t_i = \hat{t}_i + \hat{t}_j$ and $t_j = 0$ or, alternatively, equalizing attention across tasks $i$ and $j$, that is setting $t_i = t_j = (\hat{t}_i + \hat{t}_j)/2$. The intuition is identical to the intuition developed for Proposition 2. As in the two tasks case, it is optimal to either allocate a substantial amount of attention to any given task, allowing it to become very adaptive and coordinate this task *ex post* (Dessein and Santos, 2006), or force a task to largely ignore its

local information and coordinate this task *ex ante* (Dessein and Santos, 2006), which does not require any attention. The importance of coordination and the amount of attention available then determines whether it is optimal for both tasks to receive an equal amount of attention, or for one task to receive all the attention and the other none.

**Actions and performance of the $\ell-$ leader organization** Armed with Proposition 4 we can divide the organization in two groups of agents: Those who adapt more, the leaders, and those who adapt less, the followers. Without any loss of generality, let the $\ell$ leaders be the first $\ell$ agents and let the followers be agents $\{\ell+1, \ell+2, \cdots, n\}$. The primary equilibrium actions of leaders and followers are given by

$$q_{ii} = \frac{\theta_i \sigma_\theta^2}{\sigma_\theta^2 + \beta \mathsf{RV}\left(\frac{\tau}{(n-1)\ell}\right)} \quad \text{for} \quad i \leq \ell \quad \text{and} \quad q_{ii} = \frac{\theta_i}{1+\beta} \quad \text{for} \quad \ell+1 \leq i \leq n.$$

Thus leaders naturally comove more with their local information than do followers:

$$\Omega\left(\frac{\tau}{(n-1)\,\ell}\right) = \frac{\operatorname{cov}(q_{ii}, \theta_i)}{\sigma_\theta^2} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta \mathsf{RV}\left(\frac{\tau}{(n-1)\ell}\right)} \quad \text{for} \quad i \leq \ell \tag{30}$$

$$\Omega(0) = \operatorname{cov}(q_{ii}, \theta_i) = \frac{1}{1+\beta} \quad \text{for} \quad \ell+1 \leq i \leq n. \tag{31}$$

Leaders adapt more because they are paid attention by the rest of the agents in the organization, both by the followers and by the other leaders. In addition, casual inspection of (30) shows that, other things equal, as the number of leaders increase the influence of each of them decreases as now they have to share the same amount of attention $\tau$ with a larger number of other leaders. The number of leaders will only change in the presence of exogenous sources of variation, and therefore the equilibrium level of adaptation may go up or down. However, our result points out that an increase in the number of leaders can only be at the expense of the adaptiveness of the existing leadership.

When the communication network takes the form of an $\ell-$leader organization, the expression of the profit function (29) can be re-written as:

$$E\left[\pi\left(\mathsf{q}|\boldsymbol{\theta}\right)\right] = -n\sigma_\theta^2 + \sigma_\theta^2 \ell \left[\frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta \mathsf{RV}\left(\frac{\tau}{(n-1)\ell}\right)}\right] + (n-\ell)\left[\frac{\sigma_\theta^2}{1+\beta}\right], \tag{32}$$

where to obtain (32) we have made use of both Proposition 4 and (30) and (31). As in the two-agent case, we can rewrite (32) as follows

$$E\left[\pi\left(\mathsf{q}|\boldsymbol{\theta}\right)\right]=-\beta\ell\Omega\left(\frac{\tau}{\left(n-1\right)\ell}\right)\mathsf{RV}\left(\frac{\tau}{\left(n-1\right)\ell}\right)-\beta\left(n-\ell\right)\Omega\left(0\right)\mathsf{RV}\left(0\right), \tag{33}$$

highlighting the complementarity between the adaptiveness of a task $\Omega(t_i)$ and the residual variance $\mathsf{RV}(t_i)$ surrounding it. The optimal number of leaders, then, is given by

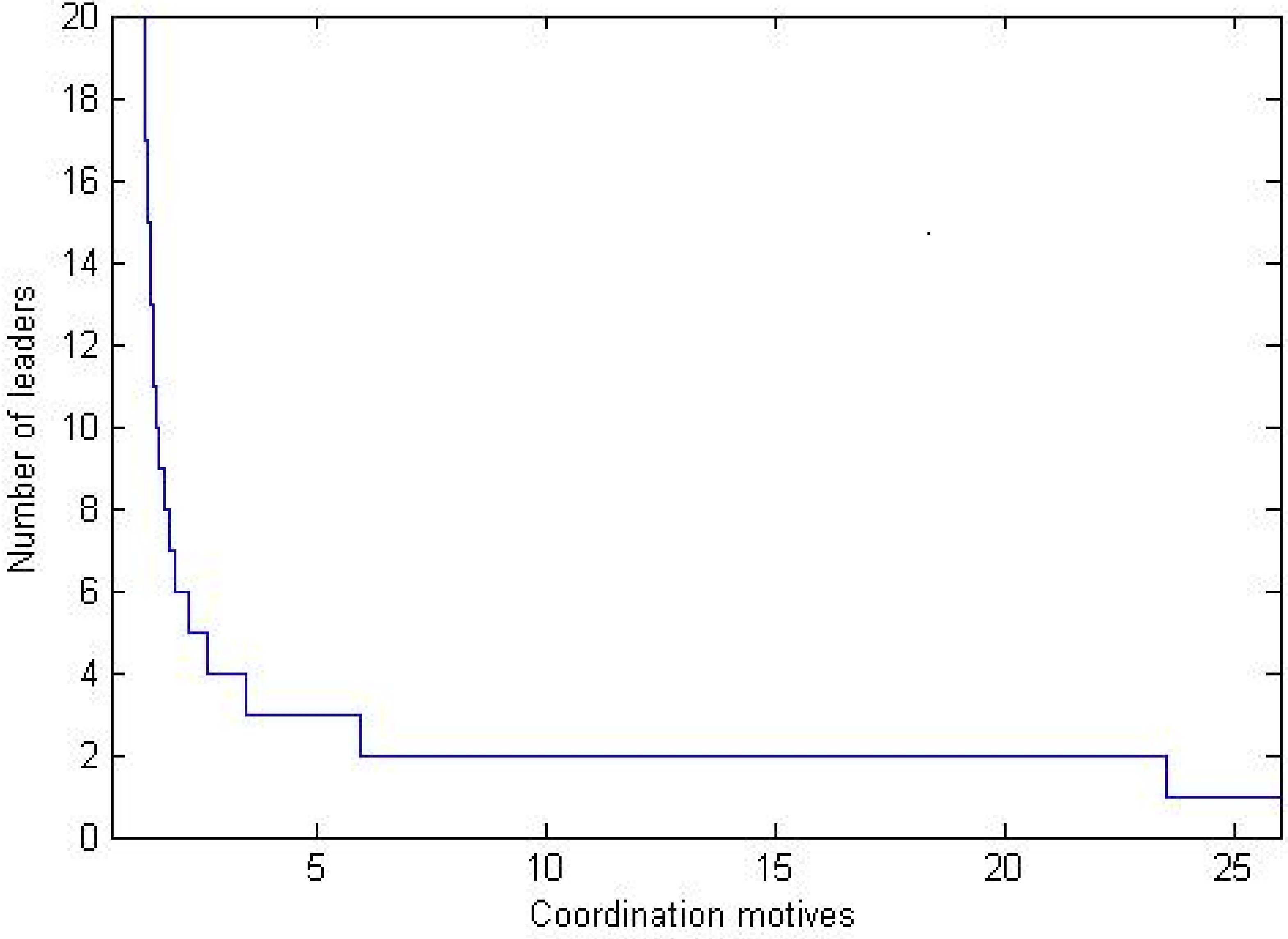
$$\ell^*=\mathsf{argmax}_{\ell\in\{1,2,\cdots,n\}}E\left[\pi\left(\mathsf{q}|\boldsymbol{\theta}\right)\right]. \tag{34}$$

**Comparative statics of the of the $\ell-$ leader organization** Armed with (34) we are able to offer a sharp characterization of the $\ell-$leader organization as a function of the organization's communication capacity $\tau$ and the task-interdepence or coordination parameter $\beta$.

**Proposition 7** *There exists* $0<\overline{\beta}\left(n\right)<\ldots<\overline{\beta}\left(\ell+1\right)<\overline{\beta}\left(\ell\right)<\ldots<\overline{\beta}\left(2\right)$ *such that*

1. $\ell^*=n$ *if* $\beta<\overline{\beta}\left(n\right)$, $\ell^*=\ell\in\{2,\cdots,n-1\}$ *if* $\beta\in(\overline{\beta}\left(\ell+1\right),\overline{\beta}\left(\ell\right))$, *and* $\ell^*=1$ *if* $\beta>\overline{\beta}\left(2\right)$
2. *For all* $\ell\in\{1,...,n\}$, $\overline{\beta}\left(\ell\right)$ *is increasing in* $\tau$ *and* $\lim_{\tau\to\infty}\overline{\beta}\left(\ell\right)=\infty$.

Figure **??** illustrates the results of Proposition 7 for a team composed of $n=20$ agents; it shows how the the optimal number of leaders changes when coordination costs increase, as captured by an increase in $\beta$. The intuition for Proposition 7 is similar to the one for Proposition 2, with the obvious difference that now there is an intermediate region where the communication network is neither entirely focused nor completely balanced. A balanced organization is optimal when coordination is sufficiently un-important, that is $\beta<\overline{\beta}\left(n\right)$. When coordination becomes more important, the communication becomes more focussed around fewer leaders. Finally, when tasks are sufficiently interdepent, the organization has a single leader. Proposition 7 further shows how an exogenous change in the availability or effectiveness of attention $\tau$, increases the number of leaders and makes the organization more balanced. Regardless of the importance of coordination, a balanced organization is always optimal if attention is sufficiently unconstrained. Again, this implies that as communication technology improves, organizations become less focussed and leadership is more broadly shared.


Number of leaders
0
2
4
6
8
10
12
14
16
18
20
5
10
15
20
25
Coordination motives

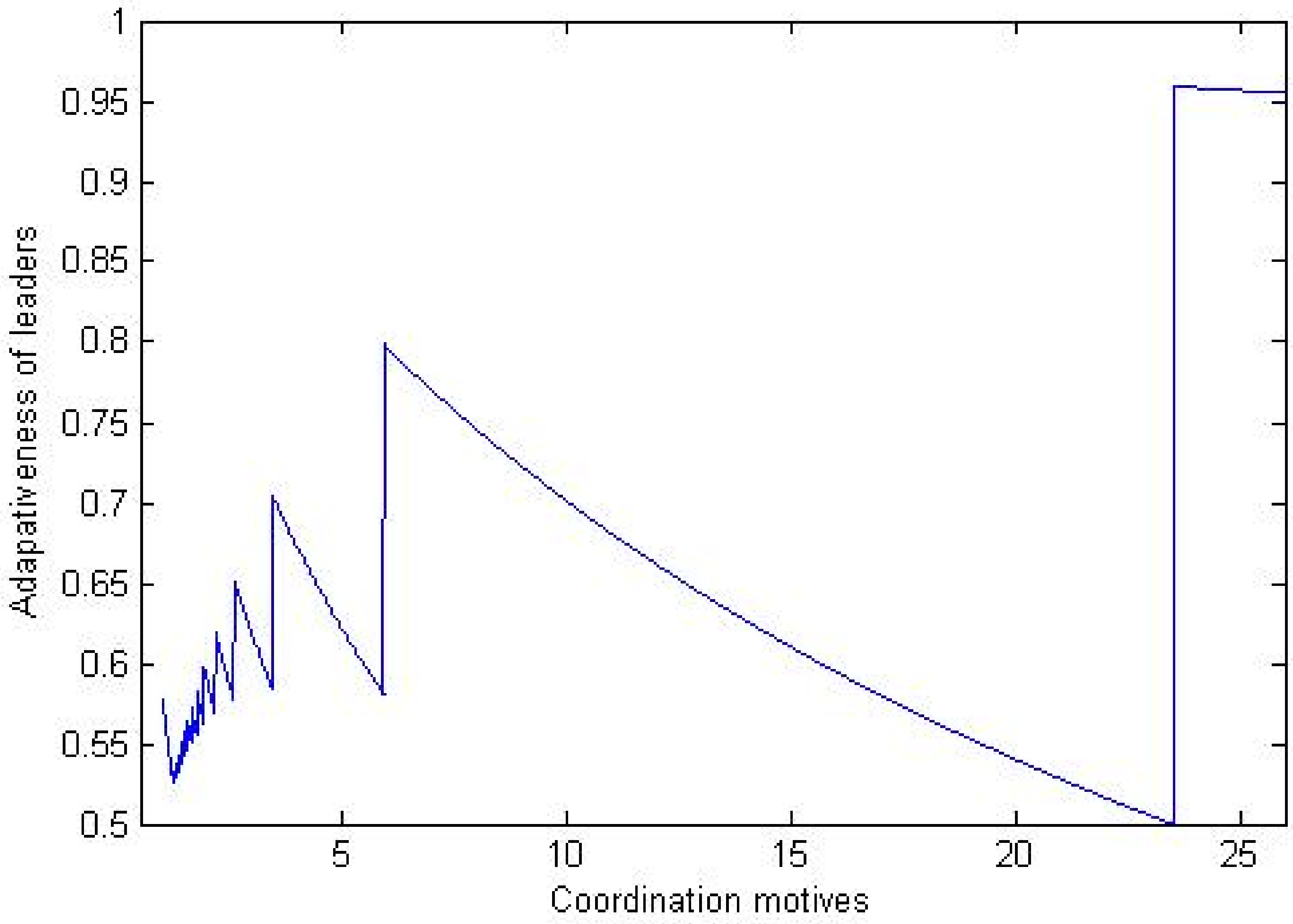

Adapativeness of leaders
Coordination motives
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
5
10
15
20
25

Figure **??** shows how the adaptiveness of each leader to his local shock ($\Omega$ as defined by expression 30) changes when tasks become more interdependent and mis-coordination costs are larger. Interestingly, leaders tend to be much more adaptive when coordination costs are higher, as they then share influence with less other leaders. For example, when $\beta = 25$, there is only one leader, but this leader is 50% more adaptive to his local information then when $\beta = 5$, and there are 3 leaders. In turn, those three leaders are much more adaptive than when $\beta$ is close to 1 and there are close to 20 leaders. Intuitively, for a given number of $\ell$ leaders, the adaptiveness of any given leader decreases as coordination becomes more important. But for $\ell < n$, this gradual decrease is more than compensated when $\beta$ passes the threshold $\overline{\beta}(\ell)$ and the number of leaders decreases to $\ell - 1$, resulting in a huge boost to the adaptiveness of the remaining leaders.

# 5 Extensions

To conclude, we extend our model in a variety of directions in order to shed a number of additional insights. For simplicity, we focus on the two-agent version of our model. We also maintain the assumption of decreasing marginal returns to attention, conveniently modelled as stemming from a finite (Shannon) attention or communication capacity.

## 5.1 Optimal division of labor and organizational structure.

A central feature of our model is that tasks are carried out by a *team* of agents. As shown in Section 3.3, without such division of labor, organizational focus is never optimal. This raises the question, however, what is the optimal degree of *labor specialization*? When is it optimal to have all tasks carried out by one agent rather than by a number of specialists?

Consistent with the literature (Becker and Murphy 1992, Bolton and Dewatripont 1994, Dessein and Santos 2006), the costs of the division of labor in our model is the need for coordination. Indeed, when tasks are carried out by a single agent, there are no coordination problems. As benefits of the division of labor, in turn, we posit that that specialized agents are better informed about task-specific shocks. Thus, a generalist in charge of both tasks can only learn about $\theta_1$ and $\theta_2$ by allocating attention to task 1 and task 2. In contrast, a specialist

in task $i$ directly observes $\theta_i$. This approach is similar to the one taken in Geanakoplos and Migrom (1991).

Let us therefore compare two organizations:

(i) Under *team production*, each task $i$ is implemented by a different specialist who perfectly observes the local shock $\theta_i$. In this organization, a finite communication capacity $\tau$ limits the ability of the team to coordinate actions. As we have done throughout this paper, we can distinguish between a *focussed team,* where all communication and attention is centered around one task, or a *balanced team,* where the communication network is symmetric and each task receives the same amount of attention.

(ii) Under *individual or centralized production*, both tasks are implemented by the same generalist who has a finite attention capacity $\tau$ to learn about those shocks, but faces no coordination problems. As shown in Section 3.3, this generalist optimally devotes an equal amount of attention to each task.

Intuitively, individual production performs well when coordination is important, just as a focussed team production but even more so. Individual production will perform very poorly, however, when attention is very constrained. Indeed, the ability to coordinate is of little benefit if the individual manager does not have the time to learn about the task-specific shocks. The benefit of task specialization and team production is exactly that specialized managers observe the local shocks of their individual tasks.

The impact of relaxing the attention constraint on the optimal division of labor is not trivial, however. On the one hand, more attention allows both for better coordination between specialized agents under team production. On the other hand, more attention allows for a generalist agent to learn more about the task-specific shocks affecting both tasks.

The following proposition shows that while a greater need for coordination favors both individual production and focussed team production – at the expense of balanced team production – limited attention unambiguously favors organizational focus. Generally speaking, individual production is more attractive as coordination becomes more important and attention is less constraint.

**Proposition 8** *There exists a* $\hat{\beta}$, $\bar{\beta} > \hat{\beta}$, *a* $\top(\beta)$ , $\hat{\top}(\beta)$, *and* $\bar{\top}(\beta)$ *such that*

1. *If* $\beta \leq \widehat{\beta}$ *then a balanced team is always optimal.*

2. *If* $\beta \in \left(\widehat{\beta}, \bar{\beta}\right)$*, then a focussed team is optimal for* $\tau \leq \top(\beta)$*, a balanced team for* $\tau \in \left(\top(\beta), \hat{\top}(\beta)\right)$ *, and individual/centralized production for* $\tau \geq \hat{\top}(\beta)$*.*

3. *If* $\beta \geq \bar{\beta}$*, then a focussed team is optimal for* $\tau < \bar{\top}(\beta)$ *and individual/centralized production if* $\tau \geq \bar{\top}(\beta)$*.*

4. *An increase in* $\beta$ *favors individual/centralized production:* $\hat{\top}(\beta)$ *and* $\hat{\top}(\beta)$ *are decreasing in* $\beta$*.*

One way to interpret the above result is in terms of centralized versus decentralized decision-making. The two specialized agents can be seen as division managers who perfectly observe the local shock affecting their division – the two tasks we have referred to so far. Under decentralized decision-making, these division managers take the relevant actions, and the communication capacity $\tau$ is used to ensure coordination. Under centralized decision-making, in contrast, both tasks are undertaken by a third manager, at headquarters. This headquarter manager does not observe the realization of the local shocks, but uses the communication capacity $\tau$ to communicate with both division managers and learn about them.

Proposition 4 then implies that centralized decision making is optimal if and only if coordination is sufficiently important *and* the headquarter's attention is not very constrained. While the first insight is well known from, say, Hart and Moore (2005), Alonso et al. (2008) and Rantakri (2008), the impact of attention capacity is novel in our view. Among other things, Proposition 4 then implies that information technologies, which relax attention constraints, should result in more centralized decision-making.

## 5.2 Endogenous attention capacity.

So far we have taken $\tau$ to be a hard constraint in the amount of time agents can devote to communication with each other. In practice this is another margin that organizations can use to improve performance, by, for example, allowing more time for meetings and communication between teams. Equivalently, the organization can increase the effective communication capacity $\tau$, by cross-training and rotating employees, by hiring employee with higher cognitive

abilities, or by investing in communication technology. Assume thus that an organization can acquire a capacity $\tau$ at a cost $C(\tau)$. $C(\tau)$ represents for example the costs of having team members engaged in communications activities rather than in production. We assume that this cost has the following properties:

$$C(0) = C'(0) = 0 \qquad C'(\tau) > 0 \qquad C''(\tau) \geq 0 \quad \text{and} \quad C'''(\tau) \geq 0.$$

The problem of organizational design is now

$$\max_{\tau, \mathbf{t}} E\pi(\mathbf{q}|\boldsymbol{\theta}) - C(\tau) \qquad \text{subject to} \quad (2). \tag{35}$$

**Proposition 9** *Assume that $\beta > \widehat{\beta}$, then*

1. *The optimal communication capacity $\tau^*$ is increasing in $\sigma_\theta^2$.*
2. *There exists a $\widehat{\sigma}_\theta^2 > 0$ such that $t_1^* \in \{0, \tau^*\}$ if, and only if, $\sigma_\theta^2 < \widehat{\sigma}_\theta^2$.*
3. *There is a discrete jump in the communication capacity $\tau^*$ at $\sigma_\theta^2 = \widehat{\sigma}_\theta^2$.*

When the coordination costs are beyond a threshold, more volatility results in more capacity though the company remains focused as long as the volatility is below a certain bound $\widehat{\sigma}_\theta^2$. The intuition is that when the volatility is below a certain bound the costs of not adapting to a particular task are small. A marginal increase in the volatility of the environment results in an increase in the capacity that gets fully allocated to communicate the realization of the particular task on which the organization is adaptive. Thus, in this case, an increase in the volatility reinforces organizational focus and leadership.

### 5.3 Technological trade-offs between adaptation and coordination

In our basic model, there is no trade-off between adaptation and coordination under perfect information. The need for coordination only constrains adaptation if information is dispersed among a group of agents, and communication is imperfect. Our insights, however, can be easily extended to models in which there is a always a trade-off between adaptation and coordination, even at the first-best.

A natural model in which is the case is one in which each agent $i$ only controls one action, $q_i$, which must be both adapted to some local information $\theta_i$ and coordinated with the actions $q_j$, $j \neq i$, undertaken by other agents. Naturally, a mechanical trade-off then arises between adapting one's action and coordinating it with actions undertaken by other. For $n = 2$, pay-offs are then equivalent those in the model considered in Alonso, Dessein, Matouschek (2008) and Rantakari (2008). For $n = 2$, pay-offs are identical to a (symmetric) version of the model considered in Calvo-Armengol et al. For conciseness, we consider the case for two agents, but everything can be generalazied to $n$ agents.

Formally, assume that each agent $i$ chooses an action $q_i$. Given a particular realization of the string of local information, $\boldsymbol{\theta} = [\theta_1, \theta_2]$, and a choice of actions, $\mathsf{q} = [q_1, q_2]$, the realized profit of the organization is:

$$\pi\left(\mathsf{q}|\boldsymbol{\theta}\right) = K - (q_1 - \theta_1)^2 - (q_2 - \theta_2)^2 - \beta(q_1 - q_2)^2, \tag{36}$$

where $\beta$ is some positive constant. A motivating example for the above payoff function are multi-national firms who sell similar products in different countries or regions. There are benefits from customizing products to local demand characteristics, but there are also gains from standardization of the product line. As in the model developed in the our paper, agent $i$ has information set $\mathcal{I}_i$ that contains local shock $\theta_i$ and a message $m_j$ about local shock $\theta_j$. The communication technology follows the description in our basic model.

We relegate the details of the analysis to the Appendix, but as in Section 3, one can show that expected profits can be expressed as

$$E\left[\pi\left(\mathsf{q}|\boldsymbol{\theta}\right)\right] = \left(\Omega\left(t_1\right) - 1\right)\sigma_\theta^2 + \left(\Omega\left(t_2\right) - 1\right)\sigma_\theta^2, \tag{37}$$

where

$$\Omega\left(t_i\right) = \frac{\operatorname{cov}\left(q_i\left(t_i\right), \theta_i\right)}{\sigma_\theta^2} = \frac{(1+\beta)\sigma_\theta^2}{\sigma_\theta^2(1+2\beta) + \beta^2 \mathsf{RV}(t_2)} \in [0,1] \tag{38}$$

captures the *adaptiveness of task $i$ to its task-specific information*. The only difference with Section 3 is that $q_i$ is less adaptive to the local information $\theta_i$, because of the technological trade-offs between adaptation and coordination. Using the monotone transformation $\tilde{\beta} = \beta^2/(1+2\beta)$, however, we can rewrite the problem of the organization designer as

$$\max_{t \in [0,\tau]} \frac{\sigma_\theta^2}{\sigma_\theta^2 + \tilde{\beta}\mathsf{RV}(t)} + \frac{\sigma_\theta^2}{\sigma_\theta^2 + \tilde{\beta}\mathsf{RV}(\tau - t)}$$

which is identical to the one studied in Section 3 . Propositions 1 and 2 of Section 2 therefore follow immediately.

# 6 Conclusions

To be completed.

## References


Alonso, Ricardo, Wouter Dessein and Niko Matouschek (2008)"When does Coordination Require Centralization?" *American Economic Review,* March 98 (1), 145-179

Arrow, Kenneth (1974) *The Limits of Organization.* W. W. Norton & Company, New York.

Bolton, Patrick and Mathias Dewatripont. 1994. "The Firm as a Communication Network." *Quarterly Journal of Economics* 109: 809-839.

Calvo-Armengol, Antonio, Joan de Marti and Andrea Prat (2011). "Communicaton and Influence." Mimeo.

Cover, Thomas M., and Joy A. Thomas (1991) *Elements of information theory.* John Wiley & Sons, New York.

Cyert, Richard M. and James G. March (1963) A Behavioral Theory of the Firm, Blackwell, Malden, MA.

Dessein, Wouter and Tano Santos (2006) "Adaptive Organization," *Journal of Political Economy,* 114 (50), 956-995.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy.* 108: 874-904.

Prahalad, C.K. and Gary Hamel. The core competence of the corporation. *Harvard Business Review*, 68, 79-91.

Rantakari, Heikki (2008) "Governing Adaptation," *Review of Economic Studies,* 75, 1257-1285.

Sims, Christopher A. (2003). "Implications of rational inattention." *Journal of Monetary Economics.* 50(3), 665-690.

Smith, Adam. [1776] (1981). *An Inquiry into theNature and Causes of the Wealth of Nations.* Indianapolis: Liberty Classics.

To be completed.

## Appendix

### Appendix A.

**Proof of Proposition 1.** Let $t_1 = t$ and $t_2 = \tau - t$; we consider, without loss of generality, that $t \in [0, \tau/2]$. Taking the derivative of the unconditional expected profit 12 with respect to $t$ we obtain

$$\frac{\partial E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial t} = -\beta\left[\Omega_1(t)\mathsf{RV}'\left(t\right) - \Omega_2(\tau - t)\mathsf{RV}'\left(\tau - t\right)\right]. \tag{39}$$

Substituting the expression for $\Omega_i(\cdot)$ given by 11, we have

$$\frac{\partial E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial t} = -\beta\left[\frac{\mathsf{RV}'(t)}{[\sigma_\theta^2 + \beta\mathsf{RV}(t)]^2} - \frac{\mathsf{RV}'(\tau - t)}{[\sigma_\theta^2 + \beta\mathsf{RV}(\tau - t)]^2}\right]. \tag{40}$$

Constant marginal returns to communication, i.e. $\mathsf{RV}''(\cdot) = 0$, implies that $\mathsf{RV}'(t) = \mathsf{RV}'(\tau - t)$. Moreover, since $\mathsf{RV}'(t) < 0$ and $t < \tau - t$, we have that $\sigma_\theta^2 + \beta\mathsf{RV}(t) > \sigma_\theta^2 + \beta\mathsf{RV}(\tau - t)$, for all $t \in [0, \tau/2]$. These two observations imply that if $\tau < \hat{\tau}$ then it is optimal to set $t = 0$; if $\tau > \hat{\tau}$, then it is optimal to set $t = \tau - \hat{\tau}$. This concludes the proof of Proposition 1. ■

**Proof of Proposition 2.** Recall that the derivative of the unconditional expected profit 12 with respect to $t$ is given by expression 40. Using that $\mathsf{RV}(t) = \sigma_\theta^2 e^{-2t}$, after some plain algebra it follows that

$$\frac{\partial E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial t} > 0 \Longleftrightarrow 1 - \beta^2 e^{-2\tau} > 0$$

Let $\hat{\beta} = 1$ and note that if $\beta \leq \hat{\beta}$ then $1 - \beta^2 e^{-2\tau} > 0$ for all $\tau \geq 0$; hence, optimality implies that $t = \tau/2$. Consider $\beta > \hat{\beta}$; define $\mathsf{T}(\beta)$ so that $1 - \beta^2 e^{-2\mathsf{T}(\beta)} = 0$. Note that $\mathsf{T}(\beta)$ is increasing in $\beta$. If $\tau < \mathsf{T}(\beta)$ then $1 - \beta^2 e^{-2\tau} < 0$ and therefore optimality implies that $t \in \{0, \tau\}$. If $\tau > \mathsf{T}(\beta)$ then $1 - \beta^2 e^{-2\tau} > 0$ and therefore optimality implies that $t = \tau/2$. This completes the proof of Proposition 2. ■

**Proof of Proposition 3.** Recall that an HQ chooses actions as in 21. Therefore, the optimal organization $(t, \tau - t)$ maximizes

$$E[\pi(\mathbf{q})|\theta] = -\left[\mathsf{RV}(t) + \mathsf{RV}(\tau - t)\right].$$

The derivative of $E[\pi(\mathbf{q})|\theta]$ with respect to $t$ is simply

$$\frac{\partial E[\pi(\mathbf{q})|\theta]}{\partial t} = -\left[\mathsf{RV}'(t) - \mathsf{RV}'(\tau - t)\right] \geq 0,$$

where the inequality follows because, by assumption, $\mathsf{RV}''(\cdot) \geq 0$ and $t < \tau - t$. Hence, the optimum is reached when $t = \tau/2$. This completes the proof of Proposition 3. ■.

**Proof of Proposition 4.** Proposition 4 follows as a consequence of the combination of Lemma 5 and Lemma 6. We now provide the proof of the two Lemmas.

**Proof of Lemma 5.** Suppose that $\mathbf{t}$ is optimal and, for a contradiction, assume that there exists some agent $i$ such that $t_{ji} > t_{ki} \geq 0$. Define a new organization $\mathbf{t}'$, which is the same as $\mathbf{t}$ with the exception that $t'_{ji} = t_{ji} - \epsilon$ and $t'_{ki} = t_{ki} + \epsilon$, for some small and positive $\epsilon$. Using the expression for expected payoffs 29 and the fact that $\mathsf{RV}(t_{sl}) = \sigma_\theta^2 e^{-2t_{sl}}$, it is easy to verify that

$$E\left[\pi\left(\mathbf{q}, \mathbf{t}|\boldsymbol{\theta}\right)\right] - E\left[\pi\left(\mathbf{q}, \mathbf{t}'|\boldsymbol{\theta}\right)\right] \geq 0,$$

if, and only if,

$$e^{-2t'_{ji}} + e^{-2t'_{ki}} \geq e^{-2t_{ji}} + e^{-2t_{ki}}.$$

Since $t'_{ji} = t_{ji} - \epsilon$ and $t'_{ki} = t_{ki} + \epsilon$, this condition is equivalent to

$$e^{-2t_{ki}} \leq e^{-2(t_{ji} - \epsilon)} \Longleftrightarrow t_{ki} \geq t_{ji} - \epsilon,$$

which, for $\epsilon$ sufficiently small, contradicts our initial hypothesis that $t_{ji} > t_{ki}$. The completes the proof of Lemma 5. ■

**Proof of Lemma 6.** Define $t_l = \sum_j t_{jl}$. Suppose that $\mathbf{t}$ is optimal and, for a contradiction, suppose that $t_i > t_j > 0$. Consider now two alternative organizations. One organization, denoted by $\mathbf{t}'$, is the same as organization $\mathbf{t}$, but $t'_i = t_i - \epsilon$ and $t'_j = t_j + \epsilon$. The second organization, denoted by $\hat{\mathbf{t}}$, is the same as organization $\mathbf{t}$, but $\hat{t}_i = t_i + \epsilon$ and $\hat{t}_j = t_j - \epsilon$. These constructions are derived for some small and positive $\epsilon$. Since the three organizations only differ in the way attention is distributed for task $i$ and task $j$, each other task $l \neq i, j$ performs

equally across the three organizations. We can then write

$$\begin{aligned}
E\left[\pi\left(\mathbf{q},\mathbf{t}|\boldsymbol{\theta}\right)\right] &= C-\sigma_\theta^2\left[\frac{1}{1+\beta e^{-2t_i}}+\frac{1}{1+\beta e^{-2t_j}}\right];\\
E\left[\pi\left(\mathbf{q},\mathbf{t}'|\boldsymbol{\theta}\right)\right] &= C-\sigma_\theta^2\left[\frac{1}{1+\beta e^{-2(t_i-\epsilon)}}+\frac{1}{1+\beta e^{-2(t_j+\epsilon)}}\right];\\
E\left[\pi\left(\mathbf{q},\hat{\mathbf{t}}|\boldsymbol{\theta}\right)\right] &= C-\sigma_\theta^2\left[\frac{1}{1+\beta e^{-2(t_i+\epsilon)}}+\frac{1}{1+\beta e^{-2(t_j-\epsilon)}}\right].
\end{aligned}$$

Since $\mathbf{t}$ is optimal, we must have that

$$E\left[\pi\left(\mathbf{q},\mathbf{t}|\boldsymbol{\theta}\right)\right]>E\left[\pi\left(\mathbf{q},\mathbf{t}'|\boldsymbol{\theta}\right)\right].$$

This is is equivalent to

$$\left[e^{-2t_j}-e^{-2(t_i-\epsilon)}\right]\left[\beta^2e^{-2(t_i+t_j)}-1\right]>0,$$

and, since $t_i>t_j$, for small $\epsilon$ we have that $e^{-2t_j}-e^{-2(t_i-\epsilon)}>0$ and therefore optimality of $\mathbf{t}$ requires that $\beta^2e^{-2(t_i+t_j)}-1>0$.

Similarly, since $\mathbf{t}$ is optimal, we must have that

$$E\left[\pi\left(\mathbf{q},\mathbf{t}|\boldsymbol{\theta}\right)\right]>E\left[\pi\left(\mathbf{q},\hat{\mathbf{t}}|\boldsymbol{\theta}\right)\right].$$

This is equivalent to

$$-\left[e^{-2(t_j-\epsilon)}-e^{-2t_i}\right]\left[\beta^2e^{-2(t_i+t_j)}-1\right]>0,$$

and, since $t_i>t_j$, we have that $e^{-2(t_j-\epsilon)}-e^{-2t_i}>0$, and therefore optimality of $\mathbf{t}$ requires that $\beta^2e^{-2(t_i+t_j)}-1<0$. We have then reached a contradiction. This completes the proof of Lemma 6. ■

The combination of Lemma 5 and Lemma 6 completes the proof of Proposition 4. ■

**Proof of Proposition 7**. Using the expression for expected payoffs 29, the fact that $\mathsf{RV}(t)=\sigma_\theta^2e^{-2t}$, and that organization $\mathbf{t}$ is an $\ell$-leader organization, we obtain that

$$\frac{dE\left[\pi\left(\mathbf{q},\mathbf{t}|\boldsymbol{\theta}\right)\right]}{d\ell}=\frac{\beta}{(1+\beta)\ell(n-1)\left(1+\beta e^{-\frac{2\tau}{(n-1)\ell}}\right)^2}\Phi(\ell,\beta,\tau,n),$$

where

$$\Phi(\ell,\beta,\tau,n) = \ell(n-1)\left[1-e^{-\frac{2\tau}{\ell(n-1)}}\right]\left[1+\beta e^{-\frac{2\tau}{\ell(n-1)}}\right] - 2\tau(\beta+1)e^{-\frac{2\tau}{\ell(n-1)}},$$

and that

$$\frac{d^2E\left[\pi\left(\mathbf{q},\mathbf{t}|\boldsymbol{\theta}\right)\right]}{d\ell d\ell} = -\frac{4\beta\tau^2 e^{-\frac{2\tau}{(n-1)\ell}}}{\ell^3(n-1)^2\left(1+\beta e^{-\frac{2\tau}{(n-1)\ell}}\right)^3}\left[1-\beta e^{-\frac{2\tau}{\ell(n-1)}}\right].$$

Observation 1. By direct verification, the function $\Phi(\ell,\beta,\tau,n)$ is decreasing in $\beta$ for all $\ell,\tau,n$. Note also that the sign of $\frac{dE[\pi(\mathbf{q},\mathbf{t}|\boldsymbol{\theta})]}{d\ell}$ is the same as the sign of $\Phi(\ell,\beta,\tau,n)$.

Denote by $\tilde{\beta}$ the solution to $1-\tilde{\beta}e^{-\frac{2\tau}{n(n-1)}}=0$. Also, denote by $\hat{\beta}$ the solution to $1-\hat{\beta}e^{-\frac{2\tau}{(n-1)}}=0$. Since $1-\beta e^{-\frac{2\tau}{\ell(n-1)}}$ is decreasing in $\beta$ and decreasing in $L$, the following observation follows:

Observation 2: 2a. $\tilde{\beta}<\hat{\beta}$ for all $\tau,n$; 2b. If $\beta<\tilde{\beta}$ then $\frac{d^2E[\pi(\mathbf{q},\mathbf{t}|\boldsymbol{\theta})]}{d\ell d\ell}<0$ for all $\ell$; 2c. If $\beta>\hat{\beta}$ then $\frac{d^2E[\pi(\mathbf{q},\mathbf{t}|\boldsymbol{\theta})]}{d\ell d\ell}>0$ for all $\ell$.

We now show that there exists a $\underline{\beta}(\tau,n)>0$ such that for all $\beta<\underline{\beta}(\tau,n)$ the number of leaders in the optimal organization is $\ell=n$. Denote by $\underline{\beta}(\tau,n)$ the solution to $\Phi(n,\underline{\beta}(\tau,n),x,n)=0$. Explicitly,

$$\underline{\beta}(\tau,n) = \frac{n(n-1)\left(1-e^{-\frac{2\tau}{n(n-1)}}\right)-2\tau e^{-\frac{2\tau}{n(n-1)}}}{2\tau-n(n-1)\left(1-e^{-\frac{2\tau}{n(n-1)}}\right)}\tilde{\beta}.$$

Observation 3: Direct verification implies 3a. $\underline{\beta}(\tau,n)<\tilde{\beta}$ for all $\tau,n$; 3b. $\underline{\beta}(\tau,n)$ is increasing in $\tau$.

Observation 3a together with observation 2b imply that $\frac{dE[\pi(\mathbf{q},\mathbf{t}|\boldsymbol{\theta})]}{d\ell}$ is declining in $\ell$ for all $\beta<\underline{\beta}(\tau,n)$. So, for all $\beta<\underline{\beta}(\tau,n)$, the lower value of $\frac{dE[\pi(\mathbf{q},\mathbf{t}|\boldsymbol{\theta})]}{dL}$ is obtained when $\ell=n$, and, at $\ell=n$ we have

$$\frac{dE\left[\pi\left(\mathbf{q},\mathbf{t}|\boldsymbol{\theta}\right)\right]}{d\ell}|_{\ell=n} = \frac{\beta}{(1+\beta)n(n-1)\left(1+\beta e^{-\frac{2\tau}{(n-1)n}}\right)^2}\Phi(n,\beta,\tau,n)>0,$$

because, by observation 1, $\Phi(n,\beta,\tau,n)>\Phi(n,\underline{\beta}(\tau,n),\tau,n)$, and, by definition, $\Phi(n,\underline{\beta}(\tau,n),\tau,n)=0$. Hence, for all $\beta<\underline{\beta}(\tau,n)$ the expected returns of an $\ell$-leader organisation are increasing in the number of leaders, which implies that the optimal organization has $\ell^*=n$ leaders.

Next, observation 3b together with the observation that $\lim_{\tau \to 0} \underline{\beta}(\tau, n) = 1$, imply that for all $\beta < 1$, the optimal organization has $\ell^* = n$ leaders, regardless of the level of $\tau$.

We now show that there exists a $\bar{\beta}(\tau, n) > \underline{\beta}(\tau, n)$ such that for all $\beta > \bar{\beta}(\tau, n)$ in the optimal organization the number of leaders is $\ell^* = 1$. Denote by $\bar{\beta}(\tau, n)$ the solution to $\Phi(1, \bar{\beta}(\tau, n), \tau, n) = 0$. Explicitly

$$\bar{\beta}(\tau, n) = \frac{(n-1)\left(1 - e^{-\frac{2\tau}{(n-1)}}\right) - 2\tau e^{-\frac{2\tau}{(n-1)}}}{2\tau - (n-1)\left(1 - e^{-\frac{2\tau}{(n-1)}}\right)} \hat{\beta}.$$

Observation 4: Direct verification shows that: 4a. $\tilde{\beta} < \bar{\beta}(\tau, n) < \hat{\beta}$, for all $\tau$ and $n$; 4b. $\bar{\beta}(\tau, n)$ is increasing in $\tau$.

Observation 1 together with $\Phi(1, \bar{\beta}(\tau, n), \tau, n) = 0$ imply that $\Phi(1, \beta, \tau, n) < 0$ for all $\beta > \bar{\beta}(\tau, n)$. Similarly, observation 1 together with $\Phi(n, \underline{\beta}(\tau, n), \tau, n) = 0$ and observation 4a, imply that $\Phi(n, \beta, \tau, n) < 0$ for all $\beta > \bar{\beta}(\tau, n)$. So, $\frac{dE[\pi(\mathbf{q}, \mathbf{t}|\boldsymbol{\theta})]}{d\ell}$ is negative at $\ell = 1$ and at $\ell = n$. Observation 4a and observation 2b implies that $\frac{dE[\pi(\mathbf{q}, \mathbf{t}|\boldsymbol{\theta})]}{d\ell}$ is either first decreasing in $\ell$ and then increasing in $\ell$ (when $\beta \in [\bar{\beta}(\tau, n), \hat{\beta}]$) or it is always increasing in $\ell$ (when $\beta > \hat{\beta}]$). Hence, the profits of the organization are decreasing in $\ell$ for all $\beta > \bar{\beta}(\tau)$ and therefore the optimal organization has $\ell^* = 1$ leader.

We now conclude by considering the case where $\beta \in (\underline{\beta}(\tau, n), \bar{\beta}(\tau, n))$. From the analysis above we infer that the marginal expected profits to $\ell$ of the organization around $\ell = 1$ are positive, because $\Phi(1, \beta, \tau, n) > 0$, and that the marginal expected profits of the organization around $\ell = n$ are negative, because $\Phi(n, \beta, \tau, n) < 0$. Furthermore, observation 2b implies that, for all $\beta \in (\underline{\beta}(\tau, n), \bar{\beta}(\tau, n))$, the marginal expected profits of the organization, $\frac{dE[\pi(\mathbf{q}, \mathbf{t}|\boldsymbol{\theta})]}{d\ell}$, are either always decreasing in $\ell$ (when $\beta \in [\underline{\beta}(\tau, n), \tilde{\beta}]$) or they are first decreasing in $\ell$ and then increasing in $\ell$ (when $\beta \in [\tilde{\beta}, \bar{\beta}(\tau, n)]$). Hence, there exists a unique $\ell^* \in [1, n]$ such that $\frac{dE[\pi(\mathbf{q}, \mathbf{t}|\boldsymbol{\theta})]}{d\ell}|_{\ell = \ell^*} = 0$; such value of $\ell^*$ is the solution to $\Phi(\ell^*, \beta, x, n) = 0$ and, $\ell^*$ maximises the expected profit of the organization. Finally, by applying the implicit function theorem, $d\ell^*/d\beta < 0$ if and only if $d\Phi(\ell^*, \beta, \tau, n)/dL < 0$. Note that this last inequality holds because the fact that there exists a unique $\ell^*$ in which $\Phi(\ell^*, \beta, \tau, n) = 0$ and the fact that $\Phi(1, \beta, \tau, n) > 0$

and $\Phi(n, \beta, \tau, n) < 0$, assure that for all $\beta \in (\underline{\beta}(\tau, n), \bar{\beta}(\tau, n))$ the function $\Phi(\ell, \beta, \tau, n)$ is decreasing around $\ell^*$.

We have therefore shown that for every $\ell \in \{1, ..., n-1\}$ there exists a $\beta(\ell+1) < \beta(\ell)$ such that: a. if $\beta = \beta(\ell+1)$ the optimal organization has $\ell^* = \ell+1$ leaders; b. if $\beta \in (\beta(\ell+1), \beta(\ell))$ the optimal organization has either $\ell^* = \ell$ leaders or $\ell^* = \ell+1$ leaders, and c. if $\beta = \beta(\ell)$ the optimal organization has $\ell^* = \ell$ leaders. We now show that the optimal number of leaders $\ell^*$ is increasing in $\beta$, which, in view of the above analysis, amounts in showing that, for every $\ell \in \{1, ..., n-1\}$ there exists a unique value of $\beta \in (\beta(\ell+1), \beta(\ell))$, say $\beta_\ell$, such that at $\beta = \beta_\ell$ the expected profit of the $\ell$-leader organization is the same as the expected profit of the $\ell+1$-leader organization. This is what we show next.

For brevity define $\hat{RV}(x) = e^{-\frac{2\tau}{(n-1)x}}$ and denote by $\Delta(\ell, \beta)$ the difference between the expected profit generated by the $\ell+1$-leader organization and the expected profit generated by the $\ell$-leader organization. Using expression 32, we obtain

$$\Delta(\ell, \beta) = \sigma_\theta^2 \left[ \frac{\ell+1}{1+\beta \hat{RV}(\ell+1)} - \frac{\ell}{1+\beta \hat{RV}(\ell)} - \frac{1}{1+\beta} \right].$$

Taking the minimum common denominator, we have that $\Delta(\ell, \beta) = 0$ if, and only if,

$$(1+\beta)\left[(\ell+1)(1+\beta \hat{RV}(\ell)) - \ell(1+\beta \hat{RV}(\ell+1))\right] - [1+\beta \hat{RV}(\ell)][1+\beta \hat{RV}(\ell+1)] = 0.$$

This is a quadratic equation in $\beta$ and therefore there are only two solutions of $\beta$. Moreover, it is immediate to check that $\beta = 0$ is one of the solution. Hence, there is only one non-zero solution. We have therefore completed the proof of the first part of proposition 7.

To complete the proof of the proposition, we show that, for every $\ell \in \{1, ..., n-1\}$, the cut off $\beta_{\ell+1}$ is increasing in $\tau$. Define $t = 2\tau/(n-1)$, then the cut off $\beta_{\ell+1}$ is the (non-zero) solution of

$$(1+\beta)\left((\ell+1)(1+\beta e^{-\frac{t}{\ell}}) - \ell(1+\beta e^{-\frac{t}{\ell+1}})\right) - \left(1+\beta e^{-\frac{t}{\ell+1}}\right)\left(1+\beta e^{-\frac{t}{\ell}}\right) = 0,$$

which, after some algebra, is

$$\beta_{\ell+1} = \frac{e^{\frac{t}{\ell+1}} + \ell e^{-\frac{t}{\ell(\ell+1)}} - (1+\ell)}{\ell + e^{-\frac{t}{\ell}} - (1+\ell)e^{-\frac{t}{\ell(\ell+1).}}}$$

Note that nominator is increasing in $t$ because

$$\frac{d\left(\ell e^{-\frac{t}{\ell(\ell+1)}}+e^{\frac{t}{\ell+1}}\right)}{dt}=\frac{1}{\ell+1}\left(e^{\frac{t}{\ell+1}}-e^{-\frac{t}{\ell^2+\ell}}\right)<0,$$

whereas the denominator is decreasing in $t$ because

$$\frac{d\left(e^{-\frac{t}{\ell}}-(1+\ell)e^{-\frac{t}{\ell(\ell+1)}}\right)}{dt}=-\frac{1}{\ell}\left(e^{-\frac{t}{\ell}}-e^{-\frac{t}{\ell^2+\ell}}\right)<0.$$

It follows that

$$\frac{d\beta_{\ell+1}}{d\tau}>0.$$

Note further that

$$\lim_{\tau\to\infty}\beta_{\ell+1}=\lim_{\tau\to\infty}\frac{1}{\ell}e^{\frac{t}{\ell+1}}=+\infty$$

This concludes the proof of proposition 7.■

**Proof of Proposition 4:** From (22), (12) and (18), balanced team production dominates individual production if and only if

$$\begin{aligned}2\beta\Omega(\tau/2)\mathsf{RV}\,(\tau/2) &\leq \mathsf{2RV}(\tau/2)\\ &\iff \mathsf{e}^{-\tau}\geq\frac{\beta-1}{\beta}\\ &\iff \tau\leq\hat{\mathsf{T}}(\beta)=\ln\frac{\beta}{\beta-1}.\end{aligned}$$

We further have already established that focused team production dominates balanced team production if and only if

$$\tau\leq\mathsf{T}(\beta)=\ln\beta. \tag{41}$$

If $\beta<\bar{\beta}=2$, then $\mathsf{T}(\beta)<\hat{\mathsf{T}}(\beta)$. Part (1) and (2) of proposition 4 then follow directly, where $\widehat{\beta}=1$. If $\beta\geq\bar{\beta}=2$, then $\mathsf{T}(\beta)>\hat{\mathsf{T}}(\beta)$, and balanced team production is always dominated. We then have that if $\tau<\hat{\mathsf{T}}(\beta)$, then focussed team production is optimal, and if $\tau>\mathsf{T}(\beta)$ then centralised production is optimal. If $\tau\in(\hat{\mathsf{T}}(\beta),\mathsf{T}(\beta))$, focussed team production dominates individual production if, and only if,

$$\frac{\beta}{1+\beta}+\frac{\beta}{1+\beta\mathsf{e}^{-2\tau}}\mathsf{e}^{-2\tau}-2\mathsf{e}^{-\tau}<0. \tag{42}$$

By evaluating the LHS of the expression above at $\tau = \hat{\mathsf{T}}(\beta)$ and at $\tau = \mathsf{T}(\beta)$ we can verify that at $\tau = \hat{\mathsf{T}}(\beta)$ the inequality holds (team production dominates individual production) whereas at $\tau = \mathsf{T}(\beta)$ the reverse holds. To conclude, we then note that the derivative of the LHS with respect to $\tau$ is

$$\frac{2e^{-\tau}}{(1+\beta e^{-2\tau})^2}\left[(1+\beta e^{-2\tau})^2 - \beta e^{-\tau}\right] > 0,$$

where the inequality follows because $\beta > \bar{\beta}$ and $\tau \in (\hat{\mathsf{T}}(\beta), \mathsf{T}(\beta))$. Hence, there exists a $\bar{\mathsf{T}}(\beta)$ so that if $\tau < \bar{\mathsf{T}}(\beta)$ focussed team production is optimal, otherwise centralised production is optimal. This implies part (3) and (4) of proposition 4. ■

**Proof of Proposition ??.** We prove each of the three parts of the proposition.

**First part.** We first show that the optimal capacity $\tau^*$ is increasing in $\sigma_\theta^2$ in the centralized organization and in the decentralized organization. This, together with Proposition 2, implies the first part of Proposition **??**: the optimal capacity $\tau^*$ is increasing in $\sigma_\theta^2$.

We consider the centralised organization first. Recall that the expected profits in the centralized organization are

$$E\left[\pi^c\left(\mathbf{q}|\boldsymbol{\theta}\right)\right] = -\beta\sigma_\theta^2\left[\frac{1}{1+\beta} + \frac{e^{-2\tau}}{1+\beta e^{-2\tau}}\right] - C(\tau).$$

Taking the derivative with respect to $\tau$ we have

$$\frac{\partial E\left[\pi^c\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial \tau} = \frac{2\beta\sigma_\theta^2 e^{-2\tau}}{\left[1+\beta e^{-2\tau}\right]^2} - C'(\tau).$$

We now observe that, since $C'(0) = 0$, it follows that $\frac{\partial E[\pi^c(\mathbf{q}|\boldsymbol{\theta})]}{\partial \tau}|_{\tau=0} > 0$, and that, since $C'(\cdot) > 0$, it follows that $\frac{\partial E[\pi^c(\mathbf{q}|\boldsymbol{\theta})]}{\partial \tau}|_{\tau=\infty} < 0$. Moreover

$$\frac{\partial^2 E\left[\pi^c\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial \tau \partial \tau} = -\left[\frac{4\beta\sigma_\theta^2 e^{-2\tau}}{\left[1+\beta e^{-2\tau}\right]^3}\left[1-\beta e^{-2\tau}\right] + C''(\tau)\right].$$

Since $C''(\cdot) \geq 0$ and since $1 - \beta e^{-2\tau}$ is negative for small value of $\tau$ (recall that $\beta > \hat{\beta} = 1$) and, as $\tau$ increases, $1 - \beta e^{-2\tau}$ becomes eventually positive, it follows that $\frac{\partial^2 E[\pi^c(\mathbf{q}|\boldsymbol{\theta})]}{\partial \tau \partial \tau}$ is either negative for all $\tau > 0$, or it is positive for small value of $\tau$ and negative otherwise. Summarizing, we have shown that the function $\frac{\partial E[\pi^c(\mathbf{q}|\boldsymbol{\theta})]}{\partial \tau}$ is (i) positive at $\tau = 0$, (ii) negative at $\tau = \infty$

and (iii) it is either decreasing in $\tau$ or it is first increasing and then decreasing in $\tau$. As a consequence of (i)-(ii) we obtain that the optimal capacity $\tau^c$ uniquely solves

$$\frac{\partial E\left[\pi^c\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial \tau} = \frac{2\beta\sigma_\theta^2 e^{-2\tau^c}}{[1+\beta e^{-2\tau^c}]^2} - C'^c) = 0.$$

Since $\frac{\partial E[\pi^c(\mathbf{q}|\boldsymbol{\theta})]}{\partial \tau}$ is increasing in $\sigma_\theta^2$ and since, from above, $\frac{\partial^2 E[\pi^c(\mathbf{q}|\boldsymbol{\theta})]}{\partial \tau \partial \tau}|_{\tau=\tau^c} < 0$, an application of the implicit function theorem implies that $\tau^c$ is an increasing function of $\sigma_\theta^2$. From investigation of the optimality condition of $\tau^c$ and the assumptions that $C'(0) = 0$, it follows that $\tau^c \to 0$ as $\sigma_\theta^2 \to 0$ and that $\tau^c \to \infty$ as $\sigma_\theta^2 \to \infty$.

We now consider the case in which the organization is decentralized. The expected profits in the decentralized organization are

$$E\left[\pi^d\left(\mathbf{q}|\boldsymbol{\theta}\right)\right] = -\frac{2\beta\sigma_\theta^2 e^{-\tau}}{1+\beta e^{-\tau}} - C(\tau).$$

Taking the derivative with respect to $\tau$ we obtain

$$\frac{\partial E\left[\pi^d\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial \tau} = \frac{2\beta\sigma_\theta^2 e^{-\tau}}{[1+\beta e^{-\tau}]^2} - C'(\tau).$$

We can now proceed in the same fashion as in the case for the centralized organization to conclude that the optimal capacity $\tau^d$ uniquely solves

$$\frac{\partial E\left[\pi^d\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial \tau} = \frac{2\beta\sigma_\theta^2 e^{-\tau^d}}{\left[1+\beta e^{-\tau^d}\right]^2} - C'^d) = 0,$$

and that $\tau^d$ is an increasing function of $\sigma_\theta^2$, $\tau^d \to 0$ as $\sigma_\theta^2 \to 0$ and $\tau^d \to \infty$ as $\sigma_\theta^2 \to \infty$.

Since the optimal capacity in the centralized and decentralised organization are both increasing in $\sigma_\theta^2$ and since, by Proposition 2, the optimal organization is either centralized or decentralized, it follows that the optimal capacity of the optimal organization is increasing in $\sigma_\theta^2$.

**Second part.** I PROVE THAT: There exists a $0 < \underline{\sigma}_\theta^2 < \bar{\sigma}_\theta^2$ such that for all $\sigma_\theta^2 < \underline{\sigma}_\theta^2$ the optimal organization is $t^* \in \{0, \tau\}$, whereas for all $\sigma_\theta^2 > \bar{\sigma}_\theta^2$ the optimal organization is $t^* = \tau/2$. NOTE THAT THIS IS A WEAKER CLAIM OF WHAT WE HAVE IN THE PROPOSITION WHERE WE ALSO CLAIM THAT $\underline{\sigma}_\theta^2 = \bar{\sigma}_\theta^2$.

First note that for a given common $\tau$

$$\frac{\partial E\left[\pi^{c}\left(\mathbf{q},\tau|\boldsymbol{\theta}\right)\right]}{\partial\tau}-\frac{\partial E\left[\pi^{d}\left(\mathbf{q},\tau|\boldsymbol{\theta}\right)\right]}{\partial\tau}>0,$$

if, and only if,

$$\frac{e^{-2\tau}}{[1+\beta e^{-2\tau}]^{2}}-\frac{e^{-\tau}}{[1+\beta e^{-\tau}]^{2}}>0,$$

and, after plain algebra, this condition is equivalent to

$$-\left[e^{-\tau}-e^{-2\tau}\right]\left[1-\beta^{2}e^{-3\tau}\right]>0 \qquad\iff\qquad 1-\beta^{2}e^{-3\tau}<0.$$

Since $\tau^{c}(\sigma_{\theta}^{2})$ is increasing in $\sigma_{\theta}^{2}$ ranging from 0 to $\infty$, there exists a unique $\hat{\sigma}_{\theta}^{2}$ that solves $1-\beta^{2}e^{-3\tau^{c}(\hat{\sigma}_{\theta}^{2})}=0$. By construction, if $\sigma_{\theta}^{2}=\hat{\sigma}_{\theta}^{2}$, then $\tau^{c}(\hat{\sigma}_{\theta}^{2})=\tau^{d}(\hat{\sigma}_{\theta}^{2})$. The next observation is used in the rest of the proof.

Observation 1. If $\sigma_{\theta}^{2}<\hat{\sigma}_{\theta}^{2}$ then $\tau^{c}(\sigma_{\theta}^{2})<\tau^{c}(\hat{\sigma}_{\theta}^{2})$. If $\sigma_{\theta}^{2}>\hat{\sigma}_{\theta}^{2}$ then $\tau^{c}(\sigma_{\theta}^{2})>\tau^{c}(\hat{\sigma}_{\theta}^{2})$.

To see the first part note that since $\tau^{c}$ is increasing in $\sigma_{\theta}^{2}$, it follows that $1-\beta^{2}e^{-3\tau^{c}(\sigma_{\theta}^{2})}<0$ for all $\sigma_{\theta}^{2}<\hat{\sigma}_{\theta}^{2}$. Hence, $\frac{\partial E\left[\pi^{d}(\mathbf{q}|\boldsymbol{\theta})\right]}{\partial\tau}|_{\tau^{c}(\sigma_{\theta}^{2})}<0$, which implies that $\tau^{d}(\sigma_{\theta}^{2})<\tau^{c}(\sigma_{\theta}^{2})$. Analogously, since $\tau$ is increasing in $\sigma_{\theta}^{2}$, it follows that $1-\beta^{2}e^{-3\tau^{c}(\sigma_{\theta}^{2})}>0$ for all $\sigma_{\theta}^{2}>\hat{\sigma}_{\theta}^{2}$. Hence, $\frac{\partial E\left[\pi^{d}(\mathbf{q}|\boldsymbol{\theta})\right]}{\partial\tau}|_{\tau^{c}(\sigma_{\theta}^{2})}>0$, which implies that $\tau^{d}(\sigma_{\theta}^{2})>\tau^{c}(\sigma_{\theta}^{2})$.

Define now $\underline{\sigma}_{\theta}^{2}$ as the solution to $1-\beta^{2}e^{-2\tau^{d}(\underline{\sigma}_{\theta}^{2})}=0$ and define $\bar{\sigma}_{\theta}^{2}$ be such that $1-\beta^{2}e^{-2\tau^{c}(\bar{\sigma}_{\theta}^{2})}=0$.

We now show that $\underline{\sigma}_{\theta}^{2}>\hat{\sigma}_{\theta}^{2}$. By definition of $\hat{\sigma}_{\theta}^{2}$ and $\underline{\sigma}_{\theta}^{2}$, we have that

$$1-\beta^{2}e^{-3\tau^{d}(\hat{\sigma}_{\theta}^{2})}=0=1-\beta^{2}e^{-2\tau^{d}(\underline{\sigma}_{\theta}^{2})},$$

which implies that $\tau^{d}(\underline{\sigma}_{\theta}^{2})>\tau^{d}(\hat{\sigma}_{\theta}^{2})$, and since $\tau^{d}$ is increasing in $\sigma_{\theta}^{2}$ it follows that $\underline{\sigma}_{\theta}^{2}>\hat{\sigma}_{\theta}^{2}$.

We now show that $\bar{\sigma}_{\theta}^{2}>\underline{\sigma}_{\theta}^{2}$. By definition of $\bar{\sigma}_{\theta}^{2}$ and $\underline{\sigma}_{\theta}^{2}$ we have that

$$1-\beta^{2}e^{-2\tau^{d}(\underline{\sigma}_{\theta}^{2})}=0=1-\beta^{2}e^{-2\tau^{c}(\bar{\sigma}_{\theta}^{2})},$$

which implies that $\tau^{d}(\underline{\sigma}_{\theta}^{2})=\tau^{c}(\bar{\sigma}_{\theta}^{2})$. Since $\underline{\sigma}_{\theta}^{2}>\hat{\sigma}_{\theta}^{2}$ and since $\tau^{d}(\sigma_{\theta}^{2})>\tau^{c}(\sigma_{\theta}^{2})$ for all $\sigma_{\theta}^{2}>\hat{\sigma}_{\theta}^{2}$, we have that $\tau^{d}(\underline{\sigma}_{\theta}^{2})>\tau^{c}(\underline{\sigma}_{\theta}^{2})$. Hence, in order for $\tau^{d}(\underline{\sigma}_{\theta}^{2})=\tau^{c}(\bar{\sigma}_{\theta}^{2})$ to hold we must have that $\bar{\sigma}_{\theta}^{2}>\underline{\sigma}_{\theta}^{2}$.

We now complete the proof of the second part of Proposition **??**. If $\sigma_\theta^2 \leq \underline{\sigma}_\theta^2$, then $1 - \beta^2 e^{-2\tau^d(\sigma_\theta^2)} \leq 0$ and $1 - \beta^2 e^{-2\tau^c(\sigma_\theta^2)} < 0$. From Proposition 2 we know that for all $\tau$ such that $1 - \beta^2 e^{-2\tau} \leq 0$ the optimal organization is centralized. Hence, if $\sigma_\theta^2 \leq \underline{\sigma}_\theta^2$ the optimal organization is centralized. Finally, if $\sigma_\theta^2 \geq \bar{\sigma}_\theta^2$, then $1-\beta^2 e^{-2\tau^c(\sigma_\theta^2)} \geq 0$ and $1-\beta^2 e^{-2\tau^d(\sigma_\theta^2)} > 0$ and therefore, in view of Proposition 2, it follows that the decentralized organization is optimal.

**Third part.** The prove of the second part of Proposition **??** implies that the organization changes from centralized to decentralized for some values of $\sigma_\theta^2 \in (\underline{\sigma}_\theta^2, \bar{\sigma}_\theta^2)$ and, $\tau^c(\sigma_\theta^2) \neq \tau^d(\sigma_\theta^2)$ for every $\sigma_\theta^2 \in (\underline{\sigma}_\theta^2, \bar{\sigma}_\theta^2)$. These two observations imply the last part of Proposition **??**. ■

**Appendix B: Technological trade-offs between adaptation and coordination**

In this Appendix we show that our insights hold in a model of coordination a la' Alonso, Dessein, Matouschek (2008) and Rantakari (2008). We consider the case for two agents, but everything can be generalazied to $n$ agents. In these class of models, instead of having the distinction between primary action and complementary action, each agent chooses one single action. We posit that agent $i$ chooses $q_i$. Given a particular realization of the string of local information, $\boldsymbol{\theta} = [\theta_1, \theta_2]$, and a choice of actions, $\mathbf{q} = [q_1, q_2]$, the realized profit of the organization is:

$$\pi(\mathbf{q}|\boldsymbol{\theta}) = K - (q_1 - \theta_1)^2 - (q_2 - \theta_2)^2 - \beta(q_1 - q_2)^2, \tag{43}$$

where $\beta$ is some positive constant. A motivating example for the above payoff function are multi-national firms who sell similar products in different countries or regions. There are benefits from customizing products to local demand characteristics, but there are also gains from standardization of the product line. This trade-off has been widely discussed in the case studies and the literature on multi-national cooperations. As in the model developed in the our paper, agent $i$ has information set $\mathcal{I}_i$ that contains local shock $\theta_i$ and a message $m_j$ about local shock $\theta_j$. The communication technology follows the description in our basic model.

Standard computation allows us to derive agents' best replies, for a given network $\mathbf{t} = (t, \tau - t)$. We obtain:

$$q_1 = \frac{1}{1+\beta}\left[\theta_1 + \beta E\left[q_2|\mathcal{I}_1\right]\right] \tag{44}$$

$$q_2 = \frac{1}{1+\beta}\left[\theta_2 + \beta E\left[q_1|\mathcal{I}_2\right]\right] \tag{45}$$

$$\tag{46}$$

We focus on characterizing equilibria in linear strategies. This is without loss of generality for the two leading examples of communication technologies. We can write (**??**) and (**??**) as

$$q_1 = a_{11}(t_1)\,\theta_1 + a_{12}(t_2)E[\theta_2|\mathcal{I}_1] \tag{47}$$

$$q_2 = a_{22}(t_2)\,\theta_1 + a_{21}(t_1)E[\theta_1|\mathcal{I}_2] \tag{48}$$

Substituting the guess (47) and (48) into (**??**) and (44), and using Assumption A, we find that the equilibrium actions are

$$q_1 = \frac{(1+\beta)\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(t_1)}\theta_1 + \frac{\beta\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(t_2)}E[\theta_2|\mathcal{I}_1] \tag{49}$$

$$q_2 = \frac{(1+\beta)\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(t_2)}\theta_2 + \frac{\beta\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(t_1)}E[\theta_1|\mathcal{I}_2] \tag{50}$$

Finally substituting (49) and (50) into (43) and taking unconditional expectations we find that the problem

$$\max_{\mathbf{t}} E\pi(\mathbf{q}|\boldsymbol{\theta}) \ s.t. t_1 + t_2 = \tau$$

is equivalent to

$$\max_{\mathbf{t}} Cov(q_1,\theta_1) + Cov(q_2,\theta_2) \ s.t. t_1 + t_2 = \tau.$$

Defining $t_1 = t$ and $t_2 = \tau - t$, and using the equilibrium action to derive the respective covariates, the problem of the designer is

$$\max_{t\in[0,\tau]} \frac{\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(t)} + \frac{\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(\tau-t)}$$

It is easy to replicate the analysis we have performed in section 3. First, when there are constant returns to communication, the same argument used in the proof of Proposition 1 applies in this new specification. Hence, under constant returns to communication the optimal organization focuses on one task.

Consider now decreasing returns to communication modelled as in section 3.3. That is $\mathsf{RV}(t) = \sigma_\theta^2 e^{-2t}$. Similarly to the proof of proposition 3, it is easy to verify that

$$\frac{\partial E\pi(\mathbf{q}|\boldsymbol{\theta})}{\partial t} > 0 \iff (1+2\beta)^2 - \beta^4 e^{-2\tau} > 0.$$

We then obtain a result that is qualitatively the same as the one stated in Proposition 3. For every $\tau$ there exists a $\beta(\tau) > 0$, so that for all $\beta < \beta(\tau)$ the optimal organization has $t = \tau/2$, whereas for every $\beta > \beta(\tau)$ the optimal organization has $t = \{0, \tau\}$. Furthermore, $\beta(\tau)$ is increasing in $\tau$.