Rational Inattention and Organizational Focus.

Wouter Dessein
Columbia University

Andrea Galeotti
University of Essex

Tano Santos
Columbia University

August 30, 2012

Abstract

We examine the allocation of scarce attention in team production. Each team member is in charge of a specialized task, which must be adapted to a privately observed shock and coordinated with other tasks. Coordination requires that agents pay attention to each other, but attention is in limited supply. We show how organizational focus and leadership naturally arise as the result of a fundamental complementarity between the attention devoted to an agent and the amount of initiative taken by that agent (the adaptiveness of his task). At the optimum, all attention is evenly allocated to a select number of focal tasks or "leaders". When attention becomes more scarce, there are fewer leaders, but they often take more initiative. The organization then excels in a small number of focal tasks at the expense of all others. Our results shed light on the importance of leadership, strategy and "core competences" in team production, as well as new trends in organization design.

1 Introduction

Ever since Adam Smith's "Wealth of Nations," ([1776] 1981), the importance of specialization has been a central idea in economics. As argued by Smith, 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 developing specialized knowledge, or how to define such organizational knowledge. The management literature, in contrast, has long argued that firms should focus and nurture a limited set of "core competences" (Prahald and Hamel 1990). Successful firms are those that "pick their strengths," and excel in a relatively small number of dimensions. Michael Porter's management classic "Competitive Advantage" (Porter 1985), for example, argued that firms should either minimize cost or produce differentiated, high-value products, but not do both. Firms that aim to be "all things to all people," will be "caught in the middle" and fail (Porter 1985;1996).[1] Classic examples include retailer Walmart and its focus on supply chain managment (and low costs), or department store Nordstorm with its excellence in customer service. Many technology firms are also reputed for their functional 'biases'. Google, for example, proclaims "[it] is and always will be an engineering company". IBM, on the other hand, used to dominate its industry with a sales-driven culture, where a massive sales force determined customer needs and drove product development.

In this paper, we show how organizational focus and knowledge specialization arise naturally in response to organizational trade-offs between coordination and adaptation. We propose a model of team production in which a number of complementary tasks, such as engineering, purchasing, manufacturing, marketing and selling must be implemented in a coordinated fashion. Each agent is in charge of one task and must adapt this task to local information. Such adaptation, however, may result in coordination failures with other tasks unless agents communicate effectively. Information flows, however, are imperfect as agents have limited attention and information-processing abilities.

[1]Both Prahalad and Hamel (1990) and Porter (1996) where recently selected among Harvard Business Review's 10 Must Read articles (HBR's 10 Must Reads: The Essentials).

In our model, there is no need for organizational focus if attention is unconstraint. All tasks can then be both very adaptive and well coordinated, and attention is evenly distributed. If attention is scarce and coordination is important, however, we show it is optimal to treat tasks asymmetrically. A few agents should then be allowed to be very responsive to their local information, and all attention should be focused on those agents and their tasks in order to avoid coordination failures. In contrast, coordination with all other tasks is achieved by limiting their adaptiveness. All tasks are then well coordinated, but only a few tasks are adaptive. Leadership, where a few agents monopolize scarce attention and take most of the initiative, arises endogenously. In contrast, a 'balanced' organization that spreads attention evenly across tasks is 'stuck in the middle': tasks are neither very adaptive nor are they very well coordinated.

The mechanism underlying the above result is a fundamental complementarity between the attention devoted to an agent, and the initiative taken by this agent. Agents take initiative by adapting their task to local information. But agents who are ignored by others are forced to also largely ignore their own private information, as taking initiative would then result in substantial coordination failures. Conversely, it is a waste of resources to allocate scarce attention to an agent who takes little or no initiative. Following the same logic, the more attention and agent receives, the more initiative this agent can take, and the more important it is to devote scarce attention to this agent in order to ensure coordination. Because of the above complementarities, members in an organization either communicate intensively about a particular task, or they ignore it. An optimal communication network equally divides all attention among a select number of tasks or agents, which we refer to as '*leaders*'. The scarcer is attention, the smaller is the number of tasks on which the organization focuses. Interestingly, those chosen tasks then often receive much more attention – and are much more adaptive – than if attention were to be abundant.

Our results shed light on the importance of leadership in team production as well as new trends in organization design. Over the last decades, there has been enormous technological innovations in communication and coordination technologies (e-mail, wireless communication and computing, intra networks). Our model predicts that as communication technology improves, more decentralized communication networks, in which there are more but less influential leaders, become optimal. The resulting organization is often less well coordinated, less cohesive, but has a broader focus – it pays attention to the task-specific information of a larger number of tasks. This is consistent with 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.[2] Such novel organizations have been documented in both case studies (for example"Proctor & Gamble Organization 2005," HBS case 9-707-519)[3] and large scale empirical studies (Guadalupe and Wulf, 2012)[4]. Our results can further be interpreted as giving credence to the importance of choosing a strategy – choosing a performance dimension or task in the value chain to focus on.[5] By the same token, we provide insights as to how focused firms should be. Our results suggest that having a narrow focus becomes less important as information technology relaxes the communication and attention constraints of organizations. Aiming to be all things to all people may be more attractive now than it used to be.

In most of our paper, tasks are ex ante symmetric, and it does not matter which tasks the organization focusses on. In reality, tasks are of course likely to differ from each other. The question, then, is not only how focussed to be, but which tasks to focus on. An interesting

---

[2] Our 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 these papers 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.

[3] In this case study, Piskorski and Spadini document how P&G has moved towards a novel organizational structure in which a separate product organization (responsible for global marketing and product development), a sales organization (responsible for delivery and customization to local markets), and a business services organization are interdependent units who are giving equal weight in decision-making processes, and achieve coordination through social networks and horizontal communication, rather than vertical authority relationships. In the past, geographically organized sales organizations had dominated P&G, which slowed down the development and roll outs of new products.

[4] Guadalupe and Wulf document how in recent decades, C-level executive teams in Fortune 500 firms have almost doubled in size, mainly because of the inclusion of more functional managers.

[5] See Van den Steen (2012) for a different view and formalization of 'what is strategy'.

asymmetry is one where some tasks impose larger coordination costs (delays, low product quality) should other tasks not be coordinated with adaptations made to it. For example, in designing a car, important changes made to how the engine works, may have important consequences for the remainder of the design. Perhaps counter-intuitively, we show that if attention is relatively scarce, it is optimal not to focus attention on highly interdependent tasks, but instead restrict their adaptiveness. It is only when attention becomes abundant, that the organization focusses on such tasks and allows them to become adaptive. Importantly, the organization then devotes most, or even all of its attention to those tasks.

Modeling attention and organizational knowledge. A necessary ingredient for our results is that attention is constraint. The specific way in which we model limits to information-processing or communication borrows from a recent literature on rational inattention (Sims 2003), which in turn is based on information theory (Cover and Thomas 1991). By virtue of carrying out a task, 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. The uncertainty regarding other tasks is expressed in terms of the *entropy* of the distribution of the local shocks. For normally distributed random 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, and can be interpreted as *organizational knowledge.* 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 this theory, organizations have a fixed communication capacity implying that the total amount of organizational knowledge that can be achieved is also 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.

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. In the absence of any complementarities induced by the need for coordination, this provides a powerful force *against* focus. In particular, when attention is abundant, there are then strongly decreasing marginal returns to focus all attention around one or a few tasks. Hence, it is only when coordination is important and attention is scarce that it is optimal to specialize organizational knowledge.

Outline. Our paper is organized as follows. After reviewing the related literature in Section 2, we describe our model in Section 3. Most of the insights and intuitions of our paper can be derived and illustrated in a simple model with two agents and two tasks, which is analyzed in Section 4. In Section 5, we generalize the model to n agents and n tasks. Section 6 considers a number of extensions: (i) Asymmetries between tasks, (ii) Centralized production, and (iii) Technological trade-offs between adaptation and coordination. Section 7 concludes. Proofs of Propositions are relegated to the Appendix.

2 Literature Review

Our paper is part of a large literature on team theory (Marschak and Radner 1972), which studies games where agents share the same objective, but have asymmetric information. Team theory has been widely used to study problems of organization design.[6] Most closely related are Dessein and Santos (2006) (DS hereafter), which introduces the organizational trade-offs between adaptation and coordination central to our paper, and Calvo-Armagenol, de Marti and Prat (2011).[7] DS studies the optimal division of labor in organizations, but restricts communication flows to be symmetric. In contrast, we take task-specialization as given and endogenize communication patterns. Calvo-Armagenal et al. also endogenizes communication patterns in a framework similar to that of DS. Their focus, however, is on how asymmetries in pay-off externalities between pairs of agents result in asymmetric communication flows and differential influence for agents. In a symmetric set-up, there are no asymmetric communication patterns

[6]See, e.g., Cremer (1980, 1993), Sah and Stiglitz (1986), Geanakoplos and Milgrom (1991), Prat (2002), and Alonso et al. (2012). See Garicano and Van Zandt (2012) for a recent survey.

[7]As an alternative to team theory, a recent literature has studied strategic communication or 'cheap talk' in hiearchies (Alonso et al. 2008, Rantakari 2008) and networks (Hagenback and Koessler 2010, Galeotti et al. 2009). As in Dessein and Santos, the trade-off between adaptation and coordination is central in those models, and pay-offs are quadratic in actions and information.

in their model: each agent is equally influential and there are no leaders. In contrast, we show how leadership and asymmetric information flows arise naturally in symmetric settings.[8]

Our model also shares similarities with beauty contests models in finance and macroeconomics. In a typical beauty contest game (see, e.g., Morris and Shin 2002), economic actors must respond to a shock, but also care about choosing similar actions as other agents in the economy. In contrast to our model, however, agents learn about a common global shock as opposed to privately observed local shocks. This has very different implications. Better public information crowds out the use of private information and there can be excessive coordination (Angeletos and Pavan 2007). In contrast, a key mechanism in our paper is that more common information allows agents to better respond to their private information. While some papers have studied optimal information acquisition strategies in this context (Hellwig and Veldkamp 2009, Myatt and Wallace 2012), the focus on a common global shock is less conducive to study communication flows inside organizations.[9]

A related theory of leadership is proposed by Bolton, Brunnermeier and Veldkamp (2008). Also in their model, coordination and adaptation play a central role, but both the communication network and the identity of the leader is exogenously given.[10] Finally, our argument in favor of organizational focus is reminiscent of at least two other literatures in organizational economics. In multitask incentive theory (Holmstrom and Milgrom, 1991,1994), a narrow task-assignment may allow a principal to provide higher-powered incentives to an agent. In multitask career concerns models, Dewatripont et al. (1999) show how incentives are impaired by an agent pursuing multiple objectives. The key insight in this literature is that it is easy to provide incentives to specialized agents. The above theories thus offer rationales for specialization at the individual level but are silent on the issue of organizational focus, which is the topic of this paper. Similarly, the literature on 'narrow business strategies' and 'vision'

[8]The main difference is that Calvo-Armagenal et al. posit a communication technology with strong decreasing marginal returns, always enough to overwhelm the convexities induced by the coordination-adaptation trade-offs. Other differences are that agents are self-interested and invest in both active and passive communication.

[9]Rather, the models are ideally suited to study the optimal provision of information to independent economic actors, e.g. by a central bank, as in Morris and Shin 2007.

[10]Hermalin (2012) provides an overview of alternative theories of leadership, such as leading by example, where leadership serves to motivate, rather than coordinate, agents.

(Rotemberger and Saloner, 1994, 2000, Van den Steen, 2005) has argued that the commitment by a principal or leader to select a certain type of projects provides strong incentives for agents to exert effort related to such projects. As in the multitask models above, 'focus' is thus again a tool to improve effort incentives.

3 The model with two agents

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. 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 and Section 3. We defer to Section 4 the analysis of $n > 2$ agents.

3.1 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 θ_i, a shock with variance σ_θ^2 and mean 0, which is relevant for the proper implementation of the assigned task. We refer to θ_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 θ_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, $\mathbf{q} = [q_1, q_2]$, the realized profit of the organization is:

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

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

3.2 The communication network

A communication network $\mathbf{t} = [t_1, t_2]$ 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 θ_j. We assume that the organization cannot devote an infinite amount of resources to communicate:

$$t_1 + t_2 \leq \tau, \quad (2)$$

where $\tau < \infty$. For example, τ 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 $t_1 = t_2 = \tau/2$.

3.3 The communication technology

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

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

Assumption A. The random variables (θ_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 θ_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).[11] 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 θ_j and m_j have zero mean. Using the law of total variance, we can then write the expected conditional variance of local shock θ_j, referred to as the *residual variance* throughout, as follows:

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

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 θ_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., $\tau < 2\hat{\tau}$.[12]

[11] 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

[12] Note that by definition of $\hat{\tau}$, we have that, for every $\mathbf{t}$, $\mathsf{RV}(t_1) + \mathsf{RV}(\tau - t_1) > 0$ if, and only if, $\tau < 2\hat{\tau}$.

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 θ_j and ϵ_j are drawn from normal distributions is sufficient for Assumption A to hold. In this case, the residual variance is

$$\mathsf{RV}(t_j) = \sigma_\theta^2 \left[1 - \frac{\sigma_\theta^2(t_j)}{\sigma_\theta^2 + \sigma_\epsilon^2(t_j)}\right]. \tag{5}$$

Assumption B is satisfied whenever $\sigma_\varepsilon^2\left(t_j\right)$ is a decreasing function of t_j, $\lim_{t_j \to 0} \sigma_\varepsilon^2\left(t_j\right) = \infty$ and $\sigma_\varepsilon^2\left(\tau/2\right) > 0$. ■

Example 2. Uniformily distributed messages and information. Assume next that θ_j is uniformly distributed on $[-1, 1]$ and that communication from agent j to agent i is successful 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 $\mathsf{RV}\left(t\right)$. We build on the literature on rational inattention (Sims, 2003), which in turn builds on 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 $\mathsf{RV}'\left(\cdot\right) < 0$ but $\mathsf{RV}''\left(\cdot\right) > 0$. 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 4.3.

3.4 Timing

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}.

4 Organizational focus with two agents

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 4.1 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, first when marginal returns to communication are constant (Section 4.2) and then when there are decreasing marginal returns (Section 4.3).

4.1 Actions and the expected profits of the organization

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

$$q_{11} = \frac{1}{1+\beta}\left[\theta_1 + \beta E\left[q_{21}|\mathcal{I}_1\right]\right] \qquad \text{and} \qquad q_{12} = E\left[q_{22}|\mathcal{I}_1\right], \tag{6}$$

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 (Examples 1 and 2 above). We can write (6) as

$$q_{11} = a_{11}(t_1)\,\theta_1 \qquad \text{and} \qquad q_{12} = a_{12}(t_2)\,E[\theta_2|\mathcal{I}_1]\,. \tag{7}$$

Substituting the guess (7) into (6), and using Assumption A, we find that the equilibrium actions for agent 1 are

$$q_{11} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta \mathsf{RV}(t_1)}\theta_1 \qquad \text{and} \qquad q_{12} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta \mathsf{RV}(t_2)} E[\theta_2|\mathcal{I}_1]\,, \tag{8}$$

and similarly for agent 2.

Note that the larger the residual variance $\mathsf{RV}(t_i)$ about task i, the less adaptive is task i to its environment. Hence, 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 θ_1. Intuitively, an agent who receives a lot of attention can respond more effectively to task-specific information, as the other agent is then able to take the appropriate coordinating action. In contrast, an agent who is ignored is forced to also largely ignore his own task-specific information, as responding to his own information would result in substantial coordination failures with the other task.

Naturally, the impact of attention on adaptation depends on the importance of coordination, β. As β goes to 0, tasks become perfectly adaptive for any level of attention t_i. In contrast, as β goes to infinity, task i becomes unresponsive to its information unless attention is perfect ($t_i \geq \hat{\tau}$) and $\mathsf{RV}(t_i) = 0$.

Substituting (8) into (1) and taking unconditional expectations we find that

$$E[\pi(\mathbf{q}|\boldsymbol{\theta})] = (\Omega(t_1) - 1)\,\sigma_\theta^2 + (\Omega(t_2) - 1)\,\sigma_\theta^2, \tag{9}$$

where

$$\Omega(t_i) = \frac{\operatorname{cov}(q_{ii}(t_i), \theta_i)}{\sigma_\theta^2} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta \mathsf{RV}(t_i)} \in [0,1] \tag{10}$$

neatly captures the *adaptiveness of task i to its task-specific information.* When the organization is fully adaptive, that is $\operatorname{cov}(q_{ii}, \theta_i) = \sigma_\theta^2$, the expected profits are maximized and

$E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right] = 0$. From (8), however, a limited attention capacity $\tau < 2\hat{\tau}$ imposes limits to adaptation such that $\text{cov}(q_{ii}, \theta_i) < \sigma_\theta^2$ and $E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right] < 0$.

An alternative representation of the expected profit function is[13]

$$E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right] = -\beta\Omega\left(t_1\right)\mathsf{RV}\left(t_1\right) - \beta\Omega\left(t_2\right)\mathsf{RV}\left(t_2\right). \tag{11}$$

Expression (11) shows how the residual variance regarding the local information of task i, as represented by $\mathsf{RV}\left(t_i\right)$, is costly to the organization only to the extent task i is adaptive to this local information, as captured by $\Omega\left(t_i\right)$. It is immediate, then, that there is a complementarity between the adaptiveness of a given task and a lower residual variance regarding the same task: One wants to reduce the residual variance of the task which is most adaptive. In turn, from expression (8), the task that receives most attention and has the lowest residual variance, is also most adaptive.

The problem of organizational design is to maximize (9) or (11) with respect to t_1 subject to $t_1 \in [0, \tau]$ and $t_2 = \tau - t_1$. Substituting $t_2 = \tau - t_1$, the derivative of the profit function with respect to t_1 is

$$\begin{aligned}\frac{\partial E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial t_1} &= \frac{\partial\Omega(t_1)}{\partial t_1}\sigma_\theta^2 + \frac{\partial\Omega(\tau - t_1)}{\partial t_1}\sigma_\theta^2 \\ &= \beta\Omega^2(t_1)\left|\mathsf{RV}'(t_1)\right| - \beta\Omega^2(t_2)\left|\mathsf{RV}'(t_2)\right|\end{aligned} \tag{12}$$

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

4.2 Constant marginal returns to communication

As a benchmark, we first consider the case of communication technologies that exhibit constant marginal returns, that is where $\mathsf{RV}''(\cdot) = 0$. For example, with uniformly distributed information and messages **(Example 2)**, constant marginal returns imply that the probability that communication is successful is linear in attention, that is $p(t) = \alpha t$ for some positive α.

[13] Expression (9) 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.

Using (12), we obtain

$$\frac{\partial E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial t_1} > 0 \qquad \Longleftrightarrow \qquad \Omega(t_1) > \Omega(t_2) \qquad \Longleftrightarrow \qquad t_1 > t_2. \tag{13}$$

It follows that the expected profits are minimized when attention is equally divided among both tasks, that is $t_1 = t_2 = \tau/2$. The following Proposition is immediate:

Proposition 1 *If there are constant marginal returns to communication, the organization focuses on one task. If $\tau < \hat{\tau}$, the organization only communicates about one task and ignores the other, that is $t_i^* \in \{0, \tau\}$. If $\tau > \hat{\tau}$ the organization perfectly learns the local shock of one task, and devotes the remaining attention to communicate about the other task, that is $t_i^* \in \{\tau - \hat{\tau}, \hat{\tau}\}$.*

Intuitively, from (11), in order to minimize coordination losses, it is optimal to devote more attention (increase t_i) and reduce the residual variance $\mathsf{RV}(t_i) = Var(\theta_i|m_i)$ of the task which is most adaptive. In turn, a task which receives more attention can afford to be more adaptive: $\Omega(t_i)$ is increasing in t_i. It follows that whenever attention is in short supply, it is optimal to either devote a lot of attention to a task or, alternatively, ignore it completely. Put differently, the organizational trade-offs between adaptation and coordination result in a profit function that is convex in the amount of attention that is devoted to a particular task. Expected profits are minimized for firms that are "stuck in the middle," and equally divide attention among both tasks.

Another way to understand the above results is through the notion 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, by means of communication, and take the appropriate coordinating actions. In Dessein and Santos (2006), this was referred to as *ex-post coordination.* 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 can be seen as *ex-ante coordination.*

While in Dessein and Santos (2006) all tasks were treated symmetrically by assumption, the insight of Proposition 1 is that when attention is scarce (that is $\tau < \hat{\tau}$), it is 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 it to be 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 plentiful, that is $\tau >> \hat{\tau}$, it is optimal for both tasks to be very adaptive, as they can be coordinated ex-post through communication. Indeed, if attention is *not* constraint, that is $\tau \geq 2\hat{\tau}$, both tasks are equally and fully adaptive to their local shock and there is no organizational focus.

4.3 Decreasing marginal returns to communication

Obviously the result in Proposition 1 holds if the communication technology displays increasing marginal returns to communication, that is $\mathsf{RV}''(\cdot) < 0$. In what follows we study the possibility of organizational focus in those contexts where communication technologies display decreasing marginal returns. A tracteable and time-tested way to do so is to draw on micro-foundations in information theory (Cover and Thomas, 1991).

4.3.1 Information Theory

Following the literature on rational inattention (Sims, 2003), we assume that the time or attention needed to communicate or process a signal about a random variable θ_i depends on the extent to which this signal reduces the differential entropy of θ_i, where this time or attention plays the role of the finite (Shannon) capacity of a noisy communication channel in information theory. Formally, the communication capacity (time, attention) needed to communicate a message m_i about θ_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 θ_i and $h(\theta_i|m_i)$ is the differential entropy of θ_i conditional on observing m_i. We then posit that[14]

$$I(\theta_1, m_1) + I(\theta_2, m_2) \leq \tau \tag{14}$$

where τ 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: probability theory, communication theory, computer science, mathematics, statistics, 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.[15]

In information theory, $I(\theta_i, m_i)$ is referred to as the *mutual information* between θ_i and m_i. While each agent is privately informed about the task-specific shock affecting his own task, the mutual information about θ_1 and θ_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 its 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 θ_i and ε_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).[16] Since the entropy of a normal variable with variance σ^2 is given by $\frac{1}{2}\ln(2\pi e\sigma^2)$,

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

[15] 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.

[16] If, as assumed in information theory, agents optimally design the distribution of $F(\theta_i|m_i)$ subject a capacity

then

$$I(\theta_i, m_i) = \frac{1}{2}\left(\ln \sigma_\theta^2 - \ln \mathrm{Var}(\theta_i|m_i)\right) \tag{15}$$

It follows that the time or attention needed to communicate a message m_i about θ_i is linear in the reduction in the log residual variance of θ_i following communication. Let t_i be the communication capacity allocated to communicate m_i, with $t_1 + t_2 = \tau$, we have that

$$\ln \mathsf{RV}\left(t_i\right) = \ln \sigma_\theta^2 - 2t_i \tag{16}$$

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

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 Δt to reduce the expected residual variance from σ_θ^2 to $\sigma_\theta^2/2$, it will take an additional Δt 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

$$\left|\mathsf{RV}'\left(t_i\right)\right| = 2\mathsf{RV}\left(t_i\right), \tag{17}$$

hence the lower the residual variance, the lower the marginal returns to further reduce this variance. While we have derived equation (16) using foundations in information theory, we believe it provides an intuitive, tractable and parsimonious way to model decreasing marginal returns to communication.[17] [18]

constraint for the communication channel, they would choose $F(\theta_i|m_i)$ to be normally distributed in order 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.

[17]Expression (16) also has a natural interpretation if θ_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.

[18]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 θ_i is distributed uniformly in $[-1, 1]$, it follows that

4.3.2 Focused versus balanced organizations

As argued above, the rationale for organizational focus relies on 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 larger is β, the stronger is this complementarity. Decreasing marginal returns to communication, however, provide 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 (12) that a focused organization with $(t_1, t_2) = (\tau, 0)$ is a *local maximum* if and only if

$$\underbrace{\Omega^2(\tau)}_{\text{Adaptiveness}} \quad \times \quad \underbrace{\left|\mathsf{RV}'(\tau)\right|}_{\text{Marg. returns to comm.}} \quad > \quad \Omega^2(0) \quad \times \quad \mathsf{RV}'(0). \tag{18}$$

As shown above, this condition is always satisfied and organizational focus is optimal if there are constant marginal returns to communication. An organization which is less focused $(0 < t_1 \leq t_2 < 1)$ may be optimal, however, when there are decreasing marginal returns to communication. Indeed, if the organization focuses 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 (β small) or attention is not very constrained (τ large), a focused organization with $(t_1, t_2) = (\tau, 0)$ is suboptimal.

Consider first the case where coordination is not very important. For β small, both tasks are then 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 τ, for β sufficiently small, inequality (18) is then violated and $(\tau, 0)$ is not a local maximum. Intuitively, the complementarity between adaptiveness and the allocation of attention relies on

the differential entropy of θ_i is $\ln(2)$ and the conditional differential entropy of θ_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.

the importance of coordination. In the limit, as β goes to zero, this complementarity and the associated increasing returns to coordination disappear.

Next, consider the case where τ is large. When attention is relatively unconstrained, there are strongly decreasing marginal returns to center all communication around one task. Hence, for τ sufficiently large, a focused organization is again not optimal. Formally, since the marginal returns to communication on task 1, $\left|\mathsf{RV}'(\tau)\right|$, go to zero as τ goes to infinity, whereas $\Omega(0)$ is strictly positive, it follows again that $(\tau, 0)$ is not a local maximum for τ sufficiently large.

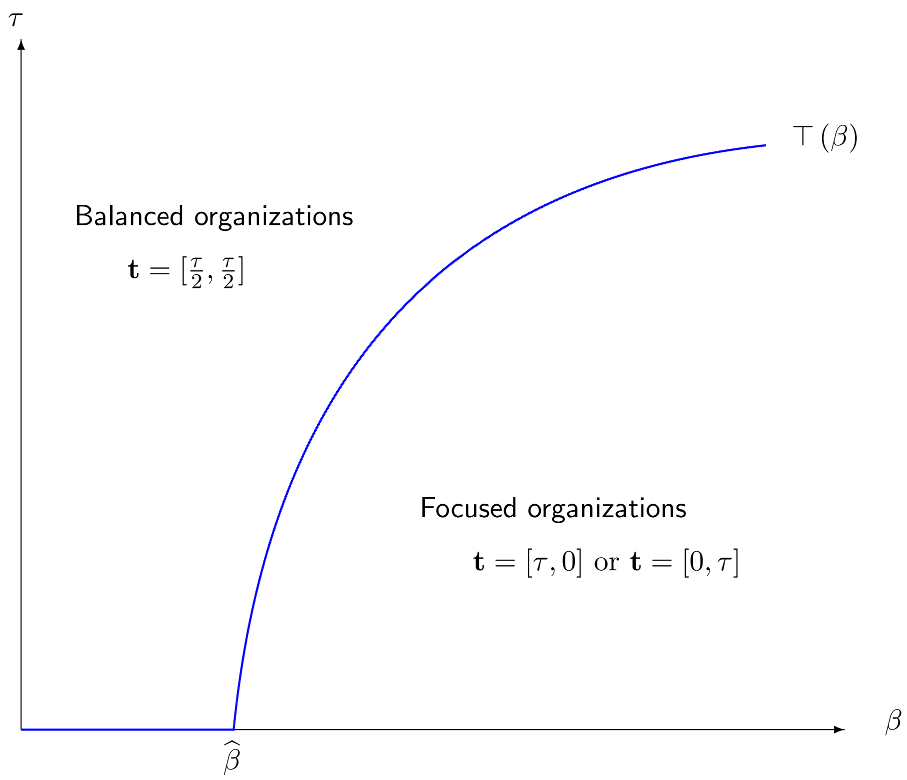
In line with the above intuitions, the following proposition shows that a fully focused organization is optimal if and only if coordination is sufficiently important and attention sufficiently scarce:

Proposition 2 *There exists a* $\widehat{\beta}$ *and* $\top(\beta)$ *such that:*
- If $\beta \leq \widehat{\beta}$ *then organizational balance is optimal:* $(t_1^*, t_2^*) = (\frac{\tau}{2}, \frac{\tau}{2})$.
- If $\beta > \widehat{\beta}$ *then*
(i) Organizational focus is optimal, $t_1^* \in \{0, \tau\}$, *if and only if* $\tau \leq \top(\beta)$
(ii) Organizational balance is optimal, $(t_1^*, t_2^*) = (\frac{\tau}{2}, \frac{\tau}{2})$, *if* $\tau > \top(\beta)$
(iii) $\top(\cdot)$ *is increasing in the importance of coordination,* β.

Figure 1 summarizes Proposition 2. As the propositions shows, organizations which are 'somewhat' focused 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 an equal amount of attention to both tasks.

Proposition 2 further yields an interesting comparative static result with respect to exogenous changes in the communication capacity τ. Improvements in the communication technology (email, wireless communication devices, intranets, ...) can be interpreted as an exogenous increase in τ. An implication of Proposition 2, therefore, is that such technological

Figure 1: **Focused and balanced organizations in the two-agent case.** The plot shows the areas where balanced and focused organizations prevail, as shown in Proposition 2.



improvements result in a shift from focused organizations which are centered around one task and excel on that task at the expense of others, towards more balanced organizations which aim to perform equally well on all tasks, but excel 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 agent who monopolizes all information flows (the leader) 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 share it equally with the other agent engaged in team production.

5 Organizational focus with many agents

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 show how the model introduced in section 2 extends to the $n > 2$ case and discuss alternative models of communication. Section 5.2 characterizes optimal actions and organizational performance and we derive, and characterize, the optimal organizational form in section 5.3. The main result of this section is that the optimal organizational form is the $\ell-$leader organization, which features a number $\ell \leq n$ of equally adaptive agents (leaders) to whom all agents in the organization devote an equal amount of attention, whereas no attention is devoted to any agent who is not a leader. Throughout, and in the interest of brevity, we assume that the communication technology features decreasing marginal returns.

5.1 The model with $n > 2$

Consider 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 θ_i with mean 0 and variance σ_θ^2. In order to implement task i, agent i chooses a primary action q_{ii}, who must be adapted to the task-specific shock θ_i, as well as $(n-1)$ coordinating actions q, who must be adapted to the primary actions chosen by the other agents $j \in \mathcal{N} \setminus \{i\}$. We denote by

$$q_i = [q_{i1}, q_{i2} ..., q_{ii}, ..., q_{in}], \tag{19}$$

the string of actions chosen 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:

$$\pi(\mathbf{q}|\boldsymbol{\theta}) = -\sum_{i \in \mathcal{N}} \left[(q_{ii} - \theta_i)^2 + \frac{\beta}{n-1} \sum_{j \in \mathcal{N} \setminus \{i\}} (q_{ii} - q_{ji})^2 \right]. \tag{20}$$

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 $m_{ij} = \theta_j + \varepsilon_{ij}$ with ε_{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 τ. Let θ_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 θ_j, then, as in (16),

$$\ln \mathsf{RV}(t_{ij}) \equiv \ln Var(\theta_j|m_{ij}) = \ln \sigma_\theta^2 - 2t_{ij}, \tag{21}$$

where it must be that

$$\sum_{j \in \mathcal{N}} \sum_{i \in N \backslash \{i\}} t_{ij} \leq \tau, \tag{22}$$

The above communication network $\mathbf{t} = \{t_{ij}\}_{i \neq j}$ is one where communication among agents is assumed to be bilateral and allows for a rich variety of asymmetries. In particular, agent j may devote more attention to agent i than another agent k, that is, $t_{ji} > t_{ki}$ and agent i may receive more attention from the organization than another agent k, that is, $\sum_j t_{ji} > \sum_j t_{jk}$.

Bilateral communication is convenient because it allows for maximum flexibility on the nature of communication flows but clearly, in the presence of $n > 2$ agents, other models of communication are reasonable alternatives. In section 5.4 we show how alternative models of communication, where communication is public or agents face individual capacity constraints, result in information structures that are equivalent to the ones that arise under the optimal bilateral communication network.

5.2 Organizational actions and performance

Having established the equivalence of these alternative communication models, we now characterize the organizational actions and performance for a given bilateral communication network. For a given network $\mathbf{t}$ and string of observed messages m_i, agent i chooses the string of actions q_i, given in (19), 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 (20) 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

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

As in the case of $n = 2$, we focus on equilibria in linear strategies, that is $q_{ii} = a_{ii}\theta_i$. Using the same method as in Section 2, the expression for the equilibrium actions can then be generalized to yield the following equilibrium actions for any $n > 1$:

$$q_{ii} = \frac{(n-1)\sigma_\theta^2\theta_i}{(n-1)\sigma_\theta^2 + \beta\sum_{j\neq i}\mathsf{RV}(t_{ji})} \qquad \text{and} \qquad q_{ji} = \frac{(n-1)\sigma_\theta^2 E\left[\theta_i|\mathcal{I}_j\right]}{(n-1)\sigma_\theta^2 + \beta\sum_{j\neq i}\mathsf{RV}(t_{ji})},$$

where $\mathsf{RV}(t_{ji}) \equiv Var(\theta_j|m_{ij})$ is given by (21). Taking into account the equilibrium actions, we find that expected profits are given by

$$\begin{aligned} E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right] &= \sum_{i\in\mathcal{N}} \text{cov}\left[(q_{ii},\theta_i) - \sigma_\theta^2\right] \\ &= -n\sigma_\theta^2 + \sigma_\theta^2\sum_{i\in\mathcal{N}}\frac{(n-1)\sigma_\theta^2}{(n-1)\sigma_\theta^2 + \beta\sum_{j\neq i}\mathsf{RV}(t_{ji})}. \end{aligned} \tag{23}$$

5.3 The $\ell-$leader organization

5.3.1 The optimality of the $\ell-$leader organization

In our analysis of optimal communication networks with two agents, we saw that organizations fluctuated between full focus, $t_1^* \in \{0, \tau,\}$ and balance $t_1^* = t_2^* = \frac{\tau}{2}$. How do the intuitions we built in the two-agent case translate to the multi-agent case? Our main result is that, as in the two-agent case, the organization optimally focuses on a limited set of tasks. That is, focus in a set of tasks arises endogenously and the agents managing those tasks, the leaders, are the focus of the attention of all agents in the organization. To show this result we start by defining the $\ell-$leader organization:

Definition: The $\ell-$leader organization. An $\ell-$leader organization is a communication network $\mathbf{t}$ where the set of agents can be partitioned in a set of leaders $\mathcal{L}(\mathbf{t})$ and followers $\mathcal{F}(\mathbf{t})$ such that

1. The number of leaders is $\ell \leq n$.

2. For each follower $i \in \mathcal{F}(\mathbf{t})$, $t_{ji} = 0$ for all $j \neq i$.

3. For each leader $j \in \mathcal{L}(\mathbf{t})$, $t_{ij} = \frac{\tau}{(n-1)\ell}$ for all $i \neq j$

An $\ell-$leader organization has the property that there is a number of agents ℓ, which we call leaders, to whom all 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 3 *The optimal communication network is an $\ell-$leader organization with $\ell \in \{1, 2, \cdots, n\}$.*

The proof of Proposition 3 follows from the next two lemmas.

Lemma 4 *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 4 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 θ_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 5 *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 5, 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 5. In the case of two tasks, it was shown (Proposition 2) that either $t_1^* \in \{0, \tau\}$, or $t_1^* = t_2^* = \tau/2$. Following the same logic, one can equally 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 $t_i = t_j = (\hat{t}_i + \hat{t}_j)/2$. 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*, or force a task to largely ignore its local information and coordinate this task with others *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.

5.3.2 Actions and performance of the $\ell-$ leader organization

Armed with Proposition 3 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 ℓ leaders be the first ℓ agents and let the followers be agents $\{\ell+1, \ell+2, \cdots, n\}$. The primary equilibrium actions of leaders and followers are given by

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

Leaders' primary actions naturally comove more with their local information than do those of followers:

$$\Omega\left(\frac{\tau}{(n-1)\,\ell}\right) = \frac{\operatorname{cov}(q_{ii}, \theta_i)}{\sigma_\theta^2} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta \mathsf{RV}\left(\frac{\tau}{(n-1)\ell}\right)} \quad \text{for} \quad i \leq \ell \tag{24}$$

$$\Omega(0) = \frac{\operatorname{cov}(q_{ii}, \theta_i)}{\sigma_\theta^2} = \frac{1}{1+\beta} \quad \text{for} \quad \ell+1 \leq i \leq n. \tag{25}$$

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 (24) 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 τ 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 (23) can be re-written as:

$$E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right] = -n\sigma_\theta^2 + \sigma_\theta^2 \ell \left[\frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta \mathsf{RV}\left(\frac{\tau}{(n-1)\ell}\right)}\right] + (n-\ell)\left[\frac{\sigma_\theta^2}{1+\beta}\right], \tag{26}$$

where to obtain (26) we have made use of both Proposition 3 and (24) and (25). As in the two-agent case, we can rewrite (26) as follows

$$E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right] = -\beta\ell\Omega\left(\frac{\tau}{(n-1)\,\ell}\right)\mathsf{RV}\left(\frac{\tau}{(n-1)\,\ell}\right) - \beta\,(n-\ell)\,\Omega\,(0)\,\mathsf{RV}\,(0)\,, \tag{27}$$

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

$$\ell^* = \mathsf{argmax}_{\ell\in\{1,2,\cdots,n\}} E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]. \tag{28}$$

5.3.3 Comparative statics of the of the $\ell-$ leader organization

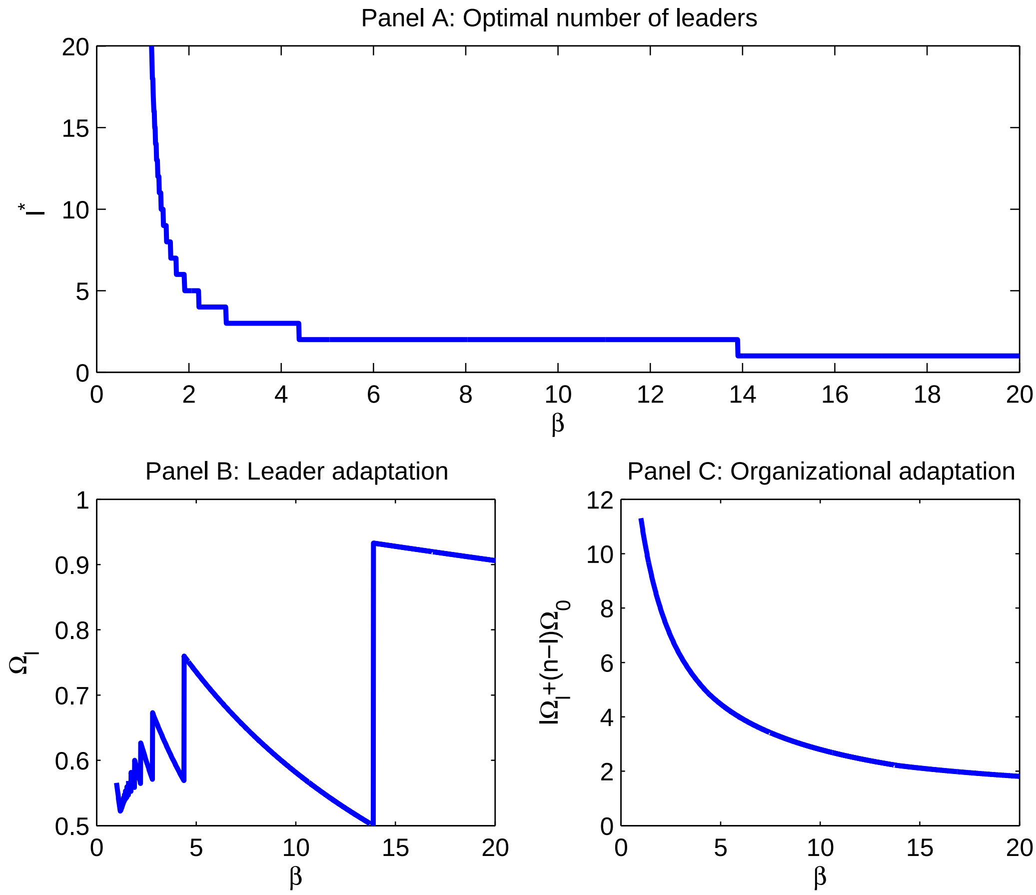
Armed with (28) we are able to offer a sharp characterization of the $\ell-$leader organization as a function of the organization's communication capacity τ and the task-interdepence or coordination parameter β.

Proposition 6 *There exists* $0 < \overline{\beta}\,(n) < \ldots < \overline{\beta}\,(\ell+1) < \overline{\beta}\,(\ell) < \ldots < \overline{\beta}\,(2)$ *such that*

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

The intuition for Proposition 6 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. Figure 2 illustrates the results of Proposition 6 for a specific numerical example ($n = 20$, $\sigma_\theta^2 = 1$ and $\tau = 50$.) Start with Panel A, which plots the optimal number of leaders ℓ^* as a function of the importance of coordination, β. A balanced organization is optimal when coordination is sufficiently un-important. In this specific example, whenever $\beta < \overline{\beta}\,(20) = 1.2$ the organization is fully balanced, that is, $\ell^* = n = 20$.

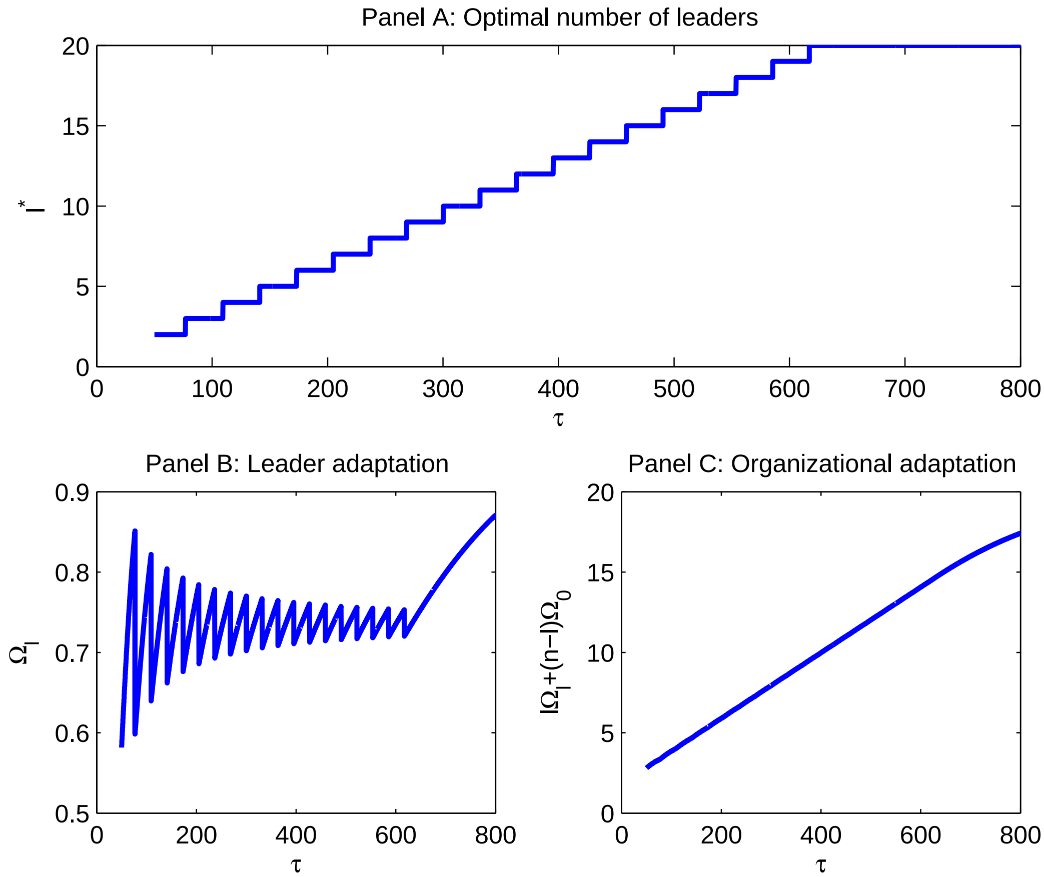
Figure 2: **Optimal number of leaders and adaptation as a function of β**
Example: $n = 20$, $\sigma_\theta^2 = 1$, $\tau = 50$, and $\beta \in [0, 20]$. Panel A: Optimal number of leaders, ℓ^*, as a function of the importance of coordination β. Panel B: Leader adaptation $\Omega_\ell = \Omega\left(\frac{\tau}{(n-1)\ell}\right)$. Panel C: Organizational adaptation, $\ell\Omega_\ell + (n - \ell)\,\Omega_0$ where $\Omega_0 = \Omega(0)$



As coordination becomes more important, the communication becomes more focused around fewer leaders. Finally, when tasks are sufficiently interdependent, when $\beta > \overline{\beta}(1) \geq 13.90$, the organization has a single leader, $\ell^* = 1$.

Panel B of Figure 2 shows how the adaptiveness of each leader to his local shock, Ω_ℓ, as defined by expression (24), changes as tasks become more interdependent as measured by β. Interestingly, leaders tend to be much more adaptive when coordination costs are higher, as they then share influence with fewer other leaders. For example, when $\beta = 14$, there is only

Figure 3: **Optimal number of leaders and adaptation as a function of τ**
Example: $n = 20$, $\sigma_\theta^2 = 1$, $\beta = 10$, and $\tau \in [50, 800]$. Panel A: Optimal number of leaders, ℓ^*, as a function of attention capacity τ. Panel B: Leader adaptation $\Omega_\ell = \Omega\left(\frac{\tau}{(n-1)\ell}\right)$. Panel C: Organizational adaptation, $\ell\Omega_\ell + (n-\ell)\,\Omega_0$ where $\Omega_0 = \Omega(0)$



one leader, but this leader is, roughly, 50% more adaptive to his local information then when $\beta = 4$ and $\ell^* = 3$. Indeed when $\beta = 14$, $\Omega_\ell = .93$ whereas when $\beta = 4$ $\Omega_\ell = .59$. Intuitively for a given number of ℓ 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 β 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. Still, as Panel C, shows, the overall adaptiveness of the organization as measured by

$$\ell^{*}\Omega\left(\frac{\tau}{(n-1)\,\ell^{*}}\right)+(n-\ell^{*})\,\Omega\,(0)$$

is a decreasing function of β. Thus an increase in the importance of coordination reduces the overall adaptiveness though it increases that of the lower number of leaders.

Proposition 6 also shows how an exogenous change in attention capacity τ, increases the number of leaders and makes the organization more balanced. Figure 3 shows the number of leaders, leader adaptiveness and organizational adaptiveness as a function of τ for a particular numerical example ($n = 20$, $\beta = 10$ and $\sigma_\theta^2 = 1$.) As capacity increases the number of leaders increases as well (Panel A), but now the adaptiveness of the leaders may decrease when their number is enlarged, as now they have to share a given capacity with an additional leader (Panel B), until the number of leaders reaches $\ell = n$ when the adaptiveness of the leader becomes monotonic in τ. The overall adaptiveness of the organization though is monotonically increasing in τ (Panel C). In general, 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 focused and leadership is more broadly shared.

5.4 Alternative models of communication

As already mentioned, we have assumed that communication is bilateral as this puts the least constraints on nature of communication flows. We consider now two alternative models of communication: Public communication, where agents can address the entire team simultaneously, and a form of communication where each agent has an individual capacity constraint. We show next that these alternatives result in information structures that are equivalent to the ones that arise under the optimal bilateral communication network.

5.4.1 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 $\mathbf{t} = \{t_1, ..., t_n\}$, 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, as shown in Lemma 4, holds for the optimal bilateral communication networks. The following equivalence result, proven in appendix, therefore follows immediately:

Result 1: *An optimal communication network $t = \{t_1, .., t_n\}$ given public communication and constraint τ 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[19] *under bilateral communication and constraint $\tau^b = (n-1)\tau$.*

Finally, notice that when $n = 2$ $\tau = \tau^b$, and there is no difference between a public and a bilateral communication network.

5.4.2 Individual communication channels

So far we have assumed that the communication constraint is determined at the organizational level. Alternatively, each agent may have a limited communication capacity τ.[20] Formally, let each agent have access to an individual communication channel, whose finite capacity τ can be

[19]We refer to 'an' optimal communication network as there are typically several optimal communication networks, where the organization focusses on the same number, but potentially different, tasks.

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

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

$$\sum_{j \in N} t_{ij} \leq \tau \qquad \forall i \in \mathcal{N}, \tag{29}$$

and where t_{ii} is the capacity devoted to broadcast information about θ_i, and t_{ij} is the capacity devoted to listen to the information broadcasted by agent $j \neq i$. The effective communication flow between agents j and i regarding θ_j then equals $\min\{t_{ij}, t_{jj}\}$ such that[21]

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

In appendix we prove the following equivalence result which again relies on Lemma 4:

Result 2: *An optimal communication network $t = \{t_{ij}\}_{i,j}$ with individual communication constraints τ 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 an organization-wide constraint $\tau^b = (n-1)\tau$.

6 Extensions

In this section we extend our model in three directions. First, we relax the assumption that all tasks are ex ante symmetric. In particular, we assume that some tasks impose larger coordination costs than others. Second, we consider individual or centralized production as an alternative to the team production. We show that organizational focus only arises under team production, and characterize when team production is optimal. Finally, we consider a variant of our model in which each agent takes only one action, which must both be adapted to his local information and coordinated with the actions of other agents. This model, versions of which can be found in the literature on organizational economics, introduces technological trade-offs

[21]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.

between adaptation and coordination. Yet we show that qualitative identical insights obtain. For simplicity, we focus on the two-agent version of our model and maintain the assumption of decreasing marginal returns to attention, as characterized by expression (16).

6.1 Asymmetries between tasks

So far we have restricted our attention to the case where all the tasks are symmetric and show that focus can arise even when all tasks are ex-ante identical. Clearly though there may be differences across tasks that renders them different ex-ante. A reasonable asymmetry to consider is one where there are some tasks that impose relatively larger costs than other tasks should actions not be coordinated with the primary action of that task. Intuitively, in the organization of production there may be some technological constraints that renders some tasks essential in that any coordination failures with that task result in delays, low product quality or any other type of cost. Should attention be focused on those highly interdependent tasks? In this section we show that this is not necessarily the case.

Consider the two task case and let the coordination parameters be β_1 and β_2 for task 1 and 2, respectively. Define $\beta = \sqrt{\beta_1 \beta_2}$, the geometric mean of β_1 and β_2 and consider situations where

$$\beta_1 = \beta\,(1+\epsilon) \qquad \text{and} \qquad \beta_2 = \beta\,(1+\epsilon)^{-1}.$$

The parameter ϵ thus determines the "spread" between the coordination costs across tasks: An increase in $\epsilon > 0$ increases the coordination costs associated with task 1 and decreases that of task 2, leaving the geometric average, a sufficient statistic for how costly lack of coordination is to the organization, unchanged. When $\epsilon = 0$ the case collapses to the one considered in Section 3. Then we can prove the following result.

Proposition 7 *Define $\widehat{\epsilon}$ as the solution to $(1+\widehat{\epsilon})^2\, e^{-2\tau} = 1$. Then*

1. *If $\tau < \ln \beta$, the optimal organization is focused on task 2, i.e., $(t_1^*, t_2^*) = (0, \tau)$.*
2. *If $\tau \geq \ln \beta$, then*
   (a) *If $\epsilon < \widehat{\epsilon}$ then attention is split among both tasks though not equally: $\tau > t_1^* > t_2^* > 0$*

(b) If $\epsilon \geq \widehat{\epsilon}$, then $(t_1^*, t_2^*) = (\tau, 0)$

If attention is limited, $\tau < \ln \beta$, then all attention is focused on task 2, which has lower coordination costs than task 1. The reason is that allocating limited attention to task 1 is essentially not worth it as it translates into little adaptation given the larger coordination costs the organization would endure: It is better to coordinate task 1 ex-ante, to use the terminology employed by Dessein and Santos (2006), and focus solely on task 2. Instead when attention capacity is larger and the asymmetry is not too large, both tasks receive attention but task 1 receives more than task 2. The reason is obvious: If both tasks are allowed to be adaptive, more attention needs to be devoted to that task that carries larger coordination costs. If asymmetries between both tasks are sufficiently large, task 2 may even receive no attention for $\tau > \ln \beta$. At the threshold $\hat{\tau} = \ln \beta$, the organization then switches from being fully focussed on task 2 to being fully focussed on task 1.

Notice thus that, perhaps counterintuitively, the task that is more interdependent is not necessarily the focus of the available attention. When attention is limited, little is gained by focusing attention on that task which the organization instead opts to coordinate ex-ante. When attention is abundant, however, the vast majority of attention is devoted to this task. Again, Proposition 7 illustrates important convexities in the allocation of attention.

6.2 Team production versus individual/centralized production

A central feature of our model is that tasks are carried out by a *team* of agents. In this section, we consider an alternative method of production, where tasks are carried out and implemented in a centralized way, by a single agent. Consistent with the literature (Becker and Murphy (1992), Bolton and Dewatripont (1994), and 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. A possible benefit of the division of labor is that specialized agents are better informed about task-specific shocks. Assume thus that a generalist in charge of both tasks can only learn about θ_1 and θ_2 by allocating attention to task

1 and task 2, whereas a specialist in task i directly observes θ_i.[22] This approach is similar to the one taken in Geanakoplos and Migrom (1991). One interpretation is that decision-making is *centralized* in head-quarters, and the headquarter manager communicates with Agents 1 and 2 who do observe respectively θ_1 and θ_2.

6.2.1 The allocation of attention under centralized production

We first show that under centralized production, there is no organizational focus. It is team production which makes coordination important, and hence organizational focus valuable. Let t_1 be the amount of attention a generalist agent allocates to learn about θ_1, and $t_2 = \tau - t_1$, the time she allocates to task 2. Let $\mathcal{I}_G$ be the information of the single agent, then she will choose

$$q_{21} = q_{11} = E\left[\theta_1|\mathcal{I}_G\right] \qquad \text{and} \qquad q_{12} = q_{22} = E\left[\theta_2|\mathcal{I}_G\right]. \tag{30}$$

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

$$E[\pi(\mathbf{q})|\theta] = -\left[\mathsf{RV}(t_1) + \mathsf{RV}(\tau - t_1)\right]. \tag{31}$$

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. The following result is then straightforward:

Proposition 8 *If all tasks are implemented by a generalist agent and there are decreasing marginal returns to attention, then this agent optimally splits attention evenly among both tasks* $t^* = \frac{\tau}{2}$.

6.2.2 The optimal division of labor

Armed with the result of Proposition, we now compare two organizational forms:

1. Under *team production*, each task i is implemented by a different specialist who perfectly observes the local shock θ_i. In this organization, a finite communication capacity τ

[22] Alternatively, a generalist might observe only a noisy signal of θ_1 and θ_2. If the single agent were to perfectly observe one or both of the random variables, the allocation of attention would be trivial.

limits the ability of the team to coordinate actions. As we have done throughout this paper, we can distinguish between a *focused 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.

2. Under *individual or centralized production*, both tasks are implemented by the same generalist who has a finite attention capacity τ to learn about those shocks, but faces no coordination problems. As shown in Proposition 8, this generalist optimally devotes an equal amount of attention to each task.

Intuitively, individual production performs well when coordination is important and poorly 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, 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 focused 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 constrained

Proposition 9 *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 focused 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 focused team is optimal for $\tau < \bar{\mathsf{T}}(\beta)$ and individual/centralized production if $\tau \geq \bar{\mathsf{T}}(\beta)$.*

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

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 in our two agent model. Under decentralized decision-making, these division managers take the relevant actions, and the communication capacity τ is used to ensure coordination. Instead under centralized decision-making, 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 τ to communicate with both division managers and learn about them.

Proposition 9 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 Rantakari (2008), the impact of attention capacity is novel in our view. Among other things, our model thus predicts that improvements in IT, which relax attention constraints, should result in more centralized decision-making.

6.3 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 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 θ_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 others.

For $n = 2$, pay-offs are then equivalent to 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. (2011). For conciseness, we consider the case of two agents, though our results can also be generalized to $n > 2$ 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, $\mathbf{q} = [q_1, q_2]$, the realized profit of the organization is:

$$\pi(\mathbf{q}|\boldsymbol{\theta}) = K - (q_1 - \theta_1)^2 - (q_2 - \theta_2)^2 - \beta(q_1 - q_2)^2, \tag{32}$$

where β 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 the local shock θ_i and a message m_j about the local shock θ_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 4, one can show that expected profits can be expressed as

$$E[\pi(\mathbf{q}|\boldsymbol{\theta})] = (\Omega(t_1) - 1)\sigma_\theta^2 + (\Omega(t_2) - 1)\sigma_\theta^2, \tag{33}$$

where

$$\Omega(t_i) = \frac{\operatorname{cov}(q_i(t_i), \theta_i)}{\sigma_\theta^2} = \frac{(1+\beta)\sigma_\theta^2}{\sigma_\theta^2(1+2\beta) + \beta^2 \mathsf{RV}(t_2)} \in [0,1] \tag{34}$$

captures the *adaptiveness of task i to its task-specific information*. The only difference with Section 4 is that q_i is less adaptive to the local information θ_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 4 . Propositions 1 and 2 of Section 4 therefore follow immediately.

7 Conclusions

We have proposed a model of team production where the specialization of organizational knowledge arises endogenously as a response to organizational trade-offs between adaptation and coordination. Our insight is that if the ability to communicate is limited, an organization wants to direct its limited attention to a few agents, referred to as leaders, who adapt their task to local shocks. All other agents are forced to largely ignore their information and implement their tasks in a rigid manner in order to maintain coordination. The organization is biased towards a few tasks, those executed by leaders whose information is the only one that gets embedded in the production process.

An interesting byproduct of our model is that it shows how an increase in communication capacity can be "traumatic" for the existing leadership of the organization. As communication capacity increases, overall adaptiveness increases, but so can be the number of leaders. Given that attention is shared equally among the leaders of the organization, this often implies that each individual leader receives less attention. This effect can be particularly pronounced when there is a move away from a one-leader organization to a setting with multiple leaders.

There are several directions for future research. For instance, in our framework organizational focus is the result of solving optimal communication flows in a team production context with adaptation-coordination trade-offs. A remaining challenge is to embed these theories in a competitive framework, where specialization can also be the result of both product market competition and the extent of the market. Also, incentives play no role in our theory but if agents have private benefits associated with adaptation and if incentives have to be provided to gather information, then organizational focus can be a tool to relax incentive problems.

References

Angeletos, George-Marios and Alessandro Pavan (2007). "Efficient Use of Information and Social Value of Information." *Econometrica*, 75 (4), 1103-1142

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

Alonso, Ricardo, Wouter Dessein and Niko Matouschek (2012). "When does Adaptation Require Decentralization?" Mimeo.

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.

Bolton, Patrick, Markus Brunnermeier and Laura Veldkamp (2008). "Leadership, Coordination and. Mission-Driven Management." Mimeo.

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.

Cremer, Jacques (1980). "A Partial Theory of the Optimal Organization of a Bureaucracy." *Bell Journal of Economics*, 11(2), 683-693.

Cremer, Jacques (1993). "Corporate Culture and Shared Knowledge." *Industrial and Corporate Change,* 2 (3), 351-386.

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

Dewatripont, Mathias, Ian Jewitt, Ian and Jean Tirole (1999). "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies." *The Review of Economic Studies*, 66 (1), 199-217.

Galeotti, Andrea, Christian Ghiglino and Francesco Squintani (2011). "Strategic Information Transmission in Networks." Mimeo.

Garicano, Luis (2000). “Hierarchies and the Organization of Knowledge in Production.,” *Journal of Political Economy,* 108, 874-904.

Garicano, Luis and Timothy Van Zandt (2012). “Hierarchy: Decentralized Coordination in Organizations.” *Handbook of Organizational Economics* (forthcoming) eds. R. Gibbons and J. Roberts. Princeton University Press.

Geanakoplos, John, and Paul Milgrom (1991). “A Theory of Hierarchies Based on Limited Managerial Attention.” *Journal of the Japanese and International Economies*, 5(3), 205–225.

Guadalupe, Maria and Julie Wulf (2012). “Who Lives in the C-Suite? Organizational Structure and the Division of Labor in Top Management.” Mimeo.

Hagenbach, Jeanne and Frederic Koessler (2010). ”Strategic Communication Networks.” *Review of Economic Studies*, 77(3), 1072-1099

Hellwig, Christian and Laura Veldkamp (2009). ”Knowing What Others Know: Coordination Motives in Information Acquisition.” *Review of Economic Studies*, 76(1), 223-251

Hermalin, Ben (2007). ”Leadership and Corporate Culture.” *Handbook of Organizational Economics* (forthcoming) eds. R. Gibbons and J. Roberts, Princeton University Press.

Holmstrom, Bengt, and Paul Milgrom (1991). ”Multitask principal-agent analysis: incentive contracts, asset ownership, and job design.’ *Journal of Law, Economics and Organization,* 7, 24-52.

Holmstrom, Bengt and Paul Milgrom (1994). ”The firm as an incentive system.” *American Economic Review,* 84, 972-991.

Marschak, Jacob and R. Radner (1972). *Economic Theory of Teams.* Yale University Press.

Morris Stephen, and Hyun Song Shin (2002). ”Social Value of Public Information.” *American Economic Review*, 92(5), 1521-1534.

Morris Stephen, and Hyun Song Shin (2007). ”Optimal Communication.” *Journal of the European Economic Association*, Vol 5 (2-3), 594-602.

Myatt, David and Chris Wallace (2012). ”Endogenous Information Acquisition in Coordination Games.” *Review of Economic Studies*, Vol 79 (1), 340-374.

Prahalad, C.K. and Gary Hamel (1990). "The core competence of the corporation." *Harvard Business Review*, 68, 79-91.

Prat, Andrea (2002). "Should a Team be Homogenous?" *European Economic Review*, 46(7), 1187–1207.

Porter, Michael (1985). *Competitive Advantage: Creating and Sustaining Superior Performance*. New York: Free Press

Porter, Michael (1996). "What is Strategy?" *Harvard Business Review*, 74, 61-78

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

Rotemberg, Julio, and Garth Saloner (1994). "Benefits of Narrow Business Strategies." *American Economic Review*, 84 (5), 1330-1349.

Rotemberg, Julio, and Garth Saloner (2000). "Visionaries, Managers, and Strategic Direction." *RAND Journal of Economics,*31 (4), 693-716

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

Sims, Christopher (2006). "Rational Inattention: Beyond the Linear-Quadratic Case." *American Economic Review*, 96 (2), 158–163.

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

Van den Steen, Eric (2005). "Organizational Beliefs and Managerial Vision." *The Journal of Law, Economics, and Organization*, 21 (1), 256-283.

Van den Steen, Eric (2012). "A Theory of Strategy and the Role of Leaders in it." Mimeo.

APPENDIX

Appendix A: Proofs of the propositions and lemmas

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 (11) with respect to t we obtain

$$\frac{\partial E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial t} = -\beta\left[\Omega_1(t)\mathsf{RV}'\left(t\right) - \Omega_2(\tau - t)\mathsf{RV}'\left(\tau - t\right)\right]. \tag{35}$$

Substituting the expression for $\Omega_i(\cdot)$ given by 10, we have

$$\frac{\partial E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial t} = -\beta\left[\frac{\mathsf{RV}'(t)}{[\sigma_\theta^2 + \beta\mathsf{RV}(t)]^2} - \frac{\mathsf{RV}'(\tau - t)}{[\sigma_\theta^2 + \beta\mathsf{RV}(\tau - t)]^2}\right]. \tag{36}$$

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 (11) with respect to t is given by expression (36). 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 β. 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 8. Recall that a single agent chooses actions as in (30). 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 8. ■.

Proof of Proposition 3. Proposition 3 follows as a consequence of the combination of Lemma 4 and Lemma 5. We now provide the proof of the two Lemmas.

Proof of Lemma 4. 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 ϵ. Using the expression for expected payoffs 23 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 ϵ sufficiently small, contradicts our initial hypothesis that $t_{ji} > t_{ki}$. The completes the proof of Lemma 4. ■

Proof of Lemma 5. 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 ϵ. 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

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

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

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

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

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 $\beta^2 e^{-2(t_i+t_j)} - 1 < 0$. We have then reached a contradiction. This completes the proof of Lemma 5. ■

The combination of Lemma 4 and Lemma 5 completes the proof of Proposition 3. ■

Proof of Proposition 6. Using the expression for expected payoffs (23), the fact that $\mathsf{RV}(t) = \sigma_\theta^2 e^{-2t}$, and that organization $\mathbf{t}$ is an ℓ-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 β for all ℓ,τ,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 β and decreasing in L, the following observation follows:

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

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 τ,n; (3b) $\underline{\beta}(\tau,n)$ is increasing in τ.

Observation 3a together with observation 2b imply that $\frac{dE[\pi(\mathbf{q},\mathbf{t}|\boldsymbol{\theta})]}{d\ell}$ is declining in ℓ 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 ℓ-leader organization 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 τ.

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 τ and n; 4b. $\bar{\beta}(\tau,n)$ is increasing in τ.

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 ℓ and then increasing in ℓ (when $\beta \in [\bar{\beta}(\tau,n),\hat{\beta}]$) or it is always increasing in ℓ (when $\beta > \hat{\beta}]$). Hence, the profits of the organization are decreasing in ℓ 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 ℓ 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 ℓ (when $\beta \in [\underline{\beta}(\tau,n),\tilde{\beta}]$) or they are first decreasing in ℓ and then increasing in ℓ (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 ℓ^* is the solution to $\Phi(\ell^*,\beta,x,n) = 0$ and, ℓ^* maximizes 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 ℓ^* 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 ℓ^*.

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 ℓ^* is increasing in β, 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 β_ℓ, such that at $\beta = \beta_\ell$ the expected profit of the ℓ-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 ℓ-leader organization. Using expression 26, 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 β and therefore there are only two solutions of β. 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 6.

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 τ. 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 6. ∎

Proof of Proposition 9: From (31), (11) and (16), balanced team production dominates individual production if and only if

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

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

$$\tau \leq \mathsf{T}(\beta) = \ln\beta. \tag{37}$$

If $\beta < \bar{\beta} = 2$, then $\mathsf{T}(\beta) < \hat{\mathsf{T}}(\beta)$. Part (1) and (2) of Proposition 9 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 centralized production is optimal. If $\tau \in (\hat{\mathsf{T}}(\beta), \mathsf{T}(\beta))$, focussed team production dominates individual production if, and only if,

$$\frac{\beta}{1+\beta} + \frac{\beta}{1+\beta\mathsf{e}^{-2\tau}}\mathsf{e}^{-2\tau} - 2\mathsf{e}^{-\tau} < 0. \tag{38}$$

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 τ 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 centralized production is optimal. This implies part (3) and (4) of Proposition 9. ■

Appendix B: Alternative communication models

Result 1. Under public communication and capacity constraint τ, an optimal communication network $\mathbf{t} = \{t_1, .., t_n\}$ satisfies

$$t_j = t_{ij}^b \textit{ foralli}, j \in \mathcal{N}$$

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

Proof of Result 1 Note that under bilateral communication and arbitrary capacity τ^b, Lemma 4 implies that the optimal network $\mathbf{t}^b$ satisfies $t_{ji}^b = t_{li}^b$ for all $j, l \neq i$. Hence, in the optimal communication network every agent $j \neq i$ devotes the same attention to agent i, that is the restriction imposed by public communication. It is immediate to see the relation between τ^b and τ. ■

Result 2. Under individual communication and individual capacity constraint τ, an optimal communication network $\mathbf{t} = \{t_i j\}_{i,j}$ satisfies

$$t_{jj} = t_{ij} = t^b_{ij} \ foralli, j \in \mathcal{N}$$

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

Proof of Result 2. Consider the case of individual communication with individual capacity constraint τ. Suppose that $\mathbf{t}$ is an optimal organization. It is immediate to see that $\mathbf{t}$ satisfies: a. $t_{ji} \leq t_i i$ for all $i, j \in \mathcal{N}$ and b. $\sum_j t_{ji} = \tau$ for all $ij \in \mathcal{N}$. Now note that if $\tau^b = (n-1)\tau$, $\mathbf{t}^b$ is an optimal organization under bilateral communication and constraint τ^b, then organization $\mathbf{t}^*$ with $t^*_{ji} = t^*_{ii} = t^b_{ji}$ is a feasible organization under individual communication and satisfies property a. and b. above. We now claim that $\mathbf{t}^*$ is optimal under individual communication and individual capacity constraint τ. Suppose there is another organization $\mathbf{t}$ that does strictly better than $\mathbf{t}^*$. First, $\mathbf{t}$ must satisfy property a and property b and therefore $\min\{t_{ji}, t_{ii}\} = t_j i$, and so the residual variance that agent j has about task i is $RV(t_{ji})$. Since $\mathbf{t}$ is strictly better than $\mathbf{t}^*$ is follows that the profile of residual variances $\{RV(t_j i)\}_{ji}$ is better than $\{RV(t^*_{ji})\}_{ji}$. But then, construct $\hat{\mathbf{t}}^b$ as follows: $\hat{t}^b_{ji} = t_{ji}$. Note that $\hat{\mathbf{t}}^b$ is feasible under bilateral communication and capacity τ. Furthermore since the profile of residual variances $\{RV(t_j i)\}_{ji}$ is better than $\{RV(t^*_{ji})\}_{ji}$, it must also be true that profile of residual variances $\{RV(\hat{t}^b_{ji})\}_{ji}$ is better than $\{RV(t^b_{ji})\}_{ji}$, and so $\hat{\mathbf{t}}^b$ must be strictly better than $\mathbf{t}^b$, which contradicts our initial hypothesis that $\mathbf{t}^b$ is an optimal network. ■

Appendix C: 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:

$$\pi(\mathbf{q}|\boldsymbol{\theta}) = K - (q_1 - \theta_1)^2 - (q_2 - \theta_2)^2 - \beta(q_1 - q_2)^2, \tag{39}$$

where β is some positive constant. As in the model developed in our paper, agent i has information set $\mathcal{I}_i$ that contains the local shock θ_i and a message m_j about local shock θ_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:

$$q_1 = \frac{1}{1+\beta}\left[\theta_1 + \beta E\left[q_2|\mathcal{I}_1\right]\right] \tag{40}$$

$$q_2 = \frac{1}{1+\beta}\left[\theta_2 + \beta E\left[q_1|\mathcal{I}_2\right]\right] \tag{41}$$

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 (40) and (41) as

$$q_1 = a_{11}(t_1)\,\theta_1 + a_{12}(t_2)E[\theta_2|\mathcal{I}_1] \tag{42}$$

$$q_2 = a_{22}(t_2)\,\theta_1 + a_{21}(t_1)E[\theta_1|\mathcal{I}_2] \tag{43}$$

Substituting the guess (42) and (43) into (40) and (41), and using Assumption A, we find that the equilibrium actions are

$$q_1 = \frac{(1+\beta)\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(t_1)}\theta_1 + \frac{\beta\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(t_2)}E[\theta_2|\mathcal{I}_1] \tag{44}$$

$$q_2 = \frac{(1+\beta)\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(t_2)}\theta_2 + \frac{\beta\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(t_1)}E[\theta_1|\mathcal{I}_2] \tag{45}$$

Finally substituting (44) and (45) into (39) 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 τ 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 τ.