# Rational Inattention and Organizational Focus

Wouter Dessein
Columbia University

Andrea Galeotti
University of Essex

Tano Santos
Columbia University

March 6, 2013

**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 that when attention is scarce, organizational focus and leadership naturally arise as a response to organizational trade-offs between coordination and adaptation. At the optimum, all attention is evenly allocated to a select number of "leaders." 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," as well as new trends in organization design. We also derive implications for the optimal size or "scope" of organizations. Surprisingly, improvements in communication technology may result in smaller but more adaptive organizations.

# 1 Introduction

Economics, according to Lionel Robbins's famous dictum, is "the science which studies human behavior as a relationship between ends and scarce means which have alternative uses." As emphasized by Herbert Simon, attention may well be the ultimate scarce resource in the economy: "a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it," (Simon 1971, pp. 40–41). This paper studies the optimal allocation of (scarce) attention in organizations. We show that when attention is scarce, organizational focus and leadership naturally arise as a response to organizational trade-offs between coordination and adaptation. As such, our results provide micro-foundations for a central idea in the management literature that firms should focus on a limited set of "core competences" (Prahald and Hamel 1990)[1] and firms that aim to be "all things to all people," will be "caught in the middle" and fail (Porter 1985;1996). We develop comparative statics as to when organizational focus is more or less important, and shed light on new trends in organizational design including more shared leadership, an increase in horizontal communication linkages, and a reduction in the overall size of the organization (downsizing, downscoping).[2]

Our starting point is that organizations have a limited communication capacity and must use this resource judiciously in order to coordinate their production processes.[3] 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. Alternatively, different tasks may correspond to different products or locations (e.g. a multi-product firm which exploits economies of scope). Each agent is in charge of one task and must adapt this task to local information or "shocks". Such adaptation, however, results in coordination failures with other tasks unless agents communicate effectively. *Organizational focus,* then, takes the form of allocating more scarce organizational attention to one task – or one agent – than another task.

[1]According to Prahalad and Hamel, such core competences represent the "collective learning in the organization, especially how to coordinate diverse production skills" (p. 82)

[2]See Whittington et al. (1999), Guadalupe et al. (2012), and Roberts and Saloner (2013).

[3]As Arrow (1974, p.53) stated "The information has to be coordinated if it is to be of any use to the organization. More formally stated, communication channels have to be created within the organization."

If attention is abundant there is no need for organizational focus in our model. All tasks can then be very adaptive and well coordinated, and it is optimal to distribute attention evenly. In contrast, if attention is scarce and coordination is important, 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.[4] 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 support the notion that firms need to have a clear strategy – they must choose

[4]This correspond to the two general ways in which organizations can be coordinated according to March and Simon (1958): "The *type of coordination* used in the organization is a function of the extent to which the situation is standardized. (...) We may label coordination based on pre-established schedules *coordination by plan*, and coordination that involves transmission of new information *coordination by feedback*. The more stable and predictable the situation, the greater the reliance on coordination by plan." (p182)

a set of performance dimensions or tasks in the value-chain to focus on.[5] By the same token, we provide insights as to how focused firms should be. Over the last decades, there has been enormous technological innovations in communication and coordination technologies (e-mail, wireless communication and computing, intra networks, etc). Our results suggest that having a narrow focus becomes less important as information technology relaxes the communication and attention constraints of organizations. 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 agents. This is consistent with new trends in organizational design towards more network-like organizations where communication flows are horizontal rather than vertical, and decision-making and influence is broadly shared in the organization. Such novel organizations have been documented in both case studies (for example "Proctor & Gamble Organization 2005," HBS case 9-707-519)[6] and large scale empirical studies (Whittington et al. 1999, Guadalupe and Wulf, 2012).[7] Conversely, in fast-moving environments where speedy decisions are important – a competitive market place where reacting quickly to a competitor's move is of the essence, or a platoon in the field of battle – there is often little time for extensive communication.[8] Our model predicts that leadership and organizational focus is more important in such environments compared to settings where extensive communication is feasible prior to taking action.

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

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

[6]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 dominated P&G, slowing down the development and roll-out of new products.

[7]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.

[8]As argued by Roberts and Saloner (2013), increased competition, through globalization of markets and industries, "requires firms to be able to change more quickly and respond faster to market developments." (p822)

that, perhaps counter-intuitively, when some tasks are more interdependent than others and attention is relatively scarce, it is optimal not to focus attention on highly interdependent tasks, but instead restrict their adaptiveness.

Finally, in our basic model, the size or "scope" of the organization is fixed. We endogenize the number of tasks by introducing economies of scale or scope: certain fixed costs can be shared among tasks (e.g. production facilities or a distribution network), yielding benefits to size. The size of organizations, however, is limited by the need for coordination and limited organizational attention. When the number of tasks is endogenous, we uncover an important trade-off between organizational *size* and organizational *focus*. As we show, smaller organizations have more leaders and, hence, the information of more agents is reflected in decision-making. As an organization grows larger, leadership becomes more concentrated as there is more need for coordination. This is consistent with the experience of many entrepreneurial firms, whose culture of joint-decision-making and open lateral communication often disappears as they grow bigger and more hierarchical.[9]

Two empirically relevant drivers of organizational size are the volatility of the environment and changes in information and communication technologies. Consistent with recent trends in "de-scoping"(Whittington et al. 1999), we show that the optimal scope of organizations *decreases* as the environment becomes more volatile and adaptation becomes more important.[10] Intuitively, by reducing the number of tasks that it undertakes, the organization reduces its coordination needs, hence allowing for better adaptation. At the same time, the number of tasks that receive attention increases. Hence, as the environment becomes more volatile, there is a move from large, focussed organizations that maximize scope economies to smaller, but more adaptive and balanced organizations. Improvements in information technology might be conjectured to always increase the size of organizations, as they allow for better coordination. Interestingly, we show that information technology has a decidely ambiguous impact on firm scope. Intuitively, information technology makes it optimal for organizations to shift towards

[9]In a classic management article, Greiner (1972) discusses how young, creative organizations often face a crisis of leadership as they grow bigger. According to Greiner, an initial *phase of creativity* must give way to a *phase of direction*, where companies that continue to grow, do so under "able, directive leadership".

[10]As noted by Siggelkow and Rivkin (2005), "rapid technological change, deregulation, and globalization have intensified competition and increased the turbulence that managers face, forcing them to adopt new, more responsive organizational forms." (p101) See also footnote 8.

a strategy that emphasizes adaptation to its environment, but smaller firms with shared leadership are better configured to do so. Hence, while for low levels of information technology, large, non-adaptive firms exploiting economies of scale are optimal, for intermediate levels of information technology, smaller, more flexible firms are often preferred.

*Modeling attention and organizational knowledge.* A necessary ingredient for our results is that attention is constrained. 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 can be expressed in terms of the *entropy* of the (posterior) distribution of a local shocks. Communication reduces this entropy and the *mutual information* regarding a particular local shock is defined as the reduction in entropy following communication. A central result in information theory is that the total mutual information that can be achieved is given by the (finite) capacity of the communication channel. Following the literature on rational inattention, we model attention constraints as the finite capacity constraint of a communication channel. The reduction in the entropy regarding a particular task-specific shock, then, is given by the attention devoted to that shock. For normal distributions, our leading case, the entropy is proportional to the log of the variance.

An important and intuitive feature of the above communication technology is that it implies *decreasing marginal returns* in reducing the residual variance of a particular shock. While only limited communication capacity (attention) may be required to reduce the residual variance of a posterior when the latter is very noisy, it becomes increasingly costly (in terms of attention required) to reduce the residual variance when the posterior becomes more 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 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.* 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 endogenizes the number of tasks, with larger organizations exploiting economies of scope but facing more daunting coordination problems. We conclude in Section 7 by discussing the implications of our model for skill-heterogeneity and the entrenchment of leadership and functional cultures in organizations, as well as some missing elements such as power struggles and conflicts of interests. Proofs of Propositions, as well as some model extensions, 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.[11] 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).[12] 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 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.[13]

[11] 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.

[12] 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.

[13] 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.

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.[14] Beauty contest models are suitable, however, to study the characteristics of successful leaders. Dewan and Myatt (2008) show how the ability of a leader to convey her information clearly to followers is often more important than the precision of her information. Bolton, Brunnermeir and Veldkamp (2012) highlight the benefits of leader resoluteness in achieving coordination.[15] The present paper does not speak to the characteristics of successful leaders. Instead, we show how and when leadership arises endogenously in a team of ex ante identical agents.

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' (Rotemberger and Saloner, 1994, 2000) has argued that the commitment by a principal or leader to select a certain type of projects provides strong incentives for

Other differences are that agents are self-interested and invest in both active and passive communication.

[14]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.

[15]Similarly, Van den Steen (2005) shows how having leaders with strong beliefs may be desirable as they give direction to the firm by affecting the employee's choice of project.

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. We start with the two agent case, which is enough to convey many of the intuitions of the model, and leave for Sections 5 and 6 the case $n > 2$.

### 3.1 Two-task Production

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

In particular, Agent $i$ observes a piece of information $\theta_i$, a shock with variance $\sigma_\theta^2$ and mean 0, which is relevant for the proper implementation of the assigned task. We refer to $\theta_i$ as the local information of agent $i$. The realization of this local information is independent across agents. In order to achieve perfect adaptation, agent $i$ should set his primary action $q_{ii}$ equal to $\theta_i$. In order to achieve perfect coordination with task $j$, agent $i$ should set his complementary action $q_{ij}$ equal to $q_{jj}$, the primary action of agent $j$. If tasks are imperfectly adapted or coordinated, the organization suffers adaptation and/or coordination losses. Formally, let $q_i = [q_{i1}, q_{i2}]$ be the actions taken by agent $i$, with $i \in \{1, 2\}$. Given a particular realization of the 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:[16]

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

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

Expression (1) captures the notion, going back to at least March and Simon (1958), that it is adaptation to unpredictable contingencies which creates coordination problems: "(D)ifficulties arise only if program execution rests on contingencies that cannot be predicted perfectly in advance. In this case, coordinating activity is required (...) to provide information to each subprogram unit about the relevant activities of the others." (p. 180).[17] Expression (1) implies that coordination problems arise only if (i) the states of nature $\theta_1$ and $\theta_2$ are unpredictable (contingencies arise) *and* (ii) communication is imperfect.

### 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 $\theta_j$. We assume that the organization cannot devote an infinite amount of resources to communicate:

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

where $\tau < \infty$.[18] For example, $\tau$ can be the length of a meeting, and $t_1$ and $t_2$ the time that agent 1 and 2 are allowed to speak. We say that an organization is focused on task 1 whenever

[16] Appendix C considers an alternative model where $q_{ii} = q_{ij} = q_i$ and, hence, each agent undertakes only one action, which now must both be adapted to the task-specific shock and coordinated with the action of the other agent. As shown in Appendix C, qualitatively identical results obtain.

[17] March and Simon also emphasize the role of 'complementary actions' in achieving coordination: "To the extent that contingencies arise, not anticipated in the schedule, coordination requires communication to give notice of deviations from planned or predicted conditions, or to give instructions for changes in activity to adjust to these deviations." (p182).

[18] Assuming $\tau$ to be exogenous simplifies the analysis substantially. In Appendix D, we derive some results for when $\tau$ is endogenous.

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, if $t_1 = t_2 = \tau/2$.

### 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, $\theta_i$, as well as the message received from the other agent $j$, $m_j$. The degree of precision of message $m_j$ depends on $t_j$, that is the time or attention agents devote to communicate about local information $\theta_j$. In particular, we assume that agent $i$ receives a noisy message $m_j$, which is a random variable with mean zero, variance $\sigma_m^2$ and correlation

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

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

Assumption A is satisfied, for example, if messages and information are normally distributed or uniformly distributed (see example 1 and 2 below). More generally, it will be satisfied whenever $f(\theta_j|m_j)$ belongs to a family of statistical structures known as the Exponential family with conjugate priors, which includes many of the most common distributions (Diaconis and Ylvisaker, 1979).[19] Assumption A implies that[20]

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

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

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

[19] The normal, lognormal, exponential, gamma, and weibull distributions all belong to this family.

[20] 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 possessed by agents

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

**Assumption B.** For every $j = 1, 2$:

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

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

3B. There are limited resources for communication in that, for every communication network $\mathbf{t}$, total residual variance is strictly positive, i.e., $\tau < 2\hat{\tau}$.[21]

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

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

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

The fact that $\theta_j$ and $\epsilon_j$ are 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 $\theta_j$ 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

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

$E[\theta_j|m_j] = p(t_j)m_j$ and $E[m_j|\theta_j] = p(t_j)\theta_j$, and hence Assumption A holds. The residual variance is

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

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

In order to characterize optimal communication networks, additional assumptions are required on the functional form of $\mathsf{RV}(t)$. 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 $\mathsf{RV}'(\cdot) < 0$ but $\mathsf{RV}''(\cdot) > 0$. To highlight the intuition behind our results, however, it will be useful to first focus on a benchmark case where there are constant marginal returns to communication: $\mathsf{RV}''(\cdot) = 0$. The case where communication displays decreasing marginal returns to communication will be addressed in Section 4.3.

### 3.4 Timing

The timing of our model goes as follows:

1. *Organizational design*: Optimal communication network $\mathbf{t}$ is chosen.

2. Local information $\{\theta_i\}_{i=1,2}$ is 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

## 4.1 Actions and the expected profits of the organization

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

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

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

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

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

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

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 $\theta_1$. 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, $\beta$. As $\beta$ goes to 0, tasks become perfectly adaptive for any level of attention $t_i$. In contrast, as $\beta$ goes to infinity, task $i$ becomes unresponsive to its information unless attention is perfect ($t_i \geq \hat{\tau}$) and $\mathsf{RV}(t_i) = 0$.

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

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

where

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

neatly captures the *adaptiveness of task $i$ to its task-specific information.* When the organization is fully adaptive, that is $\operatorname{cov}\left(q_{ii},\theta_i\right)=\sigma_\theta^2$, the expected profits are maximized and $E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]=0$. From (8), however, a limited attention capacity $\tau<2\hat{\tau}$ imposes limits to adaptation such that $\operatorname{cov}(q_{ii},\theta_i)<\sigma_\theta^2$ and $E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]<0$.

An alternative representation of the expected profit function is[22]

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

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 (10), 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 $\theta_i$ 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

[22]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.

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 $\alpha$. 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 $t_1^* \in \{0, \tau\}$ and $t_2^* = \tau - t_1^*$. 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 $t_1^* \in \{\tau - \hat{\tau}, \hat{\tau}\}$ and $t_2^* = \tau - t_1^*$.*

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.* The notions of ex-ante and ex-post coordination correspond to the two general ways in which organizations can be coordinated according to March and Simon (1958): *coordination by plan*, which requires that tasks are executed in a more or less standardized way (e.g. following standard operating procedures), and *coordination by feedback*, which involves the transmission of new information. As March and Simon note "it is possible to reduce the volume of communication required from day-to-day by substituting coordination by plan for coordination by feedback" (p.183).

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, it is optimal for both tasks to be very adaptive, as they both 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. We draw on well established ideas from Information Theory to build a tractable model.

#### 4.3.1 Information Theory

In order to micro-found our communication technology, we now posit that the quantity of information that can be conveyed by $m = (m_1, m_2)$ about the state of nature $\theta = (\theta_1, \theta_2)$ is limited by the capacity of a noisy communication channel, as in the literature on Rational Inattention (Sims 2003). Following this approach, the quantity of information conveyed by $m = (m_1, m_2)$ is measured by Shannon's (1948) concept of *mutual information*. Formally, the mutual information between $m$ and $\theta$, denoted by $I(\theta; m)$, equals the average amount by which the observation of $m$ reduces uncertainty about the state $\theta$, where the ex ante uncertainty is measured by the (differential) *entropy* of $\theta$,

$$H(\theta) = -\int f(\theta) \log f(\theta) d\theta,$$

and the uncertainty after observing $m$ is measured by the corresponding entropy

$$H(\theta|m) = -\int f(\theta|m) \log f(\theta|m) d\theta.$$

Denoting by $\tau$ the (Shannon) capacity of the communication channel, the constraint on information conveyed by $m$ about $\theta$ is given by[23]

$$I(\theta; m) = H(\theta) - H(\theta|m) \leq \tau. \tag{14}$$

Following Sims (2003) and the subsequent literature on rational inattention, we will assume that $\theta_1$ and $\theta_2$ are (independently) normally distributed, and communicated through a Guassian communication channel which contaminates its inputs with independent normally distributed noise, as in Example 1 of Section 3. As a result, also $m_1$ and $m_2$ and the conditional distributions $F(\theta_1|m_1)$ and $F(\theta_2|m_2)$ are independently normally distributed. As argued by Sims, Guassian communication channels minimize the variance of $F(\theta_i|m_i)$ given the constraint (14) on the mutual information between $\theta_i$ and $m_i$. Hence, they maximize the correlation between $m_i$ with $\theta_i$. [24] Given that $\theta_1$ and $\theta_2$ are independently distributed, we have

$$I(\theta; m) = I(\theta_1; m_1) + I(\theta_2; m_2), \tag{15}$$

where $I(\theta_i; m_i) = H(\theta_i) - H(\theta_i|m_i)$. Moreover, since the entropy of a normal variable with variance $\sigma^2$ is given by $\frac{1}{2}\ln(2\pi e \sigma^2)$, we obtain

$$I(\theta_i, m_i) = \frac{1}{2}\left(\ln \sigma_\theta^2 - \ln \mathrm{Var}(\theta_i|m_i)\right). \tag{16}$$

[23] The capacity of a channel is a measure of the maximum data rate that can be reliably transmitted over the channel. 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 an 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 (e.g. there is a finite number of bits that can be reliably transmitted).

[24] This follows from a well known result in information theory that among all distributions with the same level of entropy, the normal distribution minimizes the variance.

It follows that the constraint (14) on the mutual information between $\theta$ and $m$ can be rewritten as

$$\ln \sigma_\theta^2 - \ln \mathrm{Var}(\theta_1|m_1) + \ln \sigma_\theta^2 - \ln \mathrm{Var}(\theta_2|m_2) \leq 2\tau. \tag{17}$$

We can now re-interpret the *mutual information between* $m_i$ *and* $\theta_i$ as the *attention devoted by the organization to task* $i$. Denoting $t_1 \equiv I(\theta_1, m_1)$ and $t_2 \equiv I(\theta_2, m_2)$, the constraint on mutual information (14) imposed by the Shannon capacity becomes equivalent to our attention constraint $t_1 + t_2 \leq \tau$. Given an upperbound $\tau$ on the mutual information of $m = (m_1, m_2)$ and $\theta = (\theta_1, \theta_2)$, the organization designer then decides whether to focus the channel capacity mainly on one task, or to allocate capacity equally to both tasks. A focused organization has $t_i = I(\theta_i, m_i) > t_j = I(\theta_j, m_j)$, with complete focus being characterized by $I(\theta_i, m_i) = \tau$ and $I(\theta_j, m_j) = 0$. A balanced organization, in contrast, has $I(\theta_1, m_1) = I(\theta_2, m_2) = \tau/2$.

Using the above formalization, we obtain a tracteable expression for $\mathsf{RV}(t_i) \equiv Var(\theta_i|m_i)$. Indeed, from (16) and $t_i \equiv I(\theta_i, m_i)$, we have

$$\ln \mathsf{RV}\,(t_i) = \ln \sigma_\theta^2 - 2t_i, \qquad i = 1, 2. \tag{18}$$

or still

$$\mathsf{RV}\,(t_i) = \sigma_\theta^2 e^{-2t_i}, \qquad i = 1, 2, \tag{19}$$

where $t_1 + t_2 \leq \tau$. As noted by Sims (2003) and as is apparent from (19), scaling up or down the variance of the input does not result in a higher or lower correlation between the input and the received message $m_i$. We would therefore obtain identical results if agents were to communicate their actions $q = (q(\theta_1), q(\theta_2))$ rather than the local shocks $\theta = (\theta_1, \theta_2)$ to whom those actions are adapting. In contrast, in a traditional signal extraction model where the signal is contaminated by some exogenous noise, it would be optimal to communicate $\theta_1$ rather than $q(\theta_1)$, as $\theta_1$ has a larger variance and is therefore less distorted when communicated.

An important and intuitive feature of communication technology (18) 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. Indeed, if it takes $\Delta t$ to reduce the residual variance from $\sigma_\theta^2$ to $\sigma_\theta^2/2$, it will take an additional $\Delta t$ to reduce the 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 $\left|\mathsf{RV}'(t_i)\right| = 2\mathsf{RV}(t_i)$, hence the lower the residual variance, the lower the marginal returns to further reduce this variance.

While we have derived the communication technology (18) using foundations in information theory, it should be noted that the same expression for the residual variance can also be micro-founded using a more standard approach. Assume, for example, that a total of $\tau$ signals can be transmitted between the two agents, and let $t_i$ be the number of signals about $\theta_i$ that is sent from agent $i$ to agent $j$. If each signal about $\theta_i$ reveals the realization of $\theta_i$ with an independent probability $p$, and reveals a value of 0 otherwise, then

$$\mathsf{RV}(t_i) \equiv E\left[Var(\theta_i | t_i \text{ signals about } \theta_i)\right] = \sigma_\theta^2 (1-p)^{t_i}$$

or still

$$\mathsf{RV}(t_i) = \sigma_\theta^2 e^{-\lambda t_i}, \qquad \text{with } \lambda = \ln(1-p)^{-1} > 0,$$

Obviously, this communication technology yields the same decreasing marginal returns to communication as rational inattention, up to a scaling factor $\lambda$. Intuitively, the expected value of each individual signal about $\theta_i$ is decreasing in the total number of signals an agent will receive about $\theta_i$. If $t_i$ is continuous rather than discrete, the above communication technology can be viewed as a Poisson process where $\lambda$ is the constant hazard rate that the receiver correctly learns the local shock of the sender.[25]

#### 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. The more interdependent are tasks, that is the larger is $\beta$, 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

[25]While the residual variance is now probabilistic (it will either be 0 or $\sigma_\theta^2$), one can verify that the optimal actions and equilibrium profits are still given by expressions (8) and (9). Hence the analysis remains indentical.

$(t_1, t_2) = (\tau, 0)$ is a *local maximum* if, and only if,

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

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 ($\beta$ small) or attention is not very constrained ($\tau$ large), a focused organization with $(t_1, t_2) = (\tau, 0)$ is suboptimal.

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

Next, consider the case where $\tau$ is large. When attention is relatively unconstrained, there are strongly decreasing marginal returns to center all communication around one task. Hence, for $\tau$ 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 $\tau$ goes to infinity, whereas $\Omega(0)$ is strictly positive, it follows again that $(\tau, 0)$ is not a local maximum for $\tau$ sufficiently large.

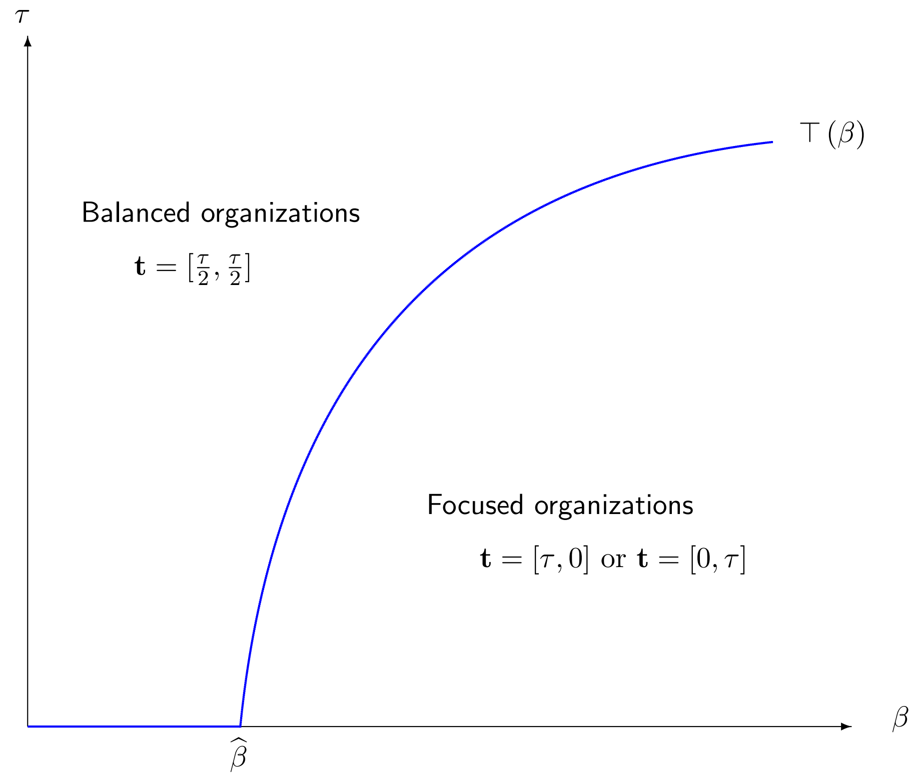
In line with the above intuitions, the following proposition shows that a fully focused organization is optimal if and only if coordination is sufficiently important and attention sufficiently scarce:

**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, $t_1^* \in \{0, \tau\}$ and $t_2^* = \tau - t_1^*$, if and only if $\tau \leq \top(\beta)$*

Figure 1: **Focused and balanced organizations in the two-agent case.**




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

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

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 $\tau$. Improvements in the communication technology (email, wireless communication devices, intranets, ...) can be interpreted as an exogenous increase in $\tau$. An implication of Proposition 2, therefore, is that such technological improvements result in a shift from 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.

### 4.4 Task Asymmetries

So far we have shown that organizational focus may be optimal even when both tasks (and both agents) are ex-ante identical. 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 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. Should attention be focused on those highly interdependent tasks? In this section we show that this is not necessarily the case.

Let the coordination parameters be $\beta_1$ and $\beta_2$ for task 1 and 2, respectively.[26] Define $\beta = \sqrt{\beta_1 \beta_2}$, the geometric mean of $\beta_1$ and $\beta_2$ and consider situations where

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

The parameter $\epsilon$ 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

[26]Hence, profits equal $-(q_{11} - \theta_1)^2 - (q_{22} - \theta_2)^2 - \beta_1(q_{21} - q_{11})^2 - \beta_2(q_{12} - q_{22})^2$

3. Maintaining the assumption of decreasing marginal returns to attention, as characterized by expression (18), we can prove the following result.

**Proposition 3** *Assume $\beta_1 > \beta_2 \geq \widehat{\beta}$, then:*

1. *If $\tau < \ln \beta$, the optimal organization is focused on task 2, i.e., $(t_1^*, t_2^*) = (0, \tau)$.*
2. *If $\tau \geq \ln \beta$, let $\widehat{\epsilon}$ be the solution to $(1+\widehat{\epsilon})^2 e^{-2\tau} = 1$ :*
   - *(a) If $\epsilon < \widehat{\epsilon}$ then $\tau > t_1^* > t_2^* > 0$.*
   - *(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 the task which is *least* interdependent: Task 2. The reason is that allocating limited attention to task 1 is essentially not worth it as it would translate into limited adaptation given the large coordination costs the organization would bear. Instead, it is better to provide all attention to task 2 and let task 2 be adaptive. Task 1 is then coordinated by restricting its adaptiveness.

Instead when the attention capacity is larger and the asymmetry $\epsilon$ is not too large, both tasks receive attention but task 1 receives more than task 2. Intuitively, if both tasks are allowed to be adaptive, more attention needs to be devoted to that task that is more interdependent. 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.

In sum, 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.

## 5 Organizational focus with many agents

We now extend our analysis to allow for an arbitrary number of agents in the team. We first characterize the optimal network when communication among agents is bilateral, as this allows for the greatest flexibility. Our main result 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. In section 5.4, we then show how we obtain the same result when communication is public. Throughout, and in the interest of brevity, we assume that the communication technology features decreasing marginal returns as characterized by expression (18).

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

$$q_i = [q_{i1}, q_{i2}..., q_{ii}, ..., q_{in}], \tag{21}$$

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 profile of actions; the realized profit of the organization is:

$$\pi(\mathbf{q}|\boldsymbol{\theta}) = -\sum_{i \in \mathcal{N}} \left[ (q_{ii} - \theta_i)^2 + \beta \sum_{j \in \mathcal{N} \setminus \{i\}} (q_{ii} - q_{ji})^2 \right]. \tag{22}$$

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 $\varepsilon_{ij}$ a random noise term. As in the two-agents case, we draw upon information theory and posit that communication constraints stem from a finite (Shannon) communication capacity $\tau$. Let $\theta_j$ and $m_{ij}$, for all $i, j \in \mathcal{N}$, be normally distributed, and let $t_{ij}$ be the mutual information between $m_{ij}$ and $\theta_j$, then, as in (18),

$$\ln \mathsf{RV}(t_{ij}) \equiv \ln Var(\theta_j | m_{ij}) = \ln \sigma_\theta^2 - 2t_{ij}, \tag{23}$$

where the communication constraint is given by

$$\sum_{j \in \mathcal{N}} \sum_{i \in N \setminus \{i\}} t_{ij} \leq \tau. \tag{24}$$

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

For a given network $\mathbf{t}$ and string of observed messages $m_i$, agent $i$ chooses the string of actions $q_i$, given in (21), 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 (22) 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

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

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 $n > 1$ :

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

where $\mathsf{RV}(t_{ji}) \equiv Var(\theta_j|m_{ij})$ is given by (23). Taking into account the equilibrium actions, we find that expected profits are given by

$$\begin{aligned} E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right] &= \sum_{i\in\mathcal{N}} \operatorname{cov}\left[(q_{ii},\theta_i) - \sigma_\theta^2\right] \\ &= -n\sigma_\theta^2 + \sigma_\theta^2\sum_{i\in\mathcal{N}}\frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta\sum_{j\neq i}\mathsf{RV}(t_{ji})}. \end{aligned} \tag{25}$$

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

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

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

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

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

To see the intuition behind Lemma 6, let $i$ and $j$ be two tasks with $\hat{t}_i = \sum_s t_{si}$ be the total attention devoted to task $i$ and $\hat{t}_j = \sum_s t_{sj}$ the total attention devoted to task $j$. Moreover, assume $\hat{t}_i > \hat{t}_j > 0$, in violation of Lemma 6. In the case of two tasks, 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*, 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 Comparative statics of the of the $\ell-$ leader organization

When the communication network takes the form of an $\ell-$leader organization, the expression of the profit function (25) can be re-written as:

$$E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right] = -n\sigma_\theta^2 + \ell * \Omega\left(\tfrac{\tau}{(n-1)\ell}\right)\sigma_\theta^2 + (n-\ell) * \Omega\left(0\right)\sigma_\theta^2, \tag{26}$$

where $\Omega(t_i)$ is the adaptiveness of task $i$ :

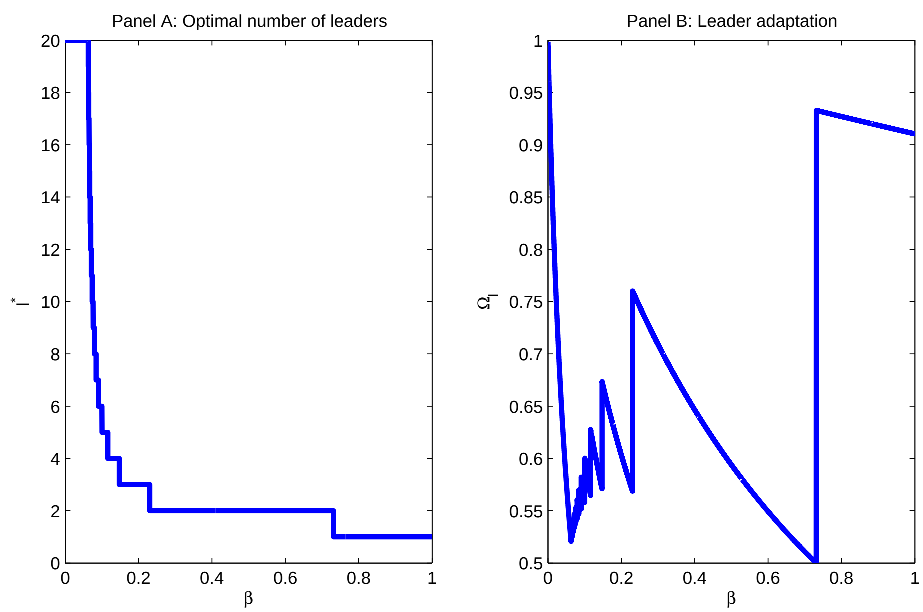
$$\Omega(t_i) = \frac{cov\left(q_{ii}, \theta_i\right)}{\sigma_\theta^2} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + (n-1)\beta \mathsf{RV}\left(t_i\right)}.$$

The optimal number of leaders, then, is given by

$$\ell^* = \mathsf{argmax}_{\ell \in \{1,2,\cdots,n\}} E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]. \tag{27}$$

Figure 2: **Optimal number of leaders and adaptation as a function of $\beta$**

Example: $n = 20$, $\sigma_\theta^2 = 1$, $\tau = 50$, and $\beta \in [0, 1]$. Panel A: Optimal number of leaders, $\ell^*$, as a function of the importance of coordination $\beta$. Panel B: Leader adaptation $\Omega_\ell$.




Armed with (27) we are able to offer a sharp characterization of the $\ell-$leader organization as a function of the organization's communication capacity $\tau$ and the task-interdepence or coordination parameter $\beta$.

**Proposition 7** *There exists* $0 < \overline{\beta}(n) < \ldots < \overline{\beta}(\ell + 1) < \overline{\beta}(\ell) < \ldots < \overline{\beta}(2)$ *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* $\ell \in \{1, \ldots, n\}$, $\overline{\beta}(\ell)$ *is increasing in* $\tau$ *and* $\lim_{\tau \to \infty} \overline{\beta}(\ell) = \infty$.

The intuition for Proposition 7 is similar to the one for Proposition 2, with the obvious difference that now there is an intermediate region where the communication network is neither entirely focused nor completely balanced. Figure 2 illustrates the results of Proposition 7 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 $\ell^*$ as a function of the importance of coordination, $\beta \in (0, 1)$. A balanced organization is optimal when coordination is sufficiently un-important. In this specific example, whenever $\beta < \overline{\beta}(20) \approx 0.06$ the organization is fully balanced, that is, $\ell^* = n = 20$. As coordination becomes more important, the communication becomes more focused around fewer leaders. Finally, when tasks are sufficiently interdependent, when $\beta > \overline{\beta}(1) \approx 0.73$, the organization has a single leader, $\ell^* = 1$.

Panel B of Figure 2 shows how the adaptiveness of each leader $j \in \mathcal{L}(\mathbf{t})$ to his local shock, $\Omega_\ell = cov(q_{jj}, \theta_j)$, changes as tasks become more interdependent as measured by $\beta$. 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 $\beta = 0.8$, there is only one leader, but this leader is, roughly, 50% more adaptive to his local information then when $\beta = 0.2$ and the number of leaders equals $\ell^* = 3$. Intuitively for a given number of $\ell$ leaders, the adaptiveness of any given leader decreases as coordination becomes more important. But for $\ell < n$, this gradual decrease is more than compensated when $\beta$ passes the threshold $\overline{\beta}(\ell)$ and the number of leaders decreases to $\ell - 1$, resulting in a huge boost to the adaptiveness of the remaining leaders.

Proposition 7 also shows how an exogenous change in attention capacity $\tau$, increases the number of leaders and makes the organization more balanced. Again, this implies that as communication technology improves, organizations become less focused and leadership is more broadly shared. As discussed in the introduction, this is consistent with new trends in organizational design towards more network-like organizations where communication flows are horizontal rather than vertical, and decision-making and influence is broadly shared in the organization, as documented in both case studies (for example "Proctor & Gamble Organization 2005," HBS case 9-707-519) and large scale empirical studies (Whittington et al. 1999, Guadalupe and Wulf, 2012).[27]

## 5.4 Public communication

As already mentioned, we have assumed that communication is bilateral as this puts the least constraints on nature of communication flows. 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"

[27] See also Roberts and Saloner (2013), Section 3.5.

or "attention" any agent $j$ receives. 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 communication network is given by $\mathbf{t} = \{t_1, ..., t_n\}$ , where $t_j$ is the is the mutual information between $m_{ij}$ and $\theta_j$ and the communication constraint is given by

$$\sum_{j \in \mathcal{N}} t_j \leq \tau^P.$$

The conditional variances are then defined by $\ln Var(\theta_j|m_{ij}) = \ln \sigma_\theta^2 - 2t_j$. Under public communication, two agents $j, k \in \mathcal{N} \setminus \{i\}$ are constrained to pay the same amount of attention to agent $i$, a property that, as shown in Lemma 5, holds for the optimal bilateral communication networks. The following equivalence result, proven in appendix, therefore follows immediately:

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

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

*where* $\mathbf{t}^b = \left\{t_{ij}^b\right\}_{i \neq j}$ *is an optimal communication network*[28] *under bilateral communication and constraint* $\tau = (n-1)\tau^P$.

It follows that also with public communication, the optimal organization is an $\ell-$leader organization with $\ell \in \{1, 2, \cdots, n\}$ leaders and the same comparative statics hold: $\ell$ is decreasing in the need for coordination, $\beta$, and an increasing in the communication capacity $\tau^P$ (Proposition 7). Only the exact number of leaders may be different . When $n = 2$, however, $\tau = \tau^P$ and there is no difference between a public and a bilateral communication network.

The two communication models analyzed so far assume that the communication constraint is determined at the organizational level. In Appendix B, we consider a model with individual attention constraints, where each agent has access to an individual communication channel, whose finite capacity $\tau^I$ can be used to broadcast information to all other agents and/or to process information broadcasted by others.[29] We prove a similar equivalence result.

[28] 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.

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

## 6 Organizational Size and Organizational Strategy

So far, we have assumed that the number of tasks that needs to be coordinated is fixed. This is a reasonable assumption if the various tasks in our model correspond to complementary functions (marketing, manufacturing, engineering) in a production process.[30] In an alternative interpretation of our model, however, each tasks corresponds to a different type of product or service that is produced or delivered by a multi-product firm. By engaging in multiple tasks, firms can spread out fixed costs and realize scope economies (Panzar and Willig, 1981). Doing so, however, increases coordination costs as now more tasks need to be coordinated. An interesting question, then, is how the optimal size (or scope) of an organization interacts with organizational focus, and how organizational size and focus are jointly optimized in response to, say, more volatile environments or improvements in communication technology.[31]

### 6.1 A model of endogenous organizational size

To study the optimal organization size, we modify our model to include fixed production costs, such as production facilities or a distribution network. The benefit of increasing organization size is that those fixed costs can be spread out over a larger number of tasks. But realizing economies of scale or scope imposes additional coordination costs on the organization. In the context of our model, sharing a plant or a distribution network makes it necessary for agents to take the appropriate coordinating actions. In contrast, if a task is executed on a stand-alone basis, there is no need for coordination. The larger the size of the organization, therefore, the more complex the coordination-adaptation trade-offs.[32]

Concretely, we posit that the realized profit of an organization which performs $n$ tasks (or

[30] Even if a task is outsourced, it must still be coordinated. In this sense, firm boundaries do not necessarily affect coordination problems.

[31] Note that we only endogenize organizational size, not firm size. We refer to Teece (1982) for a discussion of when economies of scope are optimally realized within the boundaries of a (multi-product) firm rather than across firm boundaries.

[32] See Mitchell (2002) for an alternative model of firm scope, where diseconomies of scope stem from the technological distance between tasks. Unlike in our model, communication technology or the adaptiveness of the firm (see Section 6.3) do not affect optimal firm scope.

produces $n$ products) equals

$$\pi(n) = nP - F - \sum_{i \in \mathcal{N}} \left[ (q_{ii} - \theta_i)^2 + \beta \sum_{j \in \mathcal{N} \setminus \{i\}} (q_{ii} - q_{ji})^2 \right], \tag{28}$$

where $P > 0$ represents the revenues that can be obtained per task and $F > 0$ are the fixed costs which are shared by all tasks/products. The last term in (28) is identical to the profits of the $n$ tasks model analyzed in Section 5. Regardless of the size of the organization, there is a fixed communication capacity $\tau$ which can be spent communicating about the $n$ product-lines. For simplicity, we assume that communication is public (see Section 5.4), as the assumption of a size-independent communication constraint is the most natural in this case.[33] Let $t_i$ be the attention devoted to communicating about task $i$, and $m_{ij}$ the information agent $j$ receives about $\theta_i$, then

$$E\left[Var(\theta_i | m_{ij})\right] = \sigma_\theta^2 e^{-2t_i} \qquad \text{with} \qquad \sum_i t_i \leq \tau$$

As shown in Proposition 4, for a given organizational size $n$, the optimal organization is a $\ell-$leader organization. If $\ell^*(n)$ is the optimal number of leaders given $n$ tasks, then the expected profits of an organization of size $n$ are given by

$$\begin{aligned} E\left[\pi(n)\right] &= nP - F - n\sigma_\theta^2 + \ell^*(n) \frac{\sigma_\theta^2}{1 + (n-1)\beta e^{-2\tau/\ell^*(n)}} \\ &+ (n - \ell^*(n)) \frac{\sigma_\theta^2}{1 + (n-1)\beta}. \end{aligned}$$

We posit that the size of the organization is chosen in order to maximize profits per product-line:

$$n^* = \arg\max_n \frac{1}{n} E[\pi(n)].$$

Our underlying assumption is that firms, whenever profitable, have the option to operate a set of product lines independently as a separate organization.[34]

The next proposition offers a characterization of the function $\ell^*(n)$, a useful result when investigating the comparative statics of $n^*$ with respect to the parameters of the model.

[33] With bilateral communication, it can be argued that the communication constraint should be expanded as the organization grows. We obtain identical results if the bilateral communication constraint is $(n-1)\tau$ so that each additional task/agent increases the communication capacity by $\tau$.

[34] Organizational boundaries are then based on who communicates with whom: Agents belong to the same organization if they communicate with each other.

**Proposition 8** *Larger organizations are more focussed than smaller organizations: (i) If* $\ell^*(n+1) = n+1$*, then* $\ell^*(n) = n$*. (ii) If* $\ell^*(n+1) < n+1$*, then* $\ell^*(n+1) \leq \ell^*(n)$*.*

Proposition 8 is of independent interest, as organizational size is often exogenous. It shows that, unless they are fully balanced, larger organizations have a *lower* number of leaders than smaller organizations and that the ratio of leaders to tasks is decreasing in $n$. Intuitively, as an organization grows larger, coordination-adaptation trade-offs become more pronounced, forcing the organization to become more focused and coordinate through fewer leaders. Interestingly, while the organization has more members, fewer of them receive attention. This result is consistent with the life cycle theory of organizations (Greiner 1998) according to which small entrepreneurial firms, as they grow bigger, move from an initial *phase of creativity*, where almost all employees are involved in decision-making to a *phase of direction*, where "able, directive leadership" is installed, and where "[t]he new manager and his or her key supervisors assume most of the responsibility for instituting direction." Consistent with our model, Greiner also describes how firms become less adaptive and responsive to their environment as they grow larger.

### 6.2 Monotone comparative statics

We now characterize the optimal size $n^*$ of the organization, and its interaction with organizational focus $\ell^*(n^*)$, as a function of the main parameters of our model.

**Proposition 9** *The optimal organization size* $n^*$ *is (1) decreasing in the volatility of the environment* ($\sigma_\theta^2$)*, and (2)increasing in the level of synergies/shared resources* ($F$)*.*

When choosing $n^*$, organization trades off economies of scope with coordination costs. Not surprisingly, the larger are the economies of scope, as characterized by the level of shareable fixed costs, the larger the optimal organization size $n^*$. Perhaps more interesting is the organizational response to an increase in the volatility of the environment $\sigma_\theta^2$. Management scholars have cited many reasons for the rise of new organizational forms, but one line of explanation which is especially prominent is the "increased turbulence" that managers face because of rapid technological changes, deregulation, and globalization (Rivkin and Siggelkow, 2005).[35] As $\sigma_\theta^2$

[35] The other prominent line of explanation, information technology, is addressed in the next section.

increases, so do the incentives to adapt, which in turn bring coordination costs. By narrowing firm scope (reducing $n^*$), organizations partially reduce these coordination costs, allowing for a better adaptation. Put differently, organizations trade-off economics of scale and scope with adaptation to a changing environment. Proposition 9 therefore reflects the common idea that smaller organizations are more "nimble" and "flexible."

Since for a given organizational size $n$, the number of leaders $\ell^*$ is independent of $\sigma_\theta^2$, a corollary of Propositions 8 and 9 is that a more volatile environment not only results in smaller but also more balanced organizations:

**Corollary 10** *Organizational balance, $\ell^*/n^*$, is increasing in $\sigma_\theta^2$ (decreasing in $F$). If $\ell^*(n^*) < n^*$, the number of leaders is increasing in $\sigma_\theta^2$ (decreasing in $F$).*

### 6.3 Non-monotone comparative statics

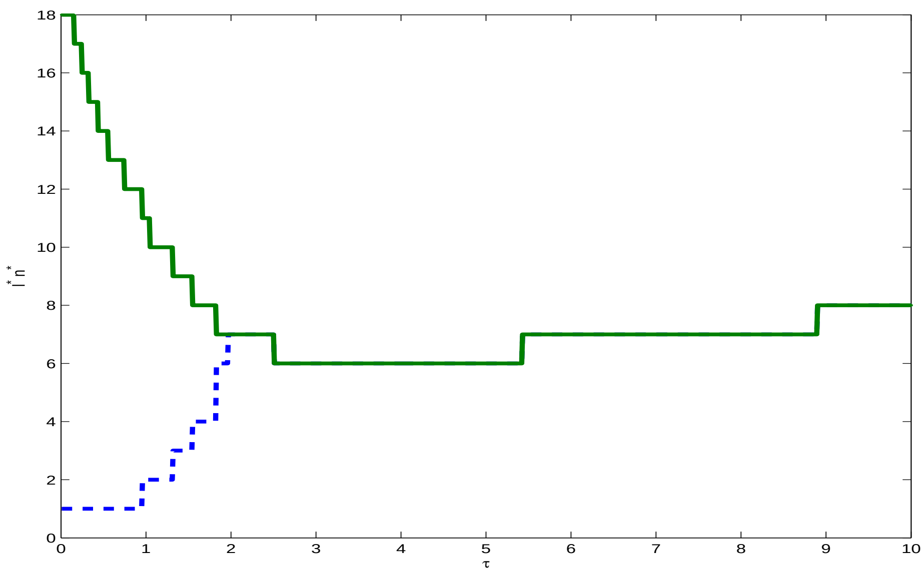
Given that limited attention and coordination costs restrict organizational size in our model, one may (naively) expect organizational size to be increasing in the communication capacity $\tau$, and to be decreasing in the interdependence of tasks, $\beta$. What truly restricts organizational size, however, is the adaptiveness of the organization. In our model, there are no constraints on organizational size as long as agents do not adapt tasks to local shocks. When communication capacity is limited, organizations may therefore pursue two distinctive organizational strategies:

- Adaptation: (Largely) give up on scope economies, and have a small, but adaptive organization.
- Economies of Scope: (Largely) give up on adapting to local shocks, and instead leverage economies of scale and scope.

When scope economies ($F$) are large, but communication capacity ($\tau$) is very limited, organizations optimally choose to minimize average fixed costs at the expense of adaptation to local shocks. Organization-wise, this strategy consists of having a large, rigid organization, and focussing all attention on one or a few leaders. Large, non-adaptive organizations with one or a few leaders are then optimal. As $\tau$ becomes larger, however, the organization may gradually want to use the extra communication capacity to become more adaptive. Doing so without incurring substantial coordination costs, however, requires reducing organizational size, often

Figure 3: **Endogenous organizational scope and focus as a function of $\tau$**

Example: maximum number of tasks $\overline{n} = 18$, $\sigma_\theta^2 = 1$, $\beta = .25$ and $F = 3$. Optimal organizational size, $n^*$ (continuous line), and of leaders, $\ell^*$ (dashed line), as a function of $\tau$.




substantially. At the same time, a smaller size allows the organization to pay attention to a larger number of tasks or leaders. It is only when the communication capacity is sufficiently large that organizations can pursue both objectives, scale economies and adaptation, in which case organizational size is increasing again in $\tau$.

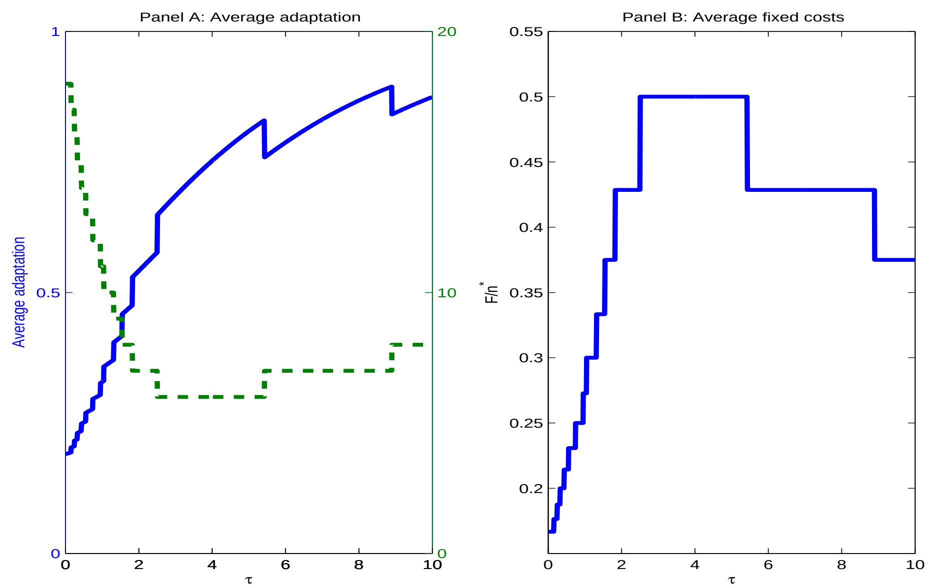
Figures 3 and 4 (Panels A and B) illustrate the above organizational strategies in response to changes in $\tau$. For simplicity, it is assumed that $n^*$ is constrained to $n^* \leq \overline{n} = 18$. For $\tau$ very small, the optimal strategy is to reduce average fixed costs and for this purpose, the organization includes a maximum number of tasks $\overline{n}$. All attention is focussed on one leader/task and, on average, the organization is very non-adaptive, as captured by the average adaptiveness across tasks in the organization

$$\overline{\Omega} = \frac{1}{n^*} \sum\nolimits_i cov(q_{ii}, \theta_i).$$

As $\tau$ increases, the organization gradually shifts from a strategy of realizing economies of scope to one of being adaptive to the environment. In the figure, the organization quickly drops in scope from eighteen to six tasks. For larger values of $\tau$, organizational size slowly increases

Figure 4: **Comparative statics with respect to $\tau$**

Example: maximum number of tasks $\overline{n} = 18$, $\sigma_\theta^2 = 1$, $\beta = .25$ and $F = 3$. Panel A: Average adaptation, $\overline{\Omega}$ (left axis, continuous line), and optimal organizational scope, $n^*$ (right axis, dashed line), as a function of $\tau$. Panel B: Average fixed costs $\frac{F}{n^*}$ as a function of $\tau$.




again with $\tau$ but now the organization is already very adaptive, and pursues both objectives simultaneously: As shown in Panels A and B of Figure 4 increases in $\tau$ beyond $\tau > 2.5$ result in both more average adaptiveness (except when the scope and the number of leaders adjust upwards) and lower average fixed costs.[36]

To the extent improvements in information and communication technology (ICT) correspond to an increase in the communication capacity, our model predicts that improvements in ICT may result in a shift from large inflexible organizations emphasizing economies of scale and scope, towards smaller, more balanced organizations, which are focused on being adaptive to external shocks and emphasize horizontal communication linkages.[37] This is largely con-

[36]Notice as well that the number of leaders is non monotonic in attention capacity: The organization is fully balanced when $n^* = 6$ but after that additional increases in $\tau$ lead to further reductions in the size, and thus in the number of leaders as well, but, as shown in Proposition 8, the organization remains balanced.

[37]This prediction stands in contrast with those of obtained in recent team-theory models that model organizations as information-processing (Bolton and Dewatripont 1994) or problem-solving institutions (Garicano,

sistent with recent trends in organization design, as described by Whittington et al. (1999) and Roberts and Saloner (2013). According to our model, only organizations that are already very adaptive, respond to ICT improvements by increasing organizational scope. Alternatively, observed trends toward de-sizing and de-scoping may have been a response to an increased variability in the environment (Proposition 9), for example because of globalization and increased competition (Roberts and Saloner, 2013).

While we have emphasized non-monotone comparative statics with respect to $\tau$, similar intuitions apply when comparative statics with respect to $\beta$ are considered. When tasks are sufficiently interdependent ($\beta$ large), an organization may give up on adaptation and maximize scale and scope economies instead, dramatically increasing organizational size in the process.

# 7 Concluding remarks

In this paper, we have studied the optimal allocation of scarce attention in organizations that face competing objectives of coordination, adaptation and, in the latter part of our paper, economies of scale and scope. We believe the proposed framework sheds light not only on the importance of organizational focus, but also on complex and often contradictory trends in organizational design, often attributed to improvements in information technology and an increasingly volatile business environment (Siggelkow and Rivkin 2005, Roberts and Saloner 2013). To conclude, we expand on three elements which are absent from our model: skill-heterogeneity among employees, conflicts of interests and coordination through hierarchies.

**Skill-heterogeneity, functional cultures and the permanence of leadership.** In order to isolate the role of leadership and focus in organizations, we considered a model in which all tasks and agents are ex ante symmetric. Indeed, when attention is not scarce, all tasks are treated symmetrically and the organization is a "team of leaders." It is informational constraints, not technological constraints, that induce the need for leadership and organizational

2000; Garicano and Rossi-Hansberg, 2006). While these papers also characterize optimal information flows in organizations, improvements in communication technology unambiguously result in larger and more centralized organizations.

focus.[38] Therefore, which tasks the organization focusses on, or who becomes a leader, is irrelevant in our model. The importance of organizational focus, however, does have implications for what type of employees an organization may want to hire. It may therefore cause asymmetries in the skill-level of employees. Higher-skilled employees may understand better how to adapt to a changing environment. If higher-skilled employees command higher wages, organizations that are focused on, say, engineering, will then optimally recruit higher-skilled engineers than organizations which are focussed on, say, marketing and sales. If communication will mainly pertain to coordinating engineering initiatives, it may further be optimal to hire employees with an engineering background even for marketing positions in order to foster communication. Over time "functional cultures" may develop, where certain parts of the organization become "backwaters," whereas other parts dominate not only decision-making but also the talent pool of the organization. Technology companies are often reputed for their "functional cultures". Google, for example, proclaims "[it] is and always will be an engineering company," and almost all employees have advanced engineering degrees even in non-engineering tasks. Walter Isaacson's biography of Steve Jobs, on the other hand, paints a picture of Apple as a company run by designers, where engineers were very often ignored. To the extent that employee turn-over is limited and hiring is costly and incremental, an initial focus on a particular set of tasks may therefore have long-term consequences. While our model is static, this suggests it may be neither feasible nor desirable for an organization to "rotate" its focus or leadership.

**Conflict, power struggles and organizational change.** Incentives play no role in our theory, but if agents care mainly about the performance of "their task" (as modeled, for example, in Alonso et al. 2008), power struggles and conflict may arise as to whom becomes the focus of the organization. A dynamic version of our model where the allocation of attention is decentralized, rather than optimally allocated by an organization designer, may investigate how agents act strategically to become the center of attention. To the extent that leadership is an endogenous and self-enforcing phenomenon, organizations may get stuck in inefficient equilibria. Indeed, agents pay attention to whomever they believe is very adaptive, and agents are

[38] Appendix C shows how our model can accommodate technological trade-offs between adaptation and coordination, and yield similar insights.

adaptive only when they are the the center of attention. Hence, multiple leadership equilibria may exist. Leadership may further become entrenched, with current leaders (agents who are currently the focus of the organization) being unwilling to give up their central position in the network. Our model further suggests that 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.

**Coordination and Hierarchies.** In our model, production is carried out by a team of agents and decision-making is de facto decentralized. As we have shown, informal leadership then arises endogenously to ensure coordination. In practice, organizations often achieve coordination by centralizing decisions and/or introducing layers of hierarchy. In the context of our model, the organization could centralize production in the hands of a headquarter manager, who takes all production decisions after communicating with agents. Centralized production has the virtue of ensuring perfectly coordinated decisions. Centralized organizations, however, are likely to be less adaptive to shocks as they must rely on (limited) communication with agents.[39] In ongoing work, we show that whenever attention is scarce, decentralized team production is then preferred over centralized production. Hence, centralized production is unlikely to be optimal in fast-moving environments where $\tau$ is small. As discussed above, empirical evidence indicates that many organizations have gone through a process of delayering in recent decades, suggesting they have come to rely less on hierarchies for coordination. To shed light on this trend, it would be interesting to introduce multi-layered hierarchies in our model, where such layers provide an alternative way of coordinating production.

[39] Alonso et al. (2008) and Rantakari (2008) also study the trade-off between centralization and decentralization in terms of an organization's ability to coordinate and adapt. Attention and communication capacity play no role in the above models. Instead, communication is strategic because conflicts of interest.

## 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 (2012). "Leadership, Coordination and Corporate Culture." *Review of Economic Studies*, forthcoming.

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.

Dewan, Torun and David Myatt (2008). "The Qualities of Leadership: Direction, Communication, and Obfuscation." *American Political Science Review.* 102 (3), 351-368,

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.

Diaconis, Persi, and Donal Ylvisaker (1979). "Conjugate priors for exponential families." *The Annals of statistics*, 7(2):269-281.

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 Esteban Rossi-Hansberg (2006). "Organization and Inequality in a Knowledge Economy." *Quarterly Journal of Economics* 121(4): 1383-1435.

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.

Greiner, Larry E. (1998) "Evolution and revolution as organizations grow. 1972." Harvard business review 76.3 : 55-68.

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.

Hart, Oliver and John Moore (2005) "On the design of hierarchies: Coordination versus specialization." *Journal of Political. Economy* 113(4): 675- 702

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 (2013) "Leadership and Corporate Culture," *Handbook of Organizational Economics,* eds. R. Gibbons and J. Roberts, Princeton University Press, 432-478.

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.

March, James G., and Herbert A. Simon (1958). *Organizations.* Wiley (USA).

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

Mitchell, Matthew (2000). "The scope and organization of production: firm dynamics over the learning curve," *Rand Journal of Economics*, 31(1): 180-205.

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.

Panzar, John C., and Robert D. Willig (1981). "Economies of scope." *American Economic Review*, 268-272.

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.

Roberts John and Garth Saloner (2013). "Strategy and Organization," *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press, 799-849.

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.

Simon, Herbert (1971)."Designing Organizations for an Information-Rich World," *Computers, Communication, and the Public Interest*, ed Martin Greenberger, Baltimore, MD: The Johns Hopkins Press.

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.

Siggelkow, N., and Rivkin, J. W. (2005). "Speed and search: Designing organizations for turbulence and complexity." *Organization Science*, 16(2): 101-122.

Teece, David J. (1982). "Towards an economic theory of the multiproduct firm," *Journal of Economic Behavior & Organization* 3(1), 39-63.

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.

Whittington, Richard, Andrew Pettigrew, Simon Peck, Evelyn Fenton, and Martin Conyon (1999). "Change and Complementarities in the New Competitive Landscape: A European Panel Study, 1992-1996." *Organization Science*, 10(5), 583-600.

# 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

$$\frac{\partial E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial t} = -\beta\left[\Omega_1(t)\mathsf{RV}'\left(t\right) - \Omega_2(\tau - t)\mathsf{RV}'\left(\tau - t\right)\right]. \tag{29}$$

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

$$\frac{\partial E\left[\pi\left(\mathbf{q}|\boldsymbol{\theta}\right)\right]}{\partial t} = -\beta\left[\frac{\mathsf{RV}'(t)}{[\sigma_\theta^2 + \beta\mathsf{RV}(t)]^2} - \frac{\mathsf{RV}'(\tau - t)}{[\sigma_\theta^2 + \beta\mathsf{RV}(\tau - t)]^2}\right]. \tag{30}$$

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 (30). Using that $\mathsf{RV}(t) = \sigma_\theta^2 e^{-2t}$, after some plain algebra it follows that

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

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

**Proof of Proposition 3.** Replicating the analysis for the model with two-tasks, by allowing for asymmetries, we obtain that equilibrium actions are

$$q_i = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta_i\mathsf{RV}(t_i)}\theta_i \qquad \text{and} \qquad q_{ij} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta_j\mathsf{RV}(t_j)}E[\theta_j|\mathcal{I}_i];$$

we can express expected profit for a given network $\mathbf{t}$ as

$$E[\pi(\mathbf{q}|\theta] = -\beta_1\Omega_1(t_1)\mathsf{RV}(t_1) - \beta_2\Omega_2(t_2)\mathsf{RV}(t_2), \tag{31}$$

where

$$\Omega_i(t_i) = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \beta_i\mathsf{RV}(t_i)}.$$

Hence, the organizational problem is to choose $t_1 = t \in [0, \tau]$ to maximize expression (31). Repeating the arguments developed for the symmetric case, we obtain that the profits of the organization are decreasing in $t$, if, and only if,

$$-[1 - \beta_1\beta_2 e^{-2\tau}][\beta_1\mathsf{RV}(t) - \beta_2\mathsf{RV}(\tau - t)] > 0, \tag{32}$$

where, we recall that, $\mathsf{RV}(x) = \sigma_\theta^2 e^{-2x}$. It is convenient to divide the analysis in two cases. Recall that we are assuming that $\beta > 1 + \epsilon$ (which is equivalent of assuming $\beta_2 > \hat{\beta} = 1$).

**Case 1.** Assume that $\beta_1 \mathsf{RV}(\tau) - \beta_2 \mathsf{RV}(0) > 0$, which is equivalent to $\beta_1 e^{-2\tau} - \beta_2 > 0$, or $\epsilon > \widehat{\epsilon}$. This assumption and the fact that $\mathsf{RV}(\cdot)$ is decreasing in $t$, implies that $\beta_1 \mathsf{RV}(t) - \beta_2 \mathsf{RV}(\tau - t) > 0$ for all $t \in [0, \tau]$. This in turn implies that the objective function is decreasing in $t$ if, and only if,

$$1 - \beta_1 \beta_2 e^{-2\tau} < 0 \Longleftrightarrow \tau < \ln \beta$$

which is always satisfied because $\beta > 1 + \epsilon$. So, if $\tau < \ln \beta$ and $\epsilon > \widehat{\epsilon}$, it is optimal to set $t = 0$ and there is focus on task 2.

**Case 2.** Assume now that $\beta_1 \mathsf{RV}(\tau) - \beta_2 \mathsf{RV}(0) < 0$, or $\epsilon < \widehat{\epsilon}$. Since $\beta_1 \mathsf{RV}(0) - \beta_2 \mathsf{RV}(\tau) > 0$ and since $\beta_1 \mathsf{RV}(t) - \beta_2 \mathsf{RV}(\tau - t)$ declines in $t$, it follows that there exists a $t^*$ so that $\beta_1 \mathsf{RV}(t^*) - \beta_2 \mathsf{RV}(\tau - t^*) = 0$. Indeed, such $t^*$ solves $\beta_1/\beta_2 = \mathsf{RV}(\tau - t^*)/\mathsf{RV}(t^*)$ and since $\beta_1 > \beta_2$ and $\mathsf{RV}(t)$ is decreasing in $t$, it follows that $t^* > \tau/2$. The next two observations complete the proof:

First, if $1 - \beta_1 \beta_2 e^{-2\tau} > 0$, or equivalently, $\tau > \ln \beta$, the objective function is increasing in $t$ for $t \leq t^*$ and it is decreasing in $t$ for all $t > t^*$. Hence, in the optimal organization $t = t^*$. Second, if $1 - \beta_1 \beta_2 e^{-2\tau} < 0$, or equivalently, $\tau > \ln \beta$, the objective function is decreasing in $t$ for all $t \leq t^*$ and increasing in $t$ for all $t \geq t^*$. Hence, there are two candidates for the minimum: either $t = 0$ or $t = \tau$. Comparing the two organizations it reveals that since $1 - \beta_1 \beta_2 e^{-2\tau} < 0$ the optimal organization has $t = 0$, and so there is focus on task 2. Note also that $1 - \beta_1 \beta_2 e^{-2\tau} > 0$ and $\beta_1 \mathsf{RV}(\tau) - \beta_2 \mathsf{RV}(0) < 0$, are mutually compatible, if and only if, $\beta > 1 + \epsilon$, which holds by assumption. This concludes the proof of Proposition 3. ■

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

**Proof of Lemma 5.** Suppose that $\mathbf{t}$ is optimal and, for a contradiction, assume that there exists some agent $i$ such that $t_{ji} > t_{ki} \geq 0$. Define a new organization $\mathbf{t}'$, which is the same as $\mathbf{t}$ with the exception that $t'_{ji} = t_{ji} - \epsilon$ and $t'_{ki} = t_{ki} + \epsilon$, for some small and positive $\epsilon$. Using the expression for expected payoffs 25 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,

$$e^{-2t'_{ji}} + e^{-2t'_{ki}} \geq e^{-2t_{ji}} + e^{-2t_{ki}}. \tag{33}$$

Since $t'_{ji} = t_{ji} - \epsilon$ and $t'_{ki} = t_{ki} + \epsilon$, after some algebra we obtain that condition 33 is equivalent to

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

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

**Proof of Lemma 6.** Suppose that $\mathbf{t}$. In view of Lemma 5 we know that for all $i$, $t_{ji} = t_i$ for all $j$. Suppose, for a contradiction, that $t_i > t_j > 0$. Consider now two alternative organizations. One organization, denoted by $\mathbf{t}'$, is the same as organization $\mathbf{t}$, but $t'_i = t_i - \epsilon$ and $t'_j = t_j + \epsilon$. The second organization, denoted by $\hat{\mathbf{t}}$, is the same as organization $\mathbf{t}$, but $\hat{t}_i = t_i + \epsilon$ and $\hat{t}_j = t_j - \epsilon$. These constructions are derived for some small and positive $\epsilon$. Since the three organizations only differ in the way attention is distributed for task $i$ and task $j$, each other task $l \neq i, j$ performs equally across the three organizations. We can then write

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

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

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

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

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

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

This is equivalent to

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

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

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

**Proof of Proposition 7**. Using the expression for expected payoffs (25), the fact that $\mathsf{RV}(t) = \sigma_\theta^2 e^{-2t}$, and that organization $\mathbf{t}$ is an $\ell$-leader organization, we obtain that

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

where

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

and that

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

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

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

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

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

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

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

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

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

because, by observation 1, $\Phi(n,\beta,\tau,n)>\Phi(n,\underline{\beta}(\tau,n),\tau,n)$, and, by definition, $\Phi(n,\underline{\beta}(\tau,n),\tau,n)=0$. Hence, for all $\beta<\underline{\beta}(\tau,n)$ the expected returns of an $\ell$-leader 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 $\tau$.

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

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

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

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

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

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

For brevity define $\widehat{\mathsf{RV}}(x) = e^{-\frac{2\tau}{(n-1)x}}$ and denote by $\Delta(\ell, \beta)$ the difference between the expected profit generated by the $\ell+1$-leader organization and the expected profit generated by the $\ell$-leader organization. Using expression 26, we obtain

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

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

$$\begin{aligned} (1 + \beta(n-1)) \left[ (\ell + 1)(1 + \beta(n-1)\widehat{\mathsf{RV}}(\ell)) - \ell(1 + \beta(n-1)\widehat{\mathsf{RV}}(\ell + 1)) \right] - \\ -[1 + \beta(n-1)\widehat{\mathsf{RV}}(\ell)][1 + \beta(n-1)\widehat{\mathsf{RV}}(\ell + 1)] \quad &= \quad 0. \end{aligned}$$

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

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

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

which, after some algebra, is

$$\beta_{\ell+1} = \frac{1}{n-1}\left[\frac{e^{\frac{t}{\ell+1}} + \ell e^{-\frac{t}{\ell(\ell+1)}} - (1+\ell)}{\ell + e^{-\frac{t}{\ell}} - (1+\ell)e^{-\frac{t}{\ell(\ell+1).}}}\right]$$

Note that nominator is increasing in $t$ because

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

whereas the denominator is decreasing in $t$ because

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

It follows that

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

Note further that

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

This concludes the proof of Proposition 7. ■

**Proof of Proposition 8**: Recall that in an organization of size $n$ and $\ell$ leaders the profits are

$$E\left[\pi(n,\ell)\right] = nP - F - n\sigma_\theta^2 + \ell\frac{\sigma_\theta^2}{1+(n-1)\beta e^{-2\tau/\ell}} + (n-\ell)\frac{\sigma_\theta^2}{1+(n-1)\beta}.$$

Similarly, in an organization of size $n+1$ and $\ell$ leaders the profits are

$$\begin{aligned} E\left[\pi(n+1,\ell)\right] &= (n+1)P - F - (n+1)\sigma_\theta^2 + \ell\frac{\sigma_\theta^2}{1+n\beta e^{-2\tau/\ell}} + (n+1-\ell)\frac{\sigma_\theta^2}{1+n\beta} \\ &= (n+1)P - F - (n+1)\sigma_\theta^2 + \frac{\sigma_\theta^2}{1+(n-1)\hat{\beta}} \\ &\quad +\ell\frac{\sigma_\theta^2}{1+(n-1)\hat{\beta}e^{-2\tau/\ell}} + (n-\ell)\frac{\sigma_\theta^2}{1+(n-1)\hat{\beta}} \end{aligned}$$

where

$$\hat{\beta} = \frac{n}{(n-1)}\beta$$

Given the above derivation we have that

$$\ell_n^*(\beta) = \arg\max_{l=1...n}\left[\ell\frac{\sigma_\theta^2}{1+(n-1)\beta e^{-2\tau/\ell}} + (n-\ell)\frac{\sigma_\theta^2}{1+(n-1)\beta}\right]$$

and

$$\ell_{n+1}^*(\hat{\beta}) = \arg\max_{l=1...n+1}\left[\ell\frac{\sigma_\theta^2}{1+(n-1)\hat{\beta}e^{-2\tau/\ell}} + (n-\ell)\frac{\sigma_\theta^2}{1+(n-1)\hat{\beta}}\right]$$

Suppose first that $\ell_{n+1}^*(\hat{\beta}) \leq n$. It the follows from Proposition 7 that since $\hat{\beta} > \beta$ then $\ell_{n+1}^*(\hat{\beta}) \leq \ell_n^*(\beta)$, which concludes the proof of the first part of the proposition.

Suppose now that $\ell_{n+1}^*(\hat{\beta}) = n+1$; From Proposition 7 we know that: A. $\ell_{n+1}^*(\hat{\beta}) = n+1$ if, and only if, $\hat{\beta} \leq \underline{\beta}(\tau, n+1)$ (where $\underline{\beta}(\tau, n+1)$ is derived in Proposition 7) and B. $\frac{dE[\pi(n+1,\ell)]}{d\ell}$ is positive and decreasing in $\ell = 1...n+1$ for all $\beta' \leq \underline{\beta}(\tau, n+1)$. Since, as the derivation above show, $\frac{dE[\pi(n+1,\ell)]}{d\ell} = \frac{dE[\pi(n,\ell)]}{d\ell}$ when evaluated at the same value of $\beta'$, we have that $\frac{dE[\pi(n,\ell)]}{d\ell}$ is positive and decreasing for all $\ell = 1...n$ when is evaluated at $\hat{\beta} \leq \underline{\beta}(\tau, n+1)$. But then, since B holds and since $\beta < \hat{\beta} \leq \underline{\beta}(\tau, n+1)$, it follows $\frac{dE[\pi(n,\ell)]}{d\ell}$ is positive and decreasing for all $\ell = 1...n$ when is evaluated at $\beta$. Hence $\ell_n^*(\beta) = n$. This completes the proof. ■

**Proof of Proposition 9**: Recall that $\ell_{n+1}^*$is the optimal number of leaders given $n+1$tasks and $\ell_n^*$ is the optimal number of leaders given $n$tasks. Then

$$\frac{E\left[\pi(n, \ell_n^*)\right]}{n} = P - \sigma_\theta^2 - F/n + \frac{1}{n}\left(\ell_n^* \frac{1}{1+(n-1)\beta e^{-2\tau/\ell_n^*}} + (n-\ell_n^*)\frac{1}{1+(n-1)\beta}\right)\sigma_\theta^2 \tag{34}$$

whereas

$$\frac{E\left[\pi(n+1, \ell_{n+1}^*)\right]}{n+1} = P - \sigma_\theta^2 - F/(n+1) + \frac{1}{n+1}\left[\frac{\ell_{n+1}^*}{1+(n-1)\hat{\beta} e^{-2\tau/\ell_{n+1}^*}} + \frac{(n-\ell_{n+1}^*)}{1+(n-1)\hat{\beta}} + \frac{1}{1+(n-1)\hat{\beta}}\right]\sigma_\theta^2, \tag{35}$$

where $\hat{\beta} = \frac{n}{(n-1)}\beta > \beta$.

Suppose first that $\ell_{n+1}^* \leq n$. Then, Proposition 8 implies that $\ell_n^* \geq \ell_{n+1}^*$. To prove the proposition is then sufficient to show that

$$\Delta \equiv \frac{E\left[\pi(n+1, \ell_{n+1}^*)\right]}{n+1} - \frac{E\left[\pi(n, \ell_n^*)\right]}{n}$$

is increasing in $F$ and is decreasing in $\sigma_\theta^2$. It is obvious to check that $\Delta$ is increasing in $F$. We now show it is decreasing in $\sigma_\theta^2$. Since $\hat{\beta} > \hat{\beta} e^{-2\tau/\ell^*}$, a sufficient condition for $\Delta$ to be decreasing in $\sigma_\theta^2$ is that

$$\ell_n^* \frac{1}{1+(n-1)\beta e^{-2\tau/\ell_n^*}} + (n-\ell_n^*)\frac{1}{1+(n-1)\beta} > \ell_{n+1}^* \frac{1}{1+(n-1)\hat{\beta} e^{-2\tau/\ell_{n+1}^*}} + (n-\ell_{n+1}^*)\frac{1}{1+(n-1)\hat{\beta}}$$

Since $\ell_n^* \geq \ell^*$ and $\hat{\beta} > \beta$, this is indeed satisfied.

Next, assume that $\ell_{n+1}^* = n+1$; Proposition 8 then implies that $\ell_n^* = n$. Hence

$$\Delta = \left[\frac{1}{1+(n-1)\hat{\beta} e^{-2\tau/(n+1)}} - \frac{1}{1+(n-1)\beta e^{-2\tau/n}}\right]\sigma_\theta^2 + F/n - F/(n+1).$$

Since $\hat{\beta} > \beta$, it follows that $\Delta$ is decreasing in $\sigma_\theta^2$ and increasing in $F$. ■

**Appendix B: Alternative communication models**.

**B.1. Public Communication.**

**Proof of Result 1** Note that under bilateral communication and arbitrary capacity $\tau$, Lemma 5 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 $\tau$ and $\tau^P$. ■

**B.2. Individual Communication Constraints**.

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

$$\sum_{j \in N} t_{ij} \leq \tau^I \qquad \forall i \in \mathcal{N}, \tag{36}$$

and where $t_{ii}$ is the capacity devoted to broadcast information about $\theta_i$, and $t_{ij}$ is the capacity devoted to listen to the information broadcasted by agent $j \neq i$. The effective communication flow between agents $j$ and $i$ regarding $\theta_j$ then equals $\min \{t_{ij}, t_{jj}\}$ such that[40]

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

We now proof the following equivalence result, which again implies that the optimal organization is an $\ell-$leader organization with $\ell \in \{1, 2, \cdots, n\}$ leaders and that the same comparative statics hold as in Proposition 7.

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

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

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

**Proof of Result 2.** Consider the case of individual communication with individual capacity constraint $\tau^I$. Suppose that $\mathbf{t}$ is an optimal organization. It is immediate to see that $\mathbf{t}$ satisfies: a. $t_{ji} \leq t_{ii}$ for all $i, j \in \mathcal{N}$ and b. $\sum_j t_{ji} = \tau^I$ for all $ij \in \mathcal{N}$. Now note that if $\tau = (n-1)\tau^I$, $\mathbf{t}^b$ is an optimal organization under bilateral communication and constraint $\tau$, 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 $\tau^I$. 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 $\min\{t_{ji}, t_{ii}\} = t_{ji}$, 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 $\tau$. 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.**

[40]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.

We show that our insights hold in a model of coordination a la Alonso, Dessein, Matouschek (2008), Rantakari (2008) and Calvo-Armengol et al (2011). 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:

$$\pi(\mathbf{q}|\boldsymbol{\theta}) = K - (q_1 - \theta_1)^2 - (q_2 - \theta_2)^2 - \beta(q_1 - q_2)^2, \tag{37}$$

where $\beta$ 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 $\theta_i$ and a message $m_j$ about local shock $\theta_j$. The communication technology follows the description in our basic model.

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

$$q_1 = \frac{1}{1+\beta}\left[\theta_1 + \beta E\left[q_2|\mathcal{I}_1\right]\right] \tag{38}$$

$$q_2 = \frac{1}{1+\beta}\left[\theta_2 + \beta E\left[q_1|\mathcal{I}_2\right]\right] \tag{39}$$

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 (38) and (39) as

$$q_1 = a_{11}(t_1)\,\theta_1 + a_{12}(t_2)E[\theta_2|\mathcal{I}_1] \tag{40}$$

$$q_2 = a_{22}(t_2)\,\theta_1 + a_{21}(t_1)E[\theta_1|\mathcal{I}_2] \tag{41}$$

Substituting the guess (40) and (41) into (38) and (39), and using Assumption A, we find that the equilibrium actions are

$$q_1 = \frac{(1+\beta)\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(t_1)}\theta_1 + \frac{\beta\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(t_2)}E[\theta_2|\mathcal{I}_1] \tag{42}$$

$$q_2 = \frac{(1+\beta)\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(t_2)}\theta_2 + \frac{\beta\sigma_\theta^2}{\sigma_\theta^2(1+2\beta)+\beta^2\mathsf{RV}(t_1)}E[\theta_1|\mathcal{I}_2] \tag{43}$$

Finally substituting (42) and (43) into (37) and taking unconditional expectations we find that the problem

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

is equivalent to

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

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

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

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

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

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

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

**Appendix D: Endogenous Attention Capacity.**

So far we have taken $\tau$ to be a hard constraint in the amount of time agents can devote to communication with each other. In practice this is another margin that organizations can use to improve performance, by, for example, allowing more time for meetings and communication between teams. Equivalently, the organization can increase the effective communication capacity $\tau$, by cross-training and rotating employees, by hiring employee with higher cognitive abilities, or by investing in communication technology. Assume thus that an organization can acquire a capacity $\tau$ at a cost $C(\tau)$. $C(\tau)$ represents for example the costs of having team members engaged in communications activities rather than in production. We assume that this cost has the following properties:

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

The problem of organizational design is now

$$\max_{\tau,\mathbf{t}} E\pi(\mathbf{q}|\boldsymbol{\theta}) - C(\tau) \qquad \text{subject to} \quad (2). \tag{44}$$

**Proposition 11** *Assume that $\beta > \widehat{\beta}$, then*

1. *The optimal communication capacity $\tau^*$ is increasing in $\sigma_\theta^2$.*

2. *There exists $\bar{\sigma}_\theta^2 > \underline{\sigma}_\theta^2 > 0$ such that $t_1^* \in \{0, \tau^*\}$ if $\sigma_\theta^2 \leq \underline{\sigma}_\theta^2$ and $t_1^* = \frac{\tau^*}{2}$ if $\sigma_\theta^2 > \bar{\sigma}_\theta^2$.*

**Proof of Proposition 11.** We prove each of the two parts of the proposition.

**First part.** We first show that the optimal capacity $\tau^*$ is increasing in $\sigma_\theta^2$ in the focused organization and in the balanced organization. This, together with Proposition 2, implies the first part of Proposition 11: the optimal capacity $\tau^*$ is increasing in $\sigma_\theta^2$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

**Second part.** We now prove the second part of the proposition. First note that for a given common $\tau$

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

if, and only if,

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

and, after plain algebra, this condition is equivalent to

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

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

Observation 1. $\tau^d(\sigma_\theta^2)<\tau^c(\hat{\sigma}_\theta^2)$ if, and only if, $\sigma_\theta^2<\hat{\sigma}_\theta^2$.

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

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

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

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

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

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

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

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

We now complete the proof of the second part of Proposition 11. If $\sigma_\theta^2\leq\underline{\sigma}_\theta^2$, then $1-\beta^2 e^{-2\tau^d(\sigma_\theta^2)}\leq 0$ and $1-\beta^2 e^{-2\tau^c(\sigma_\theta^2)}<0$. From Proposition 2 we know that for all $\tau$ such that $1-\beta^2 e^{-2\tau}\leq 0$ the optimal organization is focused. Hence, if $\sigma_\theta^2\leq\underline{\sigma}_\theta^2$ the optimal organization is focused. Finally, if $\sigma_\theta^2\geq\bar{\sigma}_\theta^2$, then $1-\beta^2 e^{-2\tau^c(\sigma_\theta^2)}\geq 0$ and $1-\beta^2 e^{-2\tau^d(\sigma_\theta^2)}>0$ and therefore, in view of Proposition 2, it follows that the balanced organization is optimal. ■

From Part 1 of the Proposition, it pays to invest more in communication capacity when the environment becomes more volatile. Intuitively, the cost of not being adapted is then larger and a better communication capacity allows for better adaptation. From Part 2, a focused organization is optimal in environments for which adaptation is not very important. Intuitively, a focused organizations is optimal when the communication capacity is limited, and the organization does not invest much in communication capacity when adaptation is not very important. Similarly, balanced organizations are optimal when adaptation to the environment is very important, and the organization invests heavily in communication capacity.