

Managerial Style and Attention*

Wouter Dessen

Tano Santos

Columbia University

Columbia University

September 1, 2016

Abstract

Is firm behavior mainly driven by its environment or rather by the characteristics of its managers? We develop a cognitive theory of manager fixed effects, where the allocation of managerial attention determines firm behavior. We show that in complex environments, the endogenous allocation of attention exacerbates manager fixed effects. Small differences in managerial expertise then may result in dramatically different firm behavior, as managers devote scarce attention in a way which amplifies initial differences. In contrast, in less complex environments, the endogenous allocation of attention mitigates manager fixed effects. Firm owners prefer managers with task-specific expertise only in complex environments.

*We thank our discussant, John Roberts, and seminar participants at Stony Brook University, USC, UCL Louvain, Boston University, Columbia GSB, the University of Utah, Queens University, the Rational Inattention Workshop in Paris, the 2014 NBER Org Econ Conference in Stanford and the 17th NBER-CCER conference on China and the World Economy in Beijing.

1 Introduction

Scholars in management have long emphasized the role of executive leadership on organizational outcomes. According to the “upper echelons theory,” as set forth by Hambrick and Mason (1984), a central requirement for understanding organizational behavior is to identify those factors that direct or orient executive attention. Organizational outcomes, such as strategies and performance, are expected to reflect the values and cognitive biases of top managers in the organization. In this view, a chief executive’s background in operations makes him more inclined to pursue a cost-reduction strategy, whereas a chief executive with a marketing and sales background is more likely to pursue growth strategies. Bounded rationality and biased information processing is seen as playing a central role in this process. According to Finkelstein, Hambrick and Cannellas (2009, p. 46):

“The logic of bounded rationality hinges on the premise that top executives are confronted with far more stimuli – both from inside and outside the organization – that they can fully possibly comprehend, and that those stimuli are often ambiguous, complex, and contradictory.”

A substantial body of empirical evidence supports this view.¹ A major concern about studies which attribute observed variation in firm behavior to manager fixed effects, however, is that managerial knowledge and expertise are largely endogenous. Managers allocate attention in order to learn about strategic choices. Similarly, boards of directors decide whether or not to hire managers with expertise in certain areas. In addition, it is unclear whether biased information processing and other ‘human’ factors are necessarily at the center of the correlation between managerial characteristics and firm behavior, as suggested by the management

¹Barker and Mueller (2002), for example, find that CEO characteristics explain a significant proportion of the variance in R&D spending even when controlling for industry and firm-level attributes. Similarly, following a seminal paper by Bertrand and Schoar (2003), a growing literature in corporate finance has shown that managerial characteristics are strongly correlated with a variety of corporate policies, such as mergers and acquisitions, debt levels and growth versus cost-cutting strategies. See Finkelstein, Hambrick and Cannellas (2009) for a comprehensive review.

literature.² Unfortunately, the development of theoretical models which study the endogenous expertise of managers, and its correlation with firm behavior, has lagged compared to the growing body of empirical work.

In this paper, we theoretically study to what extent firm behavior is driven by a firm's environment or by the characteristics of its managers. To do so, we develop a cognitive theory of manager fixed effects, where the allocation of managerial attention determines firm behavior. We show how in uncertain and complex environments the endogenous allocation of attention exacerbates manager fixed effects. Small differences in managerial expertise may then result in very different firm behavior. Moreover, firm owners (e.g. boards) then optimally hire 'managers with style' – that is managers with specialized expertise in one particular task – even when the board has no preference for a particular strategy. As a result, in complex and uncertain environments, manager fixed effects are predicted to be pervasive, even when managers are endogenously chosen by unbiased boards out of a large pool of potential candidates. In contrast, in less complex environments, the endogenous allocation of attention mitigates manager fixed effects and boards opt for generalists with a 'broad field of vision.'

In our model, a manager selectively allocates attention in order to learn about two (non-exclusive) strategic choices. Each strategic choice concerns a different task or function. For example, the two tasks may be operations and marketing, and the manager may want to learn about opportunities to reduce unit costs and grow revenues. In order to understand firm behavior, it is then important to understand how the manager allocates her attention. The role of the manager is three-fold. First, she must learn about the nature of two task-specific shocks, which inform the optimal choices pertaining to those tasks. How well she observes a particular shock depends both on her expertise (which may differ across tasks) and how much attention she devotes to each task. In our framework expertise and attention are substitutes in the learning process, not complements, which allows for a clean interpretation of our results.

²According to Hambrick (2015): "The central premise of upper echelons theory is that top executives view their situations - opportunities, threats, alternatives and likelihoods of various outcomes - through their own highly personalized lenses. These individualized construals of strategic situations arise because of executives' experiences, values, personalities and other human factors. Thus, according to the theory, organizations become reflections of their top executives."

Second, the manager makes a strategic choice for each task. Finally, she communicates the firm's strategy (the two strategic choices) to the remainder of the organization, which needs to implement the chosen strategy. We study whether it is optimal for the manager to allocate more attention to certain tasks and how this allocation is related to her expertise and the strategic choices she ends up making. We refer to *managing with style* as biasing managerial attention to one particular task, and mainly implementing strategies related to that same task.

As a benchmark, we consider the case where the manager only maximizes *external alignment*, that is how adaptive her strategic choices are to the task-specific shocks. A generalist manager who has equal expertise about both tasks then divides her attention evenly, whereas a specialist manager, who has more expertise about one task, compensates by devoting more attention to the task she is less knowledgeable about. Intuitively, additional signals about the same shock are partially substitutes, resulting in decreasing marginal returns to devoting attention to the same task. We refer to this as an “unbiased” allocation of attention.

How effective any given strategic choice is, however, also depends on how well it is executed by the organization, referred to as *internal alignment*. The importance of internal and external alignment has long been emphasized in the management literature and has also been very prominent in the recent organizational economics literature (Alonso, Dessein and Matouschek (2008), Rantakari (2008, 2013), Bolton, Brunnermeier and Veldkamp (2011), Van den Steen (2013a,b)). Internal alignment of a strategic choice depends on how well this choice is understood by the organization, who must take complementary actions to ensure effective implementation. As we show, when the need for internal alignment is sufficiently important, the manager optimally *communicates only about the task which faces the largest perceived shock*. In our example above, the manager then communicates either about a strategy for cost-reduction *or* about a strategy for revenue growth, with her choice driven by what she perceives as the biggest opportunity for the firm. By focusing all communication on one strategic choice, this choice can be very responsive to the corresponding task-specific shock (external alignment) without sacrificing internal alignment as everyone in the organization has a clear understanding of this choice. Internal alignment on the other task, in contrast, is achieved by selecting a strategy which is largely unresponsive to the relevant shock and thus

no communication is needed to achieve coordination as the strategic choice is standard.

When internal alignment is important, the organization thus only adapts to one of the two task-specific shocks. The role for the manager, then, is to identify which task should be adaptive, and communicate a strategy to the organization focused around this task. How should she allocate her scarce attention for this purpose?

Since the manager only communicates to the organization about the task affected by what she perceives to be the largest shock, attention devoted to the other task is largely wasted from an ex post perspective. From an ex ante perspective, however, the manager does not know which shock is largest and should be the focus of the organization. The more expertise she has in one task, however, or the more attention she devotes to a particular task, the more likely it is that she will perceive as largest the shock pertaining to that task. Thus, a marketing expert or a manager which devotes more attention to marketing than to operations is more likely to perceive larger opportunities for revenue growth than for cost-minimization. But a marketing expert anticipating that she is likely to implement a marketing focused strategy then realizes that any attention devoted to operations is likely to be a waste of time. Therefore, in complex environments and in contrast to our benchmark, marketing experts optimally devote all their attention to the marketing task, making it even more likely that a marketing-focused strategy is implemented. In contrast to our benchmark, the endogenous allocation of attention thus reinforces the initial ‘manager fixed effect’. Small differences in the initial expertise of two managers lead to large differences in organizational behavior, as managers specialize their allocation of attention towards areas in which they have already more expertise.

Interestingly, the same logic applies to a generalist manager who has equal expertise in both tasks. The more such a generalist manager devotes attention to, say, marketing, the more likely she is to perceive marketing shocks to be largest and, hence, the more likely she is to implement marketing-focused strategies. This in turn, makes it optimal for this manager to devote even more attention to marketing, as information learned about operations is likely to be wasted. Because of this complementarity, and in contrast to our benchmark, a generalist manager optimally focuses her scarce attention on one task in complex, uncertain environments. Thus, even if marketing and operations are equally important to the organization and have,

a priori, the same potential for profit improvement, a manager with equal expertise in both marketing and operations should focus all her attention on one task, say marketing. Ex post, such a manager mainly (but not always) selects and communicates about marketing strategies, and she *appears* to be arbitrarily and inefficiently biased towards marketing. Managers may thus have a large impact on firm behavior in a way which cannot be traced back to observable managerial characteristics. Empirical research on the role of managers may therefore underestimate the impact of managers on firm behavior.

While the endogenous allocation of attention may exacerbate manager fixed effects, this is only the case when attention is scarce and the environment is complex or uncertain. Indeed, when managerial attention is *not* scarce, it is possible for the manager to learn both task-specific shocks reasonably well. There is then little to be gained from focusing all attention on one task and the manager risks being blindsided if she neglects the task in which she has less expertise. As in our benchmark, the manager thus optimally allocates attention in a way which reduces or eliminates any differences in task knowledge. Similarly, when the environment is *not* complex or uncertain, the specific realization of the task-specific shocks (and not managerial attention) largely determines which task the manager communicates about to the organization. Because of decreasing marginal returns, a specialist manager then optimally allocates more attention to the task in which she has less expertise, again mitigating manager fixed effects. In the same vein, more capable managers, that is those with more initial expertise in both tasks, are less likely to ‘manage with style’. Indeed, more capable managers observe the environment with greater precision and it is as if the environment is less complex or uncertain. Finally, we show that ‘managing with style’ is more likely to be optimal when organizational implementation is more important. Intuitively, the more important is coordination and internal alignment, the more the manager ignores the shock which she perceives to be smaller, and the larger are the benefits of managerial focus. In sum, the extent to which firm behavior and strategic choices reflect managerial characteristics depends on the complexity of the environment and the scarcity of attention, the capability of the manager and the need for coordinated implementation inside the organization.

While most of our paper assumes that the initial expertise of managers is exogenous, we also study what happens when managers are endogenously chosen, for example by a board of directors. We show that keeping total expertise fixed, specialist managers are preferred over generalists when managing with style is optimal. Our model thus predicts generalist managers to be more prevalent in more certain and slow-moving environments (e.g. mature industries). In contrast, specialist managers are more likely to be optimal in more complex environments (e.g. in conditions faced by many start-up firms and technology companies).

Related Literature.— Following the Carnegie School (Simon and March 1958, Cyert and March 1963), a large management literature has studied limits to human cognition in order to explain organizational behavior. In particular the Upper Echelons Theory of Hambrick and Mason (1984) focuses on the biased and subjective processing of complex, ambiguous information by managers; these biases are informed by the background and values of managers.³ Instead, the present paper focuses on how managers optimally allocate scarce attention among alternative sources of information. A point of similarity is that backgrounds may matter: If the manager has some previous expertise in a particular area she is more likely to pursue strategies that emphasize those areas of expertise. Unlike the Upper Echelons Theory though we show that this is only true if the environment is sufficiently complex. Moreover, in our framework, managers matter even when the manager has no specific background in any particular area.

In the economics literature a number of papers, such as Geanakoplos and Milgrom (1991), do study the optimal allocation of attention in organizations, but almost all are focused on how hierarchies or delegation of decision-making authority can alleviate information-processing constraints or costs.⁴ An exception are a series of recent papers that study the optimal allocation of attention in networks, such as Galeotti and Goyal (2010), Akerlof and Holden (2015), Dessein, Galeotti and Santos (2015) and Herskovic and Ramos (2015). A common finding in those papers is that, even when all agents in the network are ex ante identical, attention tends to be concentrated on a select few of them. Another exception is Van den Steen (2013b), who studies a strategy formulation game in which a strategist chooses which decision to investigate – among a set of interrelated decisions – and then communicates

³See also Ocasio (1997) and the recent survey by Hambrick (2015).

⁴See Garicano and Prat (2012) and Garicano and Van Zandt (2013) for overviews of the literature.

her preferred action to a group of agents in charge of implementing those decisions. Unlike in our model, there is no trade-off between investigating several decisions versus learning more about one decision, and the strategist is never blindsided – she always communicates about the decision she investigates.⁵ In the empirical literature, Bandiera, Guiso, Prat and Sadun (2011) and Bandiera, Prat and Sadun (2013) employ time use surveys to measure how CEOs allocate their attention. Their main focus, however, is on the time CEOs spend on activities with large private benefit (such as meeting with outsiders) as opposed to activities that mainly improve firm performance (such as meeting with insiders). Finally, our paper contributes to the literature on narrow business strategies and vision (Rotemberg and Saloner 1994, 2000) and organizational focus (Dessein, Galeotti and Santos, 2015) by endogenizing to what extent selected business strategies are contingent on the organizational environment as opposed to managerial characteristics.

2 The model

2.1 The production process

We posit a team-theoretic model in which there are two tasks $i \in \{1, 2\}$, say marketing and operations, one manager and an organization consisting of a continuum of employees $j \in [0, 1]$. Profits of the organization depend on (i) *external alignment*, that is, how well is each task i adapted to an independently normally distributed task-specific shock $\theta_i \sim N(0, \sigma_\theta^2)$ and (ii) *internal alignment*, that is, how well are the two tasks implemented by the organization. Specifically, the manager selects for each task i a strategic choice a_{Mi} which must match as well as possible the task specific shock θ_i . In turn, each employee j must choose complementary actions a_{j1} and a_{j2} to match as well as possible the strategic choices selected by the

⁵Hence, strategic choices are purely a function of managerial characteristics: *ceteris paribus*, the strategist investigates and communicates about the decision about which she expects to receive the most informative signal (that is, the one she has more expertise in). In contrast, our model endogenizes to what extent strategic choices reflect managerial characteristics as opposed to the realization of environmental shocks.

manager. The realized pay-offs are given by

$$\pi \equiv \sum_{i \in \{1,2\}} \left[h(\theta_i) - (a_{Mi} - \theta_i)^2 - \beta \int_0^1 (a_{Mi} - a_{ji})^2 dj \right], \quad (1)$$

where β captures the relative importance of internal alignment.⁶ We refer to $\mathbf{a}_M = (a_{M1}, a_{M2})$ as the strategic choices or strategies for short and to $\mathbf{a}_j = (a_{j1}, a_{j2})$ as the implementation of those strategies by agent j .

Only the manager can learn about θ_1 and θ_2 , but she can communicate her strategic choices to the organization. Specifically, we assume that the manager first devotes attention to tasks 1 and 2 in order to learn about θ_1 and θ_2 , then the manager chooses her strategy $\mathbf{a}_M = (a_{M1}, a_{M2})$ and communicates those choices to the organization and, finally, the employees $j \in [0, 1]$ implement those strategies by choosing complementary actions a_{j1} and a_{j2} . Without loss of generality, we will assume that $h(\theta_i) \equiv \theta_i^2$ so that profits π are normalized to 0 whenever $\mathbf{a}_M = \mathbf{a}_j = (0, 0)$ for all $j \in [0, 1]$.⁷ Adapting to θ_i can then be interpreted as an opportunity to improve performance in task i .

We will describe in more detail the learning process and communication technology below. At this point, we want to note that whenever communication is imperfect, there is a trade-off between external and internal alignment. By selecting strategic choices which are responsive to the task-specific shocks θ_i the manager sacrifices some internal alignment as not all employees may understand her strategy. In contrast, perfect internal alignment can always be achieved by selecting the standard strategic choice $a_{Mi} = 0$. As we will show, in the absence of any communication – or when communication fails – employees optimally choose actions which are complementary to the standard strategy $a_{Mi} = 0$.

⁶The pay-off function (1) is similar to the pay-off functions considered in a series of organizational economics papers focused on coordination issues in organizations, such as Dessein and Santos (2006), Alonso, Dessein and Matouschek (2008,2015), Rantakari (2008,2013), Bolton, Brunnermeier and Veldkamp (2013), Calvo-Armengol, de Marti and Prat (2015) and Dessein, Galeotti and Santos (2015) among others. All these papers including this one view the trade-off between external and internal alignment (or adaptation and coordination) as the central trade-off in organizations.

⁷Indeed, neither the optimal choices for \mathbf{a}_M and \mathbf{a}_j , nor the optimal allocation of attention are affected by the functional form of $h(\theta_i)$.

2.2 Communication and implementation inside the organization

In order to ensure effective implementation, the manager needs to communicate her strategic choices to employees so they can take the appropriate complementary actions. Communication though is imperfect. The more attention is devoted to a strategic choice, the more likely an employee understands how to implement it. Formally, we model communication as a Poisson process with a hazard rate μ and the stochastic event corresponding to the employee ‘understanding’ a particular strategic choice. Specifically, let $r_i \geq 0$ be the amount of time devoted to process information related to strategic choice i , then an employee understands strategic choice i with probability

$$p_i = 1 - e^{-\mu r_i}, \quad (2)$$

which is independent across agents. Given this, employee j ’s choices are given by:

$$a_{ji} = \begin{cases} a_{Mi} & \text{with probability } p_i \\ 0 & \text{with probability } 1 - p_i \end{cases}$$

Thus, with probability p_i worker j understands a_{Mi} and sets the complementary action equal to this choice and with probability $1 - p_i$ the worker simply does not and sets the complementary action equal to the the mean value of the shock θ_i , which is 0. The manager controls the allocation of attention by workers r_i subject to an organizational attention constraint

$$r_1 + r_2 \leq r \quad \text{with} \quad r > 0 \quad (3)$$

Alternatively r_i can be interpreted as the time the manager devotes to communicate about strategic choice a_{Mi} . It will be useful to rewrite communication constraint (3) as follows:

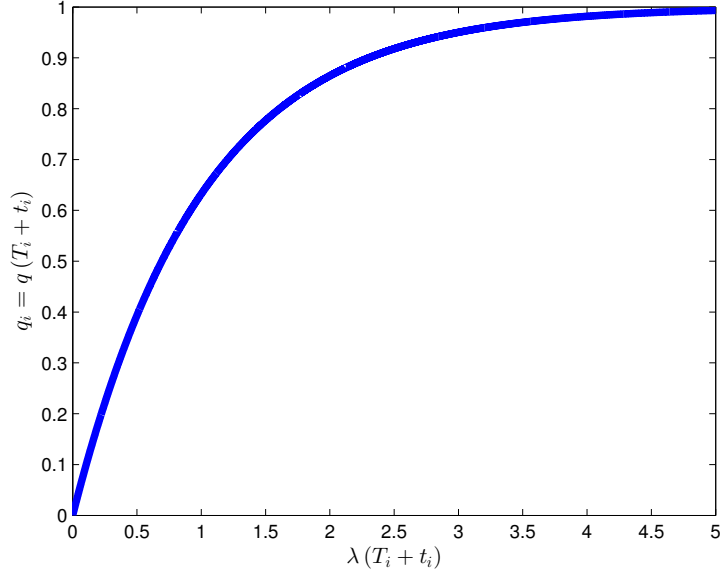
$$(1 - p_1)(1 - p_2) \geq 1 - p \quad (4)$$

where $p \equiv 1 - e^{-\mu r}$ denotes the probability of understanding a_{Mi} when an employee’s attention is fully dedicated to task i .

2.3 Allocation of attention and learning by the manager

Consider now the learning process of the manager. We assume that the manager observes an *endogenous* signal s_i about each shock θ_i whose informativeness depends on the *managerial*

Figure 1: *Managerial learning technology.* $q(T_i + t_i) = 1 - e^{-\lambda(T_i+t_i)}$ which determines the residual variance $RV(\theta_i|t_i, T_i) = \sigma_\theta^2 - q(T_i + t_i) \sigma_\theta^2$.



attention t_i devoted to task i . In addition, the manager observes an *exogenous* signal S_i about θ_i . How informative S_i is depends on her *managerial expertise* T_i in task i . In the remainder of the paper, we will assume that $F(\theta_i|S_i, s_i)$, the manager's posterior about θ_i is normally distributed with mean estimate $\hat{\theta}_i = E[\theta_i|s_i, S_i]$ and variance

$$RV(\theta_i|t_i, T_i) \equiv \sigma_\theta^2 - q_i \sigma_\theta^2 \quad \text{where} \quad q_i \equiv q(T_i + t_i) = 1 - e^{-\lambda(T_i+t_i)}. \quad (5)$$

Hence, the mean-squared error or residual variance $RV(\theta_i|t_i, T_i)$ decreases at a logarithmic rate as a function of both attention t_i and expertise T_i , where t_i and T_i are substitutes in the learning process (see Figure 1). Appendix A provides microfoundations for (5). As was the case for organizational attention (r_1, r_2) , managerial attention is also scarce in that

$$t_1 + t_2 \leq 2\tau. \quad (6)$$

We denote by Υ the set of feasible allocations of attention: $\Upsilon = \{(t_1, t_2) : t_1 + t_2 \leq 2\tau\}$. Our main propositions will deal with the optimal allocation of managerial attention (t_1, t_2) .

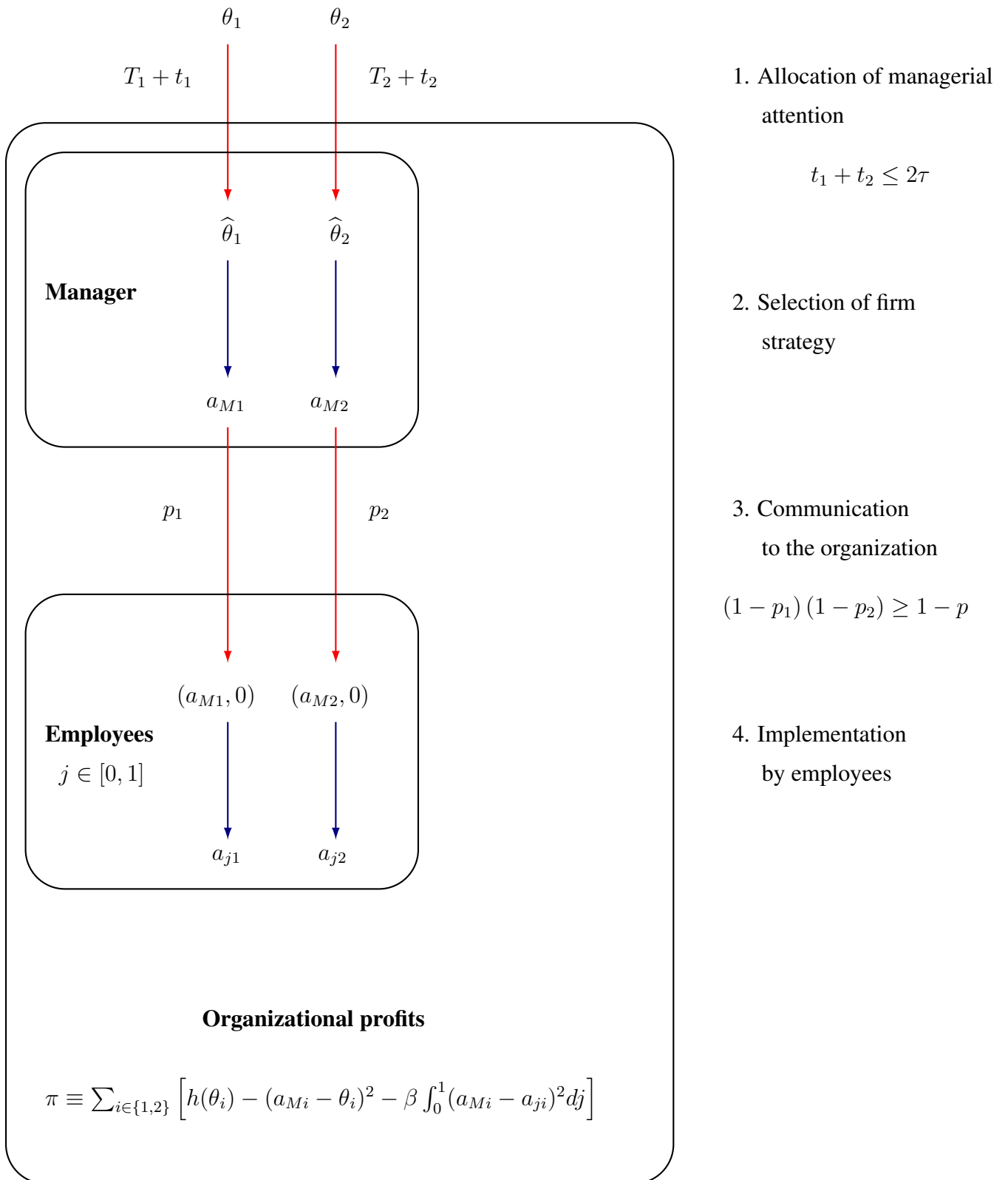
Discussion of managerial learning technology. In (5), the parameter λ characterizes the speed of learning and, hence, $1/\lambda$ the complexity of the environment. A large λ , for instance, is

consistent with situations in which ex post uncertainty is small as, say, the market in which the organization operates is mature and strategic choices are well understood. Instead a small λ is associated with environments with large ex post uncertainty, perhaps because the organization is operating in a new industry or in a period characterized by a lot of turbulence. We refer to λT_i as the manager's *effective expertise*, a combination of his managerial expertise T_i and the complexity of the environment $1/\lambda$. What determines the precision of the information on the task specific shock, (5), is the combination of the effective expertise, which is exogenous, and the *effective attention*, λt_i , which depends on the endogenous allocation of attention. In sum, attention, expertise and the complexity of the environment combine in $\lambda (T_i + t_i)$ to determine the precision of the manager's information about θ_i .

If $T_1 > T_2$ then we speak of a manager who is specialized or is an expert in task 1, whereas we refer to a manager for whom $T_1 = T_2$ as a generalist. Conceptually, we think of a manager specialized in task i as having access to more precise information about the shock pertaining to task i than a generalist or a specialist in task $j \neq i$, even before the attention decision is made. However, from the formulation of our learning technology (5), a non-specialist manager can compensate for her lack in expertise in a specific task by devoting more attention to it. For example, she can consult experts, do extensive research, or simply devote more time to analyze her options in that particular task as she cannot rely on past experience or knowledge. Thus in our model expertise and attention are substitutes: A decrease in expertise can be compensated by a corresponding increase in attention.

Our assumption that expertise and attention are substitutes is 'conservative' from a modeling perspective. Indeed, a central result in this paper is that experts often devote more attention to tasks in which they have superior expertise. By ruling out that expertise and attention are complements in the learning technology, we ensure that our main results are not driven by assumptions regarding the learning technology.

Figure 2 : A graphical representation of the model



1. Allocation of managerial attention

$$t_1 + t_2 \leq 2\tau$$

2. Selection of firm strategy

3. Communication to the organization

$$(1 - p_1)(1 - p_2) \geq 1 - p$$

4. Implementation by employees

2.4 Timing

Figure 2 illustrates the timing of our model, flow of information and actions taken by both management and employees. The timing in our model is as follows.

1. The manager allocates attention $t_i \in [0, 2\tau]$ to each task $i = 1, 2$ to learn about (θ_1, θ_2) , with $t_1 + t_2 \leq 2\tau$.
2. Having observed the corresponding signals, the manager obtains posterior estimates $\hat{\theta}_i$ for $i = 1, 2$ and selects strategic choices $\mathbf{a}_M = (a_{M1}, a_{M2})$.
3. The manager chooses the probabilities p_1 and p_2 with which employees will understand her strategic choices a_{M1} and a_{M2} with $(1 - p_1)(1 - p_2) \geq (1 - p)$.
4. Having learned about management choices, employees select their complementary actions, a_{ji} for $j \in [0, 1]$ and $i = 1, 2$.

2.5 Benchmark: Attention and External Alignment

As a benchmark, consider the case where the manager only maximizes *external alignment*

$$\sum_{i \in \{1,2\}} E [h(\theta_i) - (a_{Mi} - \theta_i)^2].$$

One interpretation of this benchmark is that there is no need for implementation ($\beta = 0$ in the pay-off function (1)). A second interpretation is that while there is a need for implementation, there are no communication frictions and the manager can perfectly communicate her strategic choices to employees. Given signals (s_1, S_1, s_2, S_2) the manager then optimally sets $(a_{M1}, a_{M2}) = (\hat{\theta}_1, \hat{\theta}_2)$ and allocates attention in order to minimize

$$\sum_{i \in \{1,2\}} E [(\hat{\theta}_i - \theta_i)^2 | t_i, T_i] = \sum_{i \in \{1,2\}} \text{RV}(\theta_i | t_i, T_i).$$

Given learning technology (5), the marginal returns to devoting attention to task i are given by

$$\frac{\partial(-\text{RV}(\theta_i | t_i, T_i))}{\partial t_i} = \lambda(1 - q_i)\sigma_\theta^2,$$

where $q_i = q(t_i + T_i)$ is increasing in both t_i and T_i . It follows that when $t_1 > t_2$, the marginal returns to devoting attention to task 2 are higher than to task 1 provided the manager is a generalist ($T_1 = T_2$). Similarly, the marginal returns to devoting attention to any given task i are decreasing in the manager's expertise T_i in that task. Intuitively, our learning technology features decreasing marginal returns to attention and, by assumption, attention and expertise are substitutes.⁸ The following result obtains:

Proposition 1 *The allocation of managerial attention $\mathbf{t} = (t_1, t_2)$ which maximizes external alignment ($\beta = 0$) is given by (1) $t_1 = t_2 = \tau$ if the manager is a generalist ($T_1 = T_2$) and (2) $0 \leq t_1 < t_2 \leq 2\tau$ if the manager has more expertise in task 1 ($T_1 > T_2$).*

2.6 A preview of the results

We graphically illustrate our main results and contrast them with the benchmark in section 2.5. Consider Figure 3, panels A and B, which show managerial attention choices when $p = 0.75$ and $\beta = 5$. There we represent managers in the space $(\lambda(T_1 + t_1), \lambda(T_2 + t_2))$. The allocation of attention can be visualized as arrows going from a point in the plane $(\lambda T_1, \lambda T_2)$, which is the manager's initial expertise, to another point $(\lambda(T_1 + t_1), \lambda(T_2 + t_2))$, her final expertise profile once the manager decides on a particular allocation of attention. The complexity of the environment, as summarized by λ , plays a critical role in what follows. The line going through points A and B in both panels corresponds to a manager with a given profile (T_1, T_2) under environments of different complexity: Point A corresponds to a complex environment (low λ) and B to a less complex environment (high λ). We refer to *managing with style* as either a situation in which a specialist manager (say with $T_1 > T_2$) devotes all his attention to the task in which she has more expertise, or a situation in which a generalist manager (for whom $T_1 = T_2$) arbitrarily biases his attention to one particular task. Note that in our benchmark

⁸Obviously, in practice there may exist settings in which there are increasing marginal returns to attention (at least for some parameter ranges) or technological complementarities between attention and expertise. As noted above, our assumptions regarding the learning technology should not be regarded as a positive statement but rather as a modeling device to highlight the organizational trade-offs that lead to 'managing with style' even in the presence of technological drivers that push against this possibility.

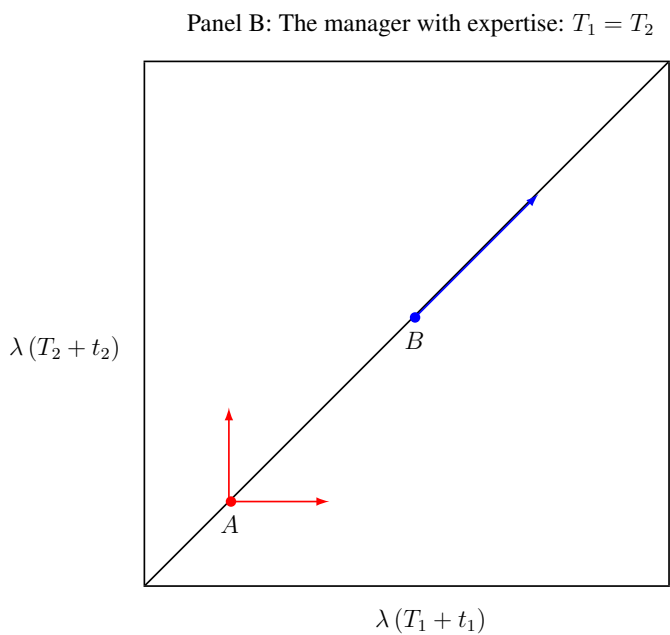
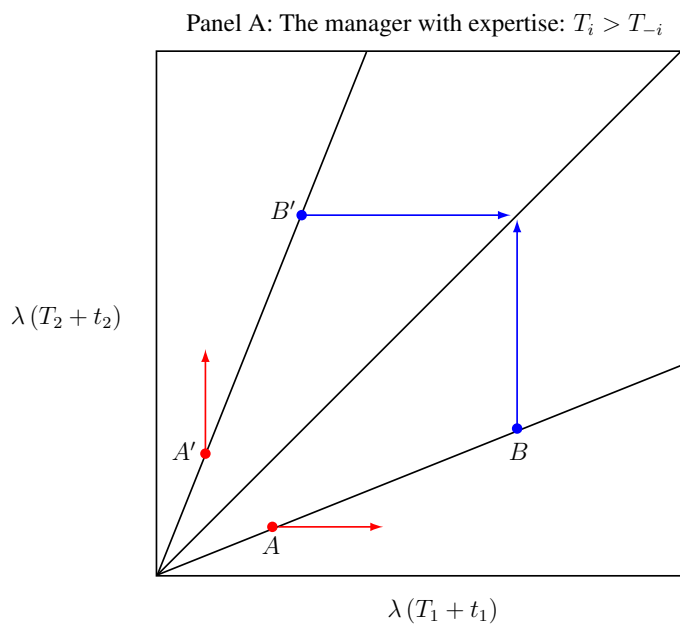
above, managing with style is always suboptimal.

Consider first the case of the specialist manager with $T_1 > T_2$ operating in a complex environment (point A) and assume that the amount of endogenous attention to be allocated, $t_1 + t_2 = 2\tau$, is not too large. Proposition 4 shows that in this case, the manager optimally “manages with style.” She will opt to allocate her attention so as to reinforce her expertise in task 1 and the manager’s precision of information in task 1 and 2 is given by $\lambda(T_1 + 2\tau)$ and λT_2 , respectively. The organization is then disproportionately responsive to shocks regarding task 1. In fact, as we will show, the manager then often only communicates to the organization about task 1. Similarly, a manager with $T_2 > T_1$ operating in a complex environment (point A’) optimally devotes all her attention to task 2 and mainly communicates to the organization about task 2. It follows that in complex environments, two identical organizations led by managers with different expertise (A versus A’) tend to respond very differently in response to identical shocks. In other words, firm behavior is largely driven by initial differences in expertise between managers, which are amplified by the allocation of managerial attention. Note that this stands in sharp contrast with the results obtained in our benchmark.

Instead when the environment is less complex (point B in panel A of Figure 3), a manager for whom $T_1 > T_2$ will instead opt to devote all her attention to task 2, in which she has less expertise, resulting in a final expertise profile which is almost identical for both tasks. In contrast to the case of a complex environment, two organizations led by managers with different initial expertise now behave very similarly in response to identical shocks. In other words, initial differences in expertise between two managers are much less predictive of organizational behavior.

Proposition 3 shows that, surprisingly, a similar result obtains for the case of the generalist manager ($T_1 = T_2$). In complex environments, a generalist manager optimally specializes in one task by devoting all her attention to one task or the other, which task specifically being a matter of indifference (see Panel B of Figure 3). In contrast, in less complex environments, a generalist manager splits her attention equally among both tasks. We now turn to a formal derivation of all these results.

Figure 3: A preview of the results. Endogenous allocation of attention. The plot shows the endogenous allocation of attention in complex (A and A') and less complex (B and B') environments for the case of the specialist and the generalist manager (Panels A and B).



3 Optimal allocation of organizational attention

The benchmark studied in section 2.5 ignored the organizational implementation of strategic choices. In this section, we exclusively focus on strategy implementation: How are strategic choices optimally communicated or, equivalently, how is organizational attention optimally allocated? And how does this affect optimal strategic choices? We answer those questions taking the allocation of managerial attention and the resulting posterior estimates $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ as given. In Section 4 we then endogenize the allocation of managerial attention.

Our aim is to characterize strategic choices (a_{M1}, a_{M2}) and organizational attention choices (p_1, p_2) given posteriors $\hat{\theta}$. Simple manipulations of the objective function (1) yield

$$\begin{aligned} E(\pi|\hat{\theta}) &= \sum_{i \in \{1,2\}} \left[E(\theta_i^2|\hat{\theta}_i) - (a_{Mi} - \hat{\theta}_i)^2 - E\left[\left(\theta_i - \hat{\theta}_i\right)^2|\hat{\theta}_i\right] - \beta(1 - p_i)a_{Mi}^2 \right] \\ &= \sum_{i \in \{1,2\}} \left[\hat{\theta}_i^2 - (a_{Mi} - \hat{\theta}_i)^2 - \beta(1 - p_i)a_{Mi}^2 \right]. \end{aligned}$$

As is standard in the team theory literature we limit our attention to equilibria in linear strategies, that is,

$$a_{Mi} = \alpha_i \hat{\theta}_i.$$

The problem for the manager is then equivalent to choosing $(p_1, p_2, \alpha_1, \alpha_2)$ in order to maximize expected profits

$$E(\pi|\hat{\theta}) = \sum_{i \in \{1,2\}} \left[\hat{\theta}_i^2 - (1 - \alpha_i)^2 \hat{\theta}_i^2 - \beta(1 - p_i) \alpha_i^2 \hat{\theta}_i^2 \right] \quad \text{subject to (4)}. \quad (7)$$

Direct inspection of (7) shows that α_i and p_i are complementary choices. The larger is α_i and the more responsive is the manager to her posterior about shock θ_i , the more organizational attention should be directed to her strategic choice i in order to ensure internal alignment. Similarly, when p_i is larger and employees are better at implementing strategic choice i , then it becomes optimal for the manager to be more responsive to her posterior $\hat{\theta}_i$. Formally, taking p_1 and p_2 as given and maximizing (7) with respect to α_1 and α_2 yields

$$\alpha_1 = \frac{1}{1 + \beta(1 - p_1)} \quad \text{and} \quad \alpha_2 = \frac{1}{1 + \beta(1 - p_2)}. \quad (8)$$

Hence, the larger is p_i , the more adaptive is the manager to the posterior $\hat{\theta}_i$. Armed with this we can now turn to the optimal choice of p_i .

Proposition 2 *Suppose $\beta > 1$. Whenever $p < \bar{p}(\beta) \equiv 1 - 1/\beta^2$, the manager directs all organizational attention to one strategic choice:*

$$(p_1^*, p_2^*) = \begin{cases} (p, 0) & \text{if } \hat{\theta}_1^2 > \hat{\theta}_2^2 \\ (0, p) & \text{if } \hat{\theta}_1^2 < \hat{\theta}_2^2 \end{cases}$$

Intuitively, when internal alignment is important (β large), external alignment is very costly unless communication is effective. The manager then optimally communicates intensively about one strategic choice, allowing that strategic choice to be responsive to its task-specific shock without compromising internal alignment (see (8)). Internal alignment on the other strategic choice is then achieved by largely giving up on external alignment regarding that task – in other words, the other task will not be very responsive to the posterior in order to avoid poor internal alignment. Naturally, it is optimal to communicate about the strategic choice which faces the largest shocks, as external alignment is most important for that task.

In contrast, when implementation/internal alignment is not very important (β is small) the manager is optimally responsive to both shocks provided they are sufficiently equal in size. Sacrificing external alignment on one task in order to improve internal alignment is then not worth it. Similarly, if attention is relatively unconstrained (p is large), then the manager can communicate effectively about both strategic choices, and there is no real trade-off between external and internal alignment. Even when implementation is very important, it is then still optimal to communicate about both tasks.⁹

⁹The result in Proposition 2 is reminiscent of the main result in Dessein, Galeotti and Santos (2015, DGS henceforth). There are though three key differences. First, in DGS, information is dispersed: For each task, there is one manager who observes information pertaining to his and his task only. In contrast, in our model, there is only one manager who decides which tasks to devote (managerial) attention to. Second, in DGS, the allocation of organizational attention *cannot* be made contingent on the realization of task specific shock. In contrast, in our model, the manager directs organizational attention *after* observing the task-specific shocks. Finally, in DGS, managers observe their task specific shock perfectly. DGS therefore cannot address the issue of managerial style and attention, the subject of this paper.

In what follows, we will assume that there is a tight bound on organizational attention:

$$p \leq \bar{p}(\beta) \tag{A1}$$

so that whenever $\hat{\theta}_i^2 > \hat{\theta}_{-i}^2$ then the manager only communicates a_{Mi} , that is $p_i = p$ and $p_{-i} = 0$, and (8) specializes to

$$a_{Mi} = \frac{\hat{\theta}_i}{1 + \beta(1 - p)} \quad \text{and} \quad a_{M-i} = \frac{\hat{\theta}_{-i}}{1 + \beta}. \tag{9}$$

We maintain (A1) throughout the paper in order to simplify the analysis. Note that (A1) is a condition that guarantees that the manager only communicates about one task, even when $|\hat{\theta}_i| = |\hat{\theta}_{-i}|$. Hence, even when $p > \bar{p}(\beta)$, the manager may communicate only about task i provided that $|\hat{\theta}_i|$ is sufficiently larger than $|\hat{\theta}_{-i}|$.

Before studying the optimal allocation of managerial attention, it is useful to consider the limit case where managerial attention is unconstrained so that $\hat{\theta}_i = \theta_i$. Given (A1), the manager then directs organizational attention to the largest realized shock. There is no sense, however, in which the organization or the manager are biased towards one particular task. Ex ante, each task is equally likely to be the focus of organizational attention. Furthermore, all organizations faced with the same environment, a particular realization of θ_1 and θ_2 , will focus attention on the same task. It follows then that scarcity of organizational attention in the absence of scarcity of managerial attention does not result in any systematic bias in organizational strategies.

4 Optimal allocation of managerial attention

Scarcity of *organizational attention* implies that the manager only communicates about the largest perceived shock and is disproportionately responsive to her posterior about this shock. Anticipating this, how does the manager optimally allocate her scarce *managerial attention*, and how does this affect firm behavior?

4.1 Expected profits

Before analyzing the optimal allocation of attention we first develop the expected profit function for a given allocation of attention $\mathbf{t} = (t_1, t_2)$. For this purpose, we first express expected profits for given posteriors $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)$ and communication choices (p_1, p_2) , and subsequently take expectations over posteriors for an allocation of managerial attention.

Given (8), the expected profit, (7), associated with task i conditional on $\hat{\boldsymbol{\theta}}$ and communication choices (p_1, p_2) , is given by

$$E \left[\pi_i | \hat{\theta}_i \right] = \hat{\theta}_i^2 - \left(\frac{\beta(1-p_i)}{1+\beta(1-p_i)} \right)^2 \hat{\theta}_i^2 - \frac{\beta(1-p_i)}{(1+\beta(1-p_i))^2} \hat{\theta}_i^2 = \frac{\hat{\theta}_i^2}{1+\beta(1-p_i)}.$$

Given that $\hat{\theta}_i^2 > \hat{\theta}_{-i}^2$ and given our assumption on scarce organizational attention, (A1), the manager only communicates about task i : $(p_i, p_{-i}) = (p, 0)$ (see Proposition 2). Hence, the expected profit of the organization conditional on posteriors $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)$ equals

$$E \left(\pi | \hat{\boldsymbol{\theta}} \right) = E \left[\pi_i | \hat{\theta}_i \right] + E \left[\pi_{-i} | \hat{\theta}_{-i} \right] = \frac{\hat{\theta}_i^2}{1+\beta(1-p)} + \frac{\hat{\theta}_{-i}^2}{1+\beta}.$$

Note that for large values of β and p , the expected profitability of task i conditional on $\hat{\theta}_i$ is much greater than that of task $-i$, even when $\hat{\theta}_i^2 \sim \hat{\theta}_{-i}^2$, as the manager only communicates about task i . Given attention allocation $\mathbf{t} = (t_1, t_2)$, the (unconditional) expected profits are given by

$$\begin{aligned} \Pi(q_1, q_2) \equiv E[\pi] &= \Pr \left(\hat{\theta}_1^2 \geq \hat{\theta}_2^2 \right) E \left[\frac{\hat{\theta}_1^2}{1+\beta(1-p)} + \frac{\hat{\theta}_2^2}{1+\beta} \mid \hat{\theta}_1^2 \geq \hat{\theta}_2^2 \right] \\ &+ \Pr \left(\hat{\theta}_1^2 < \hat{\theta}_2^2 \right) E \left[\frac{\hat{\theta}_1^2}{1+\beta} + \frac{\hat{\theta}_2^2}{1+\beta(1-p)} \mid \hat{\theta}_1^2 < \hat{\theta}_2^2 \right]. \end{aligned} \quad (10)$$

From (5), given attention allocation $\mathbf{t} = (t_1, t_2)$ we have

$$\hat{\theta}_i \sim \mathcal{N} \left(0, q_i \sigma_{\theta}^2 \right), \quad (11)$$

where $q_i = q(T_i + t_i)$. Let $F(x, y)$ denote the normal c.d.f. of random variable x with mean 0

and variance y . Then given (11), we can rewrite (10) as:¹⁰

$$\begin{aligned} & \Pi(q_1, q_2) \\ = & 4 \int_0^{+\infty} \left[\int_0^{\hat{\theta}_2} \frac{\hat{\theta}_1^2}{1+\beta} dF(\hat{\theta}_1, q_1 \sigma_\theta^2) + \int_{\hat{\theta}_2}^{+\infty} \frac{\hat{\theta}_1^2}{1+\beta(1-p)} dF(\hat{\theta}_1, q_1 \sigma_\theta^2) \right] dF(\hat{\theta}_2, q_2 \sigma_\theta^2) \\ + & 4 \int_0^{+\infty} \left[\int_0^{\hat{\theta}_1} \frac{\hat{\theta}_2^2}{1+\beta} dF(\hat{\theta}_2, q_2 \sigma_\theta^2) + \int_{\hat{\theta}_1}^{+\infty} \frac{\hat{\theta}_2^2}{1+\beta(1-p)} dF(\hat{\theta}_2, q_2 \sigma_\theta^2) \right] dF(\hat{\theta}_1, q_1 \sigma_\theta^2) \end{aligned}$$

We use this expression to investigate next the optimal allocation of attention. We proceed by distinguishing between the case of a generalist manager and a specialist manager with superior expertise in one task versus the other.

4.2 Allocation of attention by a generalist manager

Consider first a generalist manager with equal expertise about both tasks, that is $T_1 = T_2 = T$. Should the generalist manager divide her attention equally among both tasks, or should she focus her attention on one (randomly chosen) task?

From (11), when the manager allocates more attention to task 1, then she is more likely to communicate to the organization about task 1 as she is more likely to perceive the shock affecting task 1 to be the largest. Indeed, from (11), even though $\Pr(\theta_1^2 > \theta_2^2) = \Pr(\theta_2^2 > \theta_1^2)$ we have that

$$\Pr(\hat{\theta}_1^2 > \hat{\theta}_2^2) > \Pr(\hat{\theta}_2^2 > \hat{\theta}_1^2) \iff T_1 + t_1 > T_2 + t_2 \quad (12)$$

To see this more intuitively, consider the extreme case where $T_1 = T_2 = 0$ and $t_1 > t_2 = 0$. The posterior estimate of θ_2 then always equals the prior estimate, that is $\hat{\theta}_2 = 0$, so that we always have that $\hat{\theta}_1^2 > \hat{\theta}_2^2$.

The allocation of attention thus influences how adaptive the organization is to an external shock through two channels: how well the manager observes the relevant external shock and how likely the manager is to communicate about the strategic choice. It follows that the more attention a manager devotes to task 1, the more (on average) the organization is adaptive

¹⁰The integral further uses the fact that the probability density of $|\hat{\theta}_i|$ is twice the probability density of $\hat{\theta}_i = |\hat{\theta}_i|$, as reflected in the number 4 in front of the expression. For a derivation of this expression see Appendix.

to $\hat{\theta}_1$ and, hence, the more valuable it is to improve the precision of $\hat{\theta}_1$ by devoting even more attention to task 1. Formally, if $t_1 > t_2$ then $\Pr(\hat{\theta}_1^2 > \hat{\theta}_2^2) > \Pr(\hat{\theta}_2^2 > \hat{\theta}_1^2)$ and from expression (10), profits are more sensitive to $E(\hat{\theta}_1^2)$ than to $E(\hat{\theta}_2^2)$. By devoting all attention to one task, the manager is then very good at responding to the shock affecting this task and she is likely to communicate about this task.

If there were to be constant returns to attention, for example if $q_i = \delta t_i$, then using (10), it is easy to show that the manager optimally focuses all attention on one task up to the point where the shock is perfectly observed, that is $q_i = 1$. Indeed, because of the complementarity between (i) how much attention a task receives and (ii) how likely a task is to be communicated to the organization, the profit function (10) is then convex in t_i .

If instead, as in this paper, there are decreasing marginal returns to attention, then whether or not focused attention is optimal depends on the scarcity of managerial attention τ and the manager's effective expertise, λT . Proposition 3 characterizes the optimal allocation of attention, but restricts the analysis to the case where managerial attention is either abundant or scarce. To simplify the analysis, we further assume that $t_i \in \{0, \tau, 2\tau\}$ with $t_1 + t_2 \leq 2\tau$. We will revert to our standard assumption that $t_i \in [0, 2\tau]$ when studying $T_1 \neq T_2$.

Proposition 3 (Generalist Manager) *Consider a generalist manager ($T_1 = T_2 = T$) and assume (A1) and $t_i \in \{0, \tau, 2\tau\}$ with $t_1 + t_2 \leq 2\tau$.*

1. *If managerial attention is abundant (τ large), then balanced attention is optimal:*

$$(t_1^*, t_2^*) = (\tau, \tau) \tag{13}$$

2. *If managerial attention is scarce (τ small), then focused attention (managing with style) is optimal whenever the effective expertise, λT , is low enough. Specifically, there exists a $\Lambda > 0$ such that*

$$(t_1^*, t_2^*) \in \{(2\tau, 0), (0, 2\tau)\} \iff \lambda T < \Lambda. \tag{14}$$

3. *An increase in the importance of internal alignment (β) may result in a shift from balanced managerial attention to focused managerial attention (managing with style), but never the other way around.*

Proposition 3 shows that when attention is scarce (τ small) and the environment is sufficiently complex ($\lambda T < \Lambda$), a generalist manager optimally specializes in one task by devoting all her attention to one task (which task specifically being a matter of indifference). Thus, even if two tasks (say marketing and operations) are equally important to the organization and have, a priori, the same potential for profit improvement, a manager with equal expertise in both marketing and operations should focus all her attention on one task, say marketing. Ex post, such a manager mainly (but not always) selects and communicates about marketing strategies, and she *appears* to be arbitrarily and inefficiently biased towards marketing. Note that while the manager optimally focuses attention on one task, ex post she may be forced to communicate and implement a strategy about which she has ‘poor visibility’. The organization thus may be forced ex-post to do things based on very imprecise information, resulting in lower profitability and the appearance of a manager focused on the wrong things.

As discussed above, the benefits of managerial focus stem from the complementarity between (i) how much attention task i receives and (ii) how likely the shock affecting task i is perceived to be larger than the shock affecting task $-i$. When attention is not scarce (τ large) or or when the environment is not complex (λ large), however, the above complementarity is overwhelmed by decreasing marginal returns to attention. As Proposition 3 shows, balanced attention is optimal in those instances.

Intuitively, when τ is large, then by dividing her attention evenly, the manager learns both shocks with great precision. In contrast, by focusing all attention on task 1, the manager runs a high risk of being blindsided. Because of decreasing marginal returns to attention, the additional knowledge gained about task 1 is then much less than the information lost on task 2. Similarly, when the environment is not very complex so that $\lambda T > \Lambda$, the effective managerial expertise λT goes a long way in improving knowledge about the realized shocks θ_1 and θ_2 . The manager then typically identifies the largest shock correctly, and which task the manager communicates about to the organization is largely driven by the realization of θ_1 and θ_2 rather than by the allocation of managerial attention. The above complementarity between (i) how much attention a task receives and (ii) which task is communicated to the organization, is then very minor. Because of decreasing marginal returns to attention, it is then optimal to

allocate an equal amount of attention to both tasks as in our benchmark.

Beyond the complexity of the environment and scarcity of attention, Proposition 3 yields two other comparative static results: First, *ceteris paribus*, more competent managers (managers with a higher T) are less likely to focus attention on one task. This suggests a career path where managers, when they are younger, manage with style and focus attention on one area, whereas they act more as generalists when they are older and have accumulated more experience. Second, from Proposition 3.3, when the importance of internal alignment β is large, focused attention is more likely to be optimal. Indeed, when β is large, the manager is largely unresponsive to the shock affecting the task on which no communication occurs. Devoting attention to both tasks and learning both shocks is then mainly valuable to learn which shock is largest but effectively half of the information the manager collects is “wasted”. In contrast, when internal alignment is not very important (β small), the manager wants to be responsive to both shocks, even when she only communicates about one task. Learning both shocks is then much more valuable.

4.3 Allocation of attention by a specialist manager

Consider next a specialist manager, one that has more expertise about, say, task 1 than task 2, that is, $T_1 > T_2$. Formally, define

$$\rho(T_1, T_2) \equiv \frac{q(T_1)}{q(T_2)} = \frac{1 - e^{-\lambda T_1}}{1 - e^{-\lambda T_2}},$$

which we refer to as the *relative specialization ratio*. Thus ρ captures how much better the manager is at observing shock θ_1 than observing shock θ_2 . A ratio of $\rho = 1$ characterizes a generalist manager, a ratio $\rho = +\infty$ or $\rho = 0$ a fully specialized manager in task 1 and 2, respectively.

Marginal Allocations of Attention. Consider first the case where attention is scarce, that is τ is small and $t_i \in [0, 2\tau]$ with $t_1 + t_2 \leq 2\tau$. Should the specialist manager reinforce her expertise by devoting her scarce attention to the task she is already an expert on (‘managing with style’) or should she instead compensate for her lack of expertise in task 2?

The logic here is similar to the case of a generalist manager. When the manager has

more expertise in task 1, she is more likely to communicate about task 1 to the organization as she is more likely to perceive the shock affecting task 1 to be largest. Improving the precision of $\widehat{\theta}_1$ is then more valuable than improving the precision of $\widehat{\theta}_2$. Similarly, the higher is T_1 , the less likely it is that the manager will communicate about task 2 and, hence, the more likely it is that devoting scarce attention to task 2 is a waste of time. Formally, if $T_1 > T_2$, then from (11), $\Pr(\widehat{\theta}_1^2 > \widehat{\theta}_2^2) > \Pr(\widehat{\theta}_2^2 > \widehat{\theta}_1^2)$. From expression (10), profits are then more sensitive to $\widehat{\theta}_1^2$ than to $\widehat{\theta}_2^2$. It is then optimal to set $(t_1, t_2) = (2\tau, 0)$ to further improve the precision of $\widehat{\theta}_1$ provided $\Pr(\widehat{\theta}_1^2 > \widehat{\theta}_2^2) - \Pr(\widehat{\theta}_2^2 > \widehat{\theta}_1^2)$ is sufficiently large and τ is small.

In environments which are less complex (λ large), however, the realization of both shocks will be learned with great precision. In this case, the manager communicates with almost equal probability about both tasks even when T_1 is larger than T_2 . Indeed, from (11), when λ is large, $\Pr(\widehat{\theta}_1^2 > \widehat{\theta}_2^2) \approx \Pr(\widehat{\theta}_2^2 > \widehat{\theta}_1^2)$. Given that there are decreasing marginal returns to attention, the expected profit function (10) is then maximized by devoting attention to the shock in which the manager has less expertise.

Formally, fixing the relative specialization ratio $\rho > 1$ and defining

$$b \equiv \frac{\beta p}{1 + \beta(1 - p)} \in \mathbb{R}_+ \quad (15)$$

we show in Appendix that there exists a unique cut-off $\Lambda(\rho, b) > 0$ such that

$$\left. \frac{\partial}{\partial t} \Pi(q(T_1 + t), q(T_2 - t)) \right|_{t=0} > 0 \iff \lambda T_1 < \Lambda(\rho, b)$$

where $\Lambda(\rho, b)$ is a continuous function of ρ and b , implicitly defined in appendix. Hence, keeping the relative specialization ratio ρ , constant, a manager strictly prefers to ‘manage with style’ and devote her scarce attention to task 1 if and only if her effective expertise λT_1 is below a cutoff $\Lambda(\rho, b)$.

The cut-off $\Lambda(\rho, b)$ is uniquely determined by two parameters: The relative specialization ratio, ρ , and the parameter b which reflects the magnitude of the adaptation-coordination trade-off. As we show in Appendix, the cut-off $\Lambda(\rho, b)$ is strictly increasing in b . We can rewrite b as $(1 + \beta) |\alpha_1^* - \alpha_2^*|$ where α_i^* is the adaptiveness of the manager’s action to the local shock θ_i , as given by (9). Hence, an increase in the importance of internal alignment β

or an increase in the ability of the manager to communicate his strategy p (which increases $|\alpha_1^* - \alpha_2^*|$) increases $\Lambda(\rho, b)$ and makes managing with style more likely to be optimal:

Proposition 4 (Specialist Manager) *Assume (A1), and consider a specialist manager with $T_1 > T_2$. There exists a cut-off $\Lambda(\rho, b)$ such that if τ is sufficiently small, then*

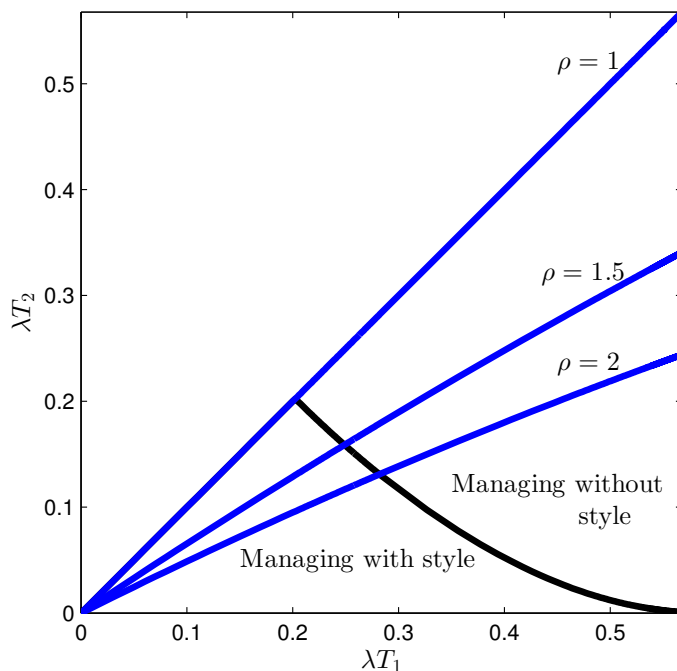
- *If $\lambda T_1 < \Lambda(\rho, b)$, managing with style is optimal: $(t_1^*, t_2^*) = (2\tau, 0)$*
- *If $\lambda T_1 > \Lambda(\rho, b)$, managing with style is suboptimal: $(t_1^*, t_2^*) = (0, 2\tau)$.*
- *The cut-off $\Lambda(\rho, b)$ is increasing in the need for internal alignment, β , and the manager's ability to communicate strategic choices, p .*

Figure 4 illustrates Proposition 4. The upward sloping (blue) lines indicate effective expertise combinations $(\lambda T_1, \lambda T_2)$ for which the relative specialization ratio ρ is constant (the plot is obviously symmetric around the 45 degree line). For each value of ρ there is a unique threshold $\Lambda(\rho, b)$ such that if $\lambda T_1 < \Lambda(\rho, b)$, then on the margin, the manager prefers to ‘manage with style’ and devote attention to task 1 rather than task 2. The downward sloping (black) line links all those points to plot the function $\Lambda(\rho, b)$ in the $(\lambda T_1, \lambda T_2)$ space. Hence, for all points $(\lambda T_1, \lambda T_2)$ below the downward sloping (black) line, the optimal allocation of the marginal unit of attention is $(t_1^*, t_2^*) = (2\tau, 0)$ whereas it is $(t_1^*, t_2^*) = (0, 2\tau)$ above it.¹¹

Proposition 4 emphasizes the interaction between managerial expertise (T_1, T_2) , the complexity of the environment $1/\lambda$ in which the manager operates, and the optimal allocation of scarce attention (t_1^*, t_2^*) . From Figure 4, for any expertise combination (T_1, T_2) with $T_1 > T_2$, we have that $(t_1^*, t_2^*) = (2\tau, 0)$ if λ is sufficiently small (complex environments), whereas $(t_1^*, t_2^*) = (0, 2\tau)$ if λ is large (less complex environments). Note that this is also the result shown in Figure 3, Panel A, which provided a preview of our results.

¹¹The numerical values used to generate this plot are the same as the ones used in the example introduced in Section 2.6, $\beta = 5$ and $p = .75$ such that $b = 5/3$. We have plotted the function $\Lambda(\rho, b)$ for $b = \beta p / (1 + \beta(1 - p))$ ranging from $b \approx 0$ to b large ($b = 100$). Notice that b is only a function of the primitives of the model. As we discuss in Appendix 6.5, for all those parameter values, we obtain similar shapes for $\Lambda(\rho, b)$ as for $b = 5/3$. In particular, the black curve representing $\Lambda(\rho, b)$ in the $(\lambda T_1, \lambda T_2)$ space is decreasing, and any linear function $\lambda T_2 = Z - \lambda T_1$, with $Z \in \mathbb{R}^+$, crosses $\Lambda(\rho, b)$ at most once, and always from above.

Figure 4: *Optimal allocation of the marginal unit of attention given effective expertise $(\lambda T_1, \lambda T_2)$ with $T_1 > T_2$. The downward sloping line is the function $\Lambda(\rho, b)$ plotted in the space $(\lambda T_1, \lambda T_2)$ (see Proposition 4). Below the downward sloping curve the manager allocates the marginal unit of attention to task 1, whereas above she allocates the marginal unit of attention to task 2. The upward sloping (blue) lines show iso- ρ , where ρ is the relative specialization ratio for $\rho \in \{1, 1.5, 2\}$.*



Intermediate Allocations of Attention. Proposition 4 also has implications for when 2τ is “intermediate” or “large”. Consider Figure 5, Panels A-C, which plots again the function $\Lambda(\rho, b)$ in the $(\lambda T_1, \lambda T_2)$ space.¹² Consider now any effective expertise configuration $(\lambda T_1, \lambda T_2)$ below $\Lambda(\rho, b)$ for which $T_1 > T_2$, such as the one represented by point $C = (\lambda T_1^C, \lambda T_2^C)$. Note that also point A in Figure 3 (preview of results) is below $\Lambda(\rho, b)$.

As long as 2τ is not too large, a manager then wants to devote all her attention to task 1, on which she is already an expert. Indeed, assume $2\tau < 2\tau_y$ as shown in Figure 5, Panel B, so

¹²Figure 5, Panels A-C, plots $\Lambda(\rho, b)$ for $b = 5/3$. Appendix 6.5 provides a more formal analysis and discussion and shows that our insights carry through for other values of b as well.

that $(\lambda T_1^C, \lambda(T_2^C + 2\tau))$ is below $\Lambda(\rho, b)$ and 45 degree line. A corollary of Proposition 4 is that for any attention allocation (t_1, t_2) with $t_2 > 0$, profits can then be strictly improved by shifting attention from task 2 to task 1. It follows that for $2\tau < 2\tau_y$, we have that $(t_1^*, t_2^*) = (2\tau, 0)$ as in Panel B.¹³

When 2τ is sufficiently large, however, it will eventually become optimal to devote all or most attention to task 2 in which the manager is not an expert. Indeed, assume that $2\tau > 2\tau_x$ as shown in Figure 5, Panel A, that is $(\lambda(T_1^C + 2\tau), \lambda T_2^C)$ is above $\Lambda(\rho, b)$. A corollary of Proposition 4 is that for any attention allocation (t_1, t_2) with $T_2^C + t_2 < T_1^C + t_1$, profits can then be marginally increased by shifting attention from task 1 to task 2. It follows that for $2\tau > 2\tau_x$ we have that either $(t_1^*, t_2^*) = (0, 2\tau)$ as in Panel C or (t_1^*, t_2^*) is such that $T_1^C + t_1^* = T_2^C + t_2^*$.

Finally, in less complex environments where $(\lambda T_1, \lambda T_2)$ is above $\Lambda(\rho, b)$ in Figure 4 (but $T_1 > T_2$), we always have that $(t_1^*, t_2^*) = (0, 2\tau)$ or $T_1^C + t_1^* = T_2^C + t_2^*$. This corresponds, for example, to the effective expertise configuration B in Figure 3 (preview of results). The logic is again that on the margin, profits can then always be improved by shifting attention away from task 1 to task 2 up to the point where $T_1^C + t_1^* = T_2^C + t_2^*$.

4.4 Endogenous managerial expertise: Specialists versus Generalists

So far, we have taken managerial expertise as given and endogenized the allocation of managerial attention. Managers though are typically appointed by boards (or, interchangeably in this paper, firm owners) who select them depending on their expertise. We are interested in the board's decision to appoint a generalist or a specialist manager, even when firm owners are indifferent about the organization's strategic direction.

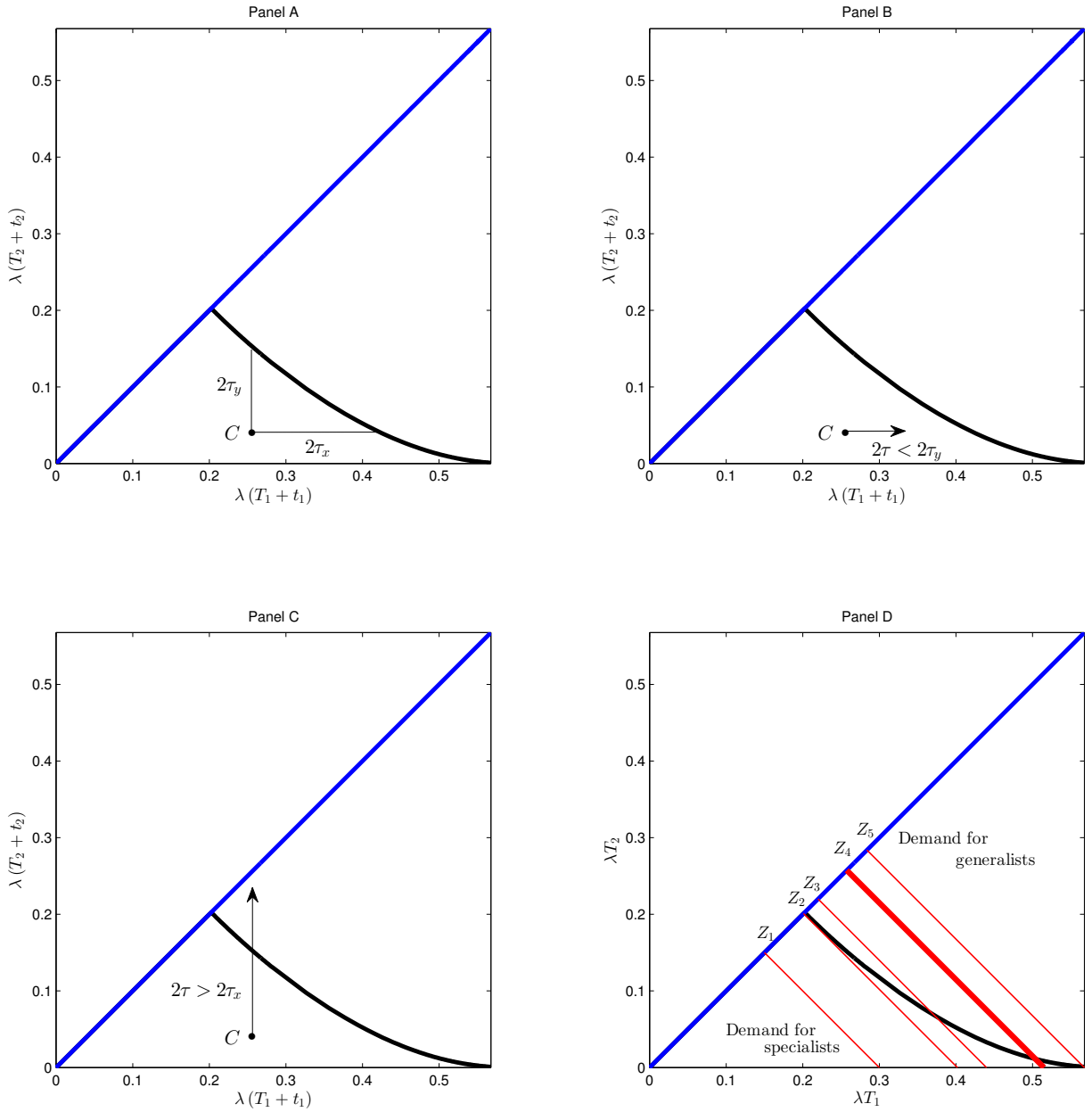
Assume therefore that a board can choose any manager whose expertise (T_1, T_2) belongs to some 'opportunity set' Γ . To fix ideas we will assume that

$$\Gamma = \Gamma(Z) \equiv \{(T_1, T_2) : T_1 + T_2 \leq Z \text{ and } T_1 \geq T_L, T_2 \geq T_L\}$$

where Z is the total 'expertise budget' and $T_L \geq 0$ is the minimum expertise of any manager.

¹³Note that $2\tau < 2\tau_x$ is not a sufficient condition for $t_1^* = 2\tau$. Indeed, if $2t_y < 2\tau < 2\tau_x$ with τ_x and τ_y defined as in Figure 5, panel A, then we cannot rule out based on Proposition 4 that $t_1^* = 0$.

Figure 5: Panel A: Attention cutoff levels, $2\tau_x$ and $2\tau_y$, given effective expertise C . Panels B-C: Endogenous allocation of attention 2τ . Panel D: Optimal choice of managerial expertise (T_1, T_2) given $T_1 + T_2 \leq Z_i$. Specialists optimal for $Z_i \leq Z_4$ and generalists optimal for $Z_i > Z_4$.



Notice that since there are no agency problems between firm owners and the manager, they agree as to the optimal allocation of managerial attention t_1 and t_2 , with $t_1 + t_2 \leq 2\tau$. We can thus think of firm owners choosing $\tilde{T}_1 = T_1 + t_1$ and $\tilde{T}_2 = T_2 + t_2$ with

$$(\tilde{T}_1, \tilde{T}_2) \in \Gamma(Z + 2\tau)$$

where \tilde{T}_i is the final expertise in task i after the optimal allocation of managerial attention t_i : An attention budget $2\tau > 0$ paired with an expertise budget Z is formally equivalent to an attention budget $2\tau = 0$ paired with an expertise budget $Z + 2\tau$. Without any loss of generality we therefore simplify the problem by setting $\tau = 0$ so that $T_i = \tilde{T}_i$. Moreover, we further posit, again without any loss in generality, that $\lambda = 1$. An increase in λ is formally equivalent with a proportional increase in the expertise budget Z . Comparative statics with respect to λ are therefore equivalent to comparative statics with respect to Z . We will thus associate a more complex environment with a smaller expertise budget for the board.

Figure 5, Panel D, is useful in illustrating board choices regarding a specialist or generalist manager.¹⁴ There in addition to the iso- ρ lines we have drawn five straight (red) iso- Z lines, where the total expertise budget $Z = T_1 + T_2$ is kept constant. Recall that the downward sloping curve (in black) represents again all expertise combinations (T_1, T_2) for which

$$\left. \frac{d\Pi(q(T_1 + t), q(T_2 - t))}{dt} \right|_{t=0} = 0$$

with this derivative being positive for values of (T_1, T_2) below this line and negative above it.

It follows from Proposition 4 that keeping the total expertise budget $Z = T_1 + T_2$ fixed, profits are decreasing in T_1 above the downward sloping curve, but increasing in T_1 below the curve. Moreover, since the curve in Figure 5 has a derivative which is larger than -1 , it follows that moderate specialists for which $0 < T_2 < T_1$ are always dominated by either extreme specialists (for which $T_2 = 0$) or by complete generalists (for which $T_1 = T_2$). Obviously, extreme specialists are preferred for the budgets Z_1 and Z_2 corresponding to the first two straight (red) lines, whereas complete generalists are preferred for the budget Z_5 corresponding to the last budget line. More generally, one can show that extreme specialists

¹⁴Figure 5, Panel D, assumes $b = 5/3$. Appendix 6.5 provides a more formal discussion and analysis and shows that our insights carry through for other values of b as well.

are preferred over complete generalists if and only if $Z \leq Z_4 \approx 0.515$, where Z_4 is the budget corresponding to the thick red line.

Sometimes, however, extreme specialists are not available to the board as any manager has some minimal expertise T_L on either task. For example, T_L can be considered as basic knowledge any manager has or, equivalently, the minimum attention a manager must devote to either task. For expertise budgets $Z < Z_2 \approx 0.402$ or budgets $Z > Z_4$ the minimal task knowledge T_L does not affect the choice between generalist and specialists. But for $Z \in (Z_2, \bar{Z})$, the board may prefer a complete generalist $(T_1, T_2) = (Z/2, Z/2)$ over a moderate specialist $(T_1, T_2) = (Z - T_L, T_L)$ when T_L is sufficiently large. Moreover, the larger is the basic knowledge T_L any manager has about both tasks, the more likely a generalist manager is optimal.¹⁵

In sum, our analysis shows that generalist managers are preferred in less complex environments – that is in environments with larger expertise budgets where even a generalist will have a precise estimate of both task-specific shocks. In contrast, when expertise is in short supply, that is for smaller expertise budgets or more complex environments, boards prefer hiring specialist managers.

An interesting extension of our model would be to endogenize the expertise budgets of firms (and, potentially, managerial wages). If, as seems intuitive, larger firms have larger expertise budgets in equilibrium, a prediction of our model would be that, all else equal, generalist managers are more likely to be hired by larger, more established firms, whereas specialist managers are more likely to be hired by smaller firms.

5 Discussion and conclusions

The goal of this paper was to shed light on when we should expect firm behavior and strategic choices to reflect managerial characteristics as opposed to the realization of environmental

¹⁵Indeed, consider the red buget line which represents an expertise budget $Z_3 = 0.44$ and which crosses the (black) downward sloping curve at $T_2 = 0.065$. Since $Z_3 < Z_4$, the board strictly prefers an extreme specialist $(Z_3, 0)$ over a complete generalist $(Z_3/2, Z_3/2)$. But whenever $T_L \geq 0.065$, profits are even minimized by choosing a manager with expertise $(Z_2 - T_L, T_L)$.

shocks. In our model, the allocation of managerial attention partially determines firm behavior, but can either exacerbate manager fixed effects or mitigate them. We now discuss our findings and their implications in more detail.

5.1 Magnitude of manager fixed effects

Performance differences between seemingly similar managers. A first implication of our model is that *small* initial differences in managerial expertise may result in dramatically different firm behavior, as managers devote their scarce attention in a way which amplifies initial differences in expertise. It follows that assessing the impact of managerial expertise on firm strategies and outcomes requires precise measures of that expertise; otherwise *small* errors in measurement can produce the standard attenuation bias and thus lead the researcher to underestimate the effect of expertise on firm strategies.

Environmental uncertainty and manager fixed effect. Ours is a theory of managerial style where differences in firm behavior are driven by differences in information processing (or cognition) between managers. An implication from cognition-based theories of managerial style is that the magnitude of manager fixed effects should depend on environmental uncertainty and complexity. As noted above, in settings with large uncertainty, the endogenous allocation of managerial attention exacerbates initial differences in task expertise. But in environments with less uncertainty (corresponding to high values of T or λ in our model), our model predicts that managers with different expertise will make similar or even identical choices as they allocate attention in a manner which reduces or eliminates differences in task expertise. The predictions of our cognition-based model can therefore be empirically distinguished from alternative theories where manager fixed effects reflect differences in the capability of managers to execute certain strategies or differences in managerial preferences. Indeed, under these alternative theories, we still expect to see strong manager fixed effects in environments with limited uncertainty, whereas this is not the case in our framework.

In the same vein, our results indicate that studies that focus on tenure and educational background (a potential proxy for T) to understand managerial strategies need to interact those

managerial characteristics with industry measures that capture the complexity of the environment in which the manager operates to explain the cross section of managerial strategies.

5.2 Managers and performance differences between firms

Persistent performance differences between seemingly similar firms. While a large literature has established the impact of managers on firm behavior, another literature has been interested in persistent performance differences between seemingly similar firms (see, for example, Gibbons and Henderson, 2013, and Li, Matouschek and Powell, 2015). In many settings, differences in behavior and performance between seemingly similar firm are arguably related to differences in managerial style. But this creates another, related, question: Why do seemingly similar firms hire managers with different managerial characteristics to begin with? Our model implies that boards – when uncertainty is large – optimally hire managers with specialized expertise, but do not necessarily care about the particular area or function the manager has expertise in (say marketing versus operations). In other words, while the area of expertise is an important predictor of firm behavior and (ex post) firm performance, it is not necessarily an important criterion of choice for a board of directors. Instead, boards are likely to choose managers based on leadership ability, general cognitive ability, availability and other factors. Different boards are therefore likely to hire very different managers (in terms of functional expertise), even when faced with an identical economic environment and even when boards are themselves very similar. Given that the average tenure of a CEO of a S&P 500 firm was 9.7 years in 2013, such ‘random’ choices may have long-lasting effects on firm behavior and performance.¹⁶

Environmental uncertainty and performance differences between firms. To the extent that differences in how managers allocate attention and process information are at the source of manager fixed effects, as posited by the present paper, we expect to see a conformity of firm

¹⁶A related result is also present in Selove (2013), who studies a dynamic investment game by firms that are initially identical. Because of increasing returns to investment, there may exist a unique equilibrium in which firms that are only slightly different focus all their investment in different market segments, causing small random differences to expand into large permanent differences.

strategies and firm behavior in low uncertainty environments, but a large dispersion of firm behavior (correlated with managerial backgrounds) in environments characterized by high uncertainty. In contrast, if differences in managerial capabilities are at the source of differences in firm behavior, one should expect to see a similar dispersion in firm behavior and firm strategies in high and low uncertainty environments.

5.3 Why do managers matter? Practices versus strategic choices

We view our paper as shedding light on the channels through which managers matter for firm performance. Following a seminal paper by Bloom and Van Reenen (2007), much of the economics' literature on this topic has been focused on 'management practices.' The premise of much of this literature is that many companies are not run efficiently, and better management can improve operational effectiveness (see, for example, Syverson (2004), Foster, Haltiwanger, and Syverson (2008), Bloom and Van Reenen (2010), Bloom et al. (2013)). Instead, much of the management literature on why managers matter has focused on the impact of managers on strategic choices as opposed to operational efficiency.¹⁷ As argued by Finkelstein et al. (2009):

'But where does the company's strategy come from? (...) To be sure, strategic actions are sometimes due to imitation, inertia, and careful, objective decision making. But a wealth of research and everyday observation indicates that strategy and other major organizational choices are made by humans who act on the basis of idiosyncratic experiences, motives and dispositions. If we want to understand strategy, we must understand strategists.'

Our view is that to assess the impact of managers on firms, empirical work on managers should move beyond management practices and focus on specific strategic choices in particular industries. For example, Kaplan, Murray and Henderson (2003) analyze the responses of 15 large, incumbent pharmaceutical firms responses to the emergence of biotechnology. Cho

¹⁷See Roberts and Saloner (2013) for a discussion of how the management literature has defined 'business strategy' and 'strategic choices'.

and Hambrick (2006) study strategic responses to airline deregulation. Finally, Kaplan (2008) studies how CEO's of 71 communications firms responded to the fiber-optic revolution in the communications technology industry. Another promising approach is to measure managerial attention to functional areas directly by employing time use surveys, as in Bandiera, Guiso, Prat and Sadun (2011), and correlate this with specific strategic choices.

While better management practices almost always improve performance, the difference between what is optimal ex ante and ex post is key when analyzing strategic choices made by managers. For example, if a firm is late to the fiber-optic revolution in telecommunications, then this is not necessarily evidence of bad management. Indeed, as our paper shows, it is optimal for managers from an ex ante perspective to focus all attention on one area. If unlucky, the manager may then be blindsided and ex post be forced to compete along dimensions to which she optimally did not devote much attention. To put it differently, management evaluation should be done along two dimensions, strategic choices and practices. Both obviously affect performance. For strategic choices the difference between ex ante and ex post is very relevant, while this is less so for the case of management practices.¹⁸ In addition, cross sectional dispersion in strategic choices within in an industry should not be taken to be a sign of inefficiency whereas diversity in practices is more likely to indicate that some firms could improve performance through the adoption of better practices. Similarly management may stick to adopted strategies particularly in the presence of temporary spikes in uncertainty (because of industry wide regulatory overhauls, for example) and volatility (because of innovations waves for instance), even at the risk of being blindsided and being forced to act on imprecise information. The empirical challenge is to distinguish between strategies and practices when assessing management.

5.4 Cognitive constraints as a source of manager fixed effects

In our model, managerial fixed effects arise because of constraints on managerial cognition (and the need for organizational alignment). Managers who face the same economic environ-

¹⁸There are of course practices, say, sound accounting, that are strategy neutral whereas other practices are naturally strategy dependent.

ment and the same facts may come to different strategic choices as they devote their scarce attention to different sources of information. Cognitive limitations are also seen as the primary source of manager fixed effects in the management literature,¹⁹ but in contrast to our model, emphasis is put on behavioral biases in decision-making as a central argument as to why managers matter. In particular, the management literature follows the logic of the Carnegie School (March and Simon 1958; Cyert and March 1963) according to which complex choices are largely determined by behavioral factors, rather than by calculations of optimal actions. According to the dominant stream in this literature:

“in arriving at their own rendition of a strategic situation, or “construed reality” (Sutton 1987), executives distill and interpret the stimuli that surround them. This occurs through a three stage filtering process. Specifically, executive orientations affect their *field of vision* (the directions in which they look and listen), *selective perception* (what they actually see and hear), and *interpretation* (how they attach meaning to what they see and hear).”²⁰

A contribution of our paper has been to show how manager fixed effects may arise even when managers optimally (and rationally) allocate attention and process information. From a normative point of view, our results thus show that managerial biases in information-processing are not necessarily pathological. Indeed, a key insight of our model is that, given the presence of cognitive limits to attend to all possible information, boards or firm owners often prefer managers whose field of vision is narrow.

From a conceptual point of view, we show that managers may matter even when they are rational optimizing creatures. Rather than exogenously posit that the ‘field of vision’ of a manager is determined by her expertise and past experiences, managers in our model optimally choose their ‘field of vision’, but behave largely as predicted (and, indeed, observed) by the management literature. We further link a manager’s ‘field of vision’ to organizational factors, such as the need for organizational alignment around simple strategies and environmental factors, such as the amount of uncertainty and the scarcity of attention.

¹⁹See Finkelstein et al. (2009), Chapter 2, for an overview.

²⁰Finkelstein et al. (2009), p.46.

References

- Akerlof, Robert and Richard Holden (2015). "Movers and Shakers," working paper, University of Warwick.
- 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 (2015). "Organizing to Adapt and Compete." *American Economic Journal: Microeconomics*, Vol. 7, 158-187
- Bandiera, Oriana, Luigi Guiso, Andrea Prat, and Raffaella Sadun (2011). "What do CEOs do?." Mimeo.
- Bandiera, Oriana, Andrea Prat, and Raffaella Sadun (2013). "Managing the family firm: evidence from CEOs at work." Mimeo, National Bureau of Economic Research.
- Barker, Vincent, and George Mueller (2002). "CEO characteristics and firm R&D spending." *Management Science* 48, no. 6 : 782-801.
- Bertrand, Marianne and Antoinette Schoar (2003). Managing with style: The effect of managers on firm policies. *The Quarterly Journal of Economics*, 1169-1208.
- Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts (2013). "Does Management Matter? Evidence from India." *Quarterly Journal of Economics* 128, no. 1
- Bloom, Nicholas, and John Van Reenen (2007). "Measuring and Explaining Management Practices Across Firms and Countries." *Quarterly journal of economics* 122 (4): 1351-1408.
- Bloom, Nicholas, and John Van Reenen. "Why do management practices differ across firms and countries?" *The Journal of Economic Perspectives* (2010): 203-224.
- Bolton, Patrick, Markus Brunnermeier and Laura Veldkamp (2013). "Leadership, Coordination and Corporate Culture." *Review of Economic Studies*, 80(2), 512-537.

- Bolton, Patrick and Mathias Dewatripont (1994). "The Firm as a Communication Network." *Quarterly Journal of Economics*, 109, 809-839.
- Cho, Theresa S., and Donald C. Hambrick (2006). "Attention as the mediator between top management team characteristics and strategic change: The case of airline deregulation." *Organization Science* 17(4): 453-469.
- Cyert, Richard, and James March (1963). *A Behavioral Theory of the Firm*. Oxford: Blackwell.
- Calvo-Armengol, Antonio, Joan de Marti and Andrea Prat (2015). "Communication and Influence." *Theoretical Economics*, 10, 649-690.
- Dessein, Wouter, Andrea Galeotti, and Tano Santos (2015). "Rational inattention and organizational focus," *American Economic Review*, forthcoming.
- Dessein, Wouter and Tano Santos (2006). "Adaptive Organizations." *Journal of Political Economy*, 114 (50), 956-995.
- Finkelstein, Sydney, Donald C. Hambrick, and Albert A. Cannella (2009). *Strategic leadership: Theory and research on executives, top management teams, and boards*. Oxford University Press.
- Ferreira, Daniel, and Marcelo Rezende. "Corporate strategy and information disclosure." *The RAND Journal of Economics* 38.1 (2007): 164-184.
- Foster, Lucia, John Haltiwanger, and Chad Syverson (2008). "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profit-ability." *American Economic Review*, 98(1): 394-425.
- Galeotti, Andrea and Sanjeev Goyal (2010). "The Law of the Few," *American Economic Review*, 100(4): 1468-92
- Garicano, Luis, and Andrea Prat (2013). "Organizational economics with cognitive costs." *Advances in Economics and Econometrics: Tenth World Congress*. Vol. 1.

- 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 Economics*, 5(3), 205–225.
- Gibbons, Robert, and Rebecca Henderson (2013). "What Do Managers Do? Exploring Persistent Performance Differences among Seemingly Similar Enterprises." *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press, 680-731.
- Hambrick, Donald C. (2015). "Upper Echelons Theory," Palgrave Macmillan. Available from:
<http://www.palgraveconnect.com/esm/doi/10.1057/9781137294678.0713>
- Hambrick, Donald C., and Phyllis A. Mason (1984). "Upper echelons: The organization as a reflection of its top managers." *Academy of Management Review* 9, no. 2 : 193-206.
- Hellwig, Christian and Laura Veldkamp (2009) "Knowing What Others Know: Coordination Motives in Information Acquisition." *Review of Economic Studies*, 76(1), 223-251.
- Herskovic, Bernard and Joao Ramos (2015), "Acquiring Information Through Peers," working paper, New York University.
- Kaplan, Sarah (2008) "Cognition, capabilities, and incentives: assessing firm response to the fiber-optic revolution." *Academy of Management Journal* 51(4): 672-695.
- Kaplan, Sarah, Fiona Murray, and Rebecca Henderson (2003). "Discontinuities and senior management: Assessing the role of recognition in pharmaceutical firm response to biotechnology." *Industrial and Corporate Change* 12(2): 203-233.
- Li, Jin, Niko Matouschek and Michael Powell (2015) "Power Dynamics in Organizations," working paper, Kellogg School of Management, Northwestern University.
- March, James, and Herbert Simon (1958). *Organizations*, John Wiley and Sons, Inc., New York.

- Ocasio, William (1997). "Towards an attention-based view of the firm." *Strategic Management Journal*, 18:187–206.
- Rantakari, Heikki (2008). "Governing Adaptation" *Review of Economic Studies*, 75, 1257-1285.
- Rantakari, Heikki (2013a). "Organizational Design and Environmental Volatility." *Journal of Law, Economics, and Organization*, 29(3), 569-607.
- Rantakari, Heikki (2013b). "Conflict Resolution and Organizational Influence." Mimeo, University of Southern California..
- 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.
- Selove, Matthew (2013). "How do firms become different? A dynamic model." *Management Science*, 60(4), 980-989.
- Syverson, Chad (2004). "Product Substitutability and Productivity Dispersion." *Review of Economics and Statistics*, 86(2): 534-50
- Van den Steen, Eric (2013a). "A Formal Theory of Strategy." Mimeo, Harvard Business School.
- Van den Steen, Eric (2013b). "Strategy and the Strategist: How it Matters Who Develops the Strategy." Mimeo, Harvard Business School.

6 Appendix

6.1 Microfoundations of learning technology

A simple microfoundation of (5) can be obtained as follows. Assume the task-specific shock θ_i is the sum of n independently distributed shocks θ_{ik} :

$$\theta_i = \sum_{k=1}^n \theta_{ik} \text{ with } \theta_{ik} \sim N(0, \sigma_\theta^2/n),$$

Each element θ_{ik} can be interpreted as an ‘‘component’’ of task i to be understood by the manager to have a complete picture of task i . The manager observes two independent signals, s_{ik} and S_{ik} , about each component θ_{ik} and signals are independent across components. Both signals have the same structure: They are either fully informative about θ_{ik} or pure noise. Signal s_{ik} is endogenous in that its precision is a function of the attention t_i that the agent devotes to task i . Specifically the manager learns θ_{ik} with probability $q(t_i)$. We assume that learning follows a Poisson process with hazard rate λ :

$$q(t_i) = 1 - e^{-\lambda t_i}.$$

S_{ik} is instead an exogenous signal. Its precision is a function of the manager’s expertise T_i , which is exogenous. As in the case of the endogenous signal the manager thus learns θ_{ik} with probability $q(T_i)$. Exogenous learning is also assumed to follow a Poisson process with hazard rate λ . The manager thus learns any given component θ_{ik} with probability:²¹

$$q_i \equiv q(t_i + T_i) = 1 - e^{-\lambda(t_i + T_i)}. \quad (16)$$

Notice thus that attention t_i and expertise T_i are substitutes in the learning process.

Denoting $s_i = [s_{i1}, \dots, s_{in}]$ and $S_i = [S_{i1}, \dots, S_{in}]$, then

$$\widehat{\theta}_i \equiv E(\theta_i | s_i, S_i) = \sum_{k=1}^n E(\theta_{ik} | s_{ik}, S_{ik})$$

In the limit as the number of components n goes to infinity, we have that

$$\text{RV}(\theta_i) = E\left(\theta_i - \widehat{\theta}_i\right)^2 = (1 - q_i)\sigma_\theta^2, \quad (17)$$

as posited in (5). Moreover, the attention constraint (6) can then be rewritten as

$$(1 - q_1)(1 - q_2) \geq e^{-\lambda(2\tau + T_1 + T_2)}.$$

²¹Notice that the probability that the manager learns θ_{ik} is given by

$$(1 - q(T_i))q(t_i) + (1 - q(t_i))q(T_i) + q(T_i)q(t_i) = q(t_i + T_i).$$

We interpret $1/\lambda$ as reflecting the uncertainty or the complexity of the environment. The larger is $1/\lambda$, the more attention and expertise are required to reduce the residual variance $RV(\theta_i)$.

6.2 Proof of Proposition 2

Without loss of generality, assume that task 1 is affected by the largest perceived shock, that is $\hat{\theta}_1^2 > \hat{\theta}_2^2$. Substituting (8) into (7) and manipulating terms, we obtain

$$E(\pi|\hat{\theta}) = \sum_{i \in \{1,2\}} \frac{\hat{\theta}_i^2}{1 + \beta(1 - p_i)} \quad (18)$$

Denoting $\hat{\theta}_2^2 = k\hat{\theta}_1^2$ with $k < 1$, we can rewrite this as

$$E(\pi|\hat{\theta}) = \left(\frac{1}{1 + \beta e^{-\mu r_1}} + k \frac{1}{1 + \beta e^{-\mu(r-r_1)}} \right) \hat{\theta}_1^2 \quad (19)$$

where $e^{-\mu r_i} = 1 - p_i$ and $r_1 + r_2 = r$. Since $k < 1$, it is easy to verify that is never optimal to set $r_1 < r/2$. Hence, let $r_1 \in [r/2, r]$. Taking the derivative of (19) with respect to r_1 we obtain

$$\frac{\partial E(\pi|\hat{\theta})}{\partial r_1} = \beta\mu \left[\frac{e^{-\mu r_1}}{(1 + \beta e^{-\mu r_1})^2} - k \frac{e^{-\mu(r-r_1)}}{(1 + \beta e^{-\mu(r-r_1)})^2} \right] \hat{\theta}_1^2. \quad (20)$$

Using plain algebra, if $k = 1$, it follows that

$$\begin{aligned} \frac{\partial E(\pi|\hat{\theta})}{\partial r_1} &> 0 \iff 1 - \beta^2 e^{-\mu r} < 0 \\ &\iff p < \bar{p}(\beta) \equiv 1 - 1/\beta^2 \end{aligned}$$

where, recall, $p = 1 - e^{-\mu r}$. Obviously, if $k \leq 1$, then $p < \bar{p}(\beta)$ is a sufficient condition for $\partial E(\pi|\hat{\theta})/\partial r_1 > 0$. Hence if $\hat{\theta}_1^2 > \hat{\theta}_2^2$, then $p < \bar{p}(\beta) = 1 - 1/\beta^2$ implies $(r_1^*, r_2^*) = (r, 0)$ and $(p_1^*, p_2^*) = (p, 0)$. Note further that $\bar{p}(\beta)$ is increasing in β .

6.3 Proof of Proposition 3

6.3.1 Preliminaries

Expected Profits conditional on q_1 and q_2 : Expected profits are given by

$$\begin{aligned} &\Pi(q_1, q_2) \\ &= 4 \int_0^{+\infty} \left[\int_0^{\hat{\theta}_2} \frac{\hat{\theta}_1^2}{1 + \beta} dF(\hat{\theta}_1, q_1 \sigma_\theta^2) + \int_{\hat{\theta}_2}^{+\infty} \frac{\hat{\theta}_1^2}{1 + \beta(1 - p)} dF(\hat{\theta}_1, q_1 \sigma_\theta^2) \right] dF(\hat{\theta}_2, q_2 \sigma_\theta^2) \\ &+ 4 \int_0^{+\infty} \left[\int_0^{\hat{\theta}_1} \frac{\hat{\theta}_2^2}{1 + \beta} dF(\hat{\theta}_2, q_2 \sigma_\theta^2) + \int_{\hat{\theta}_1}^{+\infty} \frac{\hat{\theta}_2^2}{1 + \beta(1 - p)} dF(\hat{\theta}_2, q_2 \sigma_\theta^2) \right] dF(\hat{\theta}_1, q_1 \sigma_\theta^2) \end{aligned}$$

We can make a simple change of variable $\varphi_1 \equiv \hat{\theta}_1/\sqrt{q_1}$ and $\varphi_2 \equiv \hat{\theta}_1/\sqrt{q_2}$, so that both φ_1 and φ_2 are normally distributed with variance σ_θ^2 . With some abuse of notation let $F(x) \equiv F(x, \sigma_\theta^2)$, then the expected profits can be rewritten as

$$\begin{aligned} \Pi(q_1, q_2) = & 4 \int_0^{+\infty} \left[\int_0^{\sqrt{\frac{q_2}{q_1}} \varphi_2} \frac{q_1 \varphi_1^2}{1 + \beta} dF(\varphi_1) + \int_{\sqrt{\frac{q_2}{q_1}} \varphi_2}^{+\infty} \frac{q_1 \varphi_1^2}{1 + (1-p)\beta} dF(\varphi_1) \right] dF(\varphi_2) \\ & + 4 \int_0^{+\infty} \left[\int_0^{\sqrt{\frac{q_1}{q_2}} \varphi_1} \frac{q_2 \varphi_2^2}{1 + \beta} dF(\varphi_2) + \int_{\sqrt{\frac{q_1}{q_2}} \varphi_1}^{+\infty} \frac{q_2 \varphi_2^2}{1 + (1-p)\beta} dF(\varphi_2) \right] dF(\varphi_1) \end{aligned}$$

or still

$$\Pi(q_1, q_2) = \frac{4}{1 + (1-p)\beta} \int_0^{+\infty} \left[\int_{\sqrt{\frac{q_2}{q_1}} \varphi_k}^{+\infty} q_1 \varphi_1^2 dF(\varphi_1) + \int_{\sqrt{\frac{q_1}{q_2}} \varphi_k}^{+\infty} q_2 \varphi_2^2 dF(\varphi_2) \right] dF(\varphi_k) \quad (21)$$

$$+ \frac{4}{1 + \beta} \int_0^{+\infty} \left[\int_0^{\sqrt{\frac{q_2}{q_1}} \varphi_k} q_1 \varphi_1^2 dF(\varphi_1) + \int_0^{\sqrt{\frac{q_1}{q_2}} \varphi_k} q_2 \varphi_2^2 dF(\varphi_2) \right] dF(\varphi_k), \quad (22)$$

where φ_k stands for a normally distributed random variable with mean 0 and variance σ_θ^2 .

Profits under Balanced Attention. If the generalist manager opts to balance attention evenly among tasks:

$$q_1 = q_2 \equiv q = 1 - e^{-\lambda(\tau+T)} \quad (23)$$

and the profit expression (see (21) and (22)) simplifies to

$$\Pi(q, q) = 8q \int_0^\infty \left(\frac{1}{1 + \beta} \int_0^{\varphi_k} \varphi_i^2 dF(\varphi_i) + \frac{1}{1 + \beta(1-p)} \int_{\varphi_k}^\infty \varphi_i^2 dF(\varphi_i) \right) dF(\varphi_k), \quad (24)$$

where both φ_i and φ_k are both normally distributed random variables with mean 0 and variance σ_θ^2 . (24) has the following closed form solution:

$$\Pi(q, q) = 2q \left(\frac{1}{1 + \beta} \frac{\pi - 2}{2\pi} + \frac{1}{1 + \beta(1-p)} \frac{\pi + 2}{2\pi} \right) \sigma_\theta^2,$$

which simplifies in turn to

$$\Pi(q, q) = 2qC \quad \text{with} \quad C \equiv \left(\frac{1}{1 + \beta} \right) \left[1 + \frac{\beta p}{1 + \beta(1-p)} \frac{\pi + 2}{2\pi} \right] \sigma_\theta^2. \quad (25)$$

Profits under Focused Attention. If the manager focuses all her attention on one task, say, task 1,

$$q_1 = q(2\tau + T) = 1 - e^{-\lambda(2\tau+T)} > q_2 = q(T) = 1 - e^{-\lambda T} \quad (26)$$

and we can rewrite expected profits as

$$\begin{aligned}\Pi(q_1, q_2) &= (q_1 + q_2) C \\ &+ \frac{4}{1 + (1-p)\beta} \left[\int_{\sqrt{\frac{q_2}{q_1}} \varphi_k}^{\varphi_k} q_1 \varphi_1^2 dF(\varphi_1) - \int_{\varphi_k}^{\sqrt{\frac{q_1}{q_2}} \varphi_k} q_2 \varphi_2^2 dF(\varphi_2) \right] dF(\varphi_k) \\ &- \frac{4}{1 + \beta} \left[\int_{\sqrt{\frac{q_1}{q_2}} \varphi_k}^{\varphi_k} q_1 \varphi_1^2 dF(\varphi_1) - \int_{\varphi_k}^{\sqrt{\frac{q_2}{q_1}} \varphi_k} q_2 \varphi_2^2 dF(\varphi_2) \right] dF(\varphi_k)\end{aligned}$$

or still

$$\Pi(q_1, q_2) = (q_1 + q_2) C + D \int_0^\infty [B_1(q_1, q_2) - B_2(q_1, q_2)] dF(\varphi_k) \quad (27)$$

where

$$D = \frac{4}{1 + \beta} \left(\frac{\beta p}{1 + \beta(1-p)} \right)$$

and

$$B_1(q_1, q_2) = \int_{\sqrt{\frac{q_2}{q_1}} \varphi_k}^{\varphi_k} q_1 \varphi_1^2 dF(\varphi_1) \quad \text{and} \quad B_2(q_1, q_2) = \int_{\varphi_k}^{\sqrt{\frac{q_1}{q_2}} \varphi_k} q_2 \varphi_2^2 dF(\varphi_2)$$

with

$$q_1 = 1 - \exp(-\lambda(T + 2\tau)) \quad \text{and} \quad q_2 = 1 - \exp(-\lambda T)$$

The term $\partial^2 B_2(q_1, q_2) / \partial t^2$

First notice that

$$\frac{\partial}{\partial \tau} \sqrt{\frac{q_1}{q_2}} = \frac{\lambda}{(q_1 q_2)^{\frac{1}{2}}} \exp(-\lambda(T + 2\tau))$$

Hence

$$\frac{\partial B_2(q_1, q_2)}{\partial \tau} = \lambda \left(\frac{q_1}{q_2} \right)^{\frac{1}{2}} \varphi_2^3 f \left(\sqrt{\frac{q_1}{q_2}} \varphi_2 \right) \exp(-\lambda(T + 2\tau))$$

And then

$$\begin{aligned}\frac{\partial^2 B_2}{\partial \tau^2} &= \frac{\lambda^2}{(q_1 q_2)^{\frac{1}{2}}} \varphi_2^3 f \left(\sqrt{\frac{q_1}{q_2}} \varphi_2 \right) \exp[-2\lambda(T + 2\tau)] \\ &+ \frac{\lambda^2}{q_2} \varphi_2^4 f' \left(\sqrt{\frac{q_1}{q_2}} \varphi_2 \right) \exp[-2\lambda(T + 2\tau)] \\ &- 2\lambda^2 \left(\frac{q_1}{q_2} \right)^{\frac{1}{2}} \varphi_2^3 f \left(\sqrt{\frac{q_1}{q_2}} \varphi_2 \right) \exp(-\lambda(T + 2\tau))\end{aligned}$$

It follows that

$$\begin{aligned}\left. \frac{\partial^2 B_2}{\partial \tau^2} \right|_{\tau=0} &= \frac{\lambda^2}{q_2} \varphi_2^3 f(\varphi_2) \exp(-2\lambda T) \\ &+ \frac{\lambda^2}{q_2} \varphi_2^4 f'(\varphi_2) \exp(-2\lambda T) - 2\lambda^2 \varphi_2^3 f(\varphi_2) \exp(-\lambda T)\end{aligned}$$

The term $\partial^2 B_1(q_1, q_2) / \partial t^2$

First notice that

$$\frac{\partial}{\partial \tau} \sqrt{\frac{q_2}{q_1}} = -\lambda \left(\frac{q_2^{\frac{1}{2}}}{q_1^{\frac{3}{2}}} \right) \exp(-\lambda(T + 2\tau))$$

and thus

$$\begin{aligned} \frac{\partial B_1(q_1, q_2)}{\partial \tau} &= \int_{\sqrt{\frac{q_2}{q_1}} \varphi_2}^{\varphi_2} 2\lambda \exp(-\lambda(T + 2\tau)) \varphi_1^2 dF(\varphi_1) \\ &\quad + \lambda \left(\frac{q_2}{q_1} \right)^{\frac{3}{2}} \varphi_2^3 f \left(\sqrt{\frac{q_2}{q_1}} \varphi_2 \right) \exp(-\lambda(T + 2\tau)). \end{aligned}$$

Define

$$P(q_1, q_2) = \lambda \left(\frac{q_2}{q_1} \right)^{\frac{3}{2}} \varphi_2^3 f \left(\sqrt{\frac{q_2}{q_1}} \varphi_2 \right) \exp(-\lambda(T + 2\tau))$$

Then

$$\begin{aligned} \frac{\partial^2 B_1}{\partial \tau^2} &= -4\lambda \exp(-\lambda(T + 2\tau)) \int_{\sqrt{\frac{q_2}{q_1}} \varphi_2}^{\varphi_2} \varphi_1^2 dF(\varphi_1) \\ &\quad + 2\lambda^2 \left(\frac{q_2^{\frac{3}{2}}}{q_1^{\frac{5}{2}}} \right) \exp(-2\lambda(T + 2\tau)) \varphi_2^3 f \left(\sqrt{\frac{q_2}{q_1}} \varphi_2 \right) + \frac{\partial P}{\partial \tau} \end{aligned}$$

Finally

$$\begin{aligned} \frac{\partial P}{\partial \tau} &= -3\lambda^2 \left(\frac{q_2^{\frac{3}{2}}}{q_1^{\frac{5}{2}}} \right) \varphi_2^3 f \left(\sqrt{\frac{q_2}{q_1}} \varphi_2 \right) \exp[-2\lambda(T + 2\tau)] \\ &\quad - \lambda^2 \left(\frac{q_2}{q_1} \right)^{\frac{3}{2}} \left(\frac{q_2^{\frac{1}{2}}}{q_1^{\frac{3}{2}}} \right) \varphi_2^4 f' \left(\sqrt{\frac{q_2}{q_1}} \varphi_2 \right) \exp[-2\lambda(T + 2\tau)] \\ &\quad - 2\lambda^2 \left(\frac{q_2}{q_1} \right)^{\frac{3}{2}} \varphi_2^3 f \left(\sqrt{\frac{q_2}{q_1}} \varphi_2 \right) \exp[-\lambda(T + 2\tau)] \end{aligned}$$

It follows that

$$\begin{aligned} \frac{\partial^2 B_1}{\partial \tau^2} |_{\tau=0} &= 2 \frac{\lambda^2}{q_2} \varphi_2^3 f(\varphi_2) \exp(-2\lambda T) - 3 \frac{\lambda^2}{q_2} \varphi_2^3 f(\varphi_2) \exp(-2\lambda T) \\ &\quad - \frac{\lambda^2}{q_2} \varphi_2^4 f'(\varphi_2) \exp(-2\lambda T) - 2\lambda^2 \varphi_2^3 f(\varphi_2) \exp(-\lambda T) \\ &= - \left[\frac{\lambda^2}{q_2} \varphi_2^3 f(\varphi_2) + \frac{\lambda^2}{q_2} \varphi_2^4 f'(\varphi_2) \right] \exp(-2\lambda T) - 2\lambda^2 \varphi_2^3 f(\varphi_2) \exp(-\lambda T) \end{aligned}$$

Give this we are ready to prove the following Lemma.

Lemma 5 *The profit function is such that*

$$\frac{\partial^2 \Pi(q, q)}{\partial \tau^2} |_{\tau=0} = -2\lambda^2 \exp(-\lambda T) C \quad (28)$$

and

$$\frac{\partial^2 \Pi(q_1, q_2)}{\partial \tau^2} \Big|_{\tau=0} = -4\lambda^2 \exp(-\lambda T) C + 2 \left(\frac{\lambda^2 \exp(-2\lambda T)}{1 - \exp(-\lambda T)} \right) \left(\frac{1}{4\pi} \right) D \quad (29)$$

Proof. Expression (28) follows directly from (25). From (27), we have that

$$\frac{\partial^2 \Pi(q_1, q_2)}{\partial \tau^2} \Big|_{\tau=0} = -4\lambda^2 \exp(-\lambda T) C + D \int_0^\infty \left[\frac{\partial^2 B_1}{\partial t^2} \Big|_{\tau=0} - \frac{\partial^2 B_2}{\partial t^2} \Big|_{\tau=0} \right] dF(\varphi_2)$$

where

$$\frac{\partial^2 B_1}{\partial \tau^2} \Big|_{\tau=0} - \frac{\partial^2 B_2}{\partial \tau^2} \Big|_{\tau=0} = -2 \left(\frac{\lambda^2}{q_2} \right) \exp(-2\lambda T) \varphi_2^3 [f(\varphi_2) + \varphi_2 f'(\varphi_2)]$$

and hence

$$\begin{aligned} \frac{\partial^2 \Pi(q_1, q_2)}{\partial \tau^2} \Big|_{\tau=0} &= -4\lambda^2 \exp(-\lambda T) C \\ &\quad - 2 \left(\frac{\lambda^2 \exp(-2\lambda T)}{1 - \exp(-\lambda T)} \right) D \int_0^\infty \varphi_2^3 [f(\varphi_2) + \varphi_2 f'(\varphi_2)] dF(\varphi_2) \end{aligned}$$

Since

$$\int_0^\infty \varphi_2^3 f(\varphi_2) dF(\varphi_2) = \frac{1}{4\pi} \sigma_\theta^2$$

and

$$\int_0^\infty \varphi_2^4 f'(\varphi_2) dF(\varphi_2) = \int_0^\infty x^4 \left(-\frac{1}{2} \frac{\sqrt{2}}{\sqrt{\pi}} x e^{-\frac{1}{2}x^2} \right) f(\varphi_2) d\varphi_2 = -\frac{1}{2\pi} \sigma_\theta^2$$

Hence

$$\frac{\partial^2 \Pi(q_1, q_2)}{\partial \tau^2} \Big|_{\tau=0} = -4\lambda^2 \exp(-\lambda T) C + 2 \left(\frac{\lambda^2 \exp(-2\lambda T)}{1 - \exp(-\lambda T)} \right) \left(\frac{1}{4\pi} \right) \sigma_\theta^2 D$$

which concludes the proof of Lemma 1. \square

6.3.2 Proof of Proposition 3

Proof of Proposition 3(a): In the limit as τ goes to infinity, the manager observes both θ_1 and θ_2 perfectly under balanced attention ($q_1 = q_2 = q = 1$) whereas she observes shock θ_2 imperfectly under focused attention ($q_2 < q_1 = 1$). It follows that for τ sufficiently large, balanced attention is strictly preferred over focussed attention.

Proof of Proposition 3(c): From (25) and (27), focused attention is preferred over balanced attention if and only if

$$\Pi(q, q) < \Pi(q_1, q_2) \quad (30)$$

$$\begin{aligned} \iff 2q - q_1 - q_2 &\leq \left(\frac{\beta p}{[1 + \beta(1-p) + \beta p \frac{\pi+2}{2\pi}] \sigma_\theta^2} \right) \\ &\quad \times 4 \int_0^\infty [B_1(q_1, q_2) - B_2(q_1, q_2)] dF(\varphi_k) \end{aligned} \quad (31)$$

where q , q_1 and q_2 are given by (26) and (26). 3(c) follows from the observation that the RHS of (31) is strictly increasing in β .

Proof of Proposition 3(b): We need to show that there exists a \bar{T} such that for τ sufficiently small, if $T < \bar{T}$, then

$$\Pi(q_1, q_2) = \Pi(q(T, 2\tau), q(T, 0)) > \Pi(q, q) = \Pi(q(T, \tau), q(T, \tau))$$

and if $T > \bar{T}$, then $\Pi(q_1, q_2) < \Pi(q, q)$.

First notice that

$$\Pi(q_1, q_2)|_{\tau=0} = \Pi(q, q)|_{\tau=0} \quad \text{and} \quad \frac{\partial \Pi(q_1, q_2)}{\partial \tau}|_{\tau=0} = \frac{\partial \Pi(q, q)}{\partial \tau}|_{\tau=0} = 2\lambda \exp(-\lambda T) C$$

From Lemma 1,

$$\begin{aligned} \frac{\partial^2 \Pi(q_1, q_2)}{\partial \tau^2} \Big|_{\tau=0} &= -4\lambda^2 \exp(-\lambda T) C + \frac{\lambda^2}{2\pi} \left(\frac{\exp(-2\lambda T)}{1 - \exp(-\lambda T)} \right) \sigma_\theta^2 D \\ \frac{\partial^2 \Pi(q, q)}{\partial \tau^2} \Big|_{\tau=0} &= -2\lambda^2 \exp(-\lambda T) C \end{aligned}$$

Define \bar{T} as the (unique) solution of

$$\frac{\partial^2 \Pi(q_1, q_2)}{\partial \tau^2} \Big|_{\tau=0} = \frac{\partial^2 \Pi(q, q)}{\partial \tau^2} \Big|_{\tau=0}$$

which after some trivial manipulations boils down to the solution to

$$\left(\frac{1 + \frac{\beta p}{1 + \beta(1-p)} \left(\frac{\pi+2}{2\pi} \right)}{\frac{\beta p}{1 + \beta(1-p)}} \right) \pi = \frac{\exp(-\lambda T)}{1 - \exp(-\lambda T)}.$$

Then clearly for $T < \bar{T}$

$$\frac{\partial^2 \Pi(q_1, q_2)}{\partial \tau^2} \Big|_{\tau=0} < \frac{\partial \Pi(q, q)}{\partial \tau} \Big|_{\tau=0}$$

and for $T > \bar{T}$

$$\frac{\partial^2 \Pi(q_1, q_2)}{\partial \tau^2} \Big|_{\tau=0} > \frac{\partial \Pi(q, q)}{\partial \tau} \Big|_{\tau=0}.$$

which concludes the proof. □

6.4 Proof of Proposition 4

Assume $T_1 > T_2$ and assume that $t_i \in \{0, \tau\}$ with $t_1 + t_2 = \tau$ with τ small. The proof for $t_i \in \{0, 2\tau\}$ with $t_1 + t_2 = 2\tau$ is identical, up to a transformation. Slightly abusing notation, expected profits conditional on a strategic focus (t_1, t_2) and expertise (T_1, T_2) are given by

$$\Pi(T_1 + t_1, T_2 + t_2) = (q_1 + q_2)C + D \int_0^\infty [B_1 - B_2] dF(\varphi_k)$$

where C and D are defined above,

$$q_i = q(T_i + t_i) = 1 - e^{-\lambda(T_i + t_i)}$$

and, again abusing notation,

$$\begin{aligned} B_1 &\equiv B_1(T_1 + t_1, T_2 + t_2) = \int_{\sqrt{\frac{q_2}{q_1}}\varphi_k}^{\varphi_k} q_1 \varphi_1^2 dF(\varphi_1) \\ B_1 &\equiv B_2(T_1 + t_1, T_2 + t_2) = \int_{\varphi_k}^{\sqrt{\frac{q_1}{q_2}}\varphi_k} q_2 \varphi_2^2 dF(\varphi_2) \end{aligned}$$

Note that

$$\begin{aligned} \frac{\partial B_1(T_1 + \tau, T_2)}{\partial \tau} &= \lambda \exp(-\lambda(T_1 + \tau)) \left[\int_{\sqrt{\frac{q_2}{q_1}}\varphi_k}^{\varphi_k} \varphi_1^2 dF(\varphi_1) + \frac{1}{2} \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_k\right) \right] \\ \frac{\partial B_2(T_1 + \tau, T_2)}{\partial \tau} &= \frac{\lambda}{2} \exp(-\lambda(T_1 + \tau)) \left(\frac{q_1}{q_2}\right)^{\frac{1}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_1}{q_2}}\varphi_k\right) \end{aligned}$$

and, similarly,

$$\begin{aligned} \frac{\partial B_1(T_1, T_2 + \tau)}{\partial \tau} &= -\frac{\lambda}{2} \exp(-\lambda(\tau_2 + \tau)) \left(\frac{q_2}{q_1}\right)^{\frac{1}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_k\right) \\ \frac{\partial B_2(T_1, T_2 + \tau)}{\partial \tau} &= \lambda \exp(-\lambda(T_2 + \tau)) \left[\int_{\varphi_k}^{\sqrt{\frac{q_1}{q_2}}\varphi_k} \varphi_2^2 dF(\varphi_2) - \frac{1}{2} \left(\frac{q_1}{q_2}\right)^{\frac{3}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_1}{q_2}}\varphi_k\right) \right] \end{aligned}$$

It follows that

$$\begin{aligned} \frac{\Pi(T_1 + \tau, T_2)}{\partial \tau} &= \lambda \exp(-\lambda(T_1 + \tau)) C + D \int_0^\infty \left[\frac{\partial B_1(T_1 + \tau, T_2)}{\partial \tau} - \frac{\partial B_2(T_1 + \tau, T_2)}{\partial \tau} \right] dF(\varphi_k) \\ &= \lambda \exp(-\lambda(T_1 + \tau)) C + D \lambda \exp(-\lambda(T_1 + \tau)) \\ &\quad \times \int_0^\infty \left(\frac{1}{2} \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_k\right) - \frac{1}{2} \left(\frac{q_1}{q_2}\right)^{\frac{1}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_1}{q_2}}\varphi_k\right) + \int_{\sqrt{\frac{q_2}{q_1}}\varphi_k}^{\varphi_k} \varphi_1^2 dF(\varphi_1) \right) dF(\varphi_k) \end{aligned}$$

We have that

$$\begin{aligned} \int_0^\infty \frac{1}{2} \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_k\right) dF(\varphi_k) &= \int_0^\infty \frac{1}{2} \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_k\right) f(\varphi_k) d\varphi_k \\ &= \int_0^\infty \frac{1}{2} \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \varphi_k^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{q_2}{q_1} \frac{\varphi_k^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\varphi_k^2}{2}} d\varphi_k \\ &= \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{2} \varphi_k^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{q_2+q_1}{q_1} \frac{\varphi_k^2}{2}} d\varphi_k \\ &= \frac{1}{\sqrt{2\pi}} \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \sqrt{\frac{q_1}{q_2+q_1}} \int_0^\infty \frac{1}{2} \varphi_k^3 \frac{1}{\sqrt{\frac{q_1}{q_2+q_1}} \sqrt{2\pi}} e^{-\frac{\varphi_k^2}{\frac{q_2+q_1}{q_1}}} d\varphi_k \\ &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{q_1}{q_2+q_1}} \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \int_0^\infty \frac{1}{2} \varphi_k^3 f\left(\varphi_k; 0, \sqrt{\frac{q_1}{q_2+q_1}}\right) d\varphi_k, \end{aligned}$$

where

$$f\left(\varphi_k; 0, \sqrt{\frac{q_1}{q_2 + q_1}}\right),$$

is the normal density function when the mean is 0 and the standard deviation is $\sqrt{\frac{q_1}{q_2 + q_1}}$. Since

$$\int_0^\infty \frac{1}{2} x^3 f(x; 0, \sigma) dx = \frac{\sigma^3 \sqrt{2}}{2 \sqrt{\pi}},$$

where $f(x; 0, \sigma)$ is the normal density function when the mean is 0 and the standard deviation is σ , this can be simplified to

$$\begin{aligned} \int_0^\infty \frac{1}{2} \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_2}{q_1}} \varphi_k\right) dF(\varphi_k) &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{q_1}{q_2 + q_1}} \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \sqrt{\frac{q_1}{q_2 + q_1}}^3 \frac{1}{2} \frac{\sqrt{2}}{\sqrt{\pi}} \sigma_\theta^3 \\ &= \frac{1}{2\pi} \left(\frac{q_1}{q_2 + q_1}\right)^2 \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \sigma_\theta^3 \end{aligned}$$

Similarly,

$$\begin{aligned} \int_0^\infty \frac{1}{2} \left(\frac{q_1}{q_2}\right)^{\frac{1}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_1}{q_2}} \varphi_k\right) dF(\varphi_k) &= \frac{1}{2\pi} \left(\frac{q_2}{q_2 + q_1}\right)^2 \left(\frac{q_1}{q_2}\right)^{\frac{1}{2}} \sigma_\theta^3 \\ &= \frac{1}{2\pi} \left(\frac{q_1}{q_2 + q_1}\right)^2 \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \sigma_\theta^3 \end{aligned}$$

It follows that

$$\frac{\partial E[\pi|T_1 + \tau, T_2]}{\partial \tau} = \lambda \exp(-\lambda(T_1 + \tau)) C + D \lambda \exp(-\lambda(T_1 + \tau)) \int_0^\infty \int_{\sqrt{\frac{q_2}{q_1}} \varphi_k}^{\varphi_k} \varphi_1^2 dF(\varphi_1) dF(\varphi_k)$$

Similarly,

$$\begin{aligned} \frac{\partial E[\pi|T_1, T_2 + \tau]}{\partial \tau} &= \lambda \exp(-\lambda(T_2 + \tau)) C + D \lambda \exp(-\lambda(T_2 + \tau)) \\ &\times \int_0^\infty \left[\frac{1}{2} \left(\frac{q_1}{q_2}\right)^{\frac{3}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_1}{q_2}} \varphi_k\right) - \frac{1}{2} \left(\frac{q_2}{q_1}\right)^{\frac{1}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_2}{q_1}} \varphi_k\right) - \int_{\varphi_k}^{\sqrt{\frac{q_1}{q_2}} \varphi_k} \varphi_2^2 dF(\varphi_2) \right] dF(\varphi_k) \\ &= \lambda \exp(-\lambda(T_2 + \tau)) C - D \lambda \exp(-\lambda(T_2 + \tau)) \int_0^\infty \int_{\varphi_k}^{\sqrt{\frac{q_1}{q_2}} \varphi_k} \varphi_2^2 dF(\varphi_2) dF(\varphi_k) \end{aligned}$$

In sum, we have that

$$\begin{aligned} \frac{\partial \Pi(T_1 + \tau, T_2)}{\partial \tau} \Big|_{\tau=0} &> \frac{\partial \Pi(T_1, T_2 + \tau)}{\partial \tau} \Big|_{\tau=0} \\ \Leftrightarrow \lambda \exp(-\lambda T_1) C + D \lambda \exp(-\lambda T_1) &\left[\int_0^\infty \int_{\varphi_k}^{\sqrt{\frac{q_1}{q_2}} \varphi_k} \varphi_1^2 dF(\varphi_1) dF(\varphi_k) \right] \\ &> \lambda \exp(-\lambda T_2) C - D \lambda \exp(-\lambda T_2) \left[\int_0^\infty \int_{\sqrt{\frac{q_2}{q_1}} \varphi_k}^{\varphi_k} \varphi_2^2 dF(\varphi_2) dF(\varphi_k) \right] \end{aligned}$$

where $q_1 = q(T_1) = 1 - \exp(-\lambda T_1)$ and $q_2 = q(T_2) = 1 - \exp(-\lambda T_2)$, or still

$$\Leftrightarrow \exp(-\lambda(T_1 - T_2)) > \frac{C - D \int_0^\infty \int_{\sqrt{\frac{q_2}{q_1}} \varphi_k}^{\varphi_k} \varphi_2^2 dF(\varphi_2) dF(\varphi_k)}{C + D \int_0^\infty \int_{\varphi_k}^{\sqrt{\frac{q_1}{q_2}} \varphi_k} \varphi_1^2 dF(\varphi_1) dF(\varphi_k)}$$

Define

$$\rho = \frac{q(T_1)}{q(T_2)}$$

Then

$$\begin{aligned} & \left. \frac{\partial \Pi(T_1 + \tau, T_2)}{\partial \tau} \right|_{\tau=0} > \left. \frac{\partial \Pi(T_1, T_2 + \tau)}{\partial \tau} \right|_{\tau=0} \\ \Leftrightarrow & \frac{1 - q(T_1)}{1 - q(T_1)/\rho} > \frac{C - D \int_0^\infty \int_{\varphi_k/\sqrt{\rho}}^{\varphi_k} \varphi_2^2 dF(\varphi_2) dF(\varphi_k)}{C + D \int_0^\infty \int_{\varphi_k}^{\sqrt{\rho} \varphi_k} \varphi_1^2 dF(\varphi_1) dF(\varphi_k)} \end{aligned}$$

or, substituting C and D and simplifying,

$$\Leftrightarrow \frac{1 - q(T_1)}{1 - q(T_1)/\rho} > \frac{(1 + \frac{\pi+2}{2\pi}b) \sigma_\theta^2 - 4b \int_0^\infty \int_{\varphi_k/\sqrt{\rho}}^{\varphi_k} \varphi_2^2 dF(\varphi_2) dF(\varphi_k)}{(1 + \frac{\pi+2}{2\pi}b) \sigma_\theta^2 + 4b \int_0^\infty \int_{\varphi_k}^{\sqrt{\rho} \varphi_k} \varphi_1^2 dF(\varphi_1) dF(\varphi_k)}, \quad (32)$$

where

$$b \equiv \frac{\beta p}{1 + \beta(1 - p)} \in (0, \infty),$$

Fix $\rho > 1$, then on the one hand, the RHS is strictly smaller than 1 and independent of q_1 . On the other hand, the LHS is strictly decreasing in q_1 , and equals 1 as q_1 goes to 0 and goes to 0 as q_1 goes to 1. Hence, keeping q_1/q_2 fixed, if q_1 is sufficiently small, then managing with style $((t_1^*, t_2^*) = (\tau, 0))$ is always optimal. Similarly, fixing q_1/q_2 as q_1 goes to 1, then for q_1 sufficiently large, rebalancing attention $((t_1^*, t_2^*) = (0, \tau))$ is optimal. It follows that there exists a unique cut-off q_1^* given by

$$\frac{1 - q_1^*}{1 - q_1^*/\rho} = \frac{(1 + \frac{\pi+2}{2\pi}b) \sigma_\theta^2 - 4b \int_0^\infty \left(\int_{\theta_j/\sqrt{\rho}}^{\theta_j} \theta_i^2 dF(\theta_i) \right) dF(\theta_j)}{(1 + \frac{\pi+2}{2\pi}b) \sigma_\theta^2 + 4b \int_0^\infty \left(\int_{\theta_j}^{\sqrt{\rho} \theta_j} \theta_i^2 dF(\theta_i) \right) dF(\theta_j)}, \quad (33)$$

so that if $q(T_1) < q_1^*$, we have $(t_1^*, t_2^*) = (\tau, 0)$, and for $q(T_1) > q_1^*$ we have $(t_1^*, t_2^*) = (0, \tau)$. Note further that q_1^* is continuous in ρ and continuous and increasing in b . Defining $\Lambda(\rho, b)$ as

$$q_1^* = 1 - e^{-\Lambda} \quad (34)$$

It follows that $(t_1^*, t_2^*) = (\tau, 0)$ if $\lambda T_1 < \Lambda(\rho, b)$ whereas $(t_1^*, t_2^*) = (0, \tau)$ if $\lambda T_1 > \Lambda(\rho, b)$. Moreover, $\Lambda(\rho, b)$ is continuous in ρ and continuous and increasing in b . \square

6.5 Intermediate Allocations of Attention and Endogenous Expertise

We now discuss more formally the results regarding the intermediate allocation of attention in section 4.3 as well as the endogenous choice of managerial expertise. For this purpose, it will be useful to introduce some

additional notation. Consider the curve $\xi(\cdot, b) : \rho \in (1, +\infty) \rightarrow (\lambda T_1, \lambda T_2) \in \mathbb{R}^+ \times \mathbb{R}^+$ where $\xi(\rho, b) = (\xi_1(\rho, b), \xi_2(\rho, b))$ is given by

$$\begin{aligned}\xi_1(\rho, b) &\equiv \Lambda(\rho, b) \\ \hat{q}(\xi_2(\rho, b)) &\equiv \hat{q}(\Lambda(\rho, b)) / \rho\end{aligned}$$

with $\hat{q}(x) = 1 - e^{-x}$. Since $\Lambda(\rho, b)$ is a continuous function of ρ , both $\xi_1(\rho, b)$ and $\xi_2(\rho, b)$ are continuous functions of ρ and, hence, $\xi(\cdot, b)$ is a continuous mapping from $\rho \in (1, +\infty)$ to $(\lambda T_1, \lambda T_2) \in \mathbb{R}^+ \times \mathbb{R}^+$. Abusing notation slightly, we will also use $\xi(\cdot, b)$ to refer to the image of the curve, that is the set $\xi(\cdot, b) = \{(\lambda T_1, \lambda T_2) : (\lambda T_1, \lambda T_2) = \xi(\rho, b) \text{ for } \rho > 1\}$. In Figure 4 and 5, Panels A-C, we plot $\xi(\cdot, b)$ for $b = 5/3$, as represented by the black downward-sloping curve.²²

Remark 6 For any $b > 0$, the curve $\xi(\cdot, b)$ divides the set $S = \{(\lambda T_1, \lambda T_2) : T_1 > T_2 \geq 0\}$ in two connected subsets S^- and S^+ , where $\{S^+, S^-, \xi(\cdot, b)\}$ is a partition of S , and where S^+ has no points in common with the closure of S^- (and vice versa).

Proof. Note that $\xi(\cdot, b)$ is a ‘simple’ curve which does not cross itself. Indeed, $\rho = (1 - e^{-\xi_1}) / (1 - e^{-\xi_2})$, so that $\rho' \neq \rho$ implies that $\xi(\rho, b) \neq \xi(\rho', b)$. Moreover, since $\rho = (1 - e^{-\xi_1}) / (1 - e^{-\xi_2})$, we have that

$$\xi^- = (\xi_1^-, \xi_2^-) \equiv \lim_{\rho \rightarrow 1} \xi(\rho, b) \quad (35)$$

is on the 45 degree line (where $\xi_1 = \xi_2$) and

$$\xi^+ = (\xi_1^+, \xi_2^+) \equiv \lim_{\rho \rightarrow +\infty} \xi(\rho, b) \quad (36)$$

is on the x -axis (where $\xi_2^+ = 0$). Both ξ_1^- and ξ_1^+ can be shown to be finite. \square

The following result is a corollary of Proposition 4 and Remark 6:

Corollary 7 Denote by S^- the subset of S who is to the ‘left’ of ξ (that is, whose set closure contains $(0, 0)$).

- Assume $(\lambda T_1, \lambda T_2)$ and (t_1, t_2) are such that $(\lambda(T_1 + t_1), \lambda(T_2 + t_2)) \in S^-$. If $t_2 > 0$, then profits can be increased by shifting attention from task 2 to task 1.
- Assume $(\lambda T_1, \lambda T_2)$ and (t_1, t_2) are such that $(\lambda(T_1 + t_1), \lambda(T_2 + t_2)) \in S^+$. If $t_1 > 0$, then profits can be increased by shifting attention from task 1 to task 2.

To establish the main result of this section define first the function $g(Z)$ for $Z > 0$ as

$$g(Z) \equiv \{(\lambda T_1, \lambda T_2) \in S : \lambda T_2 + \lambda T_1 = Z\}.$$

²²The plot was generated with a matlab code that is available upon request.

Assumption A2. (i) $\xi(\cdot, b)$ is downwards sloping: $\xi_2(\cdot, b)$ is decreasing in ρ and $\xi_1(\cdot, b)$ is increasing in ρ ;
(ii) the function $g(Z)$ cuts the $\xi(\cdot, b)$ at most once and always from above.²³

We have been unable to prove that Assumption A2 is met for all $b > 0$, but also unable to generate any example where the assumption is not met. Figure 6 plots the curve $\xi(\cdot, b)$ for a set of values of b , ranging from $b = .01$ to $b = 100$. All the curves in the figure (and every other we have tried) satisfy Assumption A2.

We are now ready to provide a more rigorous treatment of the results discussed in Section 4.3 regarding the intermediate allocation of attention. For a given expertise configuration $(\lambda T_1, \lambda T_2) \in S^-$, define

$$\begin{aligned}\tau_y &\equiv \sup \{ \tau \in \mathbb{R}^+ : (\lambda T_1, \lambda(T_2 + 2\tau)) \in S^- \} \\ \tau_x &\equiv \sup \{ \tau \in \mathbb{R}^+ : (\lambda(T_1 + 2\tau), \lambda T_2) \in S^- \}\end{aligned}$$

Note that if $\xi(\cdot, b)$ satisfies Assumption A2, then $0 < \tau_y < \tau_x$.

Proposition 8 *Assume that $\xi(\cdot, b)$ satisfies Assumption A2, then*

- *If $(\lambda T_1, \lambda T_2) \in S^-$ and $2\tau < 2\tau_y$, then $(t_1^*, t_2^*) = (2\tau, 0)$*
- *If $(\lambda T_1, \lambda T_2) \in S^+$ or $2\tau > 2\tau_x$, then $(t_1^*, t_2^*) = (0, 2\tau)$ if $2\tau < T_1 - T_2$, and $T_1 + t_1^* = T_2 = t_2^*$ otherwise*

Proof (i) First assume that $(\lambda T_1, \lambda T_2) \in S^-$ and $2\tau < 2\tau_y$. Then given Assumption A2, $(\lambda(T_1 + t_1), \lambda(T_2 + t_2)) \in S^-$ whenever $t_1 + t_1 \leq 2\tau$. Given Corollary 7, then $(t_1^*, t_2^*) = (2\tau, 0)$.

(ii) Next assume that $(\lambda T_1, \lambda T_2) \in S^-$ but $2\tau > 2\tau_x$. Given Assumption A2, if $2\tau < T_1 - T_2$, then $(\lambda(T_1 + t_1), \lambda(T_2 + t_2)) \in S^+$ whenever $t_1 + t_1 = 2\tau$. Given Corollary 7, it follows that $(t_1^*, t_2^*) = (0, 2\tau)$. If $(\lambda T_1, \lambda T_2) \in S^-$ but $2\tau > T_1 - T_2$, then it is possible that $(t_1, t_2) = (\hat{t}_1, \hat{t}_2)$ is such that $T_2 + \hat{t}_2 > T_1 + \hat{t}_1$. But expected profits in the latter case are identical as when $(t_1, t_2) = (\tilde{t}_1, \tilde{t}_2)$ with $(T_1 + \tilde{t}_1, T_2 + \tilde{t}_2) = (T_2 + \hat{t}_2, T_1 + \hat{t}_1)$. It is therefore without loss of generality to only consider allocations of attention (t_1, t_2) for which $T_2 + t_2 \leq T_1 + t_1$. Given $2\tau > 2\tau_x$ and given $2\tau > T_1 - T_2$, then whenever $t_1 + t_1 = 2\tau$, either $T_1 + t_1 = T_2 + t_2$ or $(\lambda(T_1 + t_1), \lambda(T_2 + t_2)) \in S^+$. But whenever $(\lambda(T_1 + t_1), \lambda(T_2 + t_2)) \in S^+$ profits can be improved by shifting attention from task 1 to task 2. It follows that (t_1^*, t_2^*) is such that $T_1 + t_1^* = T_2 + t_2^*$.

(iii) Finally, if $(\lambda T_1, \lambda T_2) \in S^+$, the same arguments apply as in (ii). \square

To conclude, assume as in section 4.4, that $2\tau = 0$, but the expertise of managers (T_1, T_2) is endogenously chosen under the constraint $\lambda(T_1 + T_2) \leq Z$. The following proposition generalizes the insights of section 4.4.

²³Formally, for any $Z > 0$, the line $g(Z)$ intersects $\xi(\cdot, b)$ at most once, that is, the set $\xi(\cdot, b) \cap g(Z)$ is empty or a singleton. Let $(\lambda \tilde{T}_1, \lambda \tilde{T}_2) = \xi(\cdot, b) \cap g(Z)$ then for all $(\lambda T_1, \lambda T_2) \in g(Z)$, (i) if $\lambda T_1 < \lambda \tilde{T}_1$, then $(\lambda T_1, \lambda T_2) \in S^+$ and (ii) if $\lambda T_1 > \lambda \tilde{T}_1$, then $(\lambda T_1, \lambda T_2) \in S^-$.

Proposition 9 *If managerial expertise (T_1, T_2) is optimally chosen under the constraint $\lambda(T_1 + T_2) \leq Z$ and $\xi(\cdot, b)$ satisfies Assumption A2, then*

- *If $Z < 2\xi_1^-$, with ξ_1^- defined in (35), then the optimal manager is a specialist: $(\lambda T_1^*, \lambda T_2^*) = (Z, 0)$ or $(\lambda T_1^*, \lambda T_2^*) = (0, Z)$.*
- *If $Z > \xi_1^+$, with ξ_1^+ defined in (36), then the optimal manager is a generalist: $(\lambda T_1^*, \lambda T_2^*) = (Z/2, Z/2)$.*

Proof. If $Z < 2\xi_1^-$, then Assumption A2 guarantees that $g(Z) \subset S^-$. Given $T_1 \geq T_2$, Corollary 7 then implies that profits can always be improved by shifting expertise from task 2 to task 1. Similarly, if $Z > \xi_1^+$, then under Assumption A2, $g(Z) \subset S^+$. Given $T_1 \geq T_2$, Corollary 7 then implies that profits can always be improved by shifting expertise from task 1 to task 2 up to the point where $T_1 = T_2$. \square

For completeness, we end this section by providing expressions for ξ_1^- and ξ_1^+ and show that we always have that $2\xi_1^- < \xi_1^+$:

(i) From (33) and (34), we have that ξ_1^+ is implicitly given by

$$e^{-\xi_1^+} = \frac{(1 + \frac{\pi+2}{2\pi}b) \sigma_\theta^2 - 4b \int_0^\infty \left(\int_0^{\theta_j} \theta_i^2 dF(\theta_i) \right) dF(\theta_j)}{(1 + \frac{\pi+2}{2\pi}b) \sigma_\theta^2 + 4b \int_0^\infty \left(\int_{\theta_j}^\infty \theta_i^2 dF(\theta_i) \right) dF(\theta_j)}, \quad (37)$$

$$= \frac{(1 + \frac{\pi+2}{2\pi}b) - \frac{\pi+2}{2\pi}b}{(1 + \frac{\pi+2}{2\pi}b) + \frac{\pi-2}{2\pi}b} = \frac{1}{1+b} \quad (38)$$

from which

$$\xi_1^+ = -\ln\left(\frac{1}{1+b}\right)$$

(ii) From Proposition 4, ξ_1^- is finite for ρ close to 1 and is continuous in ρ . Hence, by continuity, also $\xi_1^- = \lim_{\rho \rightarrow 1} \xi_1(\rho, b)$ is finite. One can further show that ξ_1^- is implicitly given by

$$\frac{e^{-\xi_1^-}}{1 - e^{-\xi_1^-}} = \left(\frac{1 + b \left(\frac{\pi+2}{2\pi} \right)}{b} \right) \pi$$

(iii) Finally, note that ξ_1^- is smaller than $\xi_1^+/2$. Indeed, we have that

$$\frac{e^{-\xi_1^+/2}}{1 - e^{-\xi_1^+/2}} = \frac{\sqrt{e^{-\xi_1^+}}}{1 - \sqrt{e^{-\xi_1^+}}} = \frac{\sqrt{\frac{1}{1+b}}}{1 - \sqrt{\frac{1}{1+b}}} < \left(\frac{1 + b \left(\frac{\pi+2}{2\pi} \right)}{b} \right) \pi = \frac{e^{-\xi_1^-}}{1 - e^{-\xi_1^-}}$$

Since $\frac{e^{-x}}{1-e^{-x}}$ is decreasing in x , it follows that $\xi_1^+/2 > \xi_1^-$.

Figure 6: The downward sloping lines are the function $\Lambda(\rho, b)$ plotted in the space $(\lambda T_1, \lambda T_2)$ for different values of b (see expression (15)), ranging from 10^{-2} (the first downward sloping curve from the right) to 10^2 . Below the downward sloping curves the manager allocates the marginal unit of attention to task 1, whereas above she allocates the marginal unit of attention to task 2.

