Managerial Style and Attention

Wouter Dessein
Columbia University

Tano Santos
Columbia University

October 28, 2014

Preliminary draft

Prepared for NBER Organizational Economics Meeting, Stanford, Fall 2014

Abstract

A large empirical literature has shown how firm behavior is correlated with the background and expertise of its managers. But managerial knowledge and expertise are largely endogenous. We develop a cognitive theory of manager fixed effects, where the allocation of managerial attention determines firm behavior. A manager learns about two strategic choices, each pertaining to a different task (e.g. operations and marketing). The manager then communicates about the chosen strategies to an organization, which must implement them in a decentralized way. The need for coordinated implementation makes it optimal to communicate 'narrow' strategies, focused on one task. This, in turn, may induce the manager to 'manage with style': despite the risk of being blindsided and forsaking valuable opportunities, the manager focuses all her attention to learn about one task and mainly communicates to the organization about this task.

We show that in uncertain, 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. Firm owners (e.g. boards) then prefer managers with task-specific expertise rather than generalist managers, even when they themselves have no preference for a particular strategy. In contrast, in less complex and more certain environments, the endogenous allocation of attention mitigates manager fixed effects and boards optimally hire generalists with equal expertise in both tasks

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] Barker 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. Among other factors, R&D expenditures are shown to be greater at firms where CEOs have significant career experience in marketing and/or engineering/R&D or at firms where CEOs have advanced science-related degrees. 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.

The main conceptual concern about the above empirical findings is that managerial knowledge and expertise are largely endogenous. Managers allocate attention – and develop expertise – in order to learn about strategic choices. Similarly, firm owners and boards of directors decide whether or not to hire managers with expertise in certain areas. In addition,

---

[1]See Finkelstein, Hambrick and Cannellas (2009) for a comprehensive review.

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

In this paper, we develop a cognitive theory of manager fixed effects, where the allocation of managerial attention determines firm behavior. In our model, 'managing with style' – defined as biasing managerial attention to one particular task or function – may result in managers being blindsided and forsaking valuable opportunities, but is often an optimal response to the need for organizational alignment and coordination around selected strategies. We show how in uncertain, complex environments, the endogenous allocation of attention exacerbates manager fixed effects. Small differences in managerial expertise may then result in dramatically 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 they themselves have 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 or firm owners out of a large pool of potential candidates. In contrast, in less complex and more certain environments, the endogenous allocation of attention mitigates manager fixed effects and boards optimally hire generalists who develop a 'broad field of vision.' We thus endogenize to what extent firm behavior and strategic choices reflect managerial characteristics as opposed to the realization of environmental shocks.

In our model, a manager selectively allocates attention – and develops expertise – 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 al-

---

[2]According 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."

locates 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. 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 are particularly interested in knowing whether it is optimal for the manager to favor certain tasks when gathering information, and how the optimal allocation of attention interacts with the expertise of the manager and the strategic choices she ends up making. In addition, we investigate whether and when specialist managers may be preferred over generalist managers.

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. In this benchmark, the manager is the only actor, and there is no need for organizational implementation. In our example above, the manager then simply aims to learn about the best strategies for cost-minimization and revenue growth, without any regard to the implementation of those strategies. As we show, a generalist manager who has equal expertise about both tasks then divides her attention evenly and a specialist manager, who has more expertise about one task, compensates by devoting more attention to the task she is less knowledgeable about. The intuition stems directly from our learning technology: We assume that 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 (e.g. Alonso, Dessein and Matouschek (2008), Rantakari (2008, 2013), Bolton, Brunnermeier and Veldkamp (2011), Van den Steen (2014)). 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. In addition, the quality of implementation also depends on

how closely a strategic choice adheres to standard operating procedures (Dessein and Santos, 2006). Formally, standard choices are choices which are adaptive to the 'average' task-specific shock. By default, agents in our organization take actions which are complementary with standard choices and no communication is required to achieve good implementation of such choices. In contrast, strategic choices which deviate substantially from standard operating procedures require intensive communication to avoid poor internal alignment. An implication of our model is thus that good external alignment may come at the expense of poor internal alignment (and vice versa).

A first insight of our model is that when the need for internal alignment is sufficiently important, then the manager *only communicates 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. Instead internal alignment on the second task is achieved by selecting a more standard (non-adaptive) strategic choice so that communicating is unnecessary. It follows that the need for internal alignment distorts the firm's strategy to be very adaptive on one dimension (one task), while being non-adaptive on the other task, even when the shocks affecting each task are very similar in size.

Our main insight is that not only are a firm's strategic choices distorted – compared to a benchmark where only external alignment matters – but so is the optimal allocation of attention. Intuitively, it is optimal for the manager to learn more precisely the shock which affects the task she is more likely to communicate about. As noted above, if internal alignment is important, the manager will only communicate about the task whose shock she perceives to be largest. Crucially, however, which shock appears to be largest not only depends on the realization of the shock, but also on the allocation of attention. Thus, a manager which devotes more attention to marketing than to operations is more likely to identify larger opportunities for revenue growth than for cost-minimization and, hence, is more likely to only communicate about marketing strategy to the organization. This, of course, makes it optimal to devote more attention to marketing to begin with. In other words, it is optimal for the manager to pay more attention to the strategic choice she is more likely to communicate about *and* the manager is

more likely to communicate about the task she pays more attention to. In contrast, whenever internal alignment is important, a manager which divides her attention equally among both task will be largely wasting half of the information learned.

Because of the above complementarity, we show how managing with style – or a manager with style – may be optimal in complex, uncertain environments. In particular, and in contrast to our benchmark, we find that a generalist manager then optimally focusses her scarce attention on one task provided that internal alignment is important. 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 then focus all her attention on of one those two functions, say marketing. Ex post, such a manager then mainly (but not always) selects and communicates about marketing strategies, and it appears as if the manager is (arbitrarily and inefficiently) biased towards marketing. Similarly, and again in contrast with our benchmark, a specialist manager then optimally focusses her scarce attention on the shock about which her expertise provides her with better information. Finally, in complex and uncertain environments, firm owners (e.g. boards) prefer managers with task-specific expertise rather than generalist managers, even when they themselves have no preference or foresight as to the best strategic choices for the firm.

We derive a number of comparative static results as to when managing with style, or a manager with style, is optimal. First, as noted above, 'managing with style' (a narrow field of vision and attention) and 'managers with style' (managers with specialized expertise) are optimal when attention is scarce and the environment is complex or uncertain. This setting corresponds, for example, to conditions faced by many start-up firms and technology companies. In contrast, in more certain and slow-moving environments (e.g. mature industries where fast decision-making is not a priority), generalist managers with a broad field of vision (balanced attention) are more likely to be optimal. Intuitively, when the environment is *not* complex or uncertain, or when managerial attention is *not* scarce, it is possible for the manager to learn the optimal strategic choice for each task reasonably well. By dividing her attention, she is then better able to identify which task she should be communicating about to the organization.

Secondly, managing with style is more likely to be optimal when organizational implementation and internal alignment is important. Intuitively, when internal alignment is *not* very

important, then either the manager communicates about both strategic choices to the organization, or she chooses strategies which are adaptive to both shocks – even when she only communicates about one strategic choice. As in our benchmark case, external alignment is then a priority and the manager optimally learns about both task-specific shocks. In sum, the above comparative statics show how the extent to which firm behavior and strategic choices reflect managerial characteristics depends both on the complexity of the environment and the need for coordinated implementation inside the organization.

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.[3] As we discuss in Section 6, the focus of this literature is on the biased and subjective processing of complex, ambiguous information, rather than on the optimal allocation of (scarce) attention. 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] Only a few papers study the optimal allocation of attention by a single manager, and show this allocation is related to firm behavior. Bandiera, Guiso, Prat and Sadun (2011) and Bandiera, Prat and Sadun (2013) employ time use surveys to measure how CEOs allocate their attention and show how it is strongly correlated to firm performance. Their main focus 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). They show, both theoretically and empircally, how the number of hours worked by the CEO and the allocation of this time to meetings with insiders are correlated with each other as well as with better firm performance and better governance. Van den Steen (2013a) studies a strategy formulation game in which a strategist investigates one decision – among a set of interrelated decisions – and then announces this decision to a group of agents in charge of implementing those decisions. The paper then analyzes the characteristics (such as irreversibility and centrality) of strategic decisions, defined as the decisions which are announced by the strategist and which guide all other decisions. Under the constraint that at most one decision can be investigated, Van den Steen (2013b) makes the ob-

---

[3]Two seminal paper are Hambrick and Mason (1984) and Ocasio (1997).

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

servation that, ceteris paribus, the strategist investigates the decision about which she expects to receive the most informative signal (that is, the one she has more expertise in). 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 announces the decision she investigates.[5] The impact of managerial expertise on firm behavior is also independent of the uncertainty of the environment. In contrast, a central prediction of our model is that manager fixed effects are more pronounced in uncertain, complex environments. 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 2014) by endogenizing to what extent selected business strategies are contingent on the organizational environment as opposed to managerial characteristics. We refer to Section 6 for a discussion of the empirical literature on why managers matter.

Outline.– Section 2 presents the model. Section 3 provides a benchmark: The allocation of managerial attention in the case where only external alignment of strategic choices matters, and shows that in this case 'managing with style' does not arise. Section 4 takes as given the allocation of managerial attention and studies managerial actions and internal communication strategies. Section 5 finally endogenizes managerial attention choices and managerial expertise. Section 6 concludes by discussing some implications of our results.

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.

---

[5]Hence, strategic choices are purely a function of managerial characteristics. In contrast, our model endogenizes to what extent strategic choices reflect managerial characteristics as opposed to the realization of environmental shocks.

Concretely, the manager must for each task i select a strategic choice a_{Mi} whose bliss-point equals the task-specific shock θ_i and each employee j must choose complementary actions a_{j1} and a_{j2} whose bliss-points equal the strategic choices selected by the manager. 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. Realized payoffs are given by

$$\pi \equiv \pi(\boldsymbol{\theta}, \mathbf{a_M}, \mathbf{a_j}) = \sum_{i \in \{1,2\}} \left[h(\theta_i) - (a_{Mi} - \theta_i)^2 - \beta \int_0^1 (a_{Mi} - a_{ji})^2 dj \right] \qquad (1)$$

where the parameter β captures the relative importance of internal alignment and $\boldsymbol{\theta} = (\theta_1, \theta_2)$.[6] 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,[7] we will assume that

$$h(\theta_i) \equiv \theta_i^2$$

so that profits $\pi(\boldsymbol{\theta}, \mathbf{a_M}, \mathbf{a_j})$ are normalized to 0 whenever $\mathbf{a_M} = \mathbf{a_j} = (0, 0)$ for all $j \in [0, 1]$. Adapting to the shock 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

---

[6]The 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 (2008,2013), Rantakari (2008,2013), Bolton, Brunnermeier and Veldkamp (2013), Calvo-Armengol, de Marti and Prat (2014) and Dessein, Galeotti and Santos (2013) 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.

[7]Indeed, 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 $h(\theta_i)$.

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 $a_{Mi} = 0$, that is $a_{ji} = 0$.

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. Roughly, the manager can direct employees' scarce attention to messages pertaining to her strategic choices. 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

$$p_i = 1 - e^{-\mu r_i}, \tag{2}$$

which is independent across agents. Given this communication technology the employee $j's$ choices are given by:

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

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 attention constraint

$$r_1 + r_2 \leq r \qquad \text{with} \qquad r > 0 \tag{3}$$

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:

$$(1 - p_1)(1 - p_2) \geq 1 - p \tag{4}$$

where $p \equiv 1 - e^{-\mu r}$ denotes the probability of understanding a_{Mi} when employees' attention is fully dedicated to task i. Finally, we define the *internal alignment of task i* as

$$E\left(a_{Mi} - a_{ji}\right)^2 = Var(a_{Mi})e^{-\mu r_i} = Var(a_{Mi})(1 - p_i)$$

2.3 Allocation of attention and learning by the manager

2.3.1 Learning technology

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 attention t_i* devoted to task i. In addition, the manager observes an *exogenous* signal S_i about θ_i whose informativeness depends on her *managerial expertise T_i* in task i. Let $\widehat{\theta}_i \equiv E\left[\theta_i | S_i, s_i\right]$, be the manager's posterior conditional on both the exogenous and the endogenous signal. Then the manager's attention choice is summarized by the mean square error or *residual variance*

$$\mathsf{RV}\left(\theta_i\right) \equiv E\left(\theta_i - \widehat{\theta}_i\right)^2,$$

which naturally should be a decreasing function of both t_i and T_i. In the remainder of the paper, we will assume that the residual variance $\mathsf{RV}\left(\theta_i\right)$ 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. In particular, we posit that

$$\mathsf{RV}\left(\theta_i\right) = \sigma_\theta^2 e^{-\lambda(t_i + T_i)}, \tag{5}$$

where λ characterizes the speed of learning and, hence, $1/\lambda$ the complexity of the environment. Microfoundations for this technology are provided below. We further posit that managerial attention is scarce and that

$$t_1 + t_2 \leq 2\tau. \tag{6}$$

We refer to the profile $\mathbf{t} = (t_1, t_2)$ as the manager's strategic focus and denote by Υ the set of feasible allocations of attention: $\Upsilon = \{(t_1, t_2) : t_1 + t_2 \leq 2\tau\}$. If $T_1 > T_2$ then we speak of a manager who is specialized 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 signals about task i than a generalist or a specialist in task $j \neq i$. For example, a manager may

10

find it much easier to assess a particular situation when the corresponding task belongs to her domain of expertise. 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, a decrease in expertise can be perfectly compensated by an increase in attention in our model. It is in this sense that expertise and attention are substitutes.

Our assumption that expertise and attention are substitutes is 'conservative' from a modelling 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.

2.3.2 Microfoundations

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:[8]

$$q_i \equiv q\left(t_i + T_i\right) = 1 - e^{-\lambda(t_i+T_i)}. \tag{7}$$

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

Denoting $s_i = [s_{i1}, ..., s_{in}]$ and $S_i = [S_{i1}, ..., S_{in}]$, 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

$$\mathsf{RV}\left(\theta_i\right) = E\left(\theta_i - \widehat{\theta}_i\right)^2 = \sigma_\theta^2(1 - q_i), \tag{8}$$

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

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

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 $\mathsf{RV}\left(\theta_i\right)$.

2.3.3 Managing with style

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. As we will show below – and as postulated by the 'upper echelon theory' of Hambrick and Mason (1984) – in order to understand firm behavior, it will be important to understand how managers allocate attention. In particular, the manager will communicate more often to the organization about the task which she devotes more managerial attention to, and the strategic choices pertaining to this task also tend to be more responsive to external shocks. It is in this sense that firm behavior is determined by the allocation of managerial attention.

---

[8]Notice that the probability that the manager learns θ_{ik} is given by

$$(1 - q\left(T_i\right)) q\left(t_i\right) + (1 - q\left(t_i\right)) q\left(T_i\right) + q\left(T_i\right) q\left(t_i\right) = q\left(t_i + T_i\right).$$

In contrast with Hambrick and Mason, however, we assume that managers *optimally* allocate attention. We thus develop a theory of managerial style which does not rely on behavioral or cognitive biases.

2.4 Timing

1. The manager allocates attention $t_i \in [0, 2\tau]$ to each task $i = 1, 2$ with $t_1 + t_2 = 2\tau$.

2. Having observed the corresponding signals, the manager obtain posteriors $\widehat{\theta}_i$ for $i = 1, 2$ and selects strategic choices $\mathbf{a_M} = (a_{M1}, a_{M2})$.

3. The manager direct the employees' attention $r_i \in [0, r]$ to strategic choice a_{Mi} for $i = 1, 2$ with $r_1 + r_2 = r$.

4. Having learned about management choices, employees select their complementary actions, $a_{ji} = E(a_{Mi})$ for $j \in [0, 1]$ and $i = 1, 2$.

3 Benchmark: Attention and External Alignment

As a benchmark, we analyze the case where the manager only cares about the *external alignment* of task 1 and 2, that is, she maximizes

$$E\left[\pi_{EA}(\boldsymbol{\theta}, \mathbf{a_M})\right] \equiv \sum_{i \in \{1,2\}} E\left[\theta_i^2 - (a_{Mi} - \theta_i)^2\right],$$

where EA stands for external alignment. We show that in this case 'managing with style' is suboptimal. 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. Finally, this benchmark could correspond to a one-man organization, where the manager takes all the complementary actions a_{ji}.

Given signals (s_1, S_1, s_2, S_2) the manager's strategic choices $\mathbf{a_M} = (a_{M1}, a_{M2})$ are then $(a_{M1}, a_{M2}) = \left(\widehat{\theta}_1, \widehat{\theta}_2\right)$ and expected payoffs are given by

$$E\left[\pi_{EA}(\theta, \mathbf{a_M})\right] = \sum_{i \in \{1,2\}} \left[\sigma_\theta^2 - \mathsf{RV}(\theta_i)\right],$$

13

It follows that the optimal allocation of attention is given by

$$\mathbf{t}^* = \arg\min_{\mathbf{t}\in\Upsilon}\{\mathsf{RV}(\theta_1) + \mathsf{RV}(\theta_2)\}$$
$$= \arg\min_{\mathbf{t}\in\Upsilon}\left\{(1-q_1)\sigma_\theta^2 + (1-q_2)\sigma_\theta^2\right\},$$

where the residual variance $\mathsf{RV}(\theta_i)$ is given by (5) and where $q_i = q(t_i + T_i)$ is given by (7). Given (7) and (8)

$$\frac{\partial\mathsf{RV}(\theta_i)}{\partial t_i} = -\lambda(1-q_i)\sigma_\theta^2 \qquad \text{and} \qquad \frac{\partial^2\mathsf{RV}(\theta_i)}{\partial t_i^2} = \lambda^2(1-q_i)\sigma_\theta^2 > 0.$$

Since q_i is increasing in attention t_i, there are *decreasing marginal returns to devoting attention* to any given task in terms of reducing the residual variance $\mathsf{RV}(\theta_i)$. Similarly, since q_i is increasing in expertise T_i, we have that $\partial^2\mathsf{RV}(\theta_i)/(\partial t_i \partial T_i) > 0$ so that the marginal returns to devoting attention to any given task are decreasing in the manager's expertise in that task.

It follows that when $t_1 > t_2$, the marginal returns to devoting attention to task 2 are higher than to to task 1 provided the manager is a generalist $(T_1 = T_2)$. Similarly, when the manager is an expert on task 1, that is $T_1 > T_2$, then the marginal returns to devoting attention to task 2 are higher than to task 1 provided that both tasks currently receive equal attention.

Proposition 1 *The optimal strategic focus* $\mathbf{t} = (t_1, t_2)$ *of a manager which maximizes external alignment (e.g.* $\beta = 0$*) is given by*

(a) $t_1 = t_2 = \tau$ *if the manager is a generalist* $(T_1 = T_2)$

(b) $0 \leq t_1 < t_2 \leq 2\tau$ *if the manager has more expertise in task 1* $(T_1 > T_2)$

The above proposition follows directly from our learning technology in which attention and expertise are substitutes and which features decreasing marginal return to attention to any given task. Obviously, 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. Such increasing marginal return technologies mechanically result in the optimality of 'managing with style'. 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.

4 Optimal allocation of organizational attention

The benchmark studied in the previous section 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 posteriors $\widehat{\boldsymbol{\theta}} = \left(\widehat{\theta}_1, \widehat{\theta}_2\right)$ as given. In Section 5 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) for given posteriors $\widehat{\boldsymbol{\theta}}$. Given the objective function (1) simple manipulations yield

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

Without loss of generality, we focus on characterizing equilibria in linear strategies, where $a_{Mi} = \alpha_i \widehat{\theta}_i$. The problem for the manager is then equivalent to choosing $(p_1, p_2, \alpha_1, \alpha_2)$ in order to maximize expected profits

$$
E(\pi|\widehat{\boldsymbol{\theta}}) = \sum_{i \in \{1,2\}} \left[\widehat{\theta}_i^2 - (1 - \alpha_i)^2 \widehat{\theta}_i^2 - \beta(1 - p_i)\alpha_i^2 \widehat{\theta}_i^2 \right] \tag{9}
$$

subject to the organizational attention constraint (4).

Direct inspection of (9) 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 $\widehat{\theta}_i$. Formally, taking p_1 and p_2 as given and maximizing (9) with respect to α_1 and α_2 yields

$$
\alpha_1 = \frac{1}{1 + \beta(1 - p_1)} \quad \text{and} \quad \alpha_2 = \frac{1}{1 + \beta(1 - p_2)}. \tag{10}
$$

Hence, the larger is p_i, the more adaptive is the manager to the posterior $\widehat{\theta}_i$.

Consider next the optimal choice of p_i and turn to organizational attention constraint (4). Notice that this constraint exhibits decreasing marginal returns to attention: If $p_1 > p_2$ then any increase in p_1 must be compensated by an even larger decrease in p_2,

$$\frac{dp_2}{dp_1} = -\frac{1 - p_2}{1 - p_1} < -1 \quad \text{whenever} \quad p_1 > p_2$$

Assume now that $\widehat{\theta}_1^2 > \widehat{\theta}_2^2$, then whenever organizational attention is *not scarce* (so that p is close to 1) then the corner solution $(p_1, p_2) = (p, 0)$ cannot be optimal. Indeed, substituting (10) and (4) into (9), and taking the derivative of expected profits with respect to p_1 yields

$$\frac{dE[\pi]}{dp_1} = \beta \left(\alpha_1 \widehat{\theta}_1 \right)^2 - \beta \left(\alpha_2 \widehat{\theta}_2 \right)^2 \left(-\frac{dp_2}{dp_1} \right), \tag{11}$$

where α_1 and α_2 are given by (10). When p is close to 1, $-dp_2/dp_1$ in (11) grows without bound so that profitability can be improved by redirecting attention to task 2 away from task 1. In contrast, even when $\widehat{\theta}_1 = \widehat{\theta}_2$, it is easy to verify that when $\beta > 1$ (implementation is important), the corner solution $(p_1, p_2) = (p, 0)$ will be a local optimum provided that organizational attention is scarce so that p is small.

The following proposition shows that directing all organizational attention to the task which faces the largest shock is indeed optimal provided that organizational attention is scarce and organizational implementation β sufficiently important. From (10), the manager is then also disproportionately responsive to the task facing the largest shock.

Proposition 2 *Suppose $\beta > 1$. There exists $\bar{p}(\beta) > 0$ with $\bar{p}'(\beta) > 0$ such that whenever $p < \bar{p}(\beta)$, the manager directs all organizational attention to one strategic choice:*

$$
\begin{aligned}
(p_1^*, p_2^*) &= (p, 0) \qquad \textit{if} \qquad \widehat{\theta}_1^2 > \widehat{\theta}_2^2 \\
(p_1^*, p_2^*) &= (0, p) \qquad \textit{if} \qquad \widehat{\theta}_1^2 < \widehat{\theta}_2^2
\end{aligned}
$$

Intuitively, when internal alignment is important, 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. Internal alignment on the other strategic choice is then achieved by largely giving up on external alignment regarding that task – in other words,

16

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. Finally, 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.

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 $\theta_i^2 > \theta_{-i}^2$ for $i \neq k$ then the manager only communicates a_{Mi}, that is $p_i = p$ and $p_{-i} = 0$, and

$$a_{Mi} = \frac{\widehat{\theta}_i}{1 + \beta(1-p)} \qquad \text{and} \qquad a_{M-i} = \frac{\widehat{\theta}_{-i}}{1+\beta} \tag{12}$$

Notice thus that as argued in subsection 2.3.3, firm behavior is largely determined by the allocation of managerial attention. When the manager allocates more attention to task i the posterior $\widehat{\theta}_i$ will have an (ex ante) distribution with higher variance. Indeed, note that, unconditionally,

$$\widehat{\theta}_i \sim \mathcal{N}\left(0, q_i \sigma_\theta^2\right), \tag{13}$$

where q_i is given in (7). If $q_i > q_{-i}$ then $\widehat{\theta}_i$ is more correlated with θ_i. In addition, as we will show, if $q_i > q_{-i}$, then it is more likely that $\left|\widehat{\theta}_i\right| > \left|\widehat{\theta}_{-i}\right|$ so that the manager is more likely to communicate about task i than about $-i$. In the latter case, the manager can afford to choose a strategic action a_{Mi} which is more responsive to $\widehat{\theta}_i$ in the knowledge that employees will be good at implementing such a choice. The allocation of attention thus influences how adaptive a strategic choice is to its 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.

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 $|\theta_i| = |\theta_{-i}|$. Hence, even when $p > \bar{p}(\beta)$, the manager may communicate only about task i provided that $|\theta_i|$ is sufficiently larger than $|\theta_{-i}|$.

Comparison with Dessein, Galeotti and Santos (2014).   The result in Proposition 2 is similar to the main result in Dessein, Galeotti and Santos (2014, DGS henceforth). One key difference is that in DGS, there are two managers who are each in charge of one task and who each only observe the shock affecting their task. As a result, organizational attention is allocated by an organizational designer before the realization of the task-specific shocks. While in DGS it is optimal to focus all organizational attention on one task under the same condition as in A1, this allocation of organizational attention cannot be made contingent on the realization of θ_1 and θ_2. In contrast, in this paper, there is only one manager who observes both shocks and selects the strategic choices for both tasks. Importantly, the manager in our model directs organizational attention *after* observing the task-specific shocks. As a result, the organization has a priori the ability to shift its focus depending on the realization of organizational shocks.

A second difference with DGS is that the information of the manager is not exogenous, but depends on her allocation of attention and on her expertise. In contrast, the managers in DGS perfectly observe their task-specific shock. The model in DGS therefore cannot speak to the issue of managerial style and attention, which is the subject of the present paper.

Benchmark with perfect managerial information.   Before studying the optimal allocation of managerial attention, it is useful to consider briefly a second benchmark where the managerial attention is unconstrained so that the she perfectly learns θ_1 and θ_2. Given A1, the manager then directs organizational attention to the largest 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 tasks.

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. Similarly,

our benchmark in Section 3 showed that if there is no scarcity in organizational attention, scarcity of managerial attention did not result in any systemic bias. The next section shows that it is the interaction of scarce managerial attention and scarce organizational attention what yields systematic organizational biases and managerial styles; both ingredients are needed for managerial styles to arise.

5   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 allocates her scarce *managerial attention*? We proceed by distinguishing between the case of a generalist managers and a manager with superior expertise in one task versus the other.

A generalist manager is one with equal expertise about both tasks, that is $T_1 = T_2 = T$. T is then related to the amount of exogenous information available to the manager. A large T, 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 T 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. Should the generalist manager divide her attention equally among both tasks, or should she focus her attention on one (randomly chosen) task? To analyze this question we assume that

$$t_i \in \{0, \tau, 2\tau\} \qquad \text{with} \qquad t_1 + t_2 \leq 2\tau. \tag{14}$$

The manager decides thus whether to obtain a signal about each task, or to obtain two signals about one task and none about the other task. Recall that in our benchmark, where only external alignment matters, the manager optimally chooses to obtain a signal about each task.

Consider next a specialist manager, one that has more expertise about, say, task 1 than task 2, that is, $T_1 > T_2$. Here, by default, the manager will obtain better exogenous information about task 1 than about task 2. The question now is whether the manager should focus her scarce attention on the margin to the task she is more familiar with, task 1, or, in contrast, compensate for her lack of expertise in task 2 and allocate the marginal unit of attention to it.

19

We assume now that

$$t_i \in \{0, \tau\} \qquad \text{with} \qquad t_1 + t_2 \leq \tau. \tag{15}$$

The manager thus needs to decide whether to draw an additional signal about task 1 or 2. Recall that in the benchmark considered in Section 3, where only external alignment mattered, the manager optimally chooses to obtain a signal about the task in which she is not an expert, which is task 2 in this case.

5.1 Expected profits

Before analyzing the optimal allocation of attention we first develop the expected profit function for a given strategic focus/allocation of attention $\mathbf{t} = (t_1, t_2)$. For this purpose, we first express expected profits for given posteriors $\widehat{\boldsymbol{\theta}} = (\theta_1, \theta_2)$ and communication choices (p_1, p_2), and subsequently take expectations over posteriors for an allocation of managerial attention.

Given (10), the expected profits associated with task i conditional on $\widehat{\boldsymbol{\theta}}$ and communication choices (p_1, p_2), (9) are given by

$$
\begin{aligned}
E\left[\pi_i | \widehat{\theta}_i\right] &= \widehat{\theta}_i^2 - \left(\frac{\beta(1 - p_i)}{1 + \beta(1 - p_i)}\right)^2 \widehat{\theta}_i^2 - \frac{\beta(1 - p_i)}{(1 + \beta(1 - p_i))^2} \widehat{\theta}_i^2 \\
&= \frac{\widehat{\theta}_i^2}{1 + \beta(1 - p_i)}
\end{aligned}
$$

Given our assumption on scarce organizational attention, A1, and given that $\widehat{\theta}_i^2 > \widehat{\theta}_{-i}^2$, the manager only communicates about task i. Expected profits of the organization conditional on posteriors $\widehat{\boldsymbol{\theta}} = \left(\widehat{\theta}_1, \widehat{\theta}_2\right)$ then equal

$$E\left(\pi | \widehat{\boldsymbol{\theta}}\right) = \frac{\widehat{\theta}_i^2}{1 + \beta(1 - p)} + \frac{\widehat{\theta}_{-i}^2}{1 + \beta}$$

Note that whenever β is large, the expected profitability conditional on $\widehat{\theta}_i$ of task i is much greater than that of task $-i$, even when $\theta_i^2 \sim \theta_{-i}^2$, as the manager only communicates about task i. Given a strategic allocation of $\mathbf{t} = (t_1, t_2)$ with $t_i \in \{0, \tau, 2\tau\}$, the (unconditional)

20

expected profits are given by

$$\Pi\left(q_1, q_2\right) \equiv E\left[\pi\right] = \Pr\left(\widehat{\theta}_1^2 \geq \widehat{\theta}_2^2\right) E\left[\frac{\widehat{\theta}_1^2}{1 + \beta(1-p)} + \frac{\widehat{\theta}_2^2}{1+\beta} \ \middle| \ \widehat{\theta}_1^2 \geq \widehat{\theta}_2^2\right] \quad (16)$$

$$+ \ \Pr\left(\widehat{\theta}_1^2 < \widehat{\theta}_2^2\right) E\left[\frac{\widehat{\theta}_1^2}{1+\beta} + \frac{\widehat{\theta}_2^2}{1 + \beta(1-p)} \ \middle| \ \widehat{\theta}_1^2 < \widehat{\theta}_2^2\right]$$

where, as seen in (13), the unconditional distribution of $\widehat{\theta}_i$ is given by $\widehat{\theta}_i \sim \mathcal{N}\left(0, q(t_i + T_i)\sigma_\theta^2\right)$ and where, recall, $q_1 \equiv q(t_1 + T_1)$ and $q_2 \equiv q(t_2 + T_2)$ capture the precision of the manager's information about shocks θ_1 and θ_2, respectively (see expression (7)).

Let $F(x, y)$ denote the normal c.d.f. of random variable x with mean 0 and variance y. Then given (13), we can rewrite the expected profits as:[9]

$$\Pi\left(q_1, q_2\right)$$

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

$$+ \ 4\int_0^{+\infty}\left[\int_0^{\widehat{\theta}_1} \frac{\widehat{\theta}_2^2}{1+\beta}dF\left(\widehat{\theta}_2, q_2\sigma_\theta^2\right) + \int_{\widehat{\theta}_1}^{+\infty} \frac{\widehat{\theta}_2^2}{1+\beta(1-p)}dF\left(\widehat{\theta}_2, q_2\sigma_\theta^2\right)\right]dF\left(\widehat{\theta}_1, q_1\sigma_\theta^2\right)$$

Having obtained a tractable expression for expected profits conditional on a strategic focus $\mathbf{t} = (t_1, t_2)$, we are now ready to analyze the optimal allocation of attention for both the generalist and the specialist manager.

5.2 Allocation of attention by a generalist manager

Recall that the generalist manager has equal expertise on both tasks, $T_1 = T_2 = T$. We ask whether and when such a generalist manager should split her attention equally among both tasks as in the benchmark case or, instead, focus all her attention on one particular task.

The manager faces the following apparent trade-off. On the one hand, when the manager spreads out her attention evenly, she risks that her attention is spread out too thinly. While she may be better at evaluating which task faces the largest opportunity, she may lack the

---

[9]The integral further uses the fact that the probability density of $\left|\widehat{\theta}_i\right|$ is twice the probability density of $\widehat{\theta}_i$ $= \left|\widehat{\theta}_i\right|$, as reflected in the number 4 in front of the expression.

knowledge to take full advantage of that opportunity. On the other hand, if she focuses all her attention on one task, she risks being 'blind-sided' on the other. For example, if the manager focuses on task 1, she may fail to notice a large shock on the second task and thus adopt the "wrong" strategy (ex post). Even when she correctly identifies θ_2 as being the largest shock, her choice for task 2 is likely to be off base given the poor precision of her posterior $\widehat{\theta}_2$. Importantly, because there are decreasing marginal returns to attention, by focussing all her attention on task 1, the additional knowledge learned about task 1 is less than the information lost on task 2, compared to a setting where attention is spread out evenly.

We now argue that in the absence of decreasing marginal returns to attention, the risks of being blind-sided are outweighed by the risks associated with a lack of focus. From profit expression (16), if the manager were to mainly communicate about task 1, it would be optimal for her to devote more attention to task 1 than task 2 since profits are then more sensitive to $\widehat{\theta}_1^2$ than to $\widehat{\theta}_2^2$. Indeed, from (13) the variance of the posterior $\widehat{\theta}_i$ equals $q_i \sigma_\theta^2$ and, hence, is increasing in the attention t_i devoted to task i. Intuitively, information is more valuable when it pertains to the task about which the manager communicates to the organization. Information about the other task is largely ignored – and, hence, wasted – whenever internal alignment is important (β is large). Crucially, which task the manager communicates about not only depends on the realization of θ_1 and θ_2, but also on the managerial attention t_1 and t_2 devoted to task 1 and task 2. In particular, a manager is more likely to perceive a shock as more important when he devotes more attention to learning about this task. From (13), if a generalist manager devotes more attention to task 1 than to task 2, it is more likely that $\widehat{\theta}_1^2 > \widehat{\theta}_2^2$ and, hence, it is optimal for the manager to mainly communicate about task 1 :

$$\Pr\left(\widehat{\theta}_1^2 > \widehat{\theta}_2^2\right) > \Pr\left(\widehat{\theta}_1^2 < \widehat{\theta}_2^2\right) \iff T_1 + t_1 > T_2 + t_2 \tag{17}$$

It follows that devoting more attention to a task and mainly communicating about the same task are complementary choices for a manager.

If there are constant marginal returns to attention, for example if $q_i = \delta t_i$, then one can show that, because of the above complementarity, it is optimal to focus all attention on one task up to the point where the shock is perfectly observed, that is $q_i = 1$. With decreasing marginal returns to attention, as assumed in this paper, whether or not focus is optimal depends on the quality of the exogenous information T and the scarcity of managerial attention. The

following proposition characterizes when managerial attention should optimally be focused or balanced, where we restrict the analysis to the case where managerial attention is either abundant (τ large) or very scarce (τ small):

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

(a) *When managerial attention is not scarce (τ sufficiently large), balanced attention is optimal:*

$$(t_1^*, t_2^*) = (\tau, \tau) \tag{18}$$

(b) *When managerial attention is scarce (τ sufficiently small), focused attention (managing with style) is optimal when the exogenous signal is not very informative: There exists a $\overline{T} > 0$ such that*

$$(t_1^*, t_2^*) \in \{(2\tau, 0), (0, 2\tau)\} \quad \Longleftrightarrow \quad \lambda T < \overline{T}. \tag{19}$$

(c) *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.*

Intuitively when managerial attention is not scarce (as in (a)), the complementarity between (i) how much attention a task receives and (ii) how likely a task is to be communicated to the organization, is overwhelmed by decreasing marginal returns to devoting attention to the same task. Even when splitting attention, the manager then learns both shocks with great precision and there is little to be gained by focussing all attention on one task. In contrast, by focussing all attention on one task, the manager runs a high risk of being blindsided. The manager therefore optimally devotes equal attention to learning θ_1 and θ_2 and almost always communicates about the 'right' strategy.

Instead, when managerial attention is scarce and the exogenous signal the manager receives is not too informative (as in (b)), decreasing marginal returns to attention are only a minor concern and focused managerial attention is optimal. Note that while the manager then

23

optimally focuses attention on one task, ex post she may be forced to communicate and implement a strategy about which she has 'poor visibility'. Ex post, organizational attention is then sometimes directed to a different task than the one on which managerial attention was focused. The better the exogenous information (the larger is T) or the less complex is the environment (the larger is λ), the larger the probability of such an 'attention reversal', which explains why focused attention is less likely to be optimal in environments with good exogenous information.

Finally, from (c), when the cost of internal alignment is large (β is large), focused attention is more likely to be optimal. The manager is then 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" as it is never communicated to the workers. In contrast, by devoting all attention to one task, the manager is very good at responding to the shock affecting this task and she is very likely to communicate about this task (provided exogenous information is not too informative). In other words, information is rarely wasted. In the other case where internal alignment is not very important (β small), however, the manager wants to be responsive to both shocks, even when she only communicates about one task, so that learning both shocks is much more valuable.

Proposition 3 only discusses very large and very small values of τ. Simulations confirm that our results generalize to intermediate values of τ. Figure 1 Panel A shows when focussed attention is preferred over balanced attention in the parameter range (τ, β) with $2\tau \in [0.5, 65]$ and $\beta \in [2, 10]$ and considers two qualities of exogenous information: $T = 0.1$ and $T = 0.15$. We further assume $p = 0.75$ and $\lambda = 1$. Assumption $A1$ is satisfied for these parameter values. As predicted by Proposition 3 (c), focused managerial attention is more likely to be optimal when internal alignment is important (β is larger). Importantly, as suggested by Proposition 3, focused managerial attention is more likely to be be optimal when managerial attention is scarce (τ is smaller) and/or the exogenous information is poorer ($T = 0.1$ rather than $T = 0.15$).

In Figure 1 Panel B we characterize both the allocation managerial attention allocation and allocation of organizational attention in the parameter range (τ, β) with $2\tau \in [0.5, 65]$ and $\beta \in [0, 10]$. In Region III the importance of internal alignment (β) is sufficiently high so

Figure 1: *Panel A: Regions for which managing with style is optimal for a generalist manager in the space $(2\tau, \beta)$ for $T = 0.1$ or $T = 0.15$ and for $p = 0.75$. Managerial attention is optimally focused on one task above the respective lines, whereas attention is equally split between both task below. Panel B: Regions for which managerial attention is focused on one task (Region III) or balanced (Regions I or II) and the regions for which organizational attention is focused on one task (Regions II and III) or more balanced (Region I). In this example $T = 0.1$ and $p = 0.75$.*
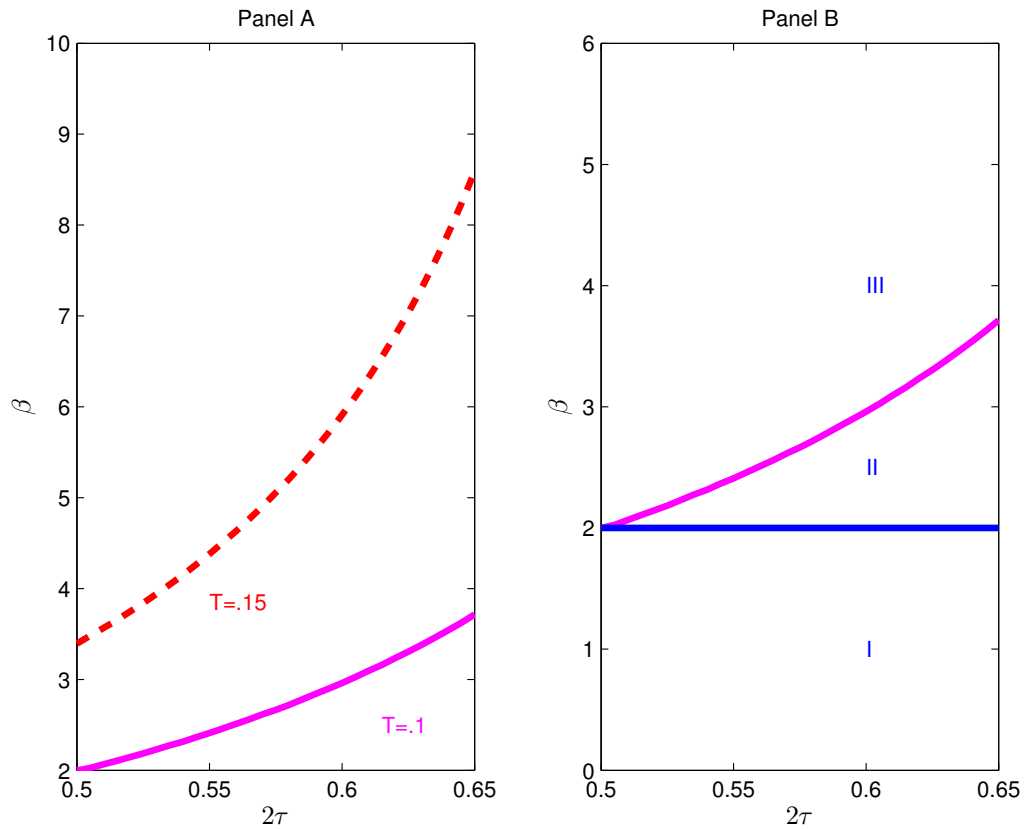
the manager focuses both her managerial attention on one task and directs the attention of the organization to the task with the largest perceived shock (typically the same task she initially focused her own attention on). In Region II, where the importance of internal alignment is intermediate, the manager still directs the attention of the organization to the task with the largest perceived shock, but her own managerial attention in now balanced: she learns about

25

both θ_1 and θ_2 but directs workers' attention to only one task. In Region I, when the need for internal alignment is limited, Assumption A1 is not met and the manager communicates about both shocks to the organization whenever they are perceived to be similar in size. Managerial attention is then again equally divided among both tasks, and organizational attention is biased towards the largest perceived shock but often both shocks will receive positive organizational attention.

5.3   Allocation of attention by a specialist manager

The previous section analyzed the case of a generalist manager, for whom $T_1 = T_2$. Here instead we assume that, in the absence of any endogenous allocation of attention, the manager has some specialized skills that gives him access to more information about one task than about the other, say $T_1 > T_2$.

In this case, will the specialist manager allocate the marginal unit of attention to the task on which she is an expert or would she try to "balance things out" in order not to be blindsided? To answer this question we assume that the manager can acquire at most one endogenous signal of quality τ. She must devote her attention to either task 1 or to task $2 : t_1 + t_2 \leq \tau$ with $t_i \in \{0, \tau\}$. We further assume that τ is small, so the question is how will the manager allocate the *marginal* unit of attention?

We capture the relative specialization of a manager in task 1 by the *relative specialization ratio*

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

where ρ captures how much better the manager is at observing shock θ_1 rather than shock θ_2. Notice that the relative specialization ratio is defined prior to the allocation of the marginal unit of attention τ. 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. Without loss of generality, we will assume that $\rho > 1$ so that the manager is better informed about task 1 than 2, absent the possibility of allocating any additional attention. As discussed above we argue that $T_1 > T_2$ reflects managerial skills specialized in matters relating to task 1. For instance, the manager may have had a career in operations and thus it is easier for him to analyze innovations in business processes and scheduling whereas he may be less well informed about other areas

such as a marketing or product development.

The following proposition characterizes the allocation of a marginal unit of attention τ as a function of the absolute level of task knowledge of the manager, $q(T_1)$, keeping the relative specialization ratio $\rho = q(T_1)/q(T_2)$ fixed. It shows how, if the exogenous level of task knowledge of a manager is limited, then on the margin, attention is better allocated to the task on which the manager has an informational advantage.

Proposition 4 (Specialist Manager) *Assume $A1$, and consider a specialist manager with $(T_1 > T_2)$ and $t_1 + t_2 \leq \tau$ with $t_i \in \{0, \tau\}$. Fixing $\rho > 1$, there exists a $T^*(\rho) > 0$ such that $(t_1^*, t_2^*) = (\tau, 0) \iff \lambda T_1 < T^*(\rho)$.*

Proposition 4 states that, keeping the level of relative specialization constant, managing with style is optimal if and only if the absolute level of task knowledge is below critical level. The same intuitions developed in the case of a generalist manager help explain the above result. As long as τ is small and since $T_1 > T_2$[10] the specialist manager is more likely to communicate about task 1 (her task of expertise) than about task 2 (see expression (17)). As shown in expression (16), profits are then more sensitive to $\widehat{\theta}_1^2$ than to $\widehat{\theta}_2^2$ and – for a given communication strategy – it is then optimal for the manager to devote the marginal unit of attention to task 1. Indeed, from (13) the variance of the posterior $\widehat{\theta}_1$ equals $q(t_1 + T_1)\sigma_\theta^2$ and, hence, is increasing in the attention t_1 devoted to task 1. In turn, this makes it even more likely that the manager will communicate about task 1. This is the basic complementarity which makes it optimal for a specialist manager to devote attention to the task on which she is already an expert. The only countervailing force is that the marginal returns to devoting attention to any given task are decreasing on the expertise that the manager already has on that task, but this effect is second order when τ is small.

Figure 2 illustrates Proposition 4 for $\lambda = 1$, $p = 0.75$ and $\beta = 5$. The upward sloping (blue) lines indicate values of (T_1, T_2) for which the relative specialization ratio ρ is constant, with $\rho \in \{1, 1.5, 2, 3, 5\}$ (the plot is obviously symmetric around the 45 degree line). For each of those lines, as shown in Proposition 4, there is a unique T^* for which if $T_1 < T^*$ the marginal unit of attention τ is allocated to task 1 and above which τ goes to task 2. The

---

[10]That is, a τ such that even if this marginal unit of attention is allocated to task 2, it remains the case that $T_1 > T_2 + \tau$.

Figure 2: *This figure plots the loci of points in (T_1, T_2) for $T_1 > T_2$ (the downward sloping black curve) below which the agent prefers to allocate the marginal unit of attention to task 1, whereas above they allocate the marginal unit of attention to task 2. The upward sloping (blue) lines show iso-ρ, where ρ is the relative specialization ratio for $\rho \in \{1, 1.5, 2, 3, 5\}$.*
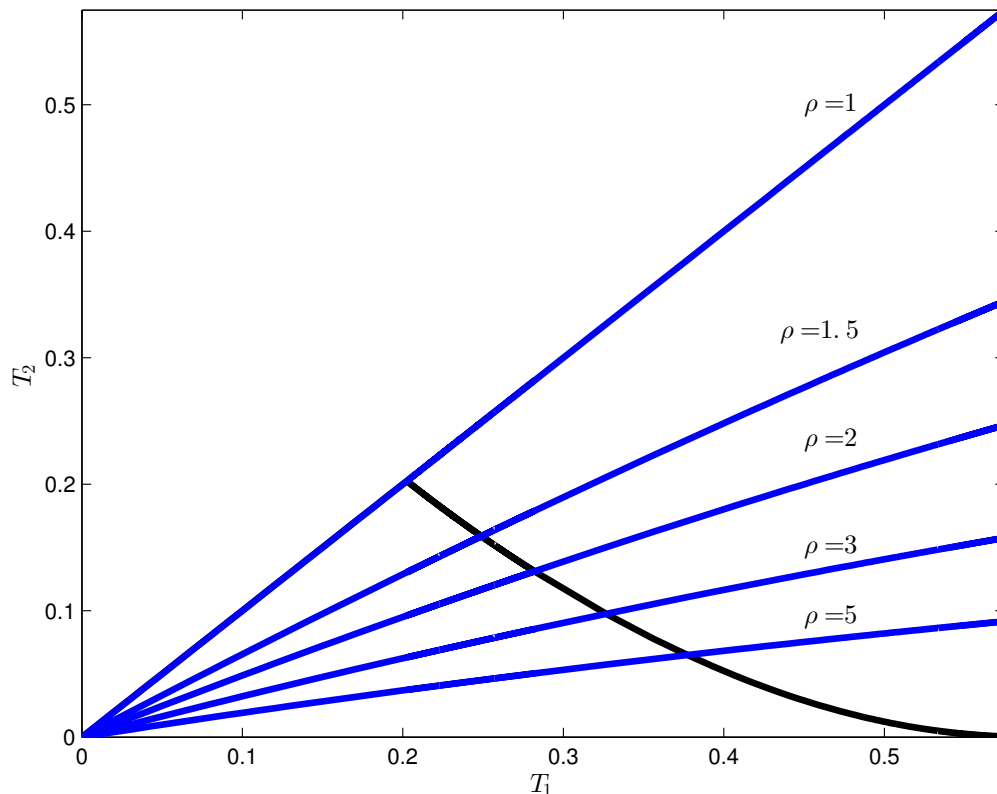
downward sloping (black) line links all those points: For those points of expertise (T_1, T_2) between the downward sloping curve and the 45 degree line the optimal allocation of the marginal unit of attention is $(t_1^*, t_2^*) = (\tau, 0)$ whereas it is $(t_1^*, t_2^*) = (0, \tau)$ above the downward sloping curve.

5.4 Specialists versus Generalists: Endogenous managerial expertise

So far, we have taken managerial expertise as given and endogenized the allocation of managerial attention. Managers though are appointed by boards 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 where $T_L \geq 0$ is the minimum expertise of a manager in any area (for example, some information may become available to any manager).

The endogenous allocation of attention can be easily incorporated in this set-up. 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 $t_1 + t_2 \leq 2\tau$. Without loss of generality 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 $\tau > 0$ paired with an expertise budget Z is formally equivalent to an attention budget $\tau = 0$ paired with an expertise budget $Z + 2\tau$. Again without loss of generality, we therefore posit that $\tau = 0$ so that $T_i = \tilde{T}_i$.

Consider again the example introduced at the end of section 5.3, in which $T_1 > T_2$ and $(\beta, p, \lambda) = (5, .75, 1)$, and turn to Figure 3 where now, instead of 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

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

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 has a positive second derivative in Figure 3, 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 $T_1 = T_2$).

Figure 3: *This figure plots the loci of points in (T_1, T_2) for $T_1 > T_2$ (the downward sloping black curve) below which the agent prefers to allocate the marginal unit of attention to task 1, whereas above they allocate the marginal unit of attention to task 2. The straight downward (red) lines represent total expertise budgets for five different values, Z_1 to Z_5.*
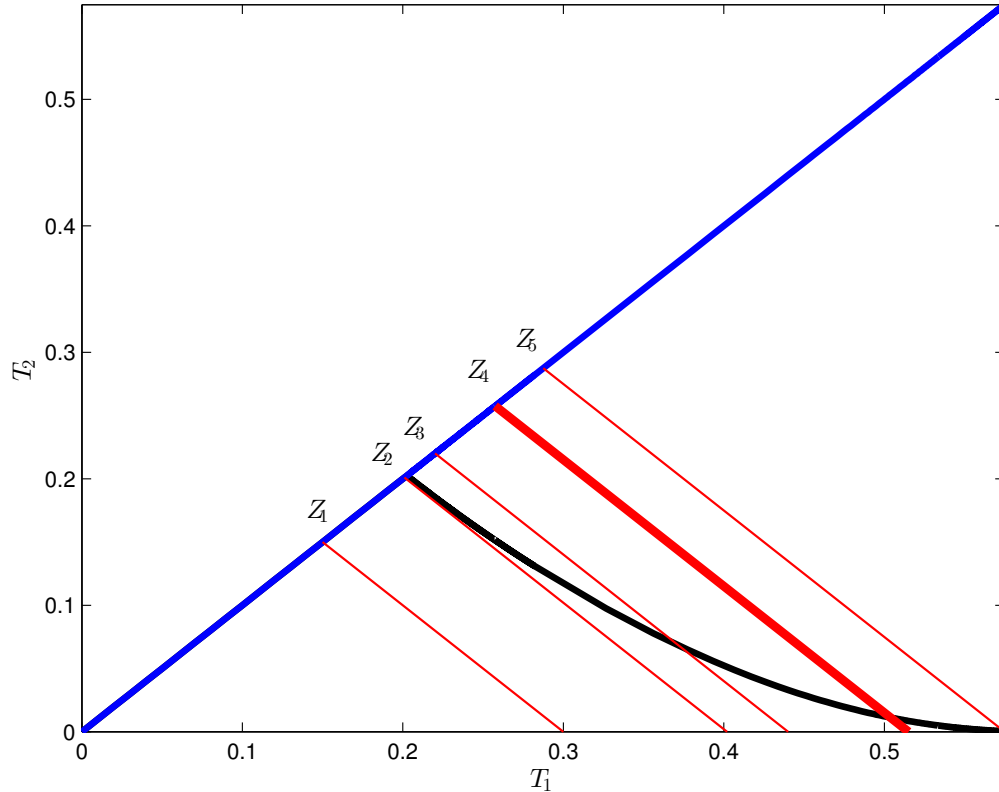
Obviously, extreme specialists are preferred for the budgets Z_1 and Z_2 corresponding to the first two straight (red) lines, whereas complete generalists are preferred for the budget Z_5 corresponding to the last budget line. More generally, one can show that extreme specialists are preferred over complete 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 mimimum attention a manager must devote

to either task. For expertise budgets $Z < Z_2 \approx 0.402$ or budgets $Z > Z_4$ the mimimal 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. Indeed, 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 $(Z_2 - T_L, T_L)$.

In sum, as in the previous subsections, our analysis suggests that generalist managers are prefered when expertise is in large supply – that is in environments where even generalist will typically have a precises estimate of both task-specific shocks. When expertise is in short supply, however, that is for smaller total expertise budgets, boards prefer hiring specialist managers.

6 Discussion

In this paper, we have shown how scarcity of organizational attention and the need for internal alignment induces managers to communicate 'simple' or 'narrow' strategies, focused on what are perceived to be the largest opportunities for the organization. In turn, when the environment is complex and uncertain, the need for simple strategies induces a manager to focus her scarce managerial attention on a few aspects (or functional areas) of this environment. Importantly, a manager's expertise then determines what will be the focus of her scarce attention. We now discuss a few implications of those findings.

6.1 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 scarce attention in a way which amplifies initial

differences in expertise. In particular, two seemingly similar managers faced with the same economic environment may nevertheless process information in a very different way – as they optimally focus their attention on different aspects – resulting in different firm strategies and outcomes. Empirical studies may therefore understate the impact of managerial characteristics 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.

6.2   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). 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 paper provides an answer by showing how 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 criterium of choice for a board of directors. Instead, boards are likely to choose managers based on leadership ability, general cognitive abililty, 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 firm performance.[11]

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 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.

6.3 Why do managers matter? Management 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)).

---

[11]A 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.

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.[12] 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 truly appreciate 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 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 may differ 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. So with strategic choices, the difference between ex ante and ex post is key, while this is less so for the case of management practices. This may create a challenge for empirical work that tries to distinguish between good and bad management.

---

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

6.4 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 environment 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,[13] 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 suround 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 *intepretation* (how they attach meaning to what they see and hear)."[14]

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 or 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,

---

[13]See Finkelstein et al. (2009), Chapter 2, for an overview.

[14]Finkelstein et al. (2009), p.46.

such as the need for organizational alignment around simple strategies and environmental factors, such as the amount of uncertainty and the scarcity of attention.

References

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 (2013). "Organizing to Adapt and Compete." American Economic Journal: Microeconomics. Forthcoming.

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.

37

Cyert, Richard, and James March (1963). A Behavioral Theory of the Firm. Oxford: Blackwell.

Calvo-Armengol, Antonio, Joan de Marti and Andrea Prat (2011). "Communicaton and Influence." manuscript, London School of Economics.

Dessein, Wouter, Andrea Galeotti, and Tano Santos (2014). "Rational inattention and organizational focus," manuscript, Columbia Business School.

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.

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/doifinder/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.

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.

March, James, and Herbert Simon (1958). Organizations, Jonh 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.

39

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.

7 Appendix

7.1 Proof of Proposition 3

7.1.1 Preliminaries

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

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

We can make a simple change of variable $\varphi_1 \equiv \widehat{\theta}_1 / \sqrt{q_1}$ and $\varphi_2 \equiv \widehat{\theta}_1 / \sqrt{q_2}$, 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

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

or still

$$
\begin{aligned}
\Pi\left(q_1, q_2\right) &= \frac{4}{1+(1-p) \beta} \int_0^{+\infty}\left[\int_{\sqrt{\frac{q_2}{q_1}} \varphi_k}^{+\infty} q_1 \varphi_1^2 dF(\varphi_1)+\int_{\sqrt{\frac{q_1}{q_2}} \varphi_k}^{+\infty} q_2 \varphi_2^2 dF(\varphi_2)\right] dF(\varphi_k) \quad (20) \\
&+ \frac{4}{1+\beta} \int_0^{+\infty}\left[\int_0^{\sqrt{\frac{q_2}{q_1}} \varphi_k} q_1 \varphi_1^2 dF(\varphi_1)+\int_0^{\sqrt{\frac{q_1}{q_2}} \varphi_k} q_2 \varphi_2^2 dF(\varphi_2)\right] dF(\varphi_k), \quad (21)
\end{aligned}
$$

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:

$$
q_1 = q_2 \equiv q = 1 - e^{-\lambda(\tau+T)} \tag{22}
$$

and the profit expression (see (20) and (21)) simplifies to

$$\Pi\left(q, q\right) = 8q \int_0^\infty \left(\frac{1}{1+\beta} \int_0^{\varphi_k} \varphi_i^2 dF(\varphi_i) + \frac{1}{1+\beta(1-p)} \int_{\varphi_k}^\infty \varphi_i^2 dF(\varphi_i) \right) dF(\varphi_k), \quad (23)$$

where both φ_i and φ_k are both normally distributed random variables with mean 0 and variance σ_θ^2. (23) has the following closed form solution:[15]

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

which simplifies in turn to

$$\Pi\left(q, q\right) = 2qC \qquad \text{with} \qquad C \equiv \left(\frac{1}{1+\beta} \right) \left[1 + \frac{\beta p}{1+\beta(1-p)} \frac{\pi+2}{2\pi} \right] \sigma_\theta^2. \qquad (24)$$

<u>Profits under Focused Attention:</u> If the manager focuses all her attention on one task, say, task 1,

$$q_1 = q\left(2\tau + T\right) = 1 - e^{-\lambda(2\tau+T)} > q_2 = q(T) = 1 - e^{-\lambda T} \qquad (25)$$

and we can rewrite expected profits as

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

or still

$$\Pi\left(q_1, q_2\right) = \left(q_1 + q_2\right) C + D \int_0^\infty \left[B_1\left(q_1, q_2\right) - B_2\left(q_1, q_2\right) \right] dF\left(\varphi_k\right) \qquad (26)$$

where

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

and

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

---

[15]This closed form solution was found using a Mathematica routine.

with

$$q_1 = 1 - \exp\left(-\lambda\left(T + 2\tau\right)\right) \quad \text{and} \quad q_2 = 1 - \exp\left(-\lambda T\right) \tag{27}$$

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

First notice that

$$\frac{\partial}{\partial \tau} \sqrt{\frac{q_1}{q_2}} = \frac{\lambda}{(q_1 q_2)^{\frac{1}{2}}} \exp\left(-\lambda\left(T + 2\tau\right)\right) \tag{28}$$

Hence

$$\frac{\partial B_2 \left(q_1, q_2\right)}{\partial \tau} = \lambda \left(\frac{q_1}{q_2}\right)^{\frac{1}{2}} \varphi_2^3 f\left(\sqrt{\frac{q_1}{q_2}}\varphi_2\right) \exp\left(-\lambda\left(T + 2\tau\right)\right) \tag{29}$$

And then

$$\frac{\partial^2 B_2}{\partial \tau^2} = \frac{\lambda^2}{(q_1 q_2)^{\frac{1}{2}}} \varphi_2^3 f\left(\sqrt{\frac{q_1}{q_2}}\varphi_2\right) \exp\left[-2\lambda\left(T + 2\tau\right)\right] \tag{30}$$

$$+ \frac{\lambda^2}{q_2} \varphi_2^4 f'\left(\sqrt{\frac{q_1}{q_2}}\varphi_2\right) \exp\left[-2\lambda\left(T + 2\tau\right)\right] \tag{31}$$

$$- 2\lambda^2 \left(\frac{q_1}{q_2}\right)^{\frac{1}{2}} \varphi_2^3 f\left(\sqrt{\frac{q_1}{q_2}}\varphi_2\right) \exp\left(-\lambda\left(T + 2\tau\right)\right) \tag{32}$$

It follows that

$$\frac{\partial^2 B_2}{\partial \tau^2}|_{\tau = 0} = \frac{\lambda^2}{q_2} \varphi_2^3 f\left(\varphi_2\right) \exp\left(-2\lambda T\right) \tag{33}$$

$$+ \frac{\lambda^2}{q_2} \varphi_2^4 f'\left(\varphi_2\right) \exp\left(-2\lambda T\right) - 2\lambda^2 \varphi_2^3 f\left(\varphi_2\right) \exp\left(-\lambda T\right) \tag{34}$$

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

First notice that

$$\frac{\partial}{\partial \tau} \sqrt{\frac{q_2}{q_1}} = -\lambda \left(\frac{q_2^{\frac{1}{2}}}{q_1^{\frac{3}{2}}}\right) \exp\left(-\lambda\left(T + 2\tau\right)\right) \tag{35}$$

43

and thus

$$\frac{\partial B_1(q_1, q_2)}{\partial \tau} = \int_{\sqrt{\frac{q_2}{q_1}}\varphi_2}^{\varphi_2} 2\lambda \exp\left(-\lambda\left(T + 2\tau\right)\right) \varphi_1^2 dF\left(\varphi_1\right) \tag{36}$$

$$+ \lambda\left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \varphi_2^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_2\right) \exp\left(-\lambda\left(T + 2\tau\right)\right). \tag{37}$$

Define

$$P\left(q_1, q_2\right) = \lambda\left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \varphi_2^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_2\right) \exp\left(-\lambda\left(T + 2\tau\right)\right) \tag{38}$$

Then

$$\frac{\partial^2 B_1}{\partial \tau^2} = -4\lambda \exp\left(-\lambda\left(T + 2\tau\right)\right) \int_{\sqrt{\frac{q_2}{q_1}}\varphi_2}^{\varphi_2} \varphi_1^2 dF\left(\varphi_1\right) \tag{39}$$

$$+ 2\lambda^2 \left(\frac{q_2^{\frac{3}{2}}}{q_1^{\frac{5}{2}}}\right) \exp\left(-2\lambda\left(T + 2\tau\right)\right) \varphi_2^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_2\right) + \frac{\partial P}{\partial \tau} \tag{40}$$

Finally

$$\frac{\partial P}{\partial \tau} = -3\lambda^2 \left(\frac{q_2^{\frac{3}{2}}}{q_1^{\frac{5}{2}}}\right) \varphi_2^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_2\right) \exp\left[-2\lambda\left(T + 2\tau\right)\right] \tag{41}$$

$$- \lambda^2 \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \left(\frac{q_2^{\frac{1}{2}}}{q_1^{\frac{3}{2}}}\right) \varphi_2^4 f'\left(\sqrt{\frac{q_2}{q_1}}\varphi_2\right) \exp\left[-2\lambda\left(T + 2\tau\right)\right] \tag{42}$$

$$- 2\lambda^2 \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \varphi_2^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_2\right) \exp\left[-\lambda\left(T + 2\tau\right)\right] \tag{43}$$

It follows that

$$\frac{\partial^2 B_1}{\partial \tau^2}|\tau = 0 = 2\frac{\lambda^2}{q_2}\varphi_2^3 f\left(\varphi_2\right) \exp\left(-2\lambda T\right) - 3\frac{\lambda^2}{q_2}\varphi_2^3 f\left(\varphi_2\right) \exp\left(-2\lambda T\right)$$

$$- \frac{\lambda^2}{q_2}\varphi_2^4 f'\left(\varphi_2\right) \exp\left(-2\lambda T\right) - 2\lambda^2 \varphi_2^3 f\left(\varphi_2\right) \exp\left(-\lambda T\right)$$

$$= -\left[\frac{\lambda^2}{q_2}\varphi_2^3 f\left(\varphi_2\right) + \frac{\lambda^2}{q_2}\varphi_2^4 f'\left(\varphi_2\right)\right] \exp\left(-2\lambda T\right) - 2\lambda^2 \varphi_2^3 f\left(\varphi_2\right) \exp\left(-\lambda T\right)$$

Second Derivative of Expected Profits

Lemma 1 (Second Derivative of Expected Profits). We have that

$$\frac{\partial^2 \Pi (q, q)}{\partial \tau^2}|\tau = 0 = -2\lambda^2 \exp\left(-\lambda T\right) C \tag{44}$$

and

$$\frac{\partial^2 \Pi (q_1, q_2)}{\partial \tau^2}|\tau = 0 = -4\lambda^2 \exp\left(-\lambda T\right) C + 2 \left(\frac{\lambda^2 \exp(-2\lambda T)}{1 - \exp(-\lambda T)}\right) \left(\frac{1}{4\pi}\right) D \tag{45}$$

Proof of Lemma 1.

Expression (44) follows directly from (24). From (26), we have that

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

where

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

and hence

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

Since

$$\int_0^\infty \varphi_2^3 f\left(\varphi_2\right) dF\left(\varphi_2\right) = \frac{1}{4\pi}$$

and

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

Hence

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

which concludes the proof of Lemma 1. □

45

7.1.2 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 (24) and (26), focused attention is preferred over balanced attention if and only if

$$\Pi(q,q) \quad < \quad \Pi(q_1, q_2) \tag{46}$$

$$\iff \quad 2q - q_1 - q_2 \le \left(\frac{\beta p}{\left[1 + \beta(1-p) + \beta p \frac{\pi+2}{2\pi}\right]\sigma_\theta^2} \right) \int_0^\infty [B_1(q_1, q_2) - B_2(q_1, q_2)]\, dF(\theta_k) \tag{47}$$

where q, q_1 and q_2 are given by (25) and (25). $3(c)$ follows from the observation that the RHS of (47) 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

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

From Lemma 1,

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

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

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

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

which after some trivial manipulations boils down to the solution to

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

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

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

and for $T > \bar{T}$

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

which concludes the proof. $\qquad\square$

7.2 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

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

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,

$$B_1 \equiv B_1\left(T_1 + t_1, T_2 + t_2\right) = \int_{\sqrt{\frac{q_2}{q_1}}\varphi_k}^{\varphi_k} q_1 \varphi_1^2 dF\left(\varphi_1\right)$$

$$B_1 \equiv B_2\left(T_1 + t_1, T_2 + t_2\right) = \int_{\varphi_k}^{\sqrt{\frac{q_1}{q_2}}\varphi_k} q_2 \varphi_2^2 dF\left(\varphi_2\right)$$

Note that

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

$$\frac{\partial B_2\left(T_1 + \tau, T_2\right)}{\partial \tau} = \frac{\lambda}{2} \exp\left(-\lambda\left(T_1 + \tau\right)\right)\left(\frac{q_1}{q_2}\right)^{\frac{1}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_1}{q_2}}\varphi_k\right)$$

and, similarly,

$$\frac{\partial B_1\left(T_1, T_2 + \tau\right)}{\partial \tau} = -\frac{\lambda}{2} \exp\left(-\lambda\left(\tau_2 + \tau\right)\right)\left(\frac{q_2}{q_1}\right)^{\frac{1}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_k\right)$$

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

It follows that

$$\frac{\Pi(T_1 + \tau, T_2)}{\partial \tau} = \lambda \exp\left(-\lambda\left(T_1 + \tau\right)\right) C + D \int_0^\infty \left[\frac{\partial B_1\left(T_1 + \tau, T_2\right)}{\partial \tau} - \frac{\partial B_2\left(T_1 + \tau, T_2\right)}{\partial \tau}\right] dF\left(\varphi_k\right)$$

$$= \lambda \exp\left(-\lambda\left(T_1 + \tau\right)\right) C + D\lambda \exp\left(-\lambda\left(T_1 + \tau\right)\right)$$
$$\times \int_0^\infty \left(\frac{1}{2}\left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_k\right) - \frac{1}{2}\left(\frac{q_1}{q_2}\right)^{\frac{1}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_1}{q_2}}\varphi_k\right) + \int_{\sqrt{\frac{q_2}{q_1}}\varphi_k}^{\varphi_k} \varphi_1^2 dF\left(\varphi_1\right)\right) dF\left(\varphi_k\right)$$

We have that

$$\int_0^\infty \frac{1}{2}\left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_k\right) dF\left(\varphi_k\right) = \int_0^\infty \frac{1}{2}\left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_k\right) f(\varphi_k) d\varphi_k$$

$$= \int_0^\infty \frac{1}{2}\left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \varphi_k^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{\frac{q_2}{q_1}\varphi_k^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\varphi_k^2}{2}} d\varphi_k$$

$$= \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{2}\varphi_k^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{\frac{q_2+q_1}{q_1}\varphi_k^2}{2}} d\varphi_k$$

$$= \frac{1}{\sqrt{2\pi}} \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \sqrt{\frac{q_1}{q_2+q_1}} \int_0^\infty \frac{1}{2}\varphi_k^3 \frac{1}{\sqrt{\frac{q_1}{q_2+q_1}}\sqrt{2\pi}} e^{-\frac{\varphi_k^2}{2\frac{q_1}{q_2+q_1}}} d\varphi_k$$

$$= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{q_1}{q_2+q_1}} \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \int_0^\infty \frac{1}{2}\varphi_k^3 f\left(\varphi_k; 0, \sqrt{\frac{q_1}{q_2+q_1}}\right) d\varphi_k,$$

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

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

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

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

$$= \frac{1}{2\pi} \left(\frac{q_1}{q_2+q_1}\right)^2 \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \sigma_\theta^3$$

Similarly,

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

$$= \frac{1}{2\pi}\left(\frac{q_1}{q_2+q_1}\right)^2 \left(\frac{q_2}{q_1}\right)^{\frac{3}{2}} \sigma_\theta^3$$

It follows that

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

Similarly,

$$\frac{\partial E\left[\pi|T_1,T_2+\tau\right]}{\partial\tau} = \lambda\exp\left(-\lambda\left(T_2+\tau\right)\right)C + D\lambda\exp\left(-\lambda\left(T_2+\tau\right)\right)$$

$$\times \int_0^\infty \left[\frac{1}{2}\left(\frac{q_1}{q_2}\right)^{\frac{3}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_1}{q_2}}\varphi_k\right) - \frac{1}{2}\left(\frac{q_2}{q_1}\right)^{\frac{1}{2}} \varphi_k^3 f\left(\sqrt{\frac{q_2}{q_1}}\varphi_k\right) - \int_{\varphi_k}^{\sqrt{\frac{q_1}{q_2}}\varphi_k} \varphi_2^2 dF(\varphi_2)\right.$$

$$= \lambda\exp\left(-\lambda\left(T_2+\tau\right)\right)C - D\lambda\exp\left(-\lambda\left(T_2+\tau\right)\right)\int_0^\infty \int_{\varphi_k}^{\sqrt{\frac{q_1}{q_2}}\varphi_k} \varphi_2^2 dF(\varphi_2)\, dF(\varphi_k)$$

In sum, we have that

$$\left.\frac{\partial\Pi(T_1+\tau,T_2)}{\partial\tau}\right|_{\tau=0} > \left.\frac{\partial\Pi(T_1,T_2+\tau)}{\partial\tau}\right|_{\tau=0}$$

$$\Leftrightarrow \lambda\exp\left(-\lambda T_1\right)C + D\lambda\exp\left(-\lambda T_1\right)\left[\int_0^\infty \int_{\varphi_k}^{\sqrt{\frac{q_1}{q_2}}\varphi_k} \varphi_1^2 dF(\varphi_1)\, dF(\varphi_k)\right]$$

$$> \lambda\exp\left(-\lambda T_2\right)C - D\lambda\exp\left(-\lambda T_2\right)\left[\int_0^\infty \int_{\sqrt{\frac{q_2}{q_1}}\varphi_k}^{\varphi_k} \varphi_2^2 dF(\varphi_2)\, dF(\varphi_k)\right]$$

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

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

Define

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

Then

$$\frac{\partial \Pi(T_1 + \tau, T_2)}{\partial \tau}\bigg|_{\tau=0} > \frac{\partial \Pi(T_1 + \tau, T_2)}{\partial \tau}\bigg|_{\tau=0}$$

$$\Leftrightarrow \frac{1 - q(T_1)}{1 - q(T_1)/\rho} > \frac{C - D \int_0^\infty \int_{\varphi_k/\sqrt{\rho}}^{\varphi_k} \varphi_2^2 dF(\varphi_2) \, dF(\varphi_k)}{C + D \int_0^\infty \int_{\varphi_k}^{\sqrt{\rho}\varphi_k} \varphi_1^2 dF(\varphi_1) \, dF(\varphi_k)}$$

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. Moreover, there exists a cut-off for q_1^* and T_1^* so that if $q_1 < q_1^*$ or $T_1 < T_1^*$, we have $(t_1^*, t_2^*) = (\tau, 0)$, and for $q_1 > q_1^*$ or or $T_1 > T_1^*$ we have $(t_1^*, t_2^*) = (0, \tau)$. This concludes the proof. $\qquad \square$