Firm Investment and Stakeholder Choices: A Top-Down Theory of Capital Budgeting*

Andres Almazan  Zhaohui Chen  Sheridan Titman
University of Texas  University of Virginia  University of Texas and NBER

18 February 2016

Abstract

This paper develops a top-down model of capital budgeting where privately informed executives make investment choices that convey information to the firm’s stakeholders (e.g., employees). Favorable information in this setting encourages stakeholders to take actions that positively contribute to the firm’s success (e.g., employees work harder). Within this framework we examine how firms may distort their investment choices to influence the transmission of information to stakeholders and show that investment rigidities and overinvestment can arise as optimal investment distortions. The analysis also examines investment distortions in multi-divisional firms and compares such distortions to those in single-division firms.

*For their comments and suggestions we thank Bruno Biais (the editor), Matthieu Bouvard, Adolfo de Motta, David Dicks, Alex Edmans, Chitru Fernando, Paolo Fulgheri, John Hatfield, Neal Galpin, and Maxwell Stinchcombe. We would like to also thank comments from referees as well as seminar participants at AFA-San Diego (2013), Australian National University, Cass Business School, DePaul University, UCLA, Universidad Carlos III, Chinese University of Hong Kong, University of Essex, University of Melbourne, Nanyang Technological University, National University of Singapore, UNC-Chapel Hill, McGill University, University of Notre Dame, University of Oklahoma, University of Pennsylvania, University of Virginia, University of Texas, and the XVIII Finance Forum at CEU (Elche-Spain) and the Finance Down Under 2011 Conference at the University of Melbourne. Corresponding author: andres.almazan@mccombs.utexas.edu.
1 Introduction

The capital investment process has been described (i.e., Brealey, et al. 2014 pp. 246) as a combination of “bottom-up” procedures, where a firm’s individual business units solicit capital from headquarters, and “top-down” procedures where headquarters use their discretion to allocate capital downstream. An extensive literature analyzes the incentive and information considerations that can emerge in “bottom-up” capital allocation processes.\(^1\) Up to now, however, the literature has not considered “top-down” procedures, where better-informed executives at a firm’s headquarters have the discretion to allocate capital to individual business units.\(^2\)

This paper develops a top-down model of capital allocation where privately informed executives make investment choices that convey information to the stakeholders of the firm. The information conveyed by their investment choices is relevant because it influences the actions of the stakeholders, which in turn, affect the likelihood of the firm’s success. Specifically, since higher investment expenditures are associated with better firm prospects, the stakeholders, who receive a share of the output, are willing to expend more effort to improve the probability that the firm will be successful. Within this setting, we examine the optimal design of capital allocation mechanisms. As we show, the process used to make investment choices affects how the information conveyed by the investment choices are interpreted by stakeholders, and this in turn influences how the capital budgeting process is designed.

To formalize these ideas our model considers a firm whose production process requires a capital investment as well as effort exerted by a stakeholder who we identify, for concreteness, as a lower level manager (also referred to as the worker). In particular, we consider

---

\(^1\)The literature has considered how distortions in capital budgeting practices can improve managerial effort (Bernardo, et al. 2001, 2004, 2006), curb managers’ empire building tendencies (Harris and Raviv 1996, 1998, Berkovitch and Israel 2004, and Marino and Matsusaka 2005) or induce managers to reveal their private information. The literature has also considered the trade-offs that arise in the decision to delegate capital budgeting decisions to a better informed agent (Aghion and Tirole 1997 and Burkart, et al. 1997).

\(^2\)Top-down procedures can also be relevant in settings in which headquarters allocate capital after receiving requests for funds from better informed downstream managers. By aggregating the information contained in such requests, headquarters end up acquiring broader information, which helps the firm determine its overall investment expenditures, (e.g., the information provided by one unit has implications for another unit’s investment). See Vayanos (2003) for a model of organizational design that abstracts from incentive considerations but considers the process of information aggregation within the different levels of a hierarchy.
a framework where the firm’s owner (also referred to as the executive) obtains private information about the firm’s prospects (which can be high or low), and chooses the level of investment and the worker’s compensation. In our framework, the worker’s effort is more productive when the future prospects of the firm are better which, in combination with the optimal compensation contract, induces the worker to exert more effort when his beliefs about the firm’s future prospects are more favorable. As a result, the owner has an incentive to overinvest relative to the case in which he has no private information, because doing so conveys favorable information.

In this setting, there is a trade-off between the benefits of using the executive’s private information to adjust its capital allocation and the inherent tendency to overinvest that arises from such discretionary capital allocation choices. As a result of this trade-off, two alternative capital allocation mechanisms emerge as potentially optimal: (i) a separating mechanism in which a high prospect firm invests more (and offers higher compensation to the worker) than a low prospect firm and (ii) a pooling mechanism in which firms pre-commit to a fixed level of investment and compensation. With the separating mechanism, the firm invests more when the marginal productivity of capital is higher but tends to overinvest because of the potential benefits associated with conveying favorable information. With the pooling mechanism, investment rigidities arise, i.e., the level of investment is independent of the executive’s information, while the firm’s overinvestment tendency is suppressed. Hence, the choice between the pooling and the separating mechanisms is determined by a trade-off between the efficiency loss of a rigid investment policy that does not incorporate information and the cost associated with overinvestment.\(^3\)

The information and incentive issues explored in this paper also have implications for how corporations are organized. To investigate this issue we extend our analysis to consider the case of a two-division firm with top executives that have superior information about the prospects of each division. The capital investment choices of each of the divisions are observable and convey information. As we show, investment distortions are less costly for the two-division firm i.e., the optimal investment policy in a two-division firm can generate more value than the optimal investment policies of two independent firms acting independently.\(^3\)

---

\(^3\)As discussed in Section 6, the analysis suggests that investment rigidities observed in practice tend to arise when there is limited uncertainty about the firm’s prospects as well as a higher likelihood of more favorable prospects (e.g., in industries in which growth is more likely).
Intuitively, although executives of a two-division firm have incentives to exaggerate the prospects of each of the divisions, they do not have an incentive to misrepresent the divisions’ relative prospects, i.e., their ranking. As a result, instead of overinvesting, firms can use asymmetric capital allocations to provide credible information about the relative prospects of the divisions.\(^4\) Thus, our analysis provides a novel rationale for conglomerates; namely the reduction in the cost of the investment distortions that arise when the firms’ stakeholders use a firm’s investment choices to make inferences about the firm’s prospects.

This paper addresses issues that have been previously considered in a number of branches of the literature. For instance, our analysis is related to papers in the leadership literature which examines how choices made by leaders influence individuals at lower levels in the organization (e.g., Rotemberg and Saloner’s 2000). In particular, our analysis is close to Hermalin (1998) and Komai, Stegeman, and Hermalin (2007) where an informed leader signals his favorable information by expending greater effort, and in doing so motivates his subordinates to work harder (i.e., leadership effects). There are, however, some key differences between our setting and the setting considered in these leadership papers. First, while the main focus of these studies is to examine how leadership effects can mitigate undereffort in teams, our analysis focuses on how the optimal design of the capital investment process can sometimes limit and other times take advantage of these leadership effects. Moreover, we consider a setting with contractible investment choices, which contrasts with the leadership models that assume that the leader’s effort is not contractible. Indeed, we find that a firm may find it optimal to commit to a rigid investment policy (i.e., pooling mechanism) that does not fully incorporate the top executives’ information.

More broadly, our paper belongs to the principal-agent literature and specifically to the branch that considers the case of an informed principal that faces an agent with a moral-hazard problem.\(^5\) Among other things, this literature analyzes the effects of the principal’s information on the optimal compensation contract (e.g., Beaudry 1994 and Inderst 2000), the value of private information to the principal (e.g., Chade and Silvers 2004) and the

\(^4\)This result is related to the findings in Chakraborty and Harbaugh (2007) which considers the case of multidimensional cheap talk and illustrates how cheap talk about rankings can sometimes convey information.

\(^5\)Maskin and Tirole (1990, 1992) study the optimal mechanism in the case of a principal who proposes the contract after being informed and faces an agent without moral hazard. In contrast, we consider the case of a principal who becomes informed after proposing the contract and faces an agent subject to moral hazard in her actions.
incentives to disclose information (i.e., provide “advice”) to the agent (e.g., Strausz 2009). While most of these papers (Strausz, 2009 is an exception) consider the case of a principal who makes choices after becoming informed (i.e., signaling) we consider a mechanism design setting in which the principal commits to a self-screening mechanism prior to acquiring information and then, after becoming informed, makes choices that convey information that influences the agent’s actions.⁶ Within this self-screening paradigm we focus on a case where the principal has a limited ability to commit to make investment choices that are suboptimal given his private information. As we show, the principal’s imperfect ability to commit implies that the optimal mechanism can be a pooling mechanism in which the principal conceals information to the agent and in which there are rigidities in the firm’s investment choices.⁷

Our contribution is also related to two other branches of the agency literature.⁸ One of them argues that since managers get private benefits from managing larger enterprises, shareholders may want to impose restrictions on managers that inhibit their incentives to overinvest (e.g., Jensen 1986, and Hart and Moore 1995). As we demonstrate, the tendency to overinvest, as well as procedures that curb this tendency, can arise within a setting with asymmetric information but without managerial private benefits. A second branch considers the investment misincentives that arise in conglomerates. In particular, this literature explores a tendency towards “corporate socialism,” what Scharfstein and Stein (2000) describe as a tendency “to engage in cross-subsidization, spending relatively too much in some divisions, and too little in others” (p. 2538). Our model also generates capital misallocation in conglomerates. However, what would be a misallocation of capital under symmetric information arises in our model as an optimal response to the informational problems that firms face with their stakeholders. Moreover, we find that the costs that

⁶See Biais and Mariotti (2005) for a mechanism design analysis with a similar timing of events, namely a commitment by the principal to narrow his choices to a pre-specified set of actions before receiving his private information. In section 3.2, we discuss further the realism and implications of our timing assumption.

⁷See Section 3 for a precise description of the limitations of the principal’s choices. See Bester and Strausz (2001, 2007) for thorough analyses of the design of optimal mechanisms when the principal has an imperfect ability to commit.

⁸There are also papers that examine how firms seek to entice their stakeholders along dimensions other than providing effort. For instance Pagano and Volpin (2005) study how managers use labor contracts to gain workers’ support against unwanted takeover threats and Cespa and Cestone (2007) show that firm value can be enhanced by introducing explicit stakeholder protections, which can have the effect of reducing the entrenchment of inefficient CEOs.
arise from capital misallocations within a conglomerate tend to be lower than the capital misallocation costs that arise if divisions are operated as independent single division firms.

It should also be noted that we are not the first to consider how capital allocations can be improved by conglomerates. Starting with Williamson (1975), researchers have considered the costs and benefits of external versus internal capital markets. The papers in this literature, such as Stein (1997), have stressed the beneficial role of internal capital markets as winner-pickers in settings with credit constraints and managerial agency conflicts. As we show, information asymmetries between firms and their stakeholders provides an additional channel that can contribute to the benefits of internal capital markets.

The rest of the paper is organized as follows. Section 2 presents the base model and Section 3 analyzes it. Section 4 extends the base model by considering the case of multi-divisional firms. Section 5 considers several extensions of our analysis and Section 6 discusses the empirical implications suggested by our results. Finally, Section 7 presents our conclusions.

2 The model

We consider a firm that operates in a risk-neutral economy. The firm is owned and run by an executive (“the principal”) and requires the input from a lower-level manager (“the agent”) who is penniless, is subject to limited liability, and has a zero reservation wage.

The firm’s technology requires making an investment choice $I \in [0, \infty)$ in order to produce a random amount of output $\tilde{q} = \tilde{\zeta}(\epsilon) \sqrt{2I}$ at a cost $I$. For notational convenience we define $k = \sqrt{2I}$ as the firm’s scale and express the firm’s cash-flow prior to paying the agent as:

$$Q(\epsilon, k) = \tilde{\zeta}(\epsilon) k - \frac{1}{2} k^2. \quad (1)$$

As (1) indicates, the firm’s output $\tilde{q} = \tilde{\zeta}(\epsilon) k$ depends on the firm’s scale $k$ and on the agent’s (unobservable) effort $\epsilon \in [0, \frac{1}{\theta}]$, which affects the capital’s (random) productivity, $\tilde{\zeta}(\epsilon)$. Specifically,

$$\tilde{q} = \begin{cases} r \cdot k & \text{with prob. } \theta \epsilon \\ 0 & \text{with prob. } (1 - \theta \epsilon) \end{cases} \quad (2)$$
where \( r > 0 \) and \( \theta \) is an exogenous productivity shock described by

\[
\theta = \begin{cases} 
\beta & \text{with prob. } \pi \\
1 & \text{with prob. } (1 - \pi)
\end{cases}
\]

(3)

where \( \beta > 1 \). Furthermore, we assume that the private cost of effort to the agent is quadratic in the level of effort \( e \) and increases linearly with the firm’s scale \( k \):

\[
h(e, k) = \frac{1}{2} ce^2 k.
\]

(4)

We consider this specific setting for economic as well as for technical reasons. First, it allows us to derive optimal compensation contracts and thus to analyze the interaction between optimal compensation and optimal investment in a relatively straightforward manner. Second, by assuming a complementarity between effort and the productivity shock (i.e., the probability of being successful is the product of \( e \) and \( \theta \)) this formulation captures the insight that inducing effort is more efficient when prospects are more favorable.\(^9\)

Finally, the formulation implies decreasing returns to scale since (i) firm revenue linearly increases with the scale while costs are quadratic in the scale and (ii) the agent experiences a quadratic cost of effort that linearly increases with firm scale. The combination of these assumptions implies that there is an interior solution to the choice of investment and effort (i.e., due to the increasing marginal costs of effort and scale) and that the worker’s moral hazard problem is independent of the firm’s scale.\(^10\)

In what follows, we restrict the analysis to mechanisms that consider compensation contracts that offer non-negative payments \( w \) to the agent (which can be contingent on the firm scale \( k \)), when the realized output is high i.e., \( q = rk \) and a zero-payment when the output is low, i.e., \( q = 0 \).\(^11\) Thus, the principal’s choice set can be described as a menu of \( n \) investment-compensation pairs, \( \{(k_i, w_i)\}_{i=1,...,n} \), which, without loss of generality, can be restricted to \( n = 2 \), i.e., mechanisms that contain, at most, two distinct pairs, \( \{(k_a, w_a), (k_b, w_b)\} \).

\(^9\)This assumption is not innocuous. Indeed, an alternative setting in which effort and firm quality are substitutes can imply that inducing effort is more efficient in situations of low firm productivity and reverse some of the results on investment described below.

\(^10\)We have investigated an alternative setting in which (i) effort costs are independent of the firm scale \( h(e) = \frac{1}{2} ce^2 \) and (ii) investment costs are cubic in the firm scale, i.e., \( g(k) \equiv k^3 \). The analysis of this alternative setting, while less tractable, produces similar results and is available upon request.

\(^11\)In the appendix we show that focusing on compensation contracts with zero rewards when the output is low is without loss of generality.
The timing of events is as follows: At \( t = 0 \), before observing \( \theta \), the principal specifies a menu of two pairs of investment (firm scale) and compensation (performance bonus) i.e., \( \{(k_a, w_a), (k_b, w_b)\} \). At \( t = 1 \), the principal privately observes \( \theta \) and chooses one of the pairs, i.e., either \( (k_a, w_a) \) or \( (k_b, w_b) \) from the pre-specified menu. At \( t = 2 \), the agent makes an unobservable effort choice \( e \). At \( t = 3 \), firm output \( \tilde{q} \) is realized and contracts are settled.

Therefore, we analyze the model as a mechanism design problem in which the optimal mechanism is determined before the principal becomes informed. It is worth mentioning that the design of the optimal mechanism accounts for the effect of the principal’s choices on the agent’s effort but, since effort is unobservable, it excludes mechanisms with contracts that specify the agent’s effort. Figure 1 summarizes the timing of events.

![Figure 1: Timing of Events](image)

### 3 Model analysis

#### 3.1 Benchmark case: Analysis when \( \theta \) is known

To gain intuition we first analyze the case in which both the principal and the agent observe the shock \( \theta \) but in which the agent’s effort choice remains unobservable. We denote as \( k = \{k_1, k_\beta\} \), \( w = \{w_1, w_\beta\} \) and \( e = \{e_1, e_\beta\} \), the investment, compensation and effort levels that correspond to the two possible realizations of \( \theta = \{1, \beta\} \). The optimal mechanism consists of finding for each \( \theta = \{1, \beta\} \) the levels of investment and compensation and the resulting agent’s effort, i.e., \( M = \{(w_1, k_1, e_1), (w_\beta, k_\beta, e_\beta)\} \), that maximize the principal’s value. By denoting \( \pi \) and \( (1 - \pi) \) as \( p_\beta \) and \( p_1 \) respectively, the principal’s problem can be compactly expressed as

\[
\max_M V = \sum_{\theta = \{1, \beta\}} p_{\theta} \left[ (r k_\theta - w_\theta)\theta e_\theta - \frac{1}{2} k_\theta^2 \right] \tag{5}
\]

s.t.:

\[
\begin{align*}
e_\theta &= \arg \max_e \{w_\theta e - \frac{1}{2} c e^2 k_\theta\}, \quad &\text{for } \theta = \{1, \beta\} \tag{6} \\
w_\theta\theta e_\theta - \frac{1}{2} c e^2 k_\theta &\geq 0, \quad &\text{for } \theta = \{1, \beta\} \tag{7}
\end{align*}
\]
For each realization of the shock, \( \theta = \{1, \beta\} \), the principal maximizes firm value subject to the corresponding agent’s optimal effort choice (6), and participation constraint (7). The problem can be simplified because the agent’s limited liability requires \( w_\theta \geq 0 \) and, since \( e = 0 \) is feasible, the constraints included in (7) always hold.\(^{12}\)

By ignoring (7) and substituting the first-order condition of (6) in (5) we get

\[
\max_{k, w} V = \sum_{\theta = \{1, \beta\}} p_\theta \left[ (r k_\theta - w_\theta) \theta \frac{w_\theta}{k_\theta} - \frac{1}{2} k_\theta^2 \right], \tag{8}
\]

the solution of which leads to the following proposition:

**Proposition 1** In the optimal mechanism \( M^* \) when \( \theta \) is observable, the compensation and investment choices are: \( w_\theta^* = \frac{r^2 \theta}{8 \pi} \) and \( k_\theta^* = \frac{r^2 \theta}{4 \pi} \) for \( \theta = \{1, \beta\} \). With these choices, the agent’s effort is \( e_\theta^* = \frac{r \theta}{\pi} \) for \( \theta = \{1, \beta\} \) and the principal’s expected payoff is

\[
V^* = \frac{1}{2} \left[ (1 - \pi) k_1^2 + \pi k_\beta^2 \right]. \tag{9}
\]

Proposition 1 indicates that in the optimal mechanism \( M^* \), compensation, investment and effort increase in \( \theta \), i.e., \( w_\theta^* > w_1^* \), \( k_\theta^* > k_1^* \) and \( e_\theta^* > e_1^* \). Intuitively, \( M^* \) implies higher investment and greater effort in state \( \beta \) because the marginal product of both capital and effort is increasing in \( \theta \). Notice that in \( M^* \), compensation is a sharing rule independent of \( \theta \), (i.e., \( \frac{w_\theta^*}{r k_\theta^*} = \frac{1}{2} \)) that is, the worker receives half of the output of the firm independently of the firm productivity \( \theta \).

### 3.2 Analysis when \( \theta \) is privately observed

In this section we examine the case where the principal but not the agent observes the shock \( \theta \). In this case the mechanism design problem must consider how the principal’s choices convey information about \( \theta \) and how the conveyed information interacts with the provision of incentives to the agent.

#### 3.2.1 Assumptions on timing and commitment

In what follows we analyze the mechanism design problem faced by an uninformed principal at \( t = 0 \) who anticipates that he will become informed about \( \theta \) at \( t = 1 \). In particular, at

\(^{12}\)We implicitly assume that the principal is the residual claimant of the output (net of wages) and, thus ignore mechanisms that consider the possibility of making payments to a third party unrelated to the production process. At the end of Section 3.2.2 we discuss how relaxing this assumption can affect the results.
At $t = 0$ the principal imposes a menu that limits his own potential choices at $t = 1$ with the objective to influence the transmission of information that occurs when he actually makes a choice after learning $\theta$.\textsuperscript{13} We assume that these imposed limits are enforceable (i.e., the executive can indeed commit to limit his choices at $t = 1$ to those included in the menu specified at $t = 0$). In practice, a number of institutional features can help executives to limit themselves to a reduced set of potential future choices. These limits can stem from the internal organization of the firm (e.g., internal procedures that restrict the firm’s future investment behavior) or from parties external to the firm (e.g., executives choose to provide ex-ante guidance to equity analysts about their upcoming investment expenditures). Although these are not binding legal commitments, to the extent that executives experience negative effects (e.g., a loss in credibility) from a lack of compliance, these institutional features can effectively provide the executive with the ability to constrain their own future choices.

While the uninformed principal can limit himself to choose from a pre-specified menu of choices, we assume that at $t = 1$, the informed principal makes the optimal choice (conditional on the observed $\theta$) among those available in the pre-specified menu.\textsuperscript{14} In other words, the principal has a limited ability to commit at $t = 1$. This limited ability to commit plays an important role in the mechanism design problem under consideration. In particular, the principal’s limited ability to commit excludes mechanisms in which the principal’s choices are delegated to a third party who may subsequently force suboptimal choices on the principal. Our focus on mechanisms in which the principal cannot be forced to take suboptimal investment and compensation decisions is appropriate in a setting like ours, which is designed to capture the interaction between executives and lower level managers in the same firm. Furthermore, as shown below, a setting based on these premises generates a number of practical implications (e.g., implications on investment distortions) on how firms allocate capital.

\textsuperscript{13}We could have considered instead the alternative timing of an informed principal that designs the mechanism at $t = 1$ after observing $\theta$ (i.e., signaling). The screening approach, which is technically convenient, help us capture the insight that firms make deliberate decisions to manage the process of information transmission.

\textsuperscript{14}In a similar vein Laffont and Tirole (1988) analyze the design of a sequence of incentive contracts in a setting in which the principal cannot commit to not use the information conveyed by the agent in the first period in the second period contract.
3.2.2 The search for the optimal mechanism

The combination of these assumptions on timing and (limited) commitment simplifies the procedure to solve for the optimal mechanism. To understand why this is the case consider as a comparison an alternative setting where the mechanism designer (i.e., the principal) has full commitment ability, but faces an agent subject to moral hazard. With full commitment ability, the optimal mechanism implied by the revelation principle can be a mediated mechanism in which the principal discloses his private information to a third-party (i.e., a mediator) who, may partially reveal the information to the agent while forcing the principal to make choices that are potentially suboptimal for the principal.\(^\text{15}\)

However, in a setting like ours in which the principal cannot commit to suboptimal choices, the search for the optimal mechanism can be limited to the class of non-mediated mechanisms, namely mechanisms in which there is no mediator but direct communication between the principal and the agent. In this case, the optimal mechanism can be derived by solving for the optimal menu of self-imposed constraints at \(\tau = 0\), while imposing the screening (i.e., incentive compatibility) constraints on the choice at \(\tau = 1\). As it turns out, the optimal mechanism does not necessarily imply full disclosure of the principal’s information to the agent either directly (i.e., by publicly revealing it) or indirectly by making choices that lead the agent to learn the principal’s information (i.e., by implementing actions that depend on the observed \(\theta\)). In other words, in our setting with limited commitment by the principal, and with the agent’s moral hazard, the revelation principle does not apply.\(^\text{16}\)

Formally, in our setting, the optimal mechanism can be derived by considering a pre-specified menu of at most two choices for the uninformed principal \(\{(k_a, w_a), (k_b, w_b)\}\) from

\(^{15}\)Specifically, the mediator produces a publicly observable signal that is a function of the principal’s announcement along with random noise. Thus, the presence of a mediator (i) allows the agent to partially learn the principal information and (ii) forces the principal to condition his choices on the signal’s realization.

\(^{16}\)Bester and Strausz (2007) analyze the issue of mechanism design (i.e., contracting) with imperfect commitment. In their setting, the principal (i) is imperfectly informed about the agent’s type (adverse selection) and (ii) cannot commit to taking certain actions (e.g., actions that are conditioned on the agent’s revelation of information about his type). Among other things, they show that the use of general communication devices can improve the design of the principal’s optimal mechanism. By contrast, we consider a setting in which the principal (i) is initially uninformed about its own type, but privately learns its type prior to taking an action, (ii) faces an agent subject to moral hazard and (iii) has limited commitment ability (i.e., the principal is constrained to choose the optimal action after learning his type). As it turns out, the interaction of (ii) and (iii) eliminates the value of communication devices in our setting and allows us to solve the principal’s problem by just considering direct communication between the principal and the agent.
which the informed principal optimally chooses either \((k_a, w_b)\) or \((k_b, w_b)\). Therefore, three types of mechanisms can potentially emerge as optimal: (i) a separating mechanism \(M^s\) in which the principal’s choices depend on, and thus reveal, the observed \(\theta\), i.e., the principal chooses \((k_1, w_1)\) after observing \(\theta = 1\) and \((k_\beta, w_\beta)\) with \(k_\beta \neq k_1\) or \(w_\beta \neq w_1\) after observing \(\theta = \beta\); (ii) a pooling mechanism \(M^p\) in which the principal’s choices do not depend on the observed \(\theta\) and thus do not reveal \(\theta\), i.e., \(k_1 = k_\beta\) and \(w_1 = w_\beta\) and (iii) a partial pooling mechanism \(M^{pp}\) in which the principal mixes between two investment-compensation pairs and thus partially reveals \(\theta\) to the agent.

In what follows, we solve for the optimal mechanism within the separating and pooling classes and compare firm values when the optimal mechanisms within these classes are implemented. As shown in the appendix, this is without loss of generality since firm value under partial pooling is lower than firm value when either the optimal separating or the optimal pooling mechanism is considered.

### 3.2.3 The optimal separating mechanism

In a separating mechanism \(M^s = \{(k_1, w_1^s, e_1^s), (k_\beta, w_\beta^s, e_\beta^s)\}\) the principal’s choices of compensation \(w^s = \{w_1^s, w_\beta^s\}\) and investment expenditures \(k^s = \{k_1^s, k_\beta^s\}\) depend on the firm’s type \(\theta\). Furthermore, since \(\theta\) cannot be observed by the agent, these choices are subject to incentive compatibility (i.e., “truth-telling”) constraints on the principal’s choices. We define \(V^s_\hat{\theta}\) as the firm value when the principal observes \(\theta\) and chooses \((w^s_\hat{\theta}, k^s_\hat{\theta})\) for \(\hat{\theta} = \{1, \beta\}\):

\[
V^s_\hat{\theta} \equiv (r k^s_\theta - w^s_\theta) \theta e^s_\theta - \frac{1}{2} k^s_\theta^2 
\]  

(10)

More generally, the principal’s choice set could include \(n\) investment-compensation pairs, \(\{(k_i, w_i)\}_{i=1,...,n}\). Bester and Strausz (2001), however, show that without loss of generality in a setting like ours the number of contracts equals the number of types, i.e., \(n = 2\). See the appendix for details.

Formally, a partial pooling mechanism would require us to define \(\{\sigma_1, \sigma_\beta\}\) as the probabilities that the principal chooses between \((k_1, w_1)\) and \((k_\beta, w_\beta)\) after observing \(\{1, \beta\}\) and to solve for the optimal level of information revelation, i.e, the optimal \(\{\sigma_1, \sigma_\beta\}\).

Intuitively, in our setting, with a continuous set of investment choices, the class of separating and pooling mechanisms is sufficiently rich so that partial pooling mechanisms are always suboptimal. This is in contrast with Strausz (2009) who considers a related setting but in which the investment choice is discrete (i.e., either a fixed amount or zero) and finds that partial pooling mechanisms can be optimal.

Intuitively, in our setting, with a continuous set of investment choices, the class of separating and pooling mechanisms is sufficiently rich so that partial pooling mechanisms are always suboptimal. This is in contrast with Strausz (2009) who considers a related setting but in which the investment choice is discrete (i.e., either a fixed amount or zero) and finds that partial pooling mechanisms can be optimal.
where \( e_\theta^* = \arg \max_e \{ w_\theta^* \hat{e} - \frac{1}{2} ce^2 k_\theta^* \} \) for \( \hat{\theta} = \{1, \beta\} \) and denote as \( e^* = \{ e_\theta^*, e_{\bar{\theta}}^* \} \) the corresponding agent’s effort. Thus, the principal’s problem can be expressed as:

\[
\max_{M^*} V_s = \sum_{\theta = \{1, \beta\}} p_\theta \left[ (r k_\theta^* - w_\theta^*) \theta e_\theta^* - \frac{1}{2} k_\theta^2 \right] \quad \text{(11)}
\]

s.t.: \( e_\theta^* = \arg \max_e \{ w_\theta^* \hat{e} - \frac{1}{2} ce^2 k_\theta^* \}, \) for \( \theta = \{1, \beta\} \)

\[
w_\theta^* \theta e_\theta^* - \frac{1}{2} c e^2 k_\theta^* \geq 0, \quad \text{for } \theta = \{1, \beta\} \quad \text{(12)}
\]

\[V_\theta^* \geq V_\theta^* \]

for \( \theta, \hat{\theta} = \{1, \beta\} \) and \( \hat{\theta} \neq \theta \)

Formally, problem (11)-(14) consists of the addition of the IC constraints (14) to the benchmark problem (5)-(7), where \( \theta \) is observable. This modified problem can be solved by (i) ignoring the IC constraint of the high productivity firm, i.e., \( V_\beta^* \geq V_\beta^* \), and assuming that the IC constraint of the low productivity firm binds, i.e., \( V_1^* = V_1^* \); (ii) substituting \( V_1^* = V_1^* \) into the objective function (11); (iii) deriving the corresponding first order conditions and (iv) checking that the ignored IC constraint indeed holds in the original problem. (See the appendix for details.) Proposition 2 describes the optimal separating mechanism:

**Proposition 2** In the optimal separating mechanism \( M^{ss} \) compensation and investment are: \( w^{ss} = \{ w_1^*, \Delta w_\beta^* \} \) and \( k^{ss} = \{ k_1^*, \Delta k_\beta^* \} \) where \( \Delta = \max \left\{ \frac{1+\sqrt{1-1/\beta^2}}{\beta}, 1 \right\} \). With these choices, the agent’s effort is \( e_\theta^{ss} = e_\theta^* = \frac{r_\theta}{2c} \) for \( \theta = \{1, \beta\} \) and the principal’s payoff is

\[
V^{ss} = \frac{1}{2} \left[ (1 - \pi) k_1^{ss^2} + \pi \Delta (2 - \Delta) k_\beta^{ss^2} \right]. \quad \text{(15)}
\]

A comparison between Propositions 1 and 2 shows that, relative to the case in which \( \theta \) is observable, compensation and investment remain unaltered for the low prospect firm, \( (k_1^{ss}, w_1^*) = (k_1^*, w_1^*) \), and increase by the factor \( \Delta \) for the high prospect firm, \( (k_\beta^{ss}, w_\beta^*) = (\Delta k_\beta^*, \Delta w_\beta^*) \). Effort turns out to be equal to the effort exerted when \( \theta \) is observable (i.e., \( e_\theta^{ss} = e_\theta^* \)). This is because the level of effort, which is proportional to the ratio of compensation to investment, turns out to be the same whether or not \( \theta \) is observable (i.e., \( e_\theta^* = \frac{w_\theta^* \theta}{c e_\theta} = \frac{w_\theta^* \theta}{c e_\theta} \)).

---

21 The (IC) constraint (14) captures the implications of the assumption that the informed principal cannot commit to make suboptimal choices at \( t = 1 \).

22 The allocation described by \( M^{ss} \) corresponds to the Cho and Kreps (1987) refined equilibrium in a signalling setting which is identical to ours in terms of technology and information but that differs from ours in the relative timing of contracts and actions. More specifically this refers to a signaling setting in which the principal: a) does not limit his choices to a pre-specified menu of choices and b) signals his information to the agent by choosing a single investment-compensation pair.
The overinvestment factor $\Delta$ summarizes the main intuition of the analysis in the optimal separating mechanism. Specifically, the high prospect firm overinvests (i.e., increases its scale relative to the case of observable $\theta$) to credibly communicate to the agent the firm’s high prospects. As Proposition 2 indicates, the overinvestment factor $\Delta$ depends on $\beta$, which measures the difference in the size of productivity shocks across firms. Specifically, as a function of $\beta$, overinvestment factor $\Delta$ has an inverted U-shape that reaches its maximum at $\hat{\beta} = \frac{2 \sqrt{3}}{3} \approx 1.15$ and a minimum (i.e., $\Delta = 1$) when $\beta \geq \beta^* \approx 1.84$. In other words, when differences in productivity are relatively small (i.e., when $\beta < \beta^*$) the optimal separating mechanism requires the high prospect firm to overinvest to convey its type to the agent. However, when the differences in productivity are large enough (i.e., when $\beta \geq \beta^*$) there is no overinvestment in the optimal separating mechanism. When this is the case, the optimal mechanism corresponds to the optimal mechanism in the benchmark case where the agent can observe $\theta$, i.e., $M^{**} = M^*$ when $\beta \geq \beta^*$.

![Figure 2: Optimal overinvestment factor $\Delta(\beta)$](image)

It is worth stressing that $M^{**}$ includes overinvestment (and overcompensation) even though it is possible to design separating mechanisms that exclusively rely on distortions in compensation. Investment distortions are included in the optimal mechanism for two reasons. First, investment distortions more efficiently convey information about a firm’s prospects because higher investment is costlier for a type 1 firm than for a type $\beta$ firm. Second, separating mechanisms that rely exclusively on distorting compensation can be
costlier for a type $\beta$ firm than a type 1 firm because effort incentives require compensation to be contingent on success, and the probability of success is higher for a type $\beta$ firm. In other words, while a higher $w_{\beta}$ may help deter the type 1 from mimicking a type $\beta$ firm, a higher $w_{\beta}$ is more costly for a type $\beta$ firm because it is more likely that the type 1 pays the high $w_{\beta}$.

It is also natural to consider whether by shaping his own compensation the principal can convey information about the firm’s type. For example, the principal might consider separating mechanisms with optimal investment in which information is conveyed by choosing a higher powered principal’s compensation when $\theta = \beta$ than when $\theta = 1$. It is straightforward to show that this cannot be the case. Since the program (11)-(14) includes an adding-up constraint (which implies that the principal is the residual claimant to the firm’s output net of wages) mechanisms that distort the principal’s payoffs are equivalent to mechanisms that distort the agent’s compensation. Therefore, our analysis summarized in Proposition 2 does in fact (implicitly) consider mechanisms in which the principal can shape his own compensation to convey the firm’s type information.

Alternatively we could have considered mechanisms in which the adding up constraint is relaxed by allowing the principal to commit to give up a share of the firm’s output in some situations (i.e., to burn money). In the appendix, we show that under a mild parametric condition, any mechanism that includes money burning is strictly dominated by the optimal mechanism described in Proposition 2. Intuitively, this occurs because any incentive effect produced by burning money can be replicated by a properly designed combination of overinvestment and compensation. This observation implies that the optimal separating mechanism derived in Proposition 2 is also the optimal mechanism in a richer setting in which the principal can make positive third-party side-payments or engage in any other form of money burning.

Finally it is worth noting that our overinvestment result is obtained relative to the benchmark case with symmetric information about $\theta$ and incentive problems due to unobservable effort $e$. We refer to our results as “overinvestment” since the marginal profit

---

23Technically, this relaxes the adding-up constraint and makes the principal the residual claimant net of wages and any money burned.

24In the appendix we provide a formal argument to support this claim and the specific parametric condition. In Section 5, we reexamine how the inclusion of third parties can affect the optimal design of capital budgeting rules.
obtained from the last unit of invested capital is negative and firms nevertheless invest to convey favorable information to their workers.\textsuperscript{25}

3.2.4 The optimal pooling mechanism

We now analyze pooling mechanisms, namely those mechanisms in which the principal commits at $t = 0$ (before observing $\hat{\theta}$) to make investment and compensation choices at $t = 1$ that are independent of the observed $\hat{\theta}$. Formally we denote a pooling mechanism as single-contract menu $M^p = \{(w^p, k^p, e^p)\}$ where $w^p$ and $k^p$ are, respectively, the compensation in case of success, the investment and the agent’s induced effort. If we define $\bar{\theta} \equiv \pi \beta + (1 - \pi)1$ as the average productivity shock, then the optimal pooling mechanism $M^p^*$ solves:

$$
\begin{align*}
\max_{M^p} V_p &= (rk^p - w^p)\bar{\theta}e^p - \frac{1}{2}k^p2 \\
\text{s.t.: } e^p &= \arg\max_e \{w^p\bar{\theta}e - \frac{1}{2}ce^2k^p\}m.
\end{align*}
$$

(16)

(17)

Proposition 3 describes the optimal pooling mechanism:

\textbf{Proposition 3} In the optimal pooling mechanism $M^p^*$ compensation and investment are:

$w^{p*} = \frac{3\theta^2}{8e}$ and $k^{p*} = \frac{3\theta^2}{4e}$. With these choices the agent’s effort is $e^{p*} = \frac{\bar{\theta}}{2e}$ and the principal’s payoff is $V_{p}^* = \frac{1}{2}k^{p*2}$.

In the optimal pooling mechanism $M^p^*$, the level of the agent’s effort is the average of the two effort levels under the benchmark and optimal separating cases, i.e., $e^{p*} = \frac{\bar{\theta}}{2e}$ and the investment choice $k^{p*}$ does not depend on the realized $\hat{\theta}$, which implies that the firm overinvests when $\hat{\theta} = 1$ and underinvests when $\hat{\theta} = \beta$. Furthermore, under $M^p^*$ compensation is the same proportion of realized output as in $M^*$ and $M^{**}$, i.e., $\frac{w^{p*}}{r k^{p*}} = \frac{1}{2}$.

The commitment to a rigid investment policy summarizes the main intuition of the analysis in the optimal pooling mechanism. Indeed, that investment rigidities can arise an element of an optimally designed investment process is a notable result that arises due to the interaction of two key ingredients in our setting: (1) the presence of a moral hazard

\textsuperscript{25}We could have considered instead an alternative benchmark in which both $\theta$ and $e$ are observable. Relative to this latter benchmark, a setting like ours, with both asymmetric information about $\theta$ and incentive problems on $e$ generates underinvestment. Intuitively, even with symmetric information about $\theta$, firms invest less when $e$ is unobservable because with unobservable effort workers obtain rents (due to their limited liability) that increase with the size of the investment. As a result, firms reduce their investment expenditures on the margin to reduce the size of these rents.
problem, which makes the principal less inclined to disclose information to the worker and (2) the inability of the principal to commit to suboptimal choices conditional on the observed $\theta$, which makes the principal unable of rely on richer contracting spaces (e.g., mediated mechanisms) that could create some investment flexibility at the expense of some ex-post investment inefficiencies.

### 3.2.5 Separating or pooling mechanism?

We now compare $M^{ps}$ with $M^{ss}$, i.e., the trade-off between investment rigidities and over-investment. Note that, in contrast to the optimal separating mechanism $M^{ss}$, with the optimal pooling mechanism $M^{ps}$ the firm does not overinvest or overcompensate managers when $\theta = \beta$. The cost, however, is that the firm does not tailor its investment expenditures to its realized marginal productivity. The choice between $M^{ss}$ and $M^{ps}$ depends on the trade-off between these costs and benefits as the next proposition states.

**Proposition 4** Depending on parameters, either $M^{ps}$ or $M^{ss}$ is the optimal mechanism. Furthermore the difference in firm value under the alternative mechanisms i.e., $(V^{ss} - V^{ps})$ is: (i) decreasing in the likelihood of the high productivity firm (i.e., decreasing in $\pi$) and (ii) increasing in the difference in productivity among firm types (i.e., increasing in $\beta$).

Figure 3 displays the regions in the space $(\beta, \pi)$ where each type of mechanism is optimal.
Proposition 4 states that both $M^{**}$ and $M^{p*}$ can be optimal mechanisms. When $M^{**}$ is the optimal mechanism, the firm exhibits overinvestment as a part their investment policy. By contrast when $M^{p*}$ is optimal, rigidities in investment and compensation are an integral part of on the optimal capital budgeting policy.

Intuitively the comparative statics results can be described by considering two limiting cases: (i) when $\pi$ is close to 1 (i.e., $\theta = \beta$ is highly likely) under $M^{p*}$ the type $\beta$ investment is close to the full-information investment level while the expected cost of overinvestment for type 1 is very small, by contrast $M^{**}$ requires the type $\beta$ firm to substantially overinvest, and thus produces a sizable expected cost of overinvestment cost.\(^{26}\) and (ii) when $\beta \to \beta^*$, the efficiency loss in $M^{p*}$ is large because the average investment is far from the full information level for both types while the overinvestment inefficiency inherent in $M^{**}$ is reduced. In fact $M^{**}$ converges to $M^*$ since the type 1 firm invests efficiently and the degree of overinvestment of the type $\beta$ firm goes to zero.

4 The two-division firm case

Up to this point we have established that firms with privately informed executives tend to either overinvest, relative to the first best, to influence the behavior of the firm’s worker (i.e., the separating mechanism) or, because of this tendency, commit to investment rigidities, (i.e., the pooling mechanism). This section extends our analysis to consider how these incentive issues influence the design of the capital budgeting process in multi-divisional firms. Doing so allows us to examine whether our results can be extended to conglomerates and more broadly to consider if the inherent tendency to overinvest to convey favorable prospects has implications for how firms are structured.

We consider a two-division extension of the framework examined earlier. Specifically, we analyze the design of the capital budgeting process for a firm that consists of two divisions that are each (ex-ante) identical to the single-division firm previously analyzed. In particular, we assume that each division $d = A, B$ employs one worker whose effort affects the profitability of his division. The firm experiences a productivity shock in each division, i.e., $\Theta = (\theta_A, \theta_B)$, where for $d = A, B$, $\theta_d \in \{1, \beta\}$ with $Pr(\theta_d = \beta) = \pi$ and

\(^{26}\)Put differently, for type 1 firms, the investment inefficiency is large under $M^{p*}$ and zero under $M^{**}$. However, when $\pi \to 1$ these effects are of second order importance relative to the trade-off that arises for a type $\beta$ firm.
1 < \beta < \beta^*$. We assume that the shocks are independent across divisions, which implies that the firm’s productivity shock can take one of four values $\Theta \in \{(1, 1), (1, \beta), (\beta, 1), (\beta, \beta)\}$ with respective probabilities $\Pr(\Theta) = \{(1 - \pi)^2, (1 - \pi)\pi, \pi(1 - \pi), \pi^2\}$.\footnote{Considering partially correlated shocks introduces a number of additional complications related to the fact that the worker of one division can learn from the investment decision of the other division. By contrast in the case of perfectly correlated shocks the implications are immediate: Perfectly negatively correlated shocks produce no investment distortion because the optimal investment for the combined divisions are constant while perfectly positively correlated shocks produce distortions identical to those described in the single-division case.} We keep our assumptions on commitment and timing of events as described in Figure 1.

In what follows, we analyze the optimal mechanism for the two-division firm and compare it with the optimal mechanism described earlier, for the case in which each division operates as an independent single-division firm.\footnote{Formally, the optimal mechanism corresponds to the solution of a two-dimensional screening problem (as in e.g., Armstrong and Rochet 1999 and Rochet and Stole 2003) in a setting in which the agent’s effort is unobservable and thus non-contractible.}

### 4.1 Pooling and separating mechanisms

As previously shown, in single-division firms the optimal mechanism is always either a pooling or a separating mechanism. In contrast, the analysis in this subsection shows that in the two-division case the optimal mechanism can also involve partial pooling. To demonstrate this, we derive the optimal pooling and separating mechanisms in the two-division firm and then examine whether or not partial pooling mechanisms can generate additional firm value relative to them.

#### 4.1.1 Pooling mechanisms in the two-division firm

In the two-division firm, a pooling mechanism $M^p_2$, consists of two contracts that represent a commitment to a worker’s compensation, an investment expenditure and induced efforts in each division that are independent of the division’s type, i.e., $M^p_2 = \{(w_A^p, k_A^p, e_A^p), (w_B^p, k_B^p, e_B^p)\}$. The following lemma describes the optimal pooling mechanism $M^{p*}_2$ which can be obtained by solving a problem analogous to that defined by (16)-(17):

**Lemma 1** The optimal pooling mechanism for the two-division firm replicates the optimal pooling mechanisms for two single-division firms, i.e., $M^{p*}_2 = \{(w^{p*}, k^{p*}, e^{p*}), (w^{p*}, k^{p*}, e^{p*})\}$.
Intuitively, since the divisions are subject to independent shocks, the investment choice of one division conveys no information about the other division’s type. As a result, a two-division firm does not benefit from implementing a pooling mechanism that differs from the mechanisms obtained for two independent single division firms.

### 4.1.2 Separating mechanisms in the two-division firm

In the two-division firm, a separating mechanism \( M^S_2 \) consists of a worker’s payoffs and an investment expenditure that depends on the firm’s type \( \Theta \), and whose implementation fully reveals the firm’s type (i.e., the divisions’ productivities). Formally a separating mechanism is defined by four contracts \( M^S_2 = \{c^1_2, c^2_1, c^1_2, c^2_2\} \) where (i) each contract \( c^\Theta_2 \equiv \{ (w^\Theta_A, k^\Theta_A, c^\Theta_A), (w^\Theta_B, k^\Theta_B, c^\Theta_B) \} \) specifies a wage \( w^\Theta \), investment level \( k^\Theta \) and induced worker’s effort \( e^\Theta \) for each division \( d = A, B \) conditional on the firm’s type \( \Theta \) and (ii) each contract \( c^\Theta_2 \) identifies the type \( \Theta \), i.e., \( c^\Theta_2 \neq c^\hat{\Theta}_2 \) for \( \Theta \neq \hat{\Theta} \). We define \( V_{\theta_d}^\hat{\Theta} \) as the division \( d \) value when the principal chooses a contract that implements \( \hat{\Theta} \) in a \( \theta_d \) division, where \( \theta_d, \hat{\theta}_d = \{1, \beta\} \):

\[
V_{\theta_d}^\hat{\Theta} \equiv (rk_{\theta_d}^s - w_{\theta_d}^s) e^\theta_{\theta_d} - \frac{1}{2} k_{\theta_d}^s
\]

where \( e^\theta_{\theta_d} = \arg \max_e \{ w^e_{\theta_d} \hat{\Theta} e - \frac{1}{2} \alpha^2 k_{\theta_d}^s \} \) for \( \theta = \{1, \beta\} \) and \( e^s = \{ e^s_1, e^s_\beta \} \).

The two-division firm’s problem can be compactly written as

\[
\max_{M^S_2} \sum_{\Theta} \Pr(\Theta)(V_{\theta_A}^\Theta + V_{\theta_B}^\Theta)
\]

s.t.:

\[
e^\Theta_d = \arg \max_e \{ w^e_{\theta_d} E[\theta_d | c^\Theta_e] e - \frac{1}{2} \alpha^2 k_{\theta_d}^\Theta \},
\]

\[
V_{\theta_A}^\Theta + V_{\theta_B}^\Theta \geq V_{\hat{\theta}_A}^\Theta + V_{\hat{\theta}_B}^\Theta \text{, for } \theta_A, \hat{\theta}_A, \theta_B, \hat{\theta}_B = \{1, \beta\}.
\]

Since types are not observable, (21) expresses the incentive compatibility constraints i.e., a firm of type \( \Theta \neq \hat{\Theta} \) must find it optimal to implement in its divisions \( c^\Theta_2 \) rather than \( c^\hat{\Theta}_2 \), for \( \hat{\Theta} \neq \Theta \).

The solution to (19)-(21) is stated in the following lemma:

\[\text{As in the single-division case, without loss of generality we can restrict the analysis to compensation contracts in which workers receive no pay if the output is zero, i.e., } w^A_0, w^B_0 \geq 0.\]
Lemma 2  The optimal separating mechanism for the two-division firm $M_2^{ss}$ replicates the optimal separating mechanism for two single-division firms. Specifically, the optimal compensation, investment and effort are respectively $w_{\theta d}^* = w^{ss}$, $k_{\theta d}^* = k^{ss}$ and $e_{\theta d}^* = e^{ss}$ for each division $d = A, B$.

Lemma 2 states that the optimal separating mechanism for a two-division firm is a perfect replica of the optimal separating mechanisms that arise if each of the divisions act as independent single-division firms. In particular, the lemma states that a low type division invests the first best level $k_1^*$ regardless of the other division’s type, while a high type division overinvests at level $k_\beta^*$ regardless of the other division’s type. Therefore, the previous lemma also implies that having two divisions with independent shocks does not reduce the cost (i.e., the required investment and compensation distortions) of fully conveying information about firm types to the worker. In particular it means that, relative to the distortions in allocations described in $M^{ss}$ for single division firms a two-division firm does not benefit from shifting investment or wages between divisions.

In summary, lemmas 1 and 2 indicate that if the optimal mechanism in the two-division firm generates either complete separation or complete pooling, there are no gains associated with combining two firms (with uncorrelated shocks) into a single two-division firm. As we show next, however, the optimal mechanism for the two-division firm can involve partial pooling, and when this is the case, the value of the two-division firm is strictly higher than the sum of the two independent firm values.

4.2 General mechanisms

4.2.1 Symmetric partial pooling mechanisms

We now examine partial pooling mechanisms. To facilitate the presentation we first introduce a particular class of mechanisms, i.e., symmetric partial pooling mechanisms (SPPM), defined as follows:

Definition A SPPM $M_2^\Lambda = \{c_{1,\beta}, c_{\beta,1}\}$ consists of two “mirror-image” contracts i.e., $c_{1,\beta} = \{(w_{1}^\Lambda, k_{1}^\Lambda, e_{1}^\Lambda),(w_{\beta}^\Lambda, k_{\beta}^\Lambda, e_{\beta}^\Lambda)\}$ and $c_{\beta,1} = \{(w_{\beta}^\Lambda, k_{\beta}^\Lambda, e_{\beta}^\Lambda),(w_{1}^\Lambda, k_{1}^\Lambda, e_{1}^\Lambda)\}$ such that:

1. A type $(1, \beta)$ firm implements $c_{1,\beta}$ and a type $(\beta, 1)$ firm implements $c_{\beta,1}$. 

2. Firms of types \((1, 1)\) and \((\beta, \beta)\) randomize among \(c_{1,\beta}\) and \(c_{\beta,1}\) and implement either contract with equal probability.

Focusing on mechanisms on the SPPM class is intuitive since, relative to the single-division case, a two-division firm that uses a SPPM mechanism is subject to less constraining incentive compatibility constraints and thus can potentially reveal information about the productivity of their divisions with fewer distortions. In particular, since \(c_{\beta,1}\) and \(c_{1,\beta}\) are mirror images of each other it follows that types \((1, 1)\) and \((\beta, \beta)\), are indifferent between \(c_{\beta,1}\) and \(c_{1,\beta}\) and that it is straightforward to design contracts in which a type \((1, \beta)\) firm has no incentive to mimic a type \((\beta, 1)\) firm and vice versa.

The optimal SPPS \(M_2^{\Lambda_*} = \{c_{\beta,1}^{\Lambda_*}, c_{1,\beta}^{\Lambda_*}\}\) solves:

\[
\max_{M_2^\Lambda} E(V_A + V_B \mid M_2^\Lambda) \tag{22}
\]

s.t.:
\[
e_1^\Lambda = \arg\max_e \{w_1^\Lambda E[\theta_1 | M_2^\Lambda]e - \frac{1}{2}c e^2 k_1^\Lambda\}, \tag{23}
\]
\[
e_\beta^\Lambda = \arg\max_e \{w_\beta^\Lambda E[\theta_\beta | M_2^\Lambda]e - \frac{1}{2}c e^2 k_\beta^\Lambda\}, \tag{24}
\]
\[
V_A^{1,c_{1,\beta}} + V_B^{1,c_{1,\beta}} = V_A^{1,c_{\beta,1}} + V_B^{1,c_{\beta,1}} \tag{25}
\]
\[
V_A^{1,c_{1,\beta}} + V_B^{1,c_{1,\beta}} \geq V_A^{1,c_{\beta,1}} + V_B^{1,c_{\beta,1}} \tag{26}
\]
\[
V_A^{\beta,c_{1,\beta}} + V_B^{\beta,c_{1,\beta}} \geq V_A^{\beta,c_{\beta,1}} + V_B^{\beta,c_{\beta,1}} \tag{27}
\]
\[
V_A^{\beta,c_{1,\beta}} + V_B^{\beta,c_{1,\beta}} = V_A^{\beta,c_{\beta,1}} + V_B^{\beta,c_{\beta,1}} \tag{28}
\]

where \(V_d^{\theta_d,c_A}\) refers to the division \(d = A, B\) value when the division type is \(\theta_d = \{1, \beta\}\) and a contract \(c_A \in \{c_{\beta,1}, c_{1,\beta}\}\) is chosen. Lemma 3 characterizes the optimal SPPM, \(M_2^{\Lambda_*} = \{c_{\beta,1}^{\Lambda_*}, c_{1,\beta}^{\Lambda_*}\}\).

**Lemma 3** The optimal SPPM \(M_2^{\Lambda_*} = \{c_{\beta,1}^{\Lambda_*}, c_{1,\beta}^{\Lambda_*}\}\) features: \(k_\beta^{\Lambda_*} = \frac{r^2 \bar{\beta}}{4c}\), \(w_{\beta}^{\Lambda_*} = \frac{r^3 \bar{\beta}}{8c}\), and \(c_{1,\beta}^{\Lambda_*} = \frac{r^2 \bar{\beta}}{8c}\) where \(\bar{\beta} \equiv E[\theta_A | c_{\beta,1}] = E[\theta_B | c_{1,\beta}]\) if \(c_{\beta,1}\) (resp. \(c_{1,\beta}\)) is chosen and \(k_1^{\Lambda_*} = \frac{r^2 \bar{\beta}}{4c}\), \(w_1^{\Lambda_*} = \frac{\bar{\beta}}{8c}\), and \(e_1^{\Lambda_*} = \frac{r^2 \beta}{2c}\) where \(\bar{\beta} \equiv E[\theta_B | c_{\beta,1}] = E[\theta_A | c_{1,\beta}]\) if \(c_{1,\beta}\) (resp. \(c_{\beta,1}\)) is chosen.

The optimal SPPM \(M_2^{\Lambda_*}\) imposes different levels of investment i.e., \(\tilde{k}_1^{\Lambda_*} < \tilde{k}_\beta^{\Lambda_*}\) (and compensations \(w_1^{\Lambda_*} < w_{\beta}^{\Lambda_*}\)) and separates \((\beta, 1)\) from \((1, \beta)\) without a cost i.e., by design \((\beta, 1)\) does not have the incentives to mimic \((1, \beta)\) and vice versa. There are still distortions
in \( M^\Lambda_{2*} \), however, since extreme types (1, 1) and (\( \beta, \beta \)) pool with either (\( \beta, 1 \)) or (1, \( \beta \)) types. In particular, there is a certain degree of overinvestment by type (1, 1) and underinvestment by (\( \beta, \beta \)) but since the \( c^*_\beta,1 \) are \( c^*_{1,\beta} \) are asymmetric they allow the asymmetric types (\( \beta, 1 \)) and (1, \( \beta \)) to self select themselves into the right contract and thus limit their investment distortions. As Proposition 5 indicates, this makes \( M^\Lambda_{2*} \) dominate \( M^P_{2*} \):

**Proposition 5** In the two-division case, the optimal SPPM \( M^\Lambda_{2*} \) generates greater firm value than the optimal pooling mechanism \( M^P_{2*} \).

Proposition 5 implies that it is never optimal for the two-division firm to use a pooling mechanism to allocate capital, and thus that the capital allocation choice of the two-division firm always reveals some information about its prospects. Intuitively this occurs because, relative to the optimal pooling \( M^P_{2*} \), the optimal SPPM \( M^\Lambda_{2*} \) allows the revelation of types (\( \beta, 1 \)) and (1, \( \beta \)) without incurring any additional efficiency loss, i.e., the IC constraints (26) and (27) do not bind in the optimal SPPM \( M^\Lambda_{2*} \). In fact, by comparing \( M^\Lambda_{2*} \) with \( M^P_{2*} \) it follows immediately that the investment, compensation and effort are better tailored to the private information i.e., the optimal values in \( M^\Lambda_{2*} \) are closer than the optimal values of \( M^P_{2*} \) to \( M^* \).

The comparison between \( M^\Lambda_{2*} \) and the optimal separating mechanism \( M^S_{2*} \) is less clear-cut. As stated in the following proposition, there exist cases in which \( M^\Lambda_{2*} \) strictly dominates \( M^S_{2*} \) and vice versa:

**Proposition 6** There are parameters under which \( M^\Lambda_{2*} \) generates more firm value than \( M^S_{2*} \) (i.e., \( V^*_{2*} - V^\Lambda_{2*} < 0 \)) and vice versa (i.e., \( V^*_{2*} - V^\Lambda_{2*} > 0 \)). The difference in firm value generated by \( M^\Lambda_{2*} \) relative to \( M^S_{2*} \), i.e., \( (V^\Lambda_{2*} - V^S_{2*}) \) is larger the smaller the difference in productivity among firm types (i.e., lower \( \beta \)).

Proposition 6 also implies that there are parameters in the space \( (\pi, \beta) \) for which \( M^{**} \) is an optimal mechanism in a single-division case and \( M^S_{2*} \) is not the optimal mechanism in the two-division case. More formally if we define \( R^*_\Lambda \) as the region in the space \( (\beta, \pi) \) for which \( V^\Lambda_{2*} \geq V^S_{2*} \) and \( R^*_p \) as the region in the space \( (\beta, \pi) \) for which \( V^P_{2*} \geq V^{**} \), it is immediate to conclude from the previous proposition that \( R^*_p \subset R^*_\Lambda \).

\(^{30}\)In particular: \((k^{\lambda'}_0, w^{\lambda'}_\beta, e^{\lambda'}_p) > (k^{\Lambda*}_\beta, w^{\Lambda*}_\beta, e^{\Lambda*}_p) > (k^{**}_0, w^{**}_\beta, e^{**}_p) > (k^{A*}_1, w^{A*}_1, e^{A*}_1) > (k^{*}_1, w^{*}_1, e^{*}_1)\).
The previous proposition also implies that conglomeration can improve firm value in the absence of technological synergies and informational links between divisions. Intuitively, the gain from conglomeration is directly related to the incentives to distort investment that arise when executives have private information about the divisions’ prospects. More specifically, note that, as in the single division case, the executive in the two-division firm has an incentive to overinvest to exaggerate each of the division’s prospects. However, in contrast to the single division case where the best the executive can do to eliminate his incentive to exaggerate is to use a pooling mechanism, in the two-division case, the executive can communicate some information by using mechanisms with asymmetric capital allocations. When properly designed, the mechanism communicates the ranking of the prospects of the two divisions without incurring overinvestment costs.

More specifically, as we show in Proposition 3, in the two-division case a pooling mechanism is dominated by a SPPM, i.e., a two-contract menu mechanism in which each contract asymmetrically allocates the same amount of total capital. By fixing the total amount of capital, the SPPM eliminates overinvestment, i.e., the incentive to exaggerate, while communicating some information, i.e., separating type (1, β) firms from type (β, 1). This insight can be related to the intuition derived in multidimensional cheap talk settings (e.g., Chakraborty and Harbaugh 2005) in which the informed party has information about multiple items and can credibly rank the items and communicate information that would not be credible if the informed party has information about a single item.

4.2.2 Optimal mechanisms

Following Bester and Strausz (2001) the search for the optimal mechanism can be restricted to mechanisms that include at most four distinct contracts $G_2 = \{c^0_1, c^0_2, c^0_{1,1}, c^0_{1,2}\}$ with $c^0_\Theta = \{(w^0_\Theta_A, k^0_\Theta_A, c^0_{\Theta,A}), (w^0_\Theta_B, k^0_\Theta_B, c^0_{\Theta,B})\}$. Potential mechanisms within this general class include (i) $M^s_2$, separating mechanisms, where each type $\Theta$ chooses a different contract $c^g_\Theta$, (ii) $M^p_2$, pooling mechanisms, where the contracts chosen by each type are the same, i.e., $c^g_\Theta = c^g_{\Theta'}$ for all $\Theta = \Theta'$ and (iii) $M^{pp}_2$, partial pooling mechanisms, where at least

---

31 Intuitively, relative to any pooling mechanism, it is possible to design a SPPM with a slight asymmetric capital allocation. Relative to any pooling, such a SPPM produces a negligible additional cost and a sizable additional benefit in effort allocation associated to the separation of types (1, β) and (β, 1).
two types choose the same contract in $G_2$ with a positive probability. Therefore, the search for the optimal mechanism requires us to obtain $G^*_2 = \{c^*_1, c^*_2, c^*_3, c^*_4\}$ and the probabilities that determine how different types mix among them. Thus, for each type $\Theta$, we define a probability simplex $\sigma_{\Theta, \Theta'} = \{p_{\Theta, \Theta'}\}$ where $p_{\Theta, \Theta'}$ represents the probability that a firm of type $\Theta$ chooses contract $c^*_d$ for all $\Theta, \Theta'$. Formally, the optimal mechanism $M^*_2 = \{G^*_2, \sigma_{\Theta, \Theta'}\}$ solves:

$$\max_{M^*_2} E(V_A + V_B \mid G_2, \sigma_{\Theta, \Theta'})$$

s.t.: $$e^*_d = \arg \max_e \{w_d^e E[\theta_d \mid M^*_2]e - \frac{1}{2} e^2 k^e_d\},$$

$$V_A^{1,c^*_1} + V_B^{1,c^*_1} \geq V_A^{1,c^*_0} + V_B^{1,c^*_0} \quad \text{for } \hat{\Theta} \neq (1, 1)$$

$$V_A^{1,c^*_3} + V_B^{1,c^*_3} \geq V_A^{1,c^*_0} + V_B^{1,c^*_0} \quad \text{for } \hat{\Theta} \neq (1, \beta)$$

$$V_A^{\beta,c^*_1} + V_B^{\beta,c^*_1} \geq V_A^{\beta,c^*_0} + V_B^{\beta,c^*_0} \quad \text{for } \hat{\Theta} \neq (\beta, 1)$$

$$V_A^{\beta,c^*_3} + V_B^{\beta,c^*_3} \geq V_A^{\beta,c^*_0} + V_B^{\beta,c^*_0} \quad \text{for } \hat{\Theta} \neq (\beta, \beta)$$

where $V_d^{\theta_d,c^*_d}$ refers to the division $d = A, B$ value when the division type is $\theta_d = \{1, \beta\}$ and a contract $c^*_d$ is chosen. For future reference we denote $M^{pp*}_2$ as the solution of the previous problem when the solution is restricted to be a partial pooling mechanism and describe the nature of the optimal mechanism $M^*_2$ in the following proposition:

**Proposition 7** There is a sizable region of parameters in the space $(\beta, \pi)$ for which the optimal mechanism $M^*_2$ is the optimal partial pooling mechanism $M^{pp*}_2$.

Proposition 7 implies that, for conglomerates, investment rigidities and limited revelation of information are an integral part of the optimal mechanism for a sizable region of parameters. In particular, if we define $R^*_A$ as the region in the space $(\beta, \pi)$ for which there are investment rigidities and limited revelation of information in two-division firms and $R^*_p$ as the region in the space $(\beta, \pi)$ for which there are rigidities and limited revelation of information in single-division firms, we obtain that $R^*_p \subset R^*_A$, i.e., some form of rigidities are more prevalent in conglomerates.

---

32 Consistent with Definition 4.2.1 a symmetric partial pooling mechanisms (SPPM) belongs to the class of partial pooling mechanisms.
Proposition 7 does not state the nature of the optimal contract for all possible combinations of parameters in the space \((\beta, \pi)\). In particular, we cannot exclude the possibility that for some parameters separation is the optimal mechanism. In particular while we can establish that in situations of costless separation (i.e., when \(\beta \geq \beta^*\)) the optimal mechanism is separating and while our simulations suggest that the optimal mechanism is also separating when \(\beta \rightarrow \beta^*\), we have been unable to analytically prove this result.

We conclude the analysis of the two-division firm by characterizing \(\pi^{\pi} \pi^2\), the optimal partial pooling mechanism when \(\pi^{\pi} \pi^2\) is indeed the optimal contract, i.e., \(\pi^{\pi} \pi^2 = \pi^{\beta^*} \pi^2\). The following proposition holds:

**Proposition 8** When \(\pi^{\pi} \pi^2\) is the optimal mechanism, a type \((1,1)\) firm: (1) is never revealed by \(\pi^{\pi} \pi^2\) and (2) overinvests relative to the full information case.

While the formal proof of Proposition 8 is in the appendix, its logic follows from the insight that the optimal separating mechanism, \(\pi^{\beta^*} \pi^2\) dominates any other mechanism in which a type \((1,1)\) firm is revealed. Intuitively this occurs because a type \((1,1)\) firm is revealed with positive probability only when there is some overinvestment buy all the other types’ contracts. Since higher investment is less costly for a type \(\beta\) division, mechanisms in which type \(\beta\) divisions overinvest are preferable to those mechanisms in which type 1 divisions overinvest. Indeed, what characterizes \(\pi^{\beta^*} \pi^2\) is precisely the fact that it is the more efficient mechanism among those in which overinvestment occurs only by type \(\beta\) divisions. As a result, a mechanism in which a type \((1,1)\) firm is revealed with positive probability would be dominated by \(\pi^{\beta^*} \pi^2\).

As established in Proposition 8, a two-division firm’s capital allocation is characterized by a specific type of distortion. In particular, the type \((1,1)\) firm pools with more productive types, and thus overinvests relative to the case with full information. In Section 6 we discuss the empirical relevance of this result.

### 5 Extensions

Our analysis can be extended in a number of directions. First, it can be extended to a setting with a continuum of types where the productivity shock \(\theta\) can take any value in
the interval $[1, \bar{\beta}]$.\textsuperscript{33} In this setting, we can show that the tendency toward overinvestment remains and that investment rigidities (i.e., pooling) are present in the optimal mechanism. An interesting property of the optimal mechanism in the continuum of types setting is that there is “pooling at the bottom” (i.e., a commitment by the principal to a fixed and identical-across-types investment expenditure i.e., $k$ by a positive measure of types at the bottom of the type distribution, i.e., $[1, \beta']$) as well as “pooling at the top” (i.e., a commitment to a fixed and identical-across-types investment expenditure i.e., $\bar{k}$, by a positive measure of types at the top of the type distribution, i.e., $[\beta'', \bar{\beta}]$). Since it can be shown that $k \leq \bar{k}$, the findings that arise in the continuum setting are consistent with the idea that firms require a positive NPV threshold before investing, but place a cap on the maximum level of investment.

The analysis can also be extended to examine how the participation of uninformed third parties can alter optimal investment rules.\textsuperscript{34} We have considered two alternative roles that third parties can play in our analysis. The first is the role of the mediator, as described in Myerson (1983), who chooses an outcome as a possibly random function of the information reported to him, and who privately and confidentially recommends a specific action to the parties involved. In this case, if a mediator restores the principal’s ability to commit in advance to investment plans that are ex-post inefficient, then the mediator’s presence leads to changes in investment allocations that improve firm value.\textsuperscript{35} In particular, it can be shown that the optimal mechanism under full commitment is as follows: The informed principal truthfully discloses his information to a mediator (but not to the agent). Conditional on the information communicated by the principal, the mediator produces a garbled signal, observed by the agent, which determines the allocation (i.e., investment level $k$ and agent compensation $w$) that is carried out by the principal. Thus when it is feasible to use a mediator in this way, investment rules will be less rigid, but will still convey a limited amount of information to the firm’s stakeholders.

\textsuperscript{33}Formally, rather than a two-type distribution the (common) prior distribution of $\theta$ is described by the density $f(\theta) > 0$, with $|f'| \leq M$ and cumulative distribution $F(\theta)$, for $\theta \in [1, \bar{\beta}]$. For brevity we omit the details of the analysis with a continuum of types which is available from the authors upon request.

\textsuperscript{34}See Baliga and Sjöström (2009) for an analysis of the effect of considering third parties in contract design.

\textsuperscript{35}Technically restoring full commitment by the principal implies that the revelation principle holds and thus that, without loss of generality, the search of optimal mechanism can be reduced to direct revelation mechanisms, i.e., a pooling mechanism cannot be optimal.
The second potential role of a third party is that of a budget breaker, i.e., a party who is uninvolved in the production or decision process, but who has a claim on the firm’s output. A budget breaker can broaden the contract space and alter the principal’s incentives in ways that can lead to more efficient investment choices.\textsuperscript{36} Specifically, the presence of the third party alters the incentive compatibility constraints of the principal in a way that allows him to commit to future actions that would not be incentive compatible without the third party.

Debtholders provide one example of a third party budget breaker that may be particularly relevant. In our unreported analysis,\textsuperscript{37} we considered a slightly modified setting in which the principal (before becoming informed) issues risky debt to finance the firm but still retains control over the firm’s investment and worker compensation. In this setting, we find that the principal’s incentives to overinvest are mitigated, leading to an optimal investment rule with fewer investment rigidities. More specifically, the separating mechanism becomes relatively more attractive since the firm’s debt obligations affect the principal’s incentive compatibility constraints by reducing the benefits of low prospect firms to mimic high prospect firms. Intuitively, there is a debt overhang effect that decreases the low type’s benefits from overinvesting more than it decreases the high type’s benefits, which, in turn, reduces the amount the high type must overinvest to separate.

The effect of debt on the design of optimal investment rules in this setting illustrates that firms that are run in the best interest of their owners can nevertheless find it useful to take on debt to curb overinvestment incentives. Specifically, it is possible to obtain similar conclusions to those in Jensen (1986), Stulz (1990), and Hart and Moore (1995) from a model in which firms are subject to information asymmetries rather than managerial agency conflicts.

6 Empirical implications

There are a number of potential avenues to examine the empirical relevance of our analysis. First, one can consider the prevalence of certain investment distortions, i.e., rigidities and overinvestment, in the capital investment process. For example, Cooper and Haltiwanger

\textsuperscript{36}Holmstrom (1982) first points out the value of third parties who by threatening to break the budget eliminate inefficiencies in team production.

\textsuperscript{37}The formal analysis is available from the authors upon request.
(2006) present evidence that firms have minimum investment thresholds and that large bursts of investment expenditures account for about 50% of total investment activity.\textsuperscript{38} Furthermore, in a survey of the capital allocation rules of manufacturing firms, Ross (1986) documents that in addition to these minimum thresholds, these firms impose caps on their expenditures.

Our model can also be used to motivate a cross-sectional analysis. Specifically, our theory predicts more investment distortions in more opaque organizations, where key employees and other stakeholders are less informed than top executives about the firm’s future outlook, and in firms whose stakeholders must take unobservable actions that are crucial to the firm’s success. We also note that these investment distortions are likely to be different in multi-segment versus single-segment firms.

The above discussion suggests that a study of investment distortions should consider firms along a number of different dimensions. For example, Bhide and Stevenson (1999) show that stakeholder relationships are particularly important for the development and survival of entrepreneurial ventures and for newly-created firms, so we might expect emerging new firms to distort their investment allocations more than mature firms. The issues we raise may also be more relevant for firms that are developing new technologies, where information asymmetries are inherently more important, and the incentives to convey information to stakeholders are higher, given the benefits of collaborating with firms with the most promising technologies.

Our model also makes predictions about the type of distortion (i.e., overinvestment versus investment rigidities) that are most likely to arise. In particular, according to the comparative statics results reported in Proposition 4, investment rigidities (i.e., pooling) tend to arise when differences between types are reduced (low $\beta$). Intuitively, if one takes the two-type case as an useful abstraction for a reality with a continuum of types (and thus the presence of types that are relatively close) this result suggests that rigidities are likely to be frequently observed. Proposition 4 also states that investment rigidities are more likely to arise when the ex-ante likelihood of a high productivity project (i.e., a high $\pi$) is higher. This is most likely to be the case for firms in industries with favorable growth opportunities.

\textsuperscript{38}Cooper and Haltiwanger (2006) show that investment at the plant level exhibits “lumpiness,” i.e, in 8% of their (plant, year) observations the investment rate (I/K) is near zero and in 18% of the sample observations the investment rate exceeds 20%.
and, more generally, during economic expansions.\textsuperscript{39}

Our model also suggests substantial differences between the investment distortions in single-division and multi-division firms. Although there are obvious endogeneity challenges, one may want to explore these issues by examining changes in firms’ investment distortions before and after mergers and spin-offs. Specifically, the analysis suggests that conglomerates should have capital budgeting processes that exhibit more abundant but less stringent distortions (i.e., partial pooling mechanisms are more likely to be optimal in conglomerates).

More specifically, the analysis predicts that after merging, firms should (i) eliminate some rigidities in investment rules if they formerly employed very rigid investment rules as single-division firms (i.e., pooling) and (ii) introduce some rigidities if they formerly employed very few rigidities in its policy (i.e., separation). Finally, the analysis suggests that multi-divisional firms would tend to avoid extreme capital allocations.

This last implication, i.e., avoidance of extreme capital allocations, is consistent with the empirical results documented in the conglomerates literature that suggest a certain degree of socialism in capital allocation.\textsuperscript{40} In particular the literature has stressed the tendency of better (lower) performing divisions in conglomerates being allocated less (more) capital than their single-divisions counterparts and has associated this behavior to potential agency problems such as managerial empire building tendencies or rent seeking behavior (see e.g., Rajan et al. 2000). As Proposition 7 suggests, when executives possess private information about firm prospects the optimal investment policy tends to overinvest when both divisions exhibit unfavorable prospects. It is worth mentioning that while the literature on investment distortions in conglomerates rely on managerial agency problems, in our setting the lack of sensitivity of investment to information arises in our setting even though firms are run in the best interest of their owners.

Testing the above predictions requires overcoming two major challenges. The first challenge is getting data on investment distortions. To address this challenge one can look at investment expenditures at the individual plant level, as in Cooper and Haltiwanger (2006),

\textsuperscript{39}As discussed in footnote 9, our results require a complementarity between the costly worker action \( \epsilon \) and firm prospects \( \theta \). An alternative model where \( \epsilon \) and \( \theta \) are substitutes can generate the opposite results. For example, a firm close to liquidation may want to underinvest to communicate to their employees that they will go out of business if workers do not contribute more effort. If this alternative interpretation is plausible, an empirical analysis of our implications should separately analyze healthy and distressed firms.

\textsuperscript{40}See Stein (2003) and references therein.
or at the firm or division level from accounting statements. The second challenge has to do with endogeneity issues. We expect firms to merge and spin-off divisions for reasons that relate to the investment process and when doing so creates value. However, this creates a potential opportunity as well as a challenge, since extensions of our model that considers the gains and costs of conglomerate are also likely to be of interest. Finally, it may be of interest to survey corporate executives about the extent to which they impose constraints on future investment choices.\footnote{We had private conversations with the CFO of a major corporation that allocates about $15 billion per year in CAPEX across a number of divisions. He stressed the fact that the firm is very cognizant of the effect of CAPEX on stakeholder perceptions, and how part of his job is to reign in the division heads that use worker motivation as an argument to get greater capital allocations. He said that for divisions with key employees that need to be retained, he takes the argument more seriously, but in most cases the perceptions are managed, but the CAPEX choice is not distorted. He said that the corporation implicitly commits to a maximum level of capital expenditures in their communications with analysts; if they invest more than the maximum, the firm would appear unreliable and their stock price will be hurt. This implicit commitment is useful because it gives him more credibility to manage the CAPEX allocations at the division level.}

7 Concluding remarks

Since the 1970s financial economists have studied the transmission of information from firms to financial markets. The focus of this paper is on the transmission of information within an organization, i.e., between the top executives of a firm, that set the firm’s overall strategy, and the lower level employees that actually do the work. In our model, strategic choices convey information, and because of this, the choices may be distorted by the firms’ executives. Specifically, in our top-down model of capital allocation, an unconstrained executive tends to overinvest, since higher levels of investment convey favorable information, which makes the firm’s employees work harder.

Of course, rational employees understand the firm’s incentive to overinvest, and can back out relevant information from the observed capital expenditures. As our model illustrates, the optimal mechanism for allocating capital is affected by the costs and benefits associated with communicating information to the workers. In some situations the benefits from a flexible investment policy, one that communicates information fully to workers, exceed the overinvestment costs. In other situations rules that limit overinvestment (and thus reduce the transmission of information to workers) can result in higher firm values.

Allocations can also be improved by expanding the scope of the firm, i.e., organizing
activities in multi-division conglomerates rather than as independent single division firms. Indeed, as we show, even when the prospects of the two entities are completely independent, firm value can be created by managing them as a single firm. The added value comes from an enhanced ability to convey information about the firm’s prospects (i.e., about the prospects of the individual divisions) without necessarily resorting to overinvestment or the implementation of rigid investment rules.

The idea that firms may want to organize themselves to influence the transmission of information to stakeholders can be extended in a number of ways. For example, rather than putting constraints on investment choices, the firm may want to obfuscate those choices, thereby reducing the incentive to overinvest. It may also be possible to improve firm value by delegating aspects of the investment decision process to parties that are less informed than the executive. For example, third parties, i.e., uninformed managers or consultants, can sometimes act as mediators, who muddle the information conveyed by the investment expenditures, or budget breakers, whose presence alters the executive’s payoffs, leading them to reduce investment distortions.

Finally, one can look beyond the employees’ effort choice and consider how strategic choices are influenced by a firm’s incentives to both retain and attract employees as well as other stakeholders, like customers and suppliers, whose perceptions are also key to the firm’s success. An important difference between these stakeholders and the workers considered in our model is that contracting on the contributions of these stakeholders are likely to be more challenging. While it seems reasonable to expect that our intuition about investment distortions will continue to hold in a setting with additional contracting imperfections, we leave an analysis of these issues for future research.
Appendix: Proofs and other technical derivations

This appendix is divided in two parts: (i) Proofs relative to Sections 2, 3 and 4; and (ii) Technical derivations relative to the analysis of the optimal mechanism.

Part 1: Propositions and results from Sections 2, 3 and 4

Proof of Proposition 1
Let $w_0 = \{w_{1,0}, w_{3,0}\}$ be the payments when $z = 0$ and $\theta = \{1, \beta\}$ and, as defined in the text, $w = \{w_1, w_3\}$ the payments when $z = r$ and $\theta = \{1, \beta\}$. Thus the principal’s problem is:

$$\max_{k,w,w_0,e} V = \sum_{\theta = \{1, \beta\}} p_\theta \left[ (r k_\theta - w_\theta) \theta e_\theta - w_{\theta,0}(1 - \theta e_\theta) - \frac{1}{2} k_\theta^2 \right]$$

s.t.: $e_\theta = \arg \max_e \{w_\theta e + w_{\theta,0}(1 - \theta e) - \frac{1}{2} e^2 k_\theta\}$, for $\theta = \{1, \beta\}$

$$w_\theta e_\theta + w_{\theta,0}(1 - \theta e_\theta) - \frac{1}{2} e_\theta^2 k_\theta \geq 0,$$

First, we prove $w_{\theta,0} = 0$ by contradiction. If $w_{\theta,0} > 0$ and $w_\theta = 0$ then setting $w_{\theta,0} = 0$ increases firm value and induces a higher level of agent’s effort. Alternatively, if $w_{\theta,0} > 0$ and $w_\theta > 0$ then reducing both $w_{\theta,0}$ and $w_\theta$ while keeping $(w_\theta - w_{\theta,0})$ constant increases firm value without affecting the agent’s effort incentives. Second, we impose $w_{\theta,0} = 0$, and obtain $k^*_\theta$, $w^*_\theta$ and $e^*_\theta$ by first substituting the first order condition of $e_\theta$ (36) and then solving in the first order conditions of $w_\theta$ and $k_\theta$.

Proof of Proposition 2
If $\beta \geq \beta^*$ then the unconstrained solution described in Proposition 1 satisfies $V^\beta_V \geq V^1_V$ (i.e., IC$\beta$) and $V^1_V \geq V^\beta_V$, i.e., IC$1$ which can be expressed as $\beta^*(2 - \beta) \geq 1$ or $\beta \geq \beta^*$. If instead $\beta < \beta^*$ the unconstrained solution does not satisfy IC$1$ which requires us to solve the general problem case in which $w_{\theta,0}^* \geq 0$. Let $w^*_\theta = \{w^*_1, w^*_3\}$ be the wage when $z = 0$ and $\theta = \{1, \beta\}$. Ignoring (by now) IC$\beta$ we get:

$$\max_{w^*_1, w^*_3, k^*_\theta, e^*_\theta} (1 - \pi)V^1_V + \pi V^\beta_V$$

s.t.: $e^*_\theta = \arg \max_e \{w^*_\theta e + w^*_{\theta,0}(1 - \theta e) - \frac{1}{2} e^2 k^*_\theta\}$ for $\theta = \{1, \beta\}$

$$w^*_\theta e^*_\theta + w^*_{\theta,0}(1 - \theta e^*_\theta) - \frac{1}{2} e^*_\theta^2 k^*_\theta \geq 0,$$

$$V^1_V \geq V^\beta_V$$

where $V^\beta_V \equiv r k^*_\theta \theta e^*_\theta - [w^*_\theta + (w^*_\theta - w^*_{\theta,0})\theta e^*_\theta] - \frac{1}{2} k^*_\theta^2$.

Notice that in the previous problem, type 1’s payoff is maximized as in the unconstrained case (i.e., $w^*_1 = 0$, $w^*_3 = w^*_1$, and $k^*_1 = k^*_1$) because any deviation (i.e., $w^*_1 \neq 0$, $w^*_1 \neq \frac{e}{2} k^*_1$, or $k^*_1 \neq k^*_1$) reduces $V^1_V$ without easing IC$1$. We impose such values and solve for type β’s optimal values:

$$\max_{w^*_{\theta,0}, w^*_1 k^*_3 e^*_\theta} (1 - \pi)V^1_V + \pi V^\beta_V$$

s.t.: $e^*_\theta = \arg \max_e \{w^*_\theta + (w^*_{\theta,0} - w^*_\theta)\beta e - \frac{1}{2} e^2 k^*_\theta\},$

$$V^1_V \geq V^\beta_V.$$
Expression (44) can be rewritten as:

$$V_1 = \frac{1}{2} k s^2 \geq \left( r k s - (w^* - w_{s,0}) \right) e_\beta - \frac{1}{2} k s^2 - w_{s,0} = V_1^\beta. \quad (45)$$

We define $\alpha, \beta \equiv \frac{w^* - w_{s,0}}{r k s}$, express (43) as $e_\beta = \frac{\alpha, r, \beta}{c}$ and plug them into the objective function:

$$\max_{w^*, k^s, \alpha, \beta} (1 - \pi) k s^2 + \pi (r k s (1 - \alpha) \beta) \frac{\alpha, r, \beta}{c} - \frac{1}{2} k s^2 - w_{s,0},$$

whose Lagrangian is:

$$L \equiv \frac{(1 - \pi) k s^2 + \pi \left\{ \frac{r k s (1 - \alpha) \beta}{c} - \frac{k s^2}{2} - w_{s,0} \right\} - \lambda \left\{ \frac{r k s (1 - \alpha) \beta}{c} - \frac{k s^2}{2} - w_{s,0} - \frac{k s^2}{2} \right\} + \mu w_{s,0}. \quad (46)$$

By Kuhn-Tucker Theorem, the FOCs with respect to $w_{s,0}, \alpha, \beta$, and $k^s$ are:

$$\lambda - \pi + \mu = 0 \quad (47)$$

$$(\pi, \beta - \lambda)(1 - 2\alpha) = 0 \quad (48)$$

$$(\pi, \beta - \lambda)r(1 - \alpha)\beta c^{-1} - (\pi, \beta - \lambda)k^s = 0 \quad (49)$$

Since by (47), $\pi, \beta - \lambda = \mu + (\beta - 1)\pi > 0$, then (48) implies $\alpha \beta = \frac{1}{2}$ and (49) $(\pi, \beta - \lambda)k^s > 0$. Thus $\pi - \lambda = \mu > 0$ which implies $w_{s,0} \geq 0$ is binding. Equation (49) implies that $k^s = \frac{(\pi, \beta - \lambda)r(1 - \alpha)\beta c^{-1}}{\pi, \beta - \lambda} > r(1 - \alpha)\beta c^{-1}$ (i.e., type $\beta$ overinvests). Solving for $k^s$ in the binding (45) after substituting $\alpha = \frac{1}{2}$ and $w_{s,0} = 0$, we get

$$rk^s \frac{r \beta}{4c} - \frac{1}{2} k^s - \frac{k^1}{2} = 0 \quad (50)$$

which has two real roots. But since $V_\beta = \beta r k^s \frac{r \beta}{4c} - \frac{1}{2} k^s = \frac{k^s}{1 + \sqrt{1 + 1/\beta}} k^s \beta - 1 + \frac{1}{2} k^s - \frac{k^1}{2}$, we get:

$$V_\beta = \frac{k^s}{1 + \sqrt{1 + 1/\beta}} (1 + \beta \beta) + \frac{k^1}{2} \quad (51)$$

which follows from (50). Since $V_\beta$ is increasing in $k^s$, the larger root is optimal i.e., $k^* = \frac{k^s}{1 + \sqrt{1 + 1/\beta}}$. Substituting $k^s = \frac{1 + \sqrt{1 + 1/\beta}}{1 + \sqrt{1 + 1/\beta}}$, we get $k^* = \frac{1 + \sqrt{1 + 1/\beta}}{1 + \sqrt{1 + 1/\beta}} k^s$. Notice that $\frac{1 + \sqrt{1 + 1/\beta}}{1 + \sqrt{1 + 1/\beta}} = 1$ has two real roots, 1 and $\beta$. However, since (i) $\beta$ is the largest real root, (ii) $\lim_{\beta \to \infty} \frac{1 + \sqrt{1 + 1/\beta}}{1 + \sqrt{1 + 1/\beta}} < 1$ and (iii) $\frac{1 + \sqrt{1 + 1/\beta}}{1 + \sqrt{1 + 1/\beta}} > 1$ for $\beta \to 1$, then $\frac{1 + \sqrt{1 + 1/\beta}}{1 + \sqrt{1 + 1/\beta}} > 1$ for $1 < \beta < \beta$. Finally, to check IC$_\beta$, notice that (51) and $V_\beta^1 = \frac{k^s}{2}$ implies that IC$_\beta$ holds if $k^s \beta \geq k^1$ which follows since $\beta > 1$ and $k^s > k^1$. The principal’s payoff (15) follows from plugging previous values in the objective function.

On the (sub)optimality of money burning mechanisms

Money burning arrangements are not optimal if and only if

$$c \leq \frac{\Delta^2 r^2}{2} \quad (52)$$
which is equivalent to $k^s_β ≤ \frac{r}{β}$, i.e., the marginal benefit of the high type firm from investing conditional on success is higher than the marginal cost.

Consider a separating mechanism $m$ include a third party payment. There are three possibilities of payment: (1) $g ≥ 0$ before investing so that $k_β = k_β^s - g$, (2) $g_τ ≥ 0$ after output success $r \cdot k_β$ and (3) $g_0 ≥ 0$ after failure $0$ . Building the Lagrangian as in (46) and taking first order condition, we get $g = 0$ and $g_τ = 0$. (Intuitively, $g$ and $g_τ$ payments cost the high type more and which make separation more costly.) The FOCs (47)-(49) in addition to the FOC for $g_0$ gives

$$(π_β - λ)e_β - (π - λ) + v = 0$$

(53)

where $e_β = \frac{αβrβ}{c}$ and $v ≥ 0$ is the multiplier for $g_0 ≥ 0$. (49) is

$$(π_β - λ)e_β(1 - α_β)r - (π - λ)k_β^s = 0.$$  

(54)

Substituting the optimal $α_β = \frac{1}{2}$ (see proof of Proposition 2), (53) and (54) imply:

$$(π - λ)(\frac{k_β}{r/2} - 1) = -v$$

(55)

There are two cases. First $v ≥ 0$, (55) implies $k_β^s ≤ \frac{r}{β}$ because (54) implies $π - λ > 0$ (otherwise $e_β ≤ 0$, which is either impossible or clearly suboptimal). In this case, $g_0 = 0$ and the solution is the same as that characterized in Proposition 2. In particular, $k_β^s = k_β^* ≤ \frac{r}{β}$, which is satisfied if (52) holds. On the other hand if (52) does not hold, we will have have $g_0 > 0$. Suppose not. We have $g_0 = 0$ and $k_β^s = k_β^*$ and thus $\frac{k_β^s}{r/2} - 1 > 0$ because (52) does not hold. This implies that the left hand side of (55) is positive, which implies $v < 0$, a contradiction. In summary, if (52) holds, our focus on the set of mechanisms considered in (11)-(14) is w.l.o.g.

**Proof of Proposition 3**

It follows directly from taking the FOC in the program (16)-(17).

**Proof of Proposition 4**

If $β ≥ β^*$ a separating mechanism dominates. By contrast, when $β < β^*$ a pooling mechanism dominates if and only if $V^*_p > V^*_s = \frac{r^2_β}{4} + π(1 - \frac{1}{β})Δk_β^2$, that is when $\frac{k_β^p}{k_β^s} - 1 ≥ π(1 - \frac{1}{β})Δk_β^2$.

Substituting $k_β^p = \frac{r^2_β}{4}β^2$, $k_β^s = \frac{r^2_β}{4}β^2$ and $Δ = \frac{β + \sqrt{β^2 - 1}}{2}$, we get $\frac{[1 + π(β - 1)^4]}{2} - 1 ≥ π(1 - \frac{1}{β})Δ^2(β + \sqrt{β^2 - 1})$ which holds when $[1 + π(β - 1)^4] - 1 - π(β - 1)[β^2 + β\sqrt{β^2 - 1}] ≥ 0$.

To sign $ζ$, fix $β$ and let $π^*(β)$ such that $ζ = 0$. Notice that $\frac{∂ζ}{π^*}\bigg|_{π^*=0} = (β - 1)[2 - β(β + \sqrt{β^2 - 1})] ≥ 0$ iff $β ≤ \frac{2}{\sqrt{3}}$. In this case, since $ζ(π)$ is convex in $π$, $\frac{∂ζ}{π^*} > 0$ for $π > β$. Since $ζ(π = 0) = 0$, it follows that $ζ > 0$ for all $π > π^*(β) = 0$ on $(1, \frac{2}{\sqrt{3}}]$.

If $β > \frac{2}{\sqrt{3}}$, then $\frac{∂ζ}{π^*}\bigg|_{π^*=0} < 0$ and thus $ζ < 0$ for $π$ close to $1$. Also $ζ(π = 1) > 0$ because $V^*_p(π = 1) = \frac{r^2_β}{4}$ achieves the full information payoff. Thus by continuity, $∃ π^*(β) > 0$ such that $ζ = 0$. Further, since $ζ(π)$ is convex, $ζ(π) ≤ 0$ for any $0 ≤ π ≤ π^*(β)$. In addition, because $0 = ζ(π) - ζ(0) = \int_0^π ζ(π) dπ < ζ(π^*) dπ = ζ(π^*) π^*$, then $\frac{∂ζ}{π^*} π^* > 0$ for all $π > π^*(β)$ by convexity and $ζ > 0$ for all $π > π^*(β)$. Thus (i) holds if $π > π^*(β)$.

34
Consider $\zeta(\beta, \pi^*(\beta)) = 0$. Simple algebra shows $\frac{\partial \zeta}{\partial \pi}|_{\pi = \pi^*(\beta)} < 0$ and the Implicit Function Theorem applied to $\zeta(\beta, \pi^*(\beta))$ gives: $\frac{d\pi^*}{d\beta} = -\frac{\partial \zeta}{\partial \beta}|_{\pi = \pi^*(\beta)} > 0$. Thus, $\pi^*(\beta)$ is strictly increasing on $(\frac{2}{V_0}, \beta^*)$.

Let $\beta^*(\pi)$: $(0,1) \rightarrow (\frac{2}{V_0}, \beta^*)$ be the inverse function of $\pi^*(\beta)$. For a fixed $\pi$, $\zeta(\beta) < 0$ iff $\beta^*(\pi) < \beta < \beta^*$. Now since $\zeta(\beta_1, \pi^*(\beta_1)) = 0$ and $\pi^*(\beta) > 0$, if $\beta_1 > \beta^*$ then $\pi^*(\beta_1) > \pi^*(\beta^*) = \pi$ and thus separating is optimal for $(\pi, \beta_1)$ by part (i). The case where $\beta_1 < \beta^*$ is similarly proved.

**Proof of Lemma 1**

It follows from the text.

**Proof of Lemma 2**

Consider the relaxed problem in which the only binding IC is the upward IC. That is, $(1,1)$ not mimicking $(1,\beta)$, $(\beta,1)$ or $(\beta,\beta)$. We will verify later that other ICs are satisfied. It follows that the type $(1,1)$ invests at the first best level as in the single division case and $w^{1,1}_d = 0$ for both divisions.

The value for type $(1,1)$ is $V^{1,1} = 2V_+^*$. We define as before $\alpha_d^\Theta = \frac{w_d^\Theta - w_{d,0}^\Theta}{r_k^d}$. The Lagrangian is

$$L = \sum_{\Theta} \Pr(\Theta)(V^\Theta_A + V^\Theta_B) + \lambda_{1,\beta} \left[ V^{1,1} - \sum_{d=A,B} \left( r_k^d (1 - \alpha_d^1 \beta) \alpha_d^1 r_{d,0} - \frac{1}{2}(k_d^1)^2 - w_{d,0}^1 \right) \right]$$

$$+ \lambda_{\beta,1} \left[ V^{1,1} - \sum_{d=A,B} \left( r_k^d (1 - \alpha_d^1 \beta) \alpha_d^1 r_{d,0} - \frac{1}{2}(k_d^1)^2 - w_{d,0}^1 \right) \right]$$

$$+ \lambda_{\beta,\beta} \left[ V^{1,1} - \sum_{d=A,B} \left( r_k^d (1 - \alpha_d^1 \beta) \alpha_d^1 r_{d,0} - \frac{1}{2}(k_d^1)^2 - w_{d,0}^1 \right) \right] + \sum_{\Theta} \mu_d^\Theta w_{d,0}$$

FOC for $w_{d,0}^{\beta,\beta}$ and $w_{d,0}^{\beta,1}$

$$- \Pr(\beta, \beta) + \lambda_{\beta,\beta} + \mu_{d}^{\beta,\beta} = 0$$

$$- \Pr(\beta, 1) + \lambda_{\beta,1} + \mu_{d}^{\beta,1} = 0,$$

$(\beta, 1)$ case is symmetrically handled. FOC for $\alpha_d^\Theta$ with $\Theta \neq (1,1)$,

$$(\Pr(\beta, \beta) - \lambda_{\beta,\beta})(1 - 2\alpha_d^{\beta,\beta}) = 0$$

$$(\Pr(\beta, 1) - \lambda_{\beta,1})(1 - 2\alpha_d^{\beta,1}) = 0,$$

$$(\Pr(\beta, 1) - \lambda_{\beta,1})(1 - 2\alpha_d^{\beta,1}) = 0,$$

which imply $\alpha_d^\Theta = \frac{1}{2}$. 

35
The FOC for \( k_d^\beta \)

\[
\begin{align*}
(Pr(\beta, \beta)\beta - \lambda_{\beta, \beta})\frac{r^2\beta}{4c} - (Pr(\beta, \beta) - \lambda_{\beta, \beta})k_d^\beta &= 0 \\
(Pr(\beta, 1)\beta - \lambda_{\beta, 1})\frac{r^2\beta}{4c} - (Pr(\beta, 1) - \lambda_{\beta, 1})k_A^1 &= 0 \\
(Pr(\beta, 1) - \lambda_{\beta, 1})\frac{r^2}{4c} - (Pr(\beta, 1) - \lambda_{\beta, 1})k_B^1 &= 0,
\end{align*}
\]

which imply \( \mu_d^\beta > 0 \), \( w_d^\beta = 0 \) and \( k_A^1 = k_B^1 = k_1^* \). In other words, the low type division should invest at its first best level. This result and the ICs thus imply that \( k_A^1 = k_B^1 = k_1^* = k_{1\beta}^* \). Finally, it is easy to check the other ICs are satisfied. 

**Proof of Lemma 3**

By Bayes rule

\[
Pr[\theta_A = \beta|c_{\beta, 1}] = \frac{\frac{1}{2}Pr(\beta, \beta) + Pr(\beta, 1)}{\frac{1}{2}Pr(\beta, \beta) + \frac{1}{2}Pr(1, 1) + Pr(\beta, 1)} = \pi(2 - \pi).
\]

(56)

\[
Pr[\theta_B = \beta|c_{\beta, 1}] = Pr[\theta_A = \beta|c_{1\beta}] = \pi^2,
\]

(57)

The case when \( c_{1\beta} \) is chosen is symmetrical. Thus we can define \( \overline{\theta}_\beta \equiv E[\theta_A|c_{\beta, 1}] = E[\theta_B|c_{1\beta}] \) and \( \overline{\theta}_1 \equiv E[\theta_B|c_{1\beta}] = E[\theta_A|c_{1\beta}] \). We now ignore all the ICs and solve for the contract that maximize each individual division value. Thus the problem becomes the optimal pooling mechanism as characterized by Proposition 3. we now show that ICs are satisfied (it is straightforward to check all the IRs are satisfied). First, it is clear that \( (\beta, \beta) \) and \( (1, 1) \) are indifferent between the two contracts by symmetry. Second, for \( (\beta, 1) \), the IC is

\[
\frac{r^2\overline{\theta}_\beta k_1^*}{4c} - \frac{k_1^*}{2}k_1^* \geq \frac{r^2\overline{\theta}_1 k_1^*}{4c} - \frac{k_1^*}{2}k_1^*.
\]

The above condition is equivalent to

\[
\frac{r^2\overline{\theta}_\beta k_1^*}{4c} - (\beta - 1) \geq \frac{r^2\overline{\theta}_1 k_1^*}{4c} - (\beta - 1),
\]

which is clearly satisfied because \( k_1^* > k_1^* \) and \( \overline{\theta}_\beta > \overline{\theta}_1 \). The case for \( (1, \beta) \) is proved symmetrically.

**Proof of Proposition 5**

Proposition 3 shows that a division value under optimal pooling mechanism is

\[
V_p^*(\overline{\theta}) = \frac{1}{2}k^2 \text{ where } k^2 = \frac{r^2\overline{\theta}^2}{4c}.
\]

Here \( \overline{\theta} = \pi\beta + (1 - \pi) \), the expected value of \( \theta \). \( V_p^*(\overline{\theta}) \) is convex in \( \overline{\theta} \). Under the partial pooling mechanism for combined divisions, each division's expected value is

\[
E[V_{d\beta}] = \frac{1}{2}V_p^*(\overline{\theta}_\beta) + \frac{1}{2}V_p^*(\overline{\theta}_1)
\]

36
Under $\Lambda_2^*$, by Bayes rule, $Pr[\theta_A = \beta | c_{\beta,1}^*] = \pi(2 - \pi)$ and $Pr[\theta_B = \beta | c_{\beta,1}^*] = \pi^2$. That is, $\frac{1}{2}\overline{\varphi}_\beta + \frac{1}{2}\overline{\varphi}_1 = \overline{\varphi}$. Jensen’s inequality yields

$$E[V_d] > V_p^*(\overline{\varphi}).$$

Note that the inequality is strict because $V_p^*(\overline{\varphi})$ is strictly concave.

**Proof of Proposition 6**

We have the expected value for each division under $\Lambda_2^*$

$$V^{\Lambda^*} = E[V_d] = \frac{1}{2}V_p^*(\overline{\varphi}_\beta) + \frac{1}{2}V_p^*(\overline{\varphi}_1) = \frac{\pi^4}{64c^2}(\overline{\varphi}_\beta^4 + \overline{\varphi}_1^4)$$

where

$$\overline{\varphi}_\beta = 1 + \pi(2 - \pi)(\beta - 1)$$
$$\overline{\varphi}_1 = 1 + \pi^2(\beta - 1)$$

Under $S_2^*$, the expected value of each division is

$$V^{**} = \frac{1}{2}[(1 - \pi)k_{1}^* + \pi\Delta(2 - \Delta)k_{\beta}^2]$$

by Proposition 2. First, because $V^{\Lambda^*} > V_p^*(\overline{\varphi})$ by Proposition 5, it is impossible for $V^{\Lambda^*} < V^{**}$ for $\beta < \beta^*$ by the proof of Proposition 4. Second, $V^{\Lambda^*} < V^{**}$ at $\beta = \beta^*$ because $V^{**}$ achieves first best level at $\beta = \beta^*$ while $V^{\Lambda^*}$ is bounded away from first best because there is non-trivial pooling between the types. To see this, note that by Jensen’s inequality, we have $\overline{\varphi}_\beta^4 < \pi(2 - \pi)\beta^4 + (1 - \pi(2 - \pi))$ and $\overline{\varphi}_1^4 < \pi^2\beta^4 + (1 - \pi^2)$. It follows that $V^{\Lambda^*} < \frac{\pi^4}{64c^2}(\pi\beta^4 + (1 - \pi))$, the first best value of a division. As a result, there must exists a $\beta_{A}^h \in (\beta^{**}, \beta^*)$ so that $V^{\Lambda^*} < V^{**}$ for $\beta > \beta_{A}^h$; and a $\beta_{A}^l \in (\beta^{**}, \beta^*)$ so that $V^{\Lambda^*} > V^{**}$ if $\beta < \beta_{A}^l$.

**Proof of Proposition 7**

It follows from Propositions 5 and 6.

**Proof of Proposition 8**

Suppose $(1, 1)$ fully reveals itself with positive probability in a mechanism $\Upsilon_2$, we will prove the proposition by showing that the payoff of all types will be (weakly) dominated by the optimal separating mechanism. Suppose $c_p$ is a contract that some or all of the types, $(\beta, 1), (1, \beta), (\beta, \beta)$ and $(1, 1)$ may choose with a positive probability. As a result, the posterior belief of each division conditional on $c_p$ is $\pi_A$ and $\pi_B$. Denote $\overline{\varphi}_d$ the expected value of $\theta_d$ of division $d$ conditional on $c_p$ been chosen. To ease notation, we drop $\Upsilon$ and $\Upsilon_2$

Now we study the following relaxed problem,

$$\max_{w_{d,0}, w_{d,1}, k_{d,e_d}} E(V_A + V_B | c_p)$$

37
subject to
\[
e_d = \arg\max_e \{ w_{d,0} + (w_d - w_{d,0})E[\theta_d|c_p]e - \frac{V}{2}ce^2k_d \},
\]
\[
V_A^1 + V_B^1 \geq V_A^{1,c_p} + V_B^{1,c_p},
\]
\[
\kappa = \pi_A + \pi_B,
\]
\[
w_d \geq 0, w_{d,0} \geq 0, \pi_A \geq 0, \pi_B \geq 0, \pi_A \leq 1, \pi_B \leq V.
\]

That is, we maximize the value of the firm conditional on \(c_p\) subject to (1, 1) has no incentive to choose \(c_p\) over revealing itself. Here we simplify notation by denoting \(V_A^1\) and \(V_B^1\) the division value of (1, 1) if it chooses to reveal itself in the mechanism. Furthermore, we allow for the additional flexibility that \(\pi_A\) and \(\pi_B\) can change but the sum is fixed. Because this problem allows more flexibility and has less constraints than in \(\Upsilon_2\), the solution should give a higher firm value than that in \(\Upsilon_2\). The Lagrangian is
\[
L = \left(\frac{r^2k_A(1 - \alpha_A)\alpha_A\overline{\pi}_A^2}{c} \frac{1}{2}(k_A)^2 - w_A,0\right) + \left(\frac{r^2k_B(1 - \alpha_B)\alpha_B\overline{\pi}_B^2}{c} \frac{1}{2}(k_B)^2 - w_A,0\right)
+ \lambda \left[V_A^1 + V_B^1 - \sum_{d=A,B} \left(\frac{r^2k_d(1 - \alpha_d)\alpha_d\overline{\pi}_d^2}{c} \frac{1}{2}(k_d)^2 - w_d,0\right)\right] + \delta(\kappa - \pi_A - \pi_B)
+ \sum \mu_d w_{d,0}
\]
The same FOC argument as before yields \(w_{d,0} = 0\) and \(\alpha_d = \frac{1}{2}\). FOC for \(k_d\) is
\[
k_d = \frac{r^2\overline{\pi}_d^2(\overline{\theta}_d - \lambda\overline{\pi}_d)}{1 - \lambda}
\]
Since by assumption (\(\beta < \beta^*\)) the IC is binding, we have \(0 < \lambda\) and \(k_d > \frac{r^2\overline{\pi}_d^2}{4c^2} = \overline{k}_d(\overline{\theta}_d)\) the optimal pooling investment level. (60) implies that if \(\overline{\theta}_A > \overline{\theta}_B\) then \(k_A > k_B\).

Now we look at FOC for \(\pi_A\) and \(\pi_B\). Note that \(\overline{\theta}_d = 1 + \pi_d(\beta - 1)\). Assuming interior solutions for \(\pi_A\) and \(\pi_B\), we have
\[
\frac{r^2k_A(\beta - 1)}{4c}(2\overline{\theta}_A - \lambda) = \frac{r^2k_B(\beta - 1)}{4c}(2\overline{\theta}_B - \lambda) = \delta
\]
Because \(k_1 > k_j\) if \(\overline{\theta}_i > \overline{\theta}_j\), the above condition can hold only if \(\pi_A = \pi_B\). We next show that the second order condition for \(\pi_A = \pi_B\) does not hold. We perturb the solution by \(\pi_A + \Delta\pi\) and \(\pi_B - \Delta\pi\) and keep \(k_d\) the same. It is clear that (59) holds. For the binding (58), first order variation for the left hand side is zero. The first order variation of the right hand side is
\[
\frac{\Delta\pi r^2k_A(\beta - 1)}{4c} - \frac{\Delta\pi r^2k_B(\beta - 1)}{4c} = 0
\]
because \(k_1 = k_2\). So (58) holds with first order approximation. By the second order necessary conditions a local maximum implies that
\[
(\Delta\pi, -\Delta\pi)D^2L(\Delta\pi, -\Delta\pi)^T \leq 0
\]
here $D^2L$ is the Hessian of the Lagrangian with respect to $\pi_1$ and $\pi_2$ at the local maximum. But

$$
(\Delta \pi, -\Delta \pi) D^2L(\Delta \pi, -\Delta \pi)^T = (\Delta \pi, -\Delta \pi) \begin{bmatrix}
\frac{\partial^2 L}{\partial \pi^2} & \frac{\partial^2 L}{\partial \pi \partial \pi_B} \\
\frac{\partial^2 L}{\partial \pi \partial \pi_B} & \frac{\partial^2 L}{\partial \pi_B^2}
\end{bmatrix} (\Delta \pi, -\Delta \pi)^T
$$

$$
= (\Delta \pi, -\Delta \pi) \begin{bmatrix}
\frac{r^2 k_A (\beta - 1)^2}{2c} & 0 \\
0 & \frac{r^2 k_B (\beta - 1)^2}{2c}
\end{bmatrix} (\Delta \pi, -\Delta \pi)^T
$$

$$
= \frac{r^2 k_A (\beta - 1)^2}{2c} \Delta \pi^2 + \frac{r^2 k_B (\beta - 1)^2}{2c} \Delta \pi^2 > 0
$$

So (??) does not hold and $\pi_A = \pi_B$ cannot be a solution. The only possibility is corner solution for $\pi_A$ and $\pi_B$. Without loss of generality, there are two cases: first, $\pi_A = 1$ and $\pi_B = 0$; Second, $\pi_A = 0$ and $\pi_B \leq 1$ (here we assume the fully-revealing division is division A).

For the first case, we fix division A allocation (and value) and try to improve division B value. That is, we focus on a single division problem. In particular, we relax the single division problem by allowing the high type in division B to separate with probability $p_\beta$ by choosing contract $c^\beta$ as long as the low type does not want to mimic. All the other types pool on contract $c_1$.

There are two cases:

- Without loss of generality, there are two cases: first, $\pi_A = 1$ and $\pi_B = 0$; Second, $\pi_A = 0$ and $\pi_B \leq 1$ (here we assume the fully-revealing division is division A).

Now we further improve the solution by letting $V_B^1$ and $V_B^{1, c_B}$ to change as long as the sum is $V_B^{1, c_B}$ to maximize all the types that choose $c_p$

$$
\max_{k_B^a, c_B^p} E[V_A^\beta + \pi_B V_B^{c_B^\beta} + (1 - \pi_B) V_B^{1, c_B}]
$$
subject to

\[ e_d = \max_e \left\{ \frac{r}{2} k_d E[\theta] e - \frac{1}{2} cc^2 k_d \right\}, \]

\[ V^{1,1} \geq V^{1,\beta}_A + \pi_B V^{1,c_B} + (1 - \pi_B) V^{1,c_B}_B \]

\[ V^{1,c_B}_B = V^{1,c_B}_B \]

Here we \( V^+_A = V^{1,\beta}_A \) and \( V^-_B = V^{1,c_B}_B \). The Lagrangian is

\[ L = \left( \frac{r^2}{4c} \beta^2 k_A - \frac{1}{2} k_A^2 \right) + \pi_2 \left( \frac{r^2}{4c} \beta^2 k_{c^B} - \frac{1}{2} k_{c^B}^2 \right) + (1 - \pi_2) \left( \frac{r^2}{4c} k_{c_B} - \frac{1}{2} k_{c_B}^2 \right) + \lambda (V^{1,1} - \left( \frac{r^2}{4c} \beta k_A - \frac{1}{2} k_A^2 \right) - \pi_B \left( \frac{r^2}{4c} \beta k_{c^B} - \frac{1}{2} k_{c^B}^2 \right) - (1 - \pi_B) \left( \frac{r^2}{4c} k_{c_B} - \frac{1}{2} k_{c_B}^2 \right) ) \]

FOC yields

\[ \frac{\partial L}{\partial k_A} = \frac{r^2}{4c} \beta^2 - k_A - \lambda \left( \frac{r^2}{4c} \beta - k_A \right) = 0 \]

\[ \frac{\partial L}{\partial k_{c^B}} = \pi_B \left( \frac{r^2}{4c} \beta^2 - k_{c^B} \right) - \pi_B \lambda \left( \frac{r^2}{4c} \beta - k_{c^B} \right) = 0 \]

\[ \frac{\partial L}{\partial k_{c_B}} = (1 - \pi_B)(1 - \lambda) \left( \frac{r^2}{4c} - k_{c_B} \right) = 0 \]

This yields \( k_A = k_{c^B} \) and \( k_{c_B} = \frac{r^2}{4c} = k_1^* \). That is, the low type division choosing \( c_p \) never has to signal and the high type division signals at the same level. Since the optimal \( V^{1,1} = 2V^*_1 \), the optimal symmetric information value of \((1, 1)\), this implies that the high type division invests at the optimal single division separating level, \( k_1^* \). As a result, the best outcome achievable in this case is dominated by the optimal separating mechanism. For \( \pi_A = 0 \) and \( \pi_B \leq 1 \), similar analysis yields that it is dominated by the optimal separating mechanism.\( \blacksquare \)

Part 2: The optimal mechanism (technical derivations)

The results in Proposition 4 confirm that the Revelation Principle (Laffont and Green, 1977, Myerson, 1979, and Dasgupta, Hammond, and Maskin, 1979) does not hold in this setting. Since the agent’s effort choice maximizes the agent’s payoff conditional on the principal’s announcements, solving for the optimal mechanism requires an additional incentive compatibility constraint which breaks the equivalence between a direct mechanism, in which the principal announces the type, and the indirect mechanism in which the principal does not. As it turns out, the optimal mechanism features either full separation (i.e., a full disclosure of the private information) or full pooling (i.e., a complete absence of disclosure). To prove this property, we first show that w.l.o.g. the mechanism design problem can be solved with a menu of two contracts. Specifically, we denote a contract as \( c_i \equiv (w_i(q), k_i) \) and consider a general menu of contracts \( \{c_i\}_{i \in I} \) where \( I \) may be infinite (and through which types may not be fully revealed). Let \( \psi(\theta) \) be the set of contracts that \( \theta \) chooses with positive probability. Let \( w_{i,0} \equiv w_i(0) \) and \( w_{i} \equiv w_i(rk_i) \) and let the probability of each type
choosing $c_i$, be $\pi_\theta(c_i)$ for $c_i \in \psi(\theta)$. Then the principal’s problem is

$$\max_{c_i, \epsilon \in I} \sum_{\epsilon \in (0, \theta)} \Pr(\theta) \sum_{c_i \in \psi(\theta)} \pi_\theta(c_i)V_\theta(c_i)$$

s.t.:

$$e_i \in \arg \max_{e} \{w_{i,0} + (w_i - w_{i,0})E[\theta|c_i] - h(e, k_i)\},$$

$$w_{i,0} + (w_i - w_{i,0})E[\theta|c_i] - h(e, k_i) \geq 0,$$

$$V_\theta(c_i) \geq V_\theta(c_j), \text{ for any } c_i \in \psi(\theta)$$

$$\sum_i \pi_\theta(c_i) = 1,$$

$$w_i \geq 0, w_{i,0} \geq 0.$$

We define $\alpha_i \equiv \frac{w_{i,0} - w_{i,0}}{w_{i,0}}$ and $\alpha_i \equiv rk_i(1 - \alpha_i)c_i$ and state the following monotonicity result:

**Lemma 4 (Monotonicity)** Any feasible mechanism requires that: (i) If $\theta > \theta'$ picks $c_i$ with positive probability and $\theta'$ picks $c_j$ with positive probability, then $a_i \geq a_j$; and (ii) If $\theta > \theta'$ weakly prefers $c_j$ to $c_i$ with $a_i > a_j$, then $\theta'$ strictly prefers $c_j$ to $c_i$; if $\theta' < \theta'$ weakly prefers $c_i$ to $c_j$ with $a_i > a_j$, then $\theta$ strictly prefers $c_j$.

**Proof.** (i): IC constraints imply: $a_i - \frac{1}{2}k_1^2 - w_{i,0} \geq a_j - \frac{1}{2}k_2^2 - w_{i,0}$ and $a_i\theta' - \frac{1}{2}k_2^2 - w_{i,0} \geq a_j\theta' - \frac{1}{2}k_1^2 - w_{i,0}$. Adding them up: $(a_i - a_j)\theta \geq 0$.

(ii): $\theta'$s preference implies: $\frac{1}{2}k_1^2 - \frac{1}{2}k_2^2 + \max(a_i - a_j) > \theta(a_i - a_j)$ which implies that $\theta'$ strictly prefers $c_j$. The other case follows by a similar argument.

**Lemma 5** For any multi-contract mechanism, there exists another mechanism with at most two contracts that is no worse for each type.

**Proof:** We distinguish two cases:

Case 1: Any multi-contract mechanism with $c_\theta$, $\theta \in \{1, \beta\}$ in which $c_\theta$ is chosen by type $\theta$ only, is no better than a two-contract separation mechanism, $c_\theta, \theta \in \{1, \beta\}$ in which $\theta$ type only chooses $c_\theta$ with probability $1$. Since for each type, $V_\theta(c_\theta)$ is the best value achieved in the original mechanism which is achieved in the separation mechanism with probability $1$. Notice that $c_\theta$ satisfies all the constraints from (66) to (70).

Case 2: A multiple contract mechanism in which one type always pool with the other (and the pooling type is $\beta$), i.e., type 1 chooses contracts in $\psi(\beta)$ with positive probability $\pi_1(\psi(\beta)) \leq 1$. By Lemma 4, if $c_1, c_j \in \psi(\beta)$, then $a_1 = a_j = a$. Consider mechanism $c'_p$ in which both types pool. Pick a contract, $c_p$, in $\psi(\beta)$ such that $E[\theta|c_p] \leq E[\theta|c_j]$, where $E[\theta|\psi(\beta)] \equiv E[\theta|c_j](\psi(\beta))$, is the conditional expectation of $\theta$ on any contract $c_j \in \psi(\beta)$ is chosen. We then modify $c_p$ to $c'_p.$ We let type $\beta$ choose $c'_p$ with probability 1 and type 1 choose it with probability $\pi_1(\psi(\beta)).$ Thus, $E[\theta|c'_p] = E[\theta|\psi(\beta)].$ We modify $k'_p$ so that $a'_p = k'_p(1 - a_p) = \frac{k_p(1 - a_p)}{\pi_1(\psi(\beta))} = k_p(1 - a_p)\pi_1(\psi(\beta)).$ Since $E[\theta|c'_p] = E[\theta|c_p], k'_p \leq k_p.$ We then increase $w_{p,0}$ to $w'_{p,0}$ so that $\frac{1}{2}k_p^2 + w_{p,0} = \frac{1}{2}k'_p^2 + w'_{p,0}.$ Thus $c'_p = \{k'_p, a'_p, w'_{p,0}\}.$ By construction, both types’ payoff from choosing $c'_p$ are the same as that of choosing a contract in $\psi(\beta)$ in the original mechanism: For $\theta = 1$: If $\pi_1(\psi(\beta)) = 1,$ then the original mechanism is the same as this one contract, $c'_p$ with both types pooling on it; If $\pi_1(\psi(\beta)) < 1,$ then there exists a contract $c_1$ chosen by type 1 with positive probability in the original mechanism. We now keep $c_1$ in the new mechanism and let type 1 chooses it with probability $1 - \pi_1(\psi(\beta)).$ The new mechanism has two contracts, $c'_p$ and $c_1$ and type 1 is indifferent between the two contracts and type $\beta$ strictly prefers $c'_p$ and each type gets the same payoff as in the original mechanism. (The case where the pooling type is 1 is similarly proved.)
**Proposition 9** The principal’s expected payoff is maximized under pooling or separating mechanism.

**Proof:** We focus on $\beta < \beta^*$ (since if $\beta \geq \beta^*$, separation achieves the full information payoff). By Lemma 5, we consider only two contracts as in the separation case. However, a principal may select the contracts randomly, i.e., revealing information partially. Denote $c_{\theta}$ the contract a type $\theta$ is more likely to choose. Now the mechanism is $(c_{\theta}, \sigma_{\theta})$ for $\theta = 1, \beta$. $\sigma_{\theta}$ is the probability that a type $\theta$ chooses $c_{\theta}$. The value of a type $\theta$ from choosing $c_{\theta}$ is $V_{\theta}^\theta$. Formally the contract design problem is

$$\max_{w_{\beta,0}, w_{\beta}, k_{\beta}, \sigma_{\theta}} \sum_{\theta=1: \beta} p_{\theta} V_{\theta}^\theta$$

s.t.: 

- $e_{\theta} \in \arg\max_e \{ w_{\theta,0} + (w_{\theta} - w^{\theta}_{\theta,0})E[\theta|\hat{\theta}]e - k_{\beta} f(e) \}$, 

- $w_{\theta,0} + (w_{\theta} - w^{\theta}_{\theta,0})E[\theta|\hat{\theta}]e_{\theta} - k_{\beta} f(e_{\theta}) \geq 0$, 

- $V_{\theta}^\theta \geq V_{\theta}^\theta$, for any $\hat{\theta} \neq \theta$ \hspace{1cm} (74)

- $(1 - \sigma_{\theta})[V_{\theta}^\theta - V_{\beta}^\theta] = 0$, for $\hat{\theta} \neq \theta$. \hspace{1cm} (75)

This problem adds two features to the separation case. First, the mechanism can be random. Second, there is a new constraint, (75), which is the complementary slackness constraint, i.e., if type $\theta$ is indifferent between the two contracts, randomization can occur. The previous program includes as particular cases full separation (i.e., $\sigma_{\theta} = 1$) and full pooling mechanisms (i.e., $c_{\theta}$ is the same for all $\theta$). There are three possible cases: (i) Only IC$\beta$ in (74) binds; (ii) Only IC$1$ in (74) binds; and (iii) Both ICs in (74) bind.

Case (i): If this case IC$1$ can be ignored which is a contradiction with the fact that the full information allocation in subsection 3.1 satisfies all the constraints of the problem.

Case (ii): In this case, the type 1 uses a mixed strategy in choosing the two contracts, and the type $\beta$ always chooses $(w_{\beta,0}, w_{\beta}, k_{\beta})$. Because when type 1 chooses $c_{1}$ she is fully revealed, the allocation in this case should be the same as in the full information case to relax IC$1$ as much as possible as argued in the proof of proposition 2. Formally, the program is

$$\max_{w_{\beta,0}, w_{\beta}, k_{\beta}, \sigma_{1}} (1 - \pi)V_{1}^1 + \pi V_{\beta}^\beta$$

s.t.: 

- $e_{\beta} \in \arg\max_e \{ w_{\beta,0} + (w_{\beta} - w^{\beta}_{\beta,0})E[\theta|\hat{\theta} = \beta]e - k_{\beta} f(e) \}$, 

- $V_{1}^1 \geq V_{\beta}^\beta$, 

- $w_{\beta} \geq 0$, $w_{\beta,0} \geq 0$ 

- $k_{\beta} \geq 0$, 

- $0 \leq \sigma_{1} \leq 1$. \hspace{1cm} (80)

By Bayes rule, $E[\theta|\hat{\theta} = \beta] = 1 + \frac{\pi}{\pi + (1 - \pi)(1 - \alpha_{\beta})}(\beta - 1)$. Let $E[\theta|\hat{\theta} = \beta] = v_{\beta}$, $v \in [\beta^{-1}, 1]$. We can change the control variable to from $\sigma_{1}$ to $v$ and consider the Lagrangian:

$$L = \frac{(1-\pi)k_{1}^2}{2} + \pi (\frac{rk_{3}(1-\alpha_{\beta})\alpha_{3}r\kappa_{3}^{2}}{c}g(k_{3})w_{\beta,0}) - \lambda (\frac{rk_{3}(1-\alpha_{\beta})\alpha_{3}r\kappa_{3}^{2}}{c}g(k_{3})w_{\beta,0} - \frac{k_{1}^{2}}{2}) + \mu w_{\beta,0} + v k_{\beta} - \xi(v-1).$$
By Kuhn-Tucker Theorem, we get the following FOCs with respect to \( w_{\beta,0}, \alpha_{\beta}, k_{\beta}, \) and \( v \):

\[
\begin{align*}
\lambda - \pi + \mu &= 0, \quad (82) \\
(\pi \beta - \lambda)(1 - 2\alpha_{\beta}) &= 0, \quad (83) \\
(\pi \beta - \lambda)r(1 - \alpha_{\beta})\frac{\alpha_{\beta}r\psi}{c} - (\pi - \lambda)k_{\beta} + v &= 0, \quad (84) \\
(\pi \beta - \lambda)rk_{\beta}(1 - \alpha_{\beta})\frac{\alpha_{\beta}r\beta}{c} - \xi &= 0. \quad (85)
\end{align*}
\]

A similar argument as in the proof of Proposition 2 implies \( \pi \beta - \lambda > 0 \) which it turns implies \( \xi > 0, v = 1 \) and \( \sigma_1 = 1 \). Hence, the optimal solution is the separation case considered in the text.

Case (iii): In this case the solution would be inferior to the optimal pooling mechanism considered section 3.2.4. Since both types are indifferent between the two contracts and the objective function only depends on the contract for type 1 (i.e., \( \max_{w_{1,0}, w_1, k_1, \sigma_0} \sum_{\theta = 1, \beta} p_\theta V_\theta^1 \)). Thus, w.l.o.g. \( E[\theta | c_1] \leq E[\theta] \) (i.e., if the type 1 contract is chosen, the agent’s posterior expectation of \( \theta \) is lowered) because there is always contract such that this holds for type 1. Now this contract is (weakly) inferior to the optimal pooling mechanism because otherwise we could define \( (w_{1,0}, w_1, k_1) \) as the optimal pooling contract. In this case, the agent would exert a (weakly) higher effort in the optimal pooling mechanism than here when \( (w_{1,0}, w_1, k_1) \) is chosen since \( E[\theta | c_1] \leq E[\theta] \). The contract satisfies the agent’s IR in the optimal pooling case too because (73) is satisfied and \( E[\theta | c_1] \leq E[\theta] \). Therefore, under \( c_1 \), the principal has a better payoff than under \( (w^{\beta*}, k^{\beta*}) \), which is a contradiction.
References


