GOVERNANCE AND THE EFFICIENCY
OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 318

Hierarchical Structures and Dynamic Incentives

Dongsoo Shin*
Roland Strausz**

*Department of Economics, Leavey School of Business, Santa Clara University, Santa Clara, CA 95053, E-mail: dshin@scu.edu.
**Humboldt-Universität zu Berlin, Institute for micro economic theory, Spandauer Str. 1, D-10178 Berlin (Germany), Email: strauszr@wiwi.hu-berlin.de.

April 2010

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.sfbtr15.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Dr. Urs Schweizer. · Department of Economics · University of Bonn · D-53113 Bonn,
Phone: +49(0228)739220 · Fax: +49(0228)739221

Hierarchical Structures and Dynamic Incentives*

Dongsoo Shin† Roland Strausz‡

April 15, 2010

*We acknowledge financial supports from the Deutsche Foschungsgemeinschaft (SFB-TR 15) and Leavey Grant. We thank Susanne Ohlendorf and Andras Niedermayer for their comments.

†Department of Economics, Leavey School of Business, Santa Clara University, Santa Clara, CA 95053, **E-mail:** dshin@scu.edu.

‡Humboldt-Universität zu Berlin, Institute for micro economic theory, Spandauer Str. 1, D-10178 Berlin (Germany), **Email:** strauszr@wiwi.hu-berlin.de.

Hierarchical Structures and Dynamic Incentives

ABSTRACT

We study the optimal hierarchical structure of an organization under limited commitment. The organization cannot make a long term commitment to wages and output levels, while it can commit to its hierarchical structure. We show that the optimal hierarchical structure is *horizontal* when it is highly likely that the employees are efficient or inefficient. By contrast, when such likelihood is intermediate or output does not expand very fast over time, the optimal hierarchical structure is *vertical* — with a vertical hierarchy, the organization can mitigate dynamic incentive problems linked to limited commitment.

JEL Classification: D82, D86

Key words: Dynamic Incentives, Organization Design

1 Introduction

An organization's efficiency depends on its hierarchical structure, which in turn defines authorities and information flows within the organization. In this paper, we develop a theory of optimal hierarchical design based on dynamic incentives. In particular, we analyze how an organization can use its hierarchical structure to mitigate ratchet effects. As identified in the literature (e.g., Laffont and Tirole 1988 amongst others), such problems arise when the organization's commitment is limited. Unlike in the case of full commitment, the organization under limited commitment must motivate its employees over the different stages of their careers in a time consistent manner. Our main result is that a vertical hierarchy, in which communication lines are restricted, helps the organization in providing such motivations.

An interplay of the following two ideas is key to the result. First, a vertical hierarchy with longer and more restricted communication lines allows senior workers to extract rents from their acquired knowledge about junior ones. Our model captures this idea by the difference in communication structures between a vertical and a horizontal hierarchy.[1] Our theory, therefore, explains explicitly how rents from promotions come about — the vertical hierarchy puts the promoted worker in the privileged position to extract rents from his acquired knowledge about his subordinates.[2] This argument is in line with the frequently heard complaint among junior member of an organization that their boss often takes credit for their subordinates' ideas and achievements.[3] Our second key idea is that firms find it easier to commit to basic organizational structures than to sophisticated wage contracts. We motivate this latter idea by noting that the commitment power of wage contracts are typically limited by law due to non-slavery clauses or other unalienable worker's rights. In addition, fully fledged wage schemes are complex, and thus often subject to changes in the future, while an organization's hierarchical structure is hard to change once set up.

[1] We follow Rajan and Zingales (2001) for the terminologies horizontal and vertical hierarchy.

[2] Hence, our explanation goes beyond the literature of "promotion tournaments" in organizations (e.g. Lazaer and Rosen 1981), which presumes but does not explain the existence of rents from promotions.

[3] See, for instance, Bartolome (1989). An internet search with the keywords "boss" "steals" "ideas" reveals over 6 millions links to websites on practical advice how to deal with this problem.

The key trade-off in our result is *control vs. dynamic incentives — the horizontal hierarchy* allows an organization a tighter control, while *the vertical hierarchy* enables the organization to mitigate dynamic incentive problems associated with its limited commitment. Using a dynamic agency framework, we analyze why and under what circumstances one hierarchical structure prevails over the other. Our analysis reveals that the horizontal hierarchy is optimal when the organization's output expands quickly over time and it is highly likely that the employees are efficient or inefficient. By contrast, when such likelihood is intermediate or output growth is relatively slow, the vertical hierarchy is optimal. Our result is consistent with stylized observations. Fast growing firms with a highly efficient workforce typically exhibit flatter, more horizontal hierarchies, whereas firms in more mature industries that grow less quickly and pick their workers from an average pool of potential employees tend to have vertical hierarchies.

To present the intuition behind our results in a nutshell, we point out that an organization with limited commitment faces a combination of two incentive problems: A *hold-up* problem associated with efficient agents and a *take-the-money-and-run*[4] problem associated with inefficient agents. The combination of both problems makes information revelation more costly to the organization. Our results show that, when inefficient workers are relatively likely, a vertical hierarchy mitigates the take–the–money–and–run problem, while it mitigates the hold–up problem, when efficient workers are relatively likely. Hence, our economic insight is that a vertical hierarchy alleviates information revelation in two different ways and is beneficial when information revelation is important. This is the case when the uncertainty about the worker's inefficiency is high or output does not expand very fast over time.

To explain this in more detail, note that, due to the firm's limited commitment, an agent who reveals himself as efficient anticipates a hold–up problem in the future. In order to reveal his efficiency, an efficient worker, therefore, requires a large payment in the beginning of his career. This payment, however, induces an inefficient worker to follow a take-the-money-and-run strategy, i.e., misrepresent his efficiency in order to receive the large payment and,

[4] See Laffont and Tirole (1993) for the terminology.

subsequently, reject the organization's offer in the next period. We show that when it is relatively likely that workers are inefficient, a vertical hierarchy allows the organization to pledge a long term rent to the inefficient worker, which is lost when he decides to "run" in the second period. This reduces the take-the-money-and-run problem. In contrast, when it is relatively likely that workers are efficient, the vertical hierarchy allows the organization to "play hard ball" more credibly. Due to the loss of control in the vertical hierarchy, the cost of operation is higher and this makes future production cuts credible. These production cuts imply that an efficient worker gains less from his higher efficiency in the second period. As a consequence, his hold–up problem becomes less severe. When the efficiency uncertainty about workers is large, the vertical hierarchy exhibits both effects and is especially effective.

Finally, we compare conditional and unconditional hierarchies. The comparison shows that conditional hierarchies that promote only efficient workers are never superior to unconditional hierarchies that promote any long term worker. This result demonstrates that, with respect to dynamic incentives, seniority rather than merit is the deciding factor. This result is consistent with the stylized fact that seniority plays an important, if not crucial, role with regard to promotions. Our explanation for this puzzling observation is that promotions based on seniority enable the organization to tie also less efficient employees to the organization and thereby reduce the take-the-money-and-run problem.

The rest of the paper is organized as follows. In the next section, we discuss the related literature. We present the formal framework in Section 3. In Section 4, we analyze the horizontal hierarchy in which each agent has direct access to the principal. The vertical hierarchy in which the principal only deals with the senior agent is discussed in Section 5. In Section 6, we extend our analysis to conditional hierarchies, where the hierarchical structure in the second period depends on the senior agent's report in the first period. Conclusion follows in Section 7. All proofs are relegated to an appendix. A second appendix demonstrates that our results are robust when considering semi-separating strategies.

2 Related Literature

Among the many contributions about hierarchical structures,[5] the following studies are more related to ours. Laffont and Martimort (1998) argue that the vertical hierarchy may be optimal when the organization cannot discriminate wage transfers to the agents at the same hierarchical level. In particular, collusion among the agents can be an issue when the agents of different types receive the same wage transfers. The authors show that the vertical hierarchy enables the organization to discriminate wage transfers, thus mitigating the incentives to collude. Rajan and Zingales (2001) demonstrate that the horizontal hierarchy prevents the employees from taking the organization's production technology and becoming a competitor, while the vertical hierarchy provides the employees with more incentive to learn the tasks their immediate superior assigns them. They show that the prevailing hierarchy is horizontal when the organization's production is labor-intensive, and vertical when production is capital–intensive. Friebel and Raith (2004) show that a senior agent may fear that a more productive agent would replace him if the latter can communicate directly with the principal. Hence, the senior agent has an incentive to hire less productive agents unless communication between the principal and the new agent is limited. They show that it can be optimal to force the new agent to go through the chain of command when communicating with the principal. Unlike these papers, our study shows how an organization uses designing its hierarchical structure to cope with the issues of limited commitment and associated dynamic incentive problems.

In a static model, Baron and Besanko (1992) shows that if top management can monitor the transaction between the subunits, then the vertical hierarchy is equivalent to the horizontal hierarchy. Melumad et al. (1992) breaks this tie between the horizontal and vertical hierarchy and show that the vertical hierarchy dominates the horizontal hierarchy

[5] See, for example, Williamson (1967) and McAfee and McMillan (1995) that advocate the horizontal hierarchy by indentifying loss of control in the vertical hierarchy. Calvo and Wellisz (1979) and Qian (1994) analyze optimal wage transfers with respect to hierarchical ranks. Rosen (1982), Harris and Raviv (2002), and Hart and Moore (2005) study coordination issues in different hierarchical structures.

in the presence of costly communication.[6] Aghion and Tirole (1997) show that providing a subordinate with authority induces the subordinate to acquire and provide useful information for the organization. Also Dessein (2002) demonstrates that delegation of control to a subordinate improves his incentives to provide information, when the organization's commitment is limited. The focus in Aghion and Tirole (1997) and Dessein (2002) is, however, on the delegation of authority rather than the optimal hierarchical structure of multi–layered organizations.

The organization literature recognizes the incentive effect of promotions mostly from a hidden action perspective. Starting with Lazaer and Rosen (1981), an extensive literature points out that promotions can serve as tournaments between agents to improve incentives. In contrast, our incentive problem originates from hidden information rather than hidden action. As a consequence, results differ in two respects. First, our framework allows us to go one step farther than the tournament literature and explain where the rents from a promotion actually come from. This is left unexplained in the tournament literature, which implicitly presumes a commitment of the organization not to reduce the rents from the promotion after the tournament ends. Second, from the tournament literature it follows that promotions are given on merit and not on seniority. This conclusion cannot explain the stylized fact that, in many organizations, seniority plays an important, if not crucial, role with regard to promotions.

3 The Framework

We consider a two period ($t = 1, 2$) model in which the top management of the firm (the principal) potentially hires two experts (the agents). We assume that the firm's revenue opportunities and available technologies grow over time. For this reason, the principal hires

[6] Melumad et al. (1995) show that an appropriate design of communicational sequence can mitigate loss of control in the vertical hierarchy. See Radner (1993) and Bolton and Dewatripont (1994) for an analysis of communication issues in an organization.

one agent (agent A) in $t = 1$, and she hires an additional agent (agent B) in $t = 2$.[7] The output level of agent A in $t = 1$ is $q_{A1} \in \{0, 1\}$.[8] Because output level is verifiable, the first period contract is a wage ω_1 from the principal to agent A contingent on his output level. In $t = 2$, agent A and B produce $q_{A2} \in \{0, \gamma\}$ and $q_{B2} \in \{0, \gamma\}$ respectively, where $\gamma > 1$ represents the firm's output expansion parameter.

The agent's individual output levels map into the organization's revenue as follows. In $t = 1$, the organization's revenue, q_1, is simply agent A's output level q_{A1}, and thus:

$$q_1 = q_{A1} \in \{0, 1\}.$$

Agents' effort levels in $t = 2$ are complementary. That is, the firm's revenue in $t = 2$, equals:

$$q_2 = \begin{cases} q_{A2} + q_{B2} = 2\gamma & \text{if } q_{A2} = q_{B2} = \gamma, \\ 0 & \text{otherwise.} \end{cases}$$

The complementarity assumption does not only keep our analysis tractable but is also natural for an expanding firm that organizes overall production in specific tasks in which different agents specialize. The gain of specialization is then that each agent can produce the output $\gamma > 1$ but final output is realized only if both tasks are performed. In contrast, in $t = 1$ agent A performs both tasks so that we can express output levels by a one-dimensional $q_1 \in \{0, 1\}$. For simplicity, we normalize the common discount factor to 1 so that the firm's total revenue over the two periods is given by:[9]

$$Q = q_1 + q_2.$$

An agent $i \in \{A, B\}$ incurs the cost of production:

$$C(q_{it}, \theta) = (1 - \theta)q_{it},$$

[7] In the conclusion, we discuss how to extend this setup and our results to a more natural overlapping generation framework.

[8] As our interest is to analyze an organization's choice of its hierarchical structure, we employ binary output levels to keep the model tractable and avoid technical difficulties in a dynamic adverse selection model with limited commitment.

[9] Our results do not change if we assume a discount rate $\delta < 1$ such that $\gamma\delta > 1$.

from an individual output level q_{it} in period $t \in \{1, 2\}$. The parameter $\theta \in \{\theta^h, \theta^l\}$ represents agent i's efficiency and $1 > \theta > 0$. An agent is efficient, θ^h, with probability μ, and inefficient, θ^l, with probability $1 - \mu$, where $\Delta\theta \equiv \theta^h - \theta^l > 0$. We denote the expected efficiency of an agent by $\widetilde{\theta} \equiv \mu\theta^h + (1 - \mu)\theta^l$. The types of the two agents are drawn independently. The prior distributions are public knowledge. Agents know their own types, while the principal does not observe the type of either agent.

In environments where agents work closely together at similar tasks, it is natural that an agent learns something about the other that is not observed by top management. Moreover, when distinguishing between experienced and unexperienced workers, there exists also an asymmetry in learning abilities in that an experienced worker is more likely to learn something about the unexperienced worker than vice versa. We model these ideas by assuming that, before production starts in $t = 2$, agent A learns agent B's type perfectly, whereas agent B does not learn anything about agent A. We stress that this assumption is an extreme, simple illustration of the idea that agent A observes with some positive probability something about agent B's type, whereas agent B learns less about agent A. In the concluding section, we discuss the extention to where agent A learns agent B's type imperfectly.

Regarding the contractual environment, we assume that the principal's long term commitment is limited in that, in $t = 1$, she cannot commit to wage transfers and output levels for $t = 2$. This captures the practical problems that prevent real life labor contracts to cover a worker's entire life span. In contrast, the principal is able to make a long term commitment to the organization's hierarchical structure. In particular, we distinguish between two organization structures denoted by $\Psi \in \{H, V\}$, which determine the contractual relationship between the principal and the two agents in $t = 2$. With the horizontal hierarchy $\Psi = H$, the principal communicates and offers contracts to both agents directly. With the vertical hierarchy $\Psi = V$, agent A becomes agent B's superior in that the principal can only communicate and interact with agent A.

Hence, our idea is that, in $t = 1$, the principal commits to the first period contract with agent A and a hierarchical structure $\Psi \in \{H, V\}$ for $t = 2$. Because agent A's type is private information, the first period contract is contingent on agent A's report on his type.

Therefore, we express the first period contract Φ_1 as:

$$\Phi_1 = \{q_1(\theta_1^a),\ \omega_1(\theta_1^a)\}^{a\in\{h,l\}}, \tag{1}$$

where θ_1^a is agent A's report. At the end of $t = 1$, the contract yields the principal's and the agent's payoffs of $q_1 - \omega_1$ and $\omega_1 - (1 - \theta^a)q_1$ respectively. The timing in $t = 1$ is:

1.1 The principal commits to $\Psi \in \{H,\, V\}$.

1.2 The principal offers Φ_1 to agent A.

1.3 Agent A rejects or accepts and, upon acceptance, reports his type.

1.4 The contract is executed for $t = 1$.

The principal's offer in $t = 2$ depends on the organization structure $\Psi \in \{H, V\}$. Due to the complementarity in production, these offers satisfy $q_{A2} = q_{B2}$. Therefore, we only need to consider second period contracts that exhibit:

$$q \equiv \frac{q_2}{2} = q_{A2} = q_{B2}. \tag{2}$$

With $\Psi = H$, the principal offers agent A and B wages ω_A and ω_B respectively, for output level q from each agent. We denote by Φ_2^H the set of contracts for $t = 2$ that specify each agent's output level q, and the wage transfers ω_A and ω_B. For $t = 2$, the principal's payoff is $2q - \omega_A - \omega_B$, and agent A's and agent B's payoffs are $\omega_A - (1 - \theta^a)q$ and $\omega_B - (1 - \theta^b)q$ respectively. With $\Psi = H$, the timing in $t = 2$ is:

2.1 Agent A learns agent B's type.

2.2 The principal offers Φ_2^H to agents A and B.

2.3 The agents accepts/rejects. Upon acceptance, the agents send reports to the principal.

2.4 Contract with agent A and B are executed for $t = 2$.

With $\Psi = V$, the principal makes an offer about total production $q_2 = 2q$ and a grand wage transfer W to agent A. In turn, agent A bargains with agent B about a wage transfer ω_B for an output q. We denote by Φ_2^V the principal's offer to agent A for $t = 2$, which specifies q_2 and W. In turn, agent A's offer to agent B specifies q and ω_B. For $t = 2$, the principal's payoff is $q_2 - W$, while agent A's and agent B's payoffs are $W - (1 - \theta^a)q - \omega_B$ and $\omega_B - (1 - \theta^b)q$ respectively. With $\Psi = V$, the timing in $t = 2$ is:

2.1 Agent A learns agent B's type.

2.2 The principal offers Φ_2^V to agent A.

2.3 Agent A accepts/rejects. Upon acceptance, agent A sends reports to the principal and make an offer to agent B.

2.4 Agent B accepts/rejects the offer of agent A.

2.5 Contracts are executed for $t = 2$.

Our approach is to analyze the implementable outcomes under a horizontal and a vertical hierarchy and then compare their optimality from the principal's perspective. We start our analysis with the horizontal hierarchy.

4 Horizontal Hierarchy ($\Psi = H$)

In the horizontal hierarchy, the principal makes a second period contract offer to both agents. Using the expression in (2), the set of the second period offers, Φ_2^H, in the horizontal hierarchy can be expressed as:

$$\Phi_2^H = \{q(\theta_A^a, \theta_A^b, \theta_B^b),\ \omega_A(\theta_A^a, \theta_A^b, \theta_B^b),\ \omega_B(\theta_A^a, \theta_A^b, \theta_B^b)\}^{a,b\in\{h,l\}}, \tag{3}$$

where θ_A^a is agent A's report on his own type, θ_A^b is agent A's report on B's type, and θ_B^b is agent B's report on his own type in $t = 2$. Note that Φ_2^H depends on the history of the game — in particular, agent A's report about his type, θ_1^a, from $t = 1$.

An advantage of the horizontal hierarchy is that the principal can directly communicate with both agents, thus can have more control. As well established in the literature, such direct communications with both agents enable the principal to elicit any shared information at no cost.[10] Because agent A and B share the information about agent B's type, the principal can elicit this information costlessly. Thus, we have the lemma below.

Lemma 1 *With $\Psi = H$, agent B's type is revealed at zero costs.*

Lemma 1 states that, in the horizontal hierarchy, only the private information about agent A matters and information revelation of agent B is not an issue. Therefore, we can treat the situation as if θ^b is publicly observable and can specify contracts as if they condition directly on agent B's true type. This allows us to express Φ_2^H as follows:

$$\Phi_2^H = \{q^{ab}, \omega_A^{ab}, \omega_B^{ab}\}^{a,b\in\{h,l\}},$$

with $q^{ab} \equiv q(\theta_A^a, \theta^b, \theta^b)$, $\omega_A^{ab} \equiv \omega_A(\theta_A^a, \theta^b, \theta^b)$, and $\omega_B^{ab} \equiv \omega_b(\theta_A^a, \theta^b, \theta^b)$.

In addition, Lemma 1 implies that agent B receives no rent. Therefore, the wage transfer to agent B in the optimal contract is:

$$\omega_B^{ab} = (1 - \theta^b)q^{ab}. \tag{4}$$

Having established these preliminary results, we turn to the issue of interim information revelation. In our dynamic framework, the degree of intermediate information revelation, determined by the principal's offer in $t = 1$, plays a crucial role. Thus, we classify the first period contract Φ_1 by the degree of information revelation it induces. For expositional purpose, it suffices to focus on the two extremes: a non-revealing (pooling) and a fully revealing (separating) offer. These extremes correspond to settings where agent A plays a deterministic reporting strategy in $t = 1$. In Appendix 2 we demonstrate that our qualitative result (in particular, the optimality of the vertical hierarchy) is robust to the consideration

[10] One way is to require both agents to report about their shared information, and severely punish them when the reports are contradictory. For more elaborate schemes that elicit common private information, see Moore and Repullo (1988) and Maskin and Tirole (1999) for example.

of semi-separating, where agent A employs a mixed reporting strategy and information revelation is only partial.

4.1 Pooling in the Horizontal Hierarchy

A first period contract that does not elicit any information from agent A induces the same outcome in $t = 1$ regardless of agent A's type. Such a contract is, therefore, equivalent to a first period contract that is independent of agent A's report about his type. Hence, we express pooling contracts by contracts with the following structure:

$$\Phi_1^{HP} = \{q_1,\ \omega_1\}.$$

By definition, a pooling contract does not lead to any information revelation. Hence, in the beginning of $t = 2$, the principal still has the prior belief μ that agent A is of type θ^h. The principal's offer in $t = 2$ is a combination:

$$\Phi_2^H = \{q^{ab},\ \omega_A^{ab},\ \omega_B^{ab}\}^{a,b\in\{h,l\}},$$

where a represents agent A's report about his type in $t = 2$ and b is, by Lemma 1, agent B's actual type. Using (4), the principal's sequentially rational offer Φ_2^H maximizes her expected payoff for $t = 2$:

$$\begin{aligned}\Pi_2^H &= \mu[\mu(q^{hh} - \omega_A^{hh} + \theta^h q^{hh}) + (1-\mu)(q^{hl} - \omega_A^{hl} + \theta^l q^{hl})] \\ &\quad +(1-\mu)[\mu(q^{lh} - \omega_A^{lh} + \theta^h q^{lh}) + (1-\mu)(q^{ll} - \omega_A^{ll} + \theta^l q^{ll})],\end{aligned}$$

subject to agent A's participation constraints,

$$\omega_A^{ab} - (1-\theta^a)q^{ab} \geq 0, \quad a,b \in \{h,l\}, \tag{5}$$

and his incentive compatibility constraints,

$$\omega_A^{ab} - (1-\theta^a)q^{ab} \geq \omega_A^{a'b} - (1-\theta^a)q^{a'b}, \quad a,a',b \in \{h,l\}. \tag{6}$$

The maximization problem illustrates that the principal faces the familiar trade-off between information rent and efficient production. Effectively, the principal has two options in $t = 2$. Her first option is to offer again a pooling contract in $t = 2$ with respect to agent A's type. Alternatively, she can offer a contract that separates the types of agent A in $t = 2$. In our setup with binary output levels, the only way to achieve separation in $t = 2$ is to abandon production depending on the agents' types. The principal's sequentially rational decision whether to separate or pool depends on her beliefs and, for convenience, we introduce the following definition.

Definition 1 *Let* $\hat{\mu}(x) \equiv \frac{\theta^l + x}{\theta^h + x}$, *where* $x \in \mathbb{R}^+$.

With the definition above, we can characterize the sequentially rational contracts in $t = 2$, after a pooling contract in period $t = 1$.

Lemma 2 *Given* Φ_1^{HP} *in* $t = 1$, *the sequentially rational contracts* Φ_2^H *in* $t = 2$ *is:*

i) *If* $\mu \leq \hat{\mu}(\theta^l)$, *then* $q^{ab} = \gamma$; $\omega_A^{ab} = (1 - \theta^l)\gamma$; $a, b \in \{h, l\}$.

ii) *If* $\mu \in (\hat{\mu}(\theta^l), \hat{\mu}(\theta^h)]$, *then* $q^{hh} = q^{hl} = q^{lh} = \gamma$, $q^{ll} = 0$; $\omega_A^{hh} = \omega_A^{lh} = (1 - \theta^l)\gamma$, $\omega_A^{hl} = (1 - \theta^h)\gamma$, $\omega_A^{ll} = 0$.

iii) *If* $\mu > \hat{\mu}(\theta^h)$, *then* $q^{hb} = \gamma$, $q^{lb} = 0$; $\omega_A^{hb} = (1 - \theta^h)\gamma$, $\omega_A^{lb} = 0$; $b \in \{h, l\}$.

The lemma shows that the principal makes a pooling offer also for $t = 2$ when her belief about agent A's type is relatively pessimistic. This is because a more pessimistic belief shifts the trade-off between pooling and separation towards the former. According to the lemma, the principal's decision to pool or to separate in $t = 2$ depends also on agent B's type. More specifically, the principal is more eager to pool when agent B is efficient. The reason is that, when agent B is efficient, production is more valuable so that the principal prefers ensuring a positive output. Therefore, the disadvantage of separation that output is sometimes abandoned becomes relatively more costly to the principal.

With Lemma 2, we can obtain the principal's optimal payoff from the pooling offer in $t = 1$.

Proposition 1 *With $\Psi = H$, the principal's maximum expected payoff from pooling equals*

$$\Pi^{HP} = \begin{cases} \theta^l + (\theta^l + \widetilde{\theta})\gamma & \text{if } \mu \leq \hat{\mu}(\theta^l), \\ \theta^l + \mu(2-\mu)(\theta^h + \theta^l)\gamma & \text{if } \mu \in (\hat{\mu}(\theta^l), \hat{\mu}(\theta^h)], \\ \theta^l + \mu(\theta^h + \widetilde{\theta})\gamma & \text{if } \mu > \hat{\mu}(\theta^h). \end{cases}$$

4.2 Separating Strategy in the Horizontal Hierarchy

A first period separation contract, Φ_1^{HS}, depends on agent A's report about his type and induces him to report his type truthfully. Thus, the first period separation contract has a structure as in (1) and satisfies incentive constraints that ensure agent A's truthful report.

Given a separation contract in $t = 1$, it is straightforward to derive the sequentially rational contract offer in $t = 2$. Separation contract are, per definition, fully revealing so that the principal learns agent A's type with certainty. This together with Lemma 1 implies that the principal makes offers in $t = 2$ as if under full information.

Lemma 3 *Given Φ_1^{HS} in $t = 1$, the sequentially rational contracts Φ_2^H in $t = 2$ is $q^{ab} = \gamma$; $\omega_A^{ab} = (1 - \theta^a)\gamma$, $\omega_B^{ab} = (1 - \theta^b)\gamma$, $a, b \in \{h, l\}$.*

Due to the binary production possibilities, a first period separating contract Φ^{HS} necessarily exhibits the output schedule $q_1^h = 1$ and $q_1^l = 0$. Together with Lemma 3, it follows that, because a separating contract has to induce agent A to report his type truthfully, the following two incentive constraints must be satisfied in $t = 1$

$$\omega_1^h - (1 - \theta^h) \geq \omega_1^l + \gamma\Delta\theta, \tag{7}$$

$$\omega_1^l \geq \omega_1^h - (1 - \theta^l). \tag{8}$$

Inequality (7) guarantees that the payoff of the efficient agent A is higher when he truthfully reports his type than when he misrepresents it. The LHS of (7) is agent A's total payoff over both periods if he reveals his type honestly in $t = 1$, because he receives no rent in $t = 2$. The RHS of (7) is agent A's aggregate payoff if he misrepresents his type as θ^l in $t = 1$. In such a case, the agent must produce $q_1^l = 0$ for a wage ω_1^l in $t = 1$, and expects a rent of $\omega_A^{lb} - (1 - \theta^h)\gamma = \gamma\Delta\theta$ from the sequential rational contract in $t = 2$.

The contract Φ_1^{HS} must also satisfy (8), the incentive constraint of the inefficient type. The LHS of (8) is agent A's total payoff over the two periods when he reports his true type in $t = 1$. In the case of truth-telling, the agent produces $q_1^l = 0$ with the wage transfer ω_1^l in $t = 1$, and receives zero rent in $t = 2$. The RHS of (8) is the agent's total payoff if he misreports his type as θ^h in $t = 1$. The misreporting agent receives ω_1^h for $q_1^h = 1$ in $t = 1$, and he is offered the contract in $t = 2$ that would yield him a negative payoff, $\omega_A^{hb} - (1 - \theta^l)\gamma = -\Delta\theta\gamma$. Hence, after misreporting his type as θ^h in $t = 1$, the agent will not take the principal's offer in $t = 2$. The agent, therefore, can guarantee himself the payoff $\omega_1^h - (1 - \theta^l)$ by reporting θ^h in $t = 1$. As mentioned in the introduction, the inefficient agent achieves this payoff by employing the *take-the-money-and-run* strategy.

Lemma 4 *With $\Psi = H$, the principal cannot implement a separating contract in $t = 1$.*

According to Lemma 4, information revelation in $t = 1$ is not possible in the horizontal hierarchy. Due to the lack of commitment, the principal cannot promise agent A any rent in $t = 2$ if he revealed his true type in $t = 1$. Hence, to induce a truthful report from the efficient type in $t = 1$, the sum of rents over two periods must be paid in $t = 1$. For this reason, the principal must offer the efficient type a large wage transfer in $t = 1$. This generous first period offer is, however, also attractive to the inefficient type if he employs the take-the-money-and-run strategy. It follows from this discussion that the impossibility of an incentive compatible separating offer in $t = 1$ results from a tension between two different problems that are due to limited commitment:

1) A "hold-up" problem associated with the efficient agent.

2) A "take-the-money-and-run" problem associated with the inefficient agent.

Due to this tension, the principal can only make the pooling offer in $t = 1$. Thus, Proposition 1 presents the principal's maximum payoff in the horizontal hierarchy.

5 Vertical Hierarchy ($\Psi = V$)

We next analyze the structure of optimal contracts in the vertical hierarchy. Here, only agent A has a communication channel to the principal. We can therefore express the principal's second period contract with $\Psi = V$ as:

$$\Phi_2^V = \{q_2(\theta_A^a, \theta_A^b),\ W(\theta_A^a, \theta_A^b)\}^{a,b \in \{h,l\}}. \tag{9}$$

where θ_A^a and θ_A^b are agent A's reports on his own and agent B's type respectively. In addition, we can restrict attention to $q_2(\theta_A^a, \theta_A^b) \in \{0, 2\gamma\}$, due to the complementary production technology. Unlike the case with $\Psi = H$, the contract is not contingent on agent B's report θ_B^b because, with $\Psi = V$, the principal can communicate only with agent A.

If agent A accepts the offer Φ_2^V, he subsequently makes a take-it-or-leave-it offer $\{q, \omega_B\}$ to agent B. Because agent A observes agent B's type, agent B has no private information vis-à-vis agent A, and hence agent A provides no rent to agent B. In particular, agent A's offer to agent B is:

$$q = q_2/2 \text{ and } \omega_B = (1 - \theta^b)q_2/2. \tag{10}$$

From (10) it follows that agent A's payoff from the principal's second period offer q_2 and W is:

$$W - (1 - \theta^a)q_2/2 - \omega_B = W - [2 - (\theta^a + \theta^b)]q_2/2.$$

Again, we distinguish two different strategies regarding information revelation in the first period: a pooling and a separating offer. We first consider the case in which the principal makes a pooling offer in $t = 1$.

5.1 Pooling in the Vertical Hierarchy

The structure of the pooling contract in $t = 1$ with $\Psi = V$ is the same as for $\Psi = H$:

$$\Phi_1^{VP} = \{q_1, \omega_1\}.$$

With pooling in $t = 1$, the principal does not learn agent A's private information. Because the principal cannot communicate with agent B, agent A has, from the principal's perspective, private information about both his own and agent B's type. Hence, the situation in $t = 2$ is equivalent to the principal facing an agent A with three possible types $\Theta^h \equiv 2\theta^h$, $\Theta^m \equiv \theta^h + \theta^l$, and $\Theta^l \equiv 2\theta^l$ with the respective probabilities $\varphi^h \equiv \mu^2$, $\varphi^m \equiv 2\mu(1-\mu)$, and $\varphi^l \equiv (1-\mu)^2$. An offer (q_2, W) yields a type Θ^j the payoff:

$$W - (2 - \Theta^j)q_2/2, \quad j \in \{h, m, l\}.$$

The second period sequential rational contract Φ_2^V maximizes the principal's payoff for $t = 2$ under the agent's private information about Θ. Thus, we can express the principal's second period contract as:

$$\Phi_2^V = \{q_2^j,\ W^j\}^{j\in\{h,m,l\}}.$$

The principal's problem in $t = 2$ is then to maximize:

$$\sum_{j\in\{h,m,l\}} \varphi^j(q_2^j - W^j),$$

subject to the participation constraints,

$$W^j - (2 - \Theta^j)q_2^j/2 \geq 0, \quad j \in \{h, m, l\}, \tag{11}$$

and the incentive constraints,

$$W^j - (2 - \Theta^j)q_2^j/2 \geq W^{j'} - (2 - \Theta^j)q_2^{j'}/2, \quad j, j' \in \{h, m, l\}. \tag{12}$$

In order to show how the solution depends on the principal's prior belief μ, we introduce the following two parameters.

Definition 2 *Let* $\underline{\mu} \equiv 1 - \sqrt{\frac{\Delta\theta}{\theta^h + \theta^l}}$ *and* $\overline{\mu} \equiv \frac{2(\theta^h + \theta^l)}{3\theta^h + \theta^l}$.

We can easily verify that $\underline{\mu} < \overline{\mu}$. With these two parameters, we can characterize the sequential rational contract Φ_2^V.

Lemma 5 *Given Φ_1^{VP} in $t=1$, the sequentially rational contract Φ_2^V in $t=2$ is:*

i) If $\mu \leq \underline{\mu}$, then $q_2^j = 2\gamma$; $W^j = (2-\Theta^l)\gamma$, $j \in \{h, m, l\}$.

ii) If $\mu \in (\underline{\mu}, \overline{\mu}]$, then $q_2^h = q_2^m = 2\gamma$, $q_2^l = 0$; $W^h = W^m = (2-\Theta^m)\gamma$ and $W^l = 0$.

iii) If $\mu > \overline{\mu}$, then $q_2^h = 2\gamma$, $q_2^m = q_2^l = 0$; $W^h = (2-\Theta^h)\gamma$ and $W^m = W^l = 0$.

The lemma shows that when μ is small, the principal is better off pooling all three types and, therefore, provide information rents to types Θ^m and Θ^h. These information rents enable the principal to ensure production for all types. This is optimal when μ is small so that the principal is likely to face type Θ^l.

When μ is large, the principal is confident that both agent A and B are efficient. As a result, the principal chooses to separate type Θ^h from the other two, thereby saving information rent completely. Although there will be no output in the case of Θ^m and Θ^l, the principal is confident enough that such cases are unlikely to be realized.

In the intermediate case, where it is likely enough that at least one of the agents is efficient, the principal separates Θ^l from types Θ^m and Θ^h. She thereby saves on paying an information rent to Θ^m at the expense of forgoing production by type Θ^l. For $\mu \in (\underline{\mu}, \overline{\mu}]$, such separation is preferable because it is unlikely that the principal will face type Θ^l, but it is still likely enough that the principal will face type Θ^m.

With Lemma 5 we can compute the principal's optimal payoff from the pooling strategy in the vertical hierarchy. We show that this payoff is always smaller than her payoff in the horizontal hierarchy in Proposition 1.

Proposition 2 *With $\Psi = V$, the principal's maximum expected payoff from pooling is:*

$$\Pi^{VP} = \begin{cases} \theta^l + 2\theta^l\gamma & \text{if } \mu \leq \underline{\mu}, \\ \theta^l + \left[1-(1-\mu)^2\right](\theta^h + \theta^l)\gamma & \text{if } \mu \in (\underline{\mu}, \overline{\mu}], \\ \theta^l + 2\mu^2\theta^h\gamma & \text{if } \mu > \overline{\mu}. \end{cases}$$

The principal's payoff from the pooling strategy with $\Psi = V$ is strictly smaller than her payoff with $\Psi = H$.

The result that the principal's payoff from the pooling strategy in the vertical hierarchy is always smaller than her optimal payoff in the horizontal hierarchy is intuitive. In the horizontal hierarchy, the principal can extract agent B's private information at no cost. In the vertical hierarchy, however, the principal cannot do so because she can communicate only with agent A. Effectively, such a restriction increases agent A's private information and makes it more costly for the principal to extract it. The result clarifies that, if the principal can benefit from the vertical hierarchy at all, then the benefit must stem from the fact that the principal may be able to induce information revelation in $t = 1$. We turn to this case next.

5.2 Separating in the Vertical Hierarchy

Similarly to the horizontal hierarchy, we can express a first period contract that fully separates the types of agent A as:

$$\Phi_1^{VS} = \{q_1^a,\ \omega_1^a)\}^{a\in\{h,l\}}.$$

The contract must be incentive compatible so that agent A reports truthfully in $t = 1$. Again, due to the binary production structure, a first period separation contract necessarily exhibits $q_1^h = 1$ and $q_1^l = 0$.

With the separating strategy, the principal receives an informative report θ_1^a about agent A's true type in $t = 1$. This report affects her beliefs and therefore her subsequent contract in $t = 2$. We can thereby view the principal's offer in $t = 2$ as contingent on agent A's report in $t = 1$. Because agent A learns agent B's type, and there is no direct communication between the principal and agent B, the second period contract also depends on a report of agent A about agent B's type in $t = 2$. Hence, we express the principal's second period contract as:

$$\Phi_2^V = \{q_2^{ab},\ W^{ab}\}^{a,b\in\{h,l\}},$$

where superscript a represents agent A's report about himself in $t = 1$ and superscript b represents his report about agent B in $t = 2$.

With the separating strategy, the principal is certain about agent's A type after the agent sends his report in $t = 1$. Thus, for a given θ^a with $a \in \{h, l\}$, the principal's sequentially rational contract, Φ_2^V, maximizes her payoff for $t = 2$

$$\Pi_2^V \equiv \mu(q_2^{ah} - W^{ah}) + (1 - \mu)(q_2^{al} - W^{al}),$$

subject to agent A's participation constraints,

$$W^{ab} - (2 - \theta^a - \theta^b)q_2^{ab}/2 \geq 0, \quad b \in \{h, l\}, \tag{13}$$

and incentive constraints,

$$W^{ab} - (2 - \theta^a - \theta^b)q_2^{ab}/2 \geq W^{ab'} - (2 - \theta^a - \theta^b)q_2^{ab'}/2, \quad b, b' \in \{h, l\}. \tag{14}$$

Again, the principal faces a trade-off between pooling and separation — ensuring output vs. extracting information rent. In this case, however, agent A's private information in $t = 2$ is only agent B's type. Hence, the question is whether the principal wants to ensure production for both types of agent B at the expense of paying an information rent to agent A. The alternative is to save on information rents but forego on output in the case agent B is inefficient. The next lemma shows how the principal's sequentially rational offer in $t = 2$ depends on her belief on the likelihood that agent B is efficient.

Lemma 6 *Given Φ_1^{VS}, the sequentially rational contract Φ_2^V in $t = 2$ is:*

i) If $\mu \leq \hat{\mu}(\theta^l)$, then $q_2^{ab} = 2\gamma$; $W^{ab} = (2 - \theta^a - \theta^l)\gamma$, $a, b \in \{h, l\}$; agent A gets a rent of $\Delta\theta\gamma$ in $t = 2$ when $\theta^b = \theta^h$.

ii) If $\mu \in (\hat{\mu}(\theta^l), \hat{\mu}(\theta^h)]$, then $q_2^{hh} = q_2^{hl} = q_2^{lh} = 2\gamma$, $q_2^{ll} = 0$; $W^{hh} = W^{hl} = W^{lh} = (2 - \theta^h - \theta^l)\gamma$, $W^{ll} = 0$; agent A gets a rent of $\Delta\theta\gamma$ in $t = 2$ when $\theta^a = \theta^b = \theta^h$.

iii) If $\mu > \hat{\mu}(\theta^h)$, then $q_2^{ah} = 2\gamma$, $q_2^{al} = 0$; $W^{ah} = (2 - \theta^a - \theta^h)\gamma$, $W^{al} = 0$; $a \in \{h, l\}$; agent A gets no rent in $t = 2$.

Lemma 6 follows a similar logic as before. The trade-off between pooling and separation shifts in favor of pooling when the likelihood that agent B is efficient is smaller. The reason is that the cost of pooling, paying an information rent, is realized with a smaller probability, while the benefit, a positive output when agent B is inefficient, becomes more important for a small μ. The cutoff point of μ at which pooling prevails over separation depends on agent A's report in $t = 1$, because it determines the principal's cost of foregone revenues from abandoning output in $t = 2$.

Crucial for our subsequent analysis is agent A's rent from the sequentially rational contract, Φ_2^V, because it determines to what extent the principal can induce agent A to reveal his private information in $t = 1$. Importantly, not only agent A's rent on the equilibrium matters, but also his potential rent off equilibrium. Therefore, we need to examine explicitly agent A's strategy for the out–off–equilibrium event that the agent misrepresents his type in $t = 1$ and faces the sequential rational contract Φ_2^V in $t = 2$.

Lemma 7 *The sequentially rational contract Φ_2^V induces agent A of type θ^l to reject the offer in the case of misreporting θ^a as θ^h in $t = 1$, but to accept the offer and truthfully report agent B's type in the case of a truthful report in $t = 1$.*

As in the horizontal hierarchy, the inefficient agent A will, in case of misreporting in $t = 1$, adopt the take–the–money–and–run strategy by rejecting the principal's offer for $t = 2$. For the efficient agent, the sequentially rational contract is individually rational and incentive compatible in that it gives the agent an incentive to report his private information about agent B honestly. In the light of Lemma 4, the question is whether there exists a contract Φ_1^{VS} that prevents agent A from misreporting in $t = 1$, and induces Φ_2^V in Lemma 7.

Lemma 8 *With $\Psi = V$, Φ_1^{VS} exists when one of the following conditions is satisfied:*

i) $\mu \leq \hat{\mu}(\theta^l)$ *and* $\gamma \leq 1/(1-\mu)$.

ii) $\mu \in (\hat{\mu}(\theta^l), \hat{\mu}(\theta^h)]$.

iii) $\mu > \hat{\mu}(\theta^h)$ *and* $\gamma \leq 1/\mu$.

Lemma 8 is the counterpart of Lemma 4 and demonstrates the paper's key insight that information revelation in $t = 1$ is easier to achieve in the vertical hierarchy than in the horizontal hierarchy. In order to interpret our subsequent results, it is important to understand the intuition behind this result. Recall that the impossibility of separation in $t = 1$ with a horizontal hierarchy was due to a tension between the hold–up problem associated with the efficient agent and the take-the-money-and-run problem associated with the inefficient agent. The three cases in Lemma 8 represent different channels by which the vertical hierarchy relaxes the tension.

First, for $\mu \leq \hat{\mu}(\theta^l)$, the sequentially rational contract Φ_2^V yields not only an efficient agent A, but also an inefficient agent A an information rent. The rent to the inefficient agent relaxes the take-the-money-and-run problem because it is lost to him if he decides to run away in $t = 2$. Hence, if the tension between the hold-up and the take-the-money-and-run problem is not too severe, which is the case when γ is small, then first period separation is possible with a vertical hierarchy. Differently put, the vertical hierarchy restores partially the principal's commitment power to pledge rents for $t = 2$, which makes information revelation easier in $t = 1$.

For $\mu > \hat{\mu}(\theta^h)$, the sequentially rational contract yields an efficient agent A no rent in $t = 2$. Therefore, the agent faces a similar hold–up problem as in the horizontal hierarchy so that the sum of rents over two periods must be paid in $t = 1$. However, the efficient agent's expected rent for $t = 2$ is smaller because, when agent B is inefficient, the sequentially rational contract induces no production in $t = 2$. Accordingly, the wage transfer that induces the efficient agent to be truthful in $t = 1$ is lower than in the horizontal hierarchy. This, in turn, discourages the inefficient agent to play the take–the–money–and–run strategy. Hence, when γ is small so that the tension between the hold–up and the take-the-money-and-run problem is not too severe, separation in $t = 1$ is possible with the vertical hierarchy, but not in the horizontal strategy..

Finally, for the intermediate case $\mu \in (\hat{\mu}(\theta^l), \hat{\mu}(\theta^h)]$, the vertical hierarchy increases the principal's credibility for both rent provision and termination of production in $t = 2$. As a result, the dynamic incentive problems are fully mitigated and the principal can always achieve separation in $t = 1$. Again, second period production in the vertical hierarchy, compared to that in the horizontal hierarchy, is more costly to the principal, and this is the source for restoring the principal's commitment power.

With Lemma 8, we can now derive, Π^{VS}, the principal's overall payoff from the separating strategy in the vertical hierarchy.

Proposition 3 *With $\Psi = V$, the principal's payoff from the separating strategy is:*

$$\Pi^{VS} = \begin{cases} \mu\theta^h + (\widetilde{\theta} + \theta^l)\gamma & \text{if } \mu \leq \hat{\mu}(\theta^l) \text{ and } \gamma < \frac{1}{1-\mu}, \\ \mu\theta^h + 2\mu(\theta^h + (1-\mu)\theta^l)\gamma & \text{if } \mu \in (\hat{\mu}(\theta^l), \hat{\mu}(\theta^h)], \\ \mu\theta^h + \mu(\theta^h + \theta^l)\gamma & \text{if } \mu > \hat{\mu}(\theta^h) \text{ and } \gamma < \frac{1}{\mu}. \end{cases}$$

Before we discuss the optimality of the horizontal and the vertical hierarchy, we first provide the following definition.

Definition 3 *Let $\widehat{\gamma} \equiv \frac{\mu\theta^h - \theta^l}{\mu^2 \Delta\theta}$.*

Due to Proposition 2, we only need to compare Π^{HP} in Proposition 1 with Π^{VS} in Proposition 3. This straightforward comparison yields our main proposition.

Proposition 4 *The optimal hierarchical structure is characterized as follows:*

i) For $\mu \leq \hat{\mu}(0)$, the optimal hierarchy is $\Psi = H$.

ii) For $\mu \in (\hat{\mu}(0), \hat{\mu}(\theta^l)]$ and $\gamma \geq 1/(1-\mu)$, the optimal hierarchy is $\Psi = H$.

iii) For $\mu \in (\hat{\mu}(0), \hat{\mu}(\theta^l)]$ and $\gamma < 1/(1-\mu)$, the optimal hierarchy is $\Psi = V$.

iv) For $\mu \in (\hat{\mu}(\theta^l), \hat{\mu}(\theta^h)]$, the optimal hierarchy is $\Psi = V$.

v) For $\mu > \hat{\mu}(\theta^h)$ and $\gamma < \widehat{\gamma}$, the optimal hierarchy is $\Psi = V$.

vi) For $\mu > \hat{\mu}(\theta^h)$ and $\gamma \geq \widehat{\gamma}$, the optimal hierarchy is $\Psi = H$.

Figure 1 illustrates our main result. While the horizontal hierarchy provides the principal with more control, the vertical hierarchy enables her to mitigate the dynamic incentive problems due to limited commitment. The proposition shows that the principal prefers the vertical hierarchy when uncertainty about an agent's type is large. On the other hand, the principal prefers to retain direct control and not leave any rent in $t = 2$, when this uncertainty is small so that the agents are likely to be efficient or inefficient. This demonstrates that the key trade-off in choosing a hierarchical structure is control vs. dynamic incentives.

Note also that the optimal hierarchical structure depends, in general, on the output expansion parameter γ. For example, when $\mu \in (\hat{\mu}(0), \hat{\mu}(\theta^l)]$ the horizontal hierarchy prevails if γ is too large. The intuition is that the parameter γ measures the tension between the hold–up problem and the take–the–money–and–run strategy. The wage transfer to the efficient agent in $t = 1$ increases as γ becomes larger, and for γ large enough, the wage payment to the efficient agent makes the take-the-money-and-run profitable to the inefficient agent. As a result, separating the types of agent A in $t = 1$ becomes impossible even in a vertical hierarchy so that the vertical hierarchy has no merit over the horizontal hierarchy.

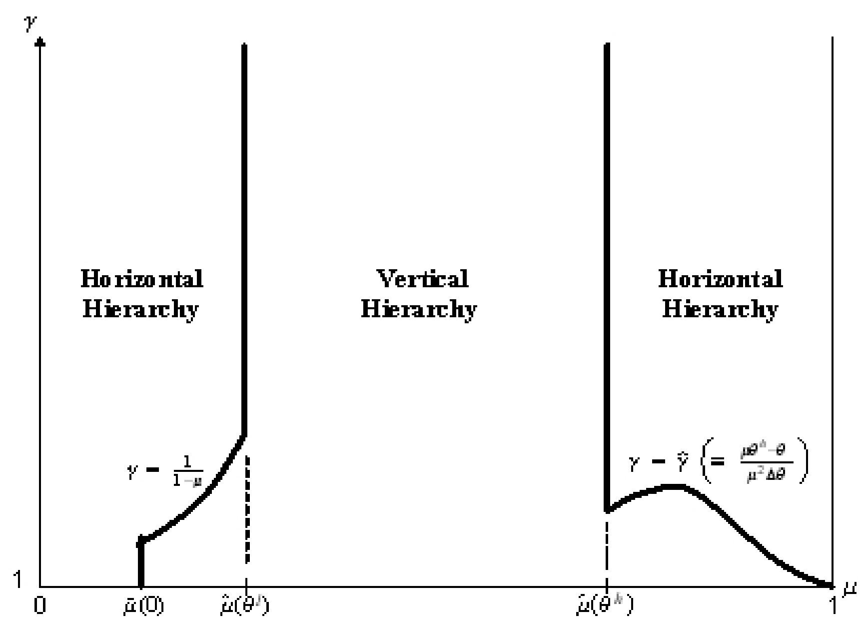

Fig 1. Horizontal vs. Vertical Hierarchy

6 Conditional Hierarchical Structures

So far we have investigated the cases in which the organization in $t = 1$ either commits to the horizontal or the vertical hierarchy. Yet, if we view the organization's long term hierarchical structure as fully contractible, then the principal could actually condition the hierarchical structure on the agent A's report (or output) in $t = 1$. In this section we study to what extend the principal benefits from using conditional rather than unconditional hierarchies. We first show that the principal does not benefit from conditional hierarchies that promote agent A only when he is efficient. Then, we demonstrate the more counter–intuitive result that for relatively small μ the principal benefits from promoting agent A only when he is inefficient.

We express conditional hierarchies by the first period offer of the following structure:

$$\Phi_1^{\Psi} = \{q_1^a,\ \omega_1^a,\ \Psi^a\}^{a\in\{h,l\}},$$

with the interpretation that when agent A reports type θ^a in $t = 1$, he then has to produce the output q_1^a, receives a wage ω_1^a in $t = 1$, and the hierarchical structure in $t = 2$ is $\Psi^a \in \{H, L\}$.

In principle, there are four possible combinations of which our previous analyses have covered the two unconditional hierarchies $\{\Psi^h = H, \Psi^l = H\}$ and $\{\Psi^h = V, \Psi^l = V\}$. Hence, we are left to consider the two conditional combinations:

$$\Psi_{VH} \equiv \{\Psi^h = V, \Psi^l = H\} \text{ and } \Psi_{HV} \equiv \{\Psi^h = H, \Psi^l = V\}.$$

The conditional hierarchy Ψ_{HV} represents a contract that promotes agent A to middle manager only when he is efficient. By contrast, the conditional hierarchy Ψ_{VH} represents a contract that promotes the agent A only when he is inefficient.

Note that an incentive compatible conditional hierarchy necessarily implies separation so that agent A's type in $t = 1$ is fully revealing. Hence, after agent A makes a report in $t = 1$, the principal is convinced that the agent is of type θ^a. It follows that the principal's offer in

$t = 2$ depends on agent A's first period report θ_A^a as follows:

$$\Phi_2^\Psi = \begin{cases} \{q^{ab}, \omega_A^{ab}, \omega_B^{ab}\}^{b\in\{h,l\}} & \text{if } \Psi^a = H, \\ \{q_2^{ab},\ W^{ab}\}^{b\in\{h,l\}} & \text{if } \Psi^a = V. \end{cases}$$

We first study the optimality of the conditional hierarchy Ψ_{VH}. With Ψ_{VH}, the principal faces the vertical hierarchy in $t = 2$ if agent A reports θ^h in $t = 1$. In this case, the principal in $t = 2$ offers a menu $\{q_2^{hb}, W^{hb}\}^{b\in\{h,l\}}$ with the belief that agent A is efficient. The situation is, therefore, similar to the vertical hierarchy with separation, and the principal's offer for $t = 2$ coincides with the one presented in Lemma 6. If, instead, agent A reports θ^l in $t = 1$, then the principal in $t = 2$ offers a menu $\{q^{lb}, \omega_A^{lb}, \omega_B^{lb}\}^{b\in\{h,l\}}$ with the belief that agent A is inefficient. Hence, the situation is similar to the horizontal hierarchy with separation, and the principal's offer for $t = 2$ coincides with the one presented in Lemma 3. This leads to the following lemma.

Lemma 9 *The sequentially rational contract Φ_2^Ψ in $t = 2$ for a conditional hierarchy Ψ_{VH} is $q^{lh} = q^{ll} = \gamma$; $\omega_A^{lh} = \omega_A^{ll} = \omega_B^{ll} = (1-\theta^l)\gamma$, $\omega_B^{lh} = (1-\theta^h)\gamma$ and*

$$\begin{cases} q_2^{hh} = q_2^{hl} = 2\gamma;\ W^{hh} = W^{hl} = (2-\theta^h-\theta^l)\gamma & \text{if } \mu \le \widehat{\mu}(\theta^h), \\ q_2^{hh} = 2\gamma,\ q_2^{hl} = 0;\ W^{hh} = (2-\theta^h-\theta^h)\gamma,\ W^{hl} = 0 & \text{if } \mu > \widehat{\mu}(\theta^h). \end{cases}$$

Whether or not the conditional hierarchy Ψ_{VH} is actually implementable depends on the existence of wage transfers ω_1^l and ω_1^h that provide agent A with an incentive to reveal his type truthfully, when he anticipates Φ_2^Ψ as in Lemma 9. The following proposition demonstrates that this is the case, when μ and γ are small enough, but the principal's payoff with Ψ_{VH} inferior to the principal's optimal payoff in the unconditional hierarchy $\Psi = V$.

Proposition 5 *The principal can implement Ψ_{VH} iff $\mu < \widehat{\mu}(\theta^h)$ and $\gamma < 1/(1-\mu)$. When implementable, Ψ_{VH} yields the principal a payoff $\Pi^{VH} \le \Pi^V$, where $\Pi^{VH} \equiv \mu\theta^h + (\widetilde{\theta} + \theta^l)\gamma$..*

Proposition 5 implies that a selective promotion of the efficient agent does not help the principal in providing incentives for information revelation.

We next turn to the conditional hierarchy Ψ_{HV}. With a conditional hierarchy Ψ_{HV}, the principal faces a horizontal hierarchy in $t = 2$ if agent A reports θ^h in $t = 1$. In this case, the principal in $t = 2$ offers a contract $\{q^{hb}, \omega_A^{hb}, \omega_B^{hb}\}^{b\in\{h,l\}}$ with the belief that agent A is efficient. Hence, the situation is similar to the horizontal hierarchy with separation, and the principal's offer for $t = 2$ coincides with the one presented in Lemma 3. If, instead, agent A reports θ^l, the principal in $t = 2$ offers a contract $\{q_2^{lb}, W^{lb}\}^{b\in\{h,l\}}$ with the belief that agent A is inefficient. The situation is similar to the vertical hierarchy with separation, and the principal's offer for $t = 2$ coincides with the one presented in Lemma 6. These considerations lead to the following result.

Lemma 10 *The sequentially rational contract Φ_2^{Ψ} in $t = 2$ for a conditional hierarchy Ψ_{HV} is $q^{hb} = \gamma$; $\omega_A^{hb} = (1 - \theta^h)\gamma$, $\omega_B^{hb} = (1 - \theta^b)\gamma$; $b \in \{h, l\}$, and*

$$\begin{cases} q_2^{lb} = 2\gamma;\ W^{lb} = (2 - \theta^l - \theta^l)\gamma;\ b \in \{h,l\} & \text{if } \mu \leq \widehat{\mu}(\theta^l), \\ q_2^{lh} = 2\gamma,\ q_2^{ll} = 0;\ W^{lh} = (2 - \theta^l - \theta^h)\gamma,\ W^{ll} = 0 & \text{if } \mu > \widehat{\mu}(\theta^l). \end{cases}$$

As before, implementability of Ψ_{HV} depends on the existence of wage transfers ω_1^l and ω_1^h so that agent A has an incentive to reveal his type truthfully, when he anticipates Φ_2^{Ψ} as in Lemma 10. The following proposition shows that this is the case, when μ is large and γ is small enough.

Proposition 6 *The principal can implement Ψ_{HV} iff $\mu > \widehat{\mu}(\theta^l)$ and $\gamma < 1/\mu$. When implementable, Ψ_{HV} yields the principal a payoff Π^{HV} which exceeds Π^V for $\mu > \hat{\mu}(\theta^h)$, where $\Pi^{HV} = \mu\theta^h + \mu(2 - \mu)(\theta^h + \theta^l)\gamma$.*

Hence, for μ large enough, the conditional hierarchy Ψ_{HV} outperforms the unconditional vertical hierarchy. To understand this result it is helpful to reconsider from the previous section, why for $\mu > \hat{\mu}(\theta^h)$, the vertical hierarchy ($\Psi = V$) performs better than the horizontal hierarchy ($\Psi = H$). Lemma 6 shows that, in the vertical hierarchy with μ large enough, production in $t = 2$ takes place only if agent B is efficient. This feature of the sequentially rational contract has a negative and a positive effect as compared to the horizontal hierarchy. The negative effect is that there is simply less production for the principal in $t = 2$. The positive effect is that the principal has to pay less rents for information revelation. As explained, it is exactly due to this smaller rent that the principal is able to induce separation in $t = 1$ and that the vertical hierarchy outperforms the horizontal one. Note, however, that this positive effect of reduced production benefits the principal only when agent A is inefficient, whereas the reduced production only hurts the principal when agent A is efficient. The principal would, therefore, benefit when she could reduce production only when agent A is inefficient. The conditional hierarchy Ψ_{HV} allows her to achieve this selective reduction in production. This explains why the conditional hierarchy can outperform the unconditional vertical hierarchy.

7 Conclusion

Dynamic incentive problems arise when organizations have limited commitment. In this paper, we propose an explanation of different hierarchical structures based on such dynamic incentive problems. We show that information revelation in the first period is easier to achieve in the vertical hierarchy, because it increases the principal's credibility to give rents and/or cut production in the future. Consequently, the optimal hierarchical structure is *vertical* when information revelation is more important, i.e., when uncertainty about the types of the agents is large. When such uncertainty is small so that information revelation is less important, the optimal hierarchy is *horizontal*. As Mintzberg (1979) points out, many professional organizations have a tendency to adopt the horizontal hierarchical hierarchy. In traditional manufacturing industries, on the other hand, the variations in skill sets and efficiencies of administrative staff and line workers are often large. Our result suggest that an organization in such industries benefit from vertical hierarchies.

To keep things tractable, we assumed that the senior agent perfectly observes the private information of the junior agent. Our results, however, are robust to a more realistic situation, where the senior agent imperfectly observes the junior's information. In such a case, the contract that the senior agent offers to the junior agent becomes more complex, and leaves the junior agent a rent for his remaining private information. However, as long as the senior agent learns something about the junior agent while the principal does not, the vertical hierarchy still mitigates the dynamic incentive problem.

Such a partial information revelation would allow us to extend our framework to an overlapping generation model, where employees first enter the organization as subordinates and, after becoming seniors over the next generation of young employees in a second stage, retire from the firm entirely.[11] As a subordinate, an employee loses part of his private information, and the associated rent, to his superior. He, however, regains the lost amount of rent when he becomes senior later. As shown in our formal framework, the postponement of rents reduces hold–up problems and facilitates early information revelation. This reflects well a typical career path of employees in many organizations.

Appendix 1: Proofs

This appendix collects the formal proofs of the lemmas and propositions.

Proof of Lemma 2 First note that solving the overall maximization problem follows from combining the two solutions of the submaximization problems indexed by $b \in \{h, l\}$.

$$P^b : \max_{q^{hb}, q^{lb}, \omega_A^{hb}, \omega_A^{lb}} \mu(q^{hb} - \omega_A^{hb} + \theta^b q^{hb}) + (1-\mu)(q^{lb} - \omega_A^{lb} + \theta^b q^{lb}), \text{ s.t. (5) and (6).}$$

As is standard, only the participation constraint in (5) with respect to the inefficient type θ_A^l and the incentive constraint in (6) with respect to the efficient type θ_A^h are binding. These constraints yield $\omega_A^{lb} = (1-\theta^l)q^{lb}$ and $\omega_A^{hb} = (1-\theta^h)q^{hb} + \Delta\theta q^{lb}$. After substituting these wages into the objective function, we are left to solve:

$$\max_{q^{hb}, q^{lb}} \mu(\theta^h + \theta^b)q^{hb} + (1-\mu)(\theta^l + \theta^b - [\mu/(1-\mu)]\Delta\theta)q^{lb}.$$

The objective function is increasing in q^{hb}, and hence $q^{hb} = \gamma$. Likewise, the objective function is non-decreasing in q^{lb} if $\mu \leq \hat{\mu}(\theta^b)$. Thus, $q^{lb} = \gamma$ if $\mu \leq \hat{\mu}(\theta^b)$, and $q^{lb} = 0$ otherwise. Because $\hat{\mu}(\theta^l) < \hat{\mu}(\theta^h)$ we have the three cases as specified in the lemma. ■

[11] We refrain from a fully fledged analysis of imperfect information revelation becasuse it yields an intractable framework.

Proof of Proposition 1 For $\mu \leq \hat{\mu}(\theta^l)$, it follows from Lemma 2 that the sequential rational contract yields the principal a second period payoff of $\Pi_2^H = (\theta^l + \widetilde{\theta})\gamma$. For $\mu \in (\hat{\mu}(\theta^l), \hat{\mu}(\theta^h)$ this payoff is $\Pi_2^H \equiv \mu(2-\mu)(\theta^h + \theta^l)\gamma$. For $\mu > \hat{\mu}(\theta^h)$, her payoff is $\Pi_2^H = \mu(\theta^h + \widetilde{\theta})\gamma$. These payoffs are independent of the first period contract Φ_2^{HP}. The principal's optimal first period pooling contract Φ_2^{HP}, therefore, simply maximizes $\Pi_1^{HP} \equiv q_1 - \omega_1$ subject to the participation constraints of agent A over both periods. Lemma 2 shows that the inefficient agent A does not get a rent from the sequential rational contract. His participation constraint is, therefore, $\omega_1 - (1-\theta^l)q_1 \leq 0$. Given this constraint, the contract $q_1 = 1$ and $\omega_1 = 1 - \theta^l$ maximizes Π_1^{HP} and, because this contract automatically satisfies also the participation of the efficient agent A, it is optimal. It yields the principal the payoff θ^l in $t = 1$. A simple summation of $\Pi_1^{HP} + \Pi_2^H$ yields the expression for Π^{HP}. ■

Proof of Lemma 3 From (4), $\omega_B^{ab} = (1-\theta^b)q^{ab}$. Given the report θ_1^a, the principal believes to face an agent A of type θ_1^a with certainty. Consequently, she believes that agent A's participation constraint is $\omega_A^{ab} - (1-\theta_1^a)q^{ab} \geq 0$. Maximizing the principal's profits $2q^{ab} - \omega_A^{ab} - (1-\theta^b)q^{ab}$ under the agent's participation constraint yields the result. ■

Proof of Lemma 4 Suppose separation is implementable with some first period contract with wages ω_1^h and ω_1^l. The incentive constraints (7) and (8) imply that:

$$1 - \theta^l \geq \omega_1^h - \omega_1^l \geq 1 - \theta^h + \gamma\Delta\theta.$$

To have $\omega_1^h$and ω_1^l that satisfy the inequality, $1 - \theta^l > 1 - \theta^h + \gamma\Delta\theta$ must hold, but this inequality contradicts with $\gamma > 1$. ■

Proof of Lemma 5 As usual in the model of this type, the participation constraint (11) for the most inefficient type Θ^l, and the incentive constraints (12) for the more efficient adjacent types, Θ^h (for Θ^m) and Θ^m (for Θ^l), are binding provided that this solution yields a non-decreasing schedule $q_2^h \geq q_2^m \geq q_2^l$.

The binding constraints imply, successively, $W^l = (2-\Theta^l)q_2^l/2$, $W^m = (2-\Theta^m)q_2^m/2 + (\Theta^m - \Theta^l)q_2^l/2$, and $W^h = (2-\Theta^h)q_2^h/2 + (\Theta^h - \Theta^m)q_2^m/2 + (\Theta^m - \Theta^l)q_2^l/2$. Substitution into the objective function implies that we maximize:

$$\begin{aligned}
& \varphi^h\left[q_2^h-(2-\Theta^h)q_2^h/2-(\Theta^h-\Theta^m)q_2^m/2-(\Theta^m-\Theta^l)q_2^l/2\right] \\
& +\varphi^m\left[q_2^m-(2-\Theta^h)q_2^h/2-(\Theta^m-\Theta^l)q_2^l/2\right]+\varphi^l\left[q_2^l-(2-\Theta^l)q_2^l/2\right] \\
= & \varphi^h\left[\Theta^h q_2^h-\Delta\theta q_2^m/2-\Delta\theta q_2^l/2\right]+\varphi^m\left[(\Theta^h+\Theta^l)q_2^m/2-\Delta\theta q_2^l/2\right]+\varphi^l\Theta^l q_2^l \\
= & q_2^h\varphi^h\Theta^h+q_2^m\left[\varphi^m(\Theta^h+\Theta^l)-\varphi^h\Delta\theta\right]/2+q_2^l\left[\varphi^l\Theta^l-\varphi^m\Delta\theta/2-\varphi^l\Delta\theta/2\right].
\end{aligned}$$

Maximizing the expression with respect to $q_2^h \in \{0, 2\gamma\}$ yields $q_2^h = 2\gamma$. Maximizing the expression with respect to $q_2^m \in \{0, 2\gamma\}$, we get $q_2^m = 2\gamma$ if $2\mu(1-\mu)(\theta^h + \theta^l) \geq \mu^2\Delta\theta$, which is equivalent to $\mu \leq \bar{\mu}$, and $q_2^m = 0$ otherwise. Finally, maximizing the expression with respect to $q_2^l \in \{0, 2\gamma\}$, we get $q_2^l = 2\gamma$ if $(1-\mu)^2\theta^l \geq 2\mu(1-\mu)\Delta\theta/2 + \mu^2\Delta\theta/2$, which is equivalent to $\mu \leq \underline{\mu}$, and $q_2^l = 0$ otherwise.

The schedule q_2^h, q_2^m, q_2^l is monotone if $\overline{\mu} \geq \underline{\mu}$, because it then follows that $q_2^l = 2\gamma$ implies $q_2^m = 2\gamma$. To see that $\bar{\mu} \geq \underline{\mu}$, note that it is, per definition, equivalent to $2(\theta^h+\theta^l)/(3\theta^h+\theta^l) \geq 1 - \sqrt{\Delta\theta/(\theta^h+\theta^l)}$, which is equivalent to $\sqrt{\Delta\theta/(\theta^h+\theta^l)} \geq \Delta\theta/(3\theta^h+\theta^l)$. But the last inequality holds due to $\sqrt{\Delta\theta/(\theta^h+\theta^l)} > \Delta\theta/(\theta^h+\theta^l) > \Delta\theta/(3\theta^h+\theta^l)$. ■

Proof of Proposition 2 For $\mu \leq \underline{\mu}$, it follows from Lemma 5 that the sequential rational contract yields the principal a second period payoff of $\Pi_2^{VP} = 2\theta^l\gamma$ and the inefficient agent A expects a payoff $\mu\Delta\theta\gamma$, because he becomes type Θ^m with probability μ. For $\mu \in (\underline{\mu}, \overline{\mu}]$ this payoff is $\Pi_2^{VP} = \left[1-(1-\mu)^2\right](\theta^h+\theta^l)\gamma$ and the inefficient agent A expects a zero payoff. For $\mu > \overline{\mu}$, her payoff is $\Pi_2^{VP} = 2\mu^2\theta^h\gamma$ and the inefficient agent A expects a zero payoff. These payoffs are independent of the first period contract Φ_2^V. The principal's optimal first period pooling contract Φ_2^V, therefore, simply maximizes $\Pi_1^{VP} \equiv q_1 - \omega_1$ subject to the participation constraints of agent A over both periods. The inefficient agent A receive a rent in $t = 2$ only for $\mu \leq \underline{\mu}$. In this case, his participation constraint is $\omega_1 - (1-\theta^l)q_1 + \mu\Delta\theta\gamma \leq 0$ so that the optimal first period contract is $q_1 = 1$ and $\omega_1 = (1-\theta^l)q_1 - \mu\Delta\theta\gamma \leq 0$. It yields the principal the payoff $\theta^l + \mu\Delta\theta\gamma$ in $t = 1$. For $\mu > \underline{\mu}$, the inefficient agent A does not expect a rent from the sequential rational contract so that his participation constraint is $\omega_1 - (1-\theta^l)q_1 \leq 0$. Hence, the optimal first period contract is $q_1 = 1$ and $\omega_1 = 1 - \theta^l$. It yields the principal the payoff θ^l in $t = 1$. Adding the payoffs over both period yields the expression for Π^{VP} in the Proposition.

Next, we show that $\Delta\Pi^P \equiv \Pi^H - \Pi^{VP} \geq 0$ for all $\mu \in [0, 1]$. In order to compute $\Delta\Pi^P$ we first claim that $\underline{\mu} < \hat{\mu}(\theta^l)$ and $\overline{\mu} > \hat{\mu}(\theta^h)$. The first claim follows because

$$\hat{\mu}(\theta^l) - \underline{\mu} = \frac{2\theta^l}{\theta^h + \theta^l} - 1 + \sqrt{\frac{\Delta\theta}{\theta^h + \theta^l}} = -\frac{\Delta\theta}{\theta^h + \theta^l} + \sqrt{\frac{\Delta\theta}{\theta^h + \theta^l}},$$

which is positive because the fraction is smaller than 1. The second claim follows from

$$\overline{\mu} - \hat{\mu}(\theta^h) = \frac{2(\theta^h + \theta^l)}{3\theta^h + \theta^l} - \frac{\theta^l + \theta^h}{2\theta^h} = \frac{{\theta^h}^2 - {\theta^l}^2}{2\theta^h(3\theta^h + \theta^l)} > 0.$$

Consequently,

$$\Delta\Pi^P = \begin{cases} 0 & \text{if } \mu \leq \underline{\mu} \\ (1-\mu)(2\theta^l - \mu(\theta^h + \theta^l))\gamma & \text{if } \underline{\mu} < \mu \leq \hat{\mu}(\theta^l) \\ 0 & \text{if } \hat{\mu}(\theta^l) < \mu \leq \hat{\mu}(\theta^h) \\ \mu(2\mu\theta^h - (\theta^h - \theta^l))\gamma & \text{if } \hat{\mu}(\theta^h) < \mu \leq \overline{\mu} \\ (1-\mu)\mu(\theta^h + \theta^l)\gamma & \text{if } \mu > \overline{\mu}. \end{cases}$$

These terms are all non-negative, because $\mu \leq \hat{\mu}(\theta^l)$ implies $2\theta^l \geq \mu(\theta^h + \theta^l)$ and $\mu > \hat{\mu}(\theta^h)$ implies $2\mu\theta^h > \theta^h - \theta^l$. ■

Proof of Lemma 6 Given the report θ_1^a in $t = 1$, the maximization problem is equivalent to a static contracting problem where there is private information about type $\theta^b \in \{\theta^h, \theta^l\}$. As in Lemma 2, the participation of the inefficient type and the incentive compatibility constraint of the efficient type are binding. This yields wages:

$$W^{al} = (2 - \theta^a - \theta^l)q_2^{al}/2 \text{ and } W^{ah} = (2 - \theta^a - \theta^h)q_2^{ah}/2 + \Delta\theta q_2^{al}/2.$$

Substituting these variables in the principal's objective function yields:

$$\begin{aligned} & \mu[q_2^{ah} - (2 - \theta^a - \theta^h)q_2^{ah}/2 - \Delta\theta q_2^{al}/2] + (1-\mu)[q_2^{al} - (2 - \theta^a - \theta^l)q_2^{al}/2] \\ = \; & \mu q_2^{ah}[\theta^a + \theta^h]/2 + q_2^{al}[(1-\mu)(\theta^a + \theta^l) - \mu\Delta\theta]/2. \end{aligned}$$

Maximizing this expression for $q_2^{ah} \in \{0, 2\gamma\}$ yields $q_2^{ah} = 2\gamma$. Maximizing the expression for $q_2^{al} \in \{0, 2\gamma\}$ yields $q_2^{al} = 2\gamma$ if $(1-\mu)(\theta^a + \theta^l) \geq \mu\Delta\theta$, which is equivalent to $\mu \leq \hat{\mu}(\theta^a)$, and $q_2^{al} = 0$ otherwise. Because $\hat{\mu}(\theta^h) > \hat{\mu}(\theta^l)$ the lemma follows. ■

Proof of Lemma 7 After reporting $\theta^a = \theta^h$ in $t = 1$, the inefficient agent A receives, as specified in Lemma 6, the contract $\{q_2^{hb}, W^{hb}\}^{b\in\{h,l\}}$ when he reports agent B's type in $t = 2$. That is, he receives the payoff $W^{lb} - (1-\theta^l)q_2^{lb}/2 - (1-\theta^{b'})q_2^{lb}/2$ from accepting the contract when agent B is actually of type $\theta^{b'}$ (which may differ from the reported θ^b). It is straightforward to check that, for the contracts in Lemma 6 these payoffs are all non-positive for all combinations (b, b') and any $\mu \in [0, 1]$ so that it is optimal for the inefficient agent to reject the contract.

After reporting $\theta^a = \theta^l$ in $t = 1$, the efficient agent A receives the contract $\{q_2^{lb}, W^{lb}\}^{b\in\{h,l\}}$ as specified in Lemma 6. That is, he receives the payoff $W^{lb} - (1-\theta^h)q_2^{lb}/2 - (1-\theta^{b'})q_2^{lb}/2$ from accepting the contract when agent B is actually of type $\theta^{b'}$. It is straightforward to check that, for the contracts in Lemma 6, the efficient agent A always has a weakly higher payoff from reporting agent B's type $\theta^{b'}$ honestly and this payoff is non–negative. ■

Proof of Lemma 8 We distinguish the three different cases of Lemma 6.

Case i) $\mu \leq \hat{\mu}(\theta^l)$: From Lemma 6 it follows that the efficient agent A receives an expected information rent $\mu\Delta\gamma$ if he reports θ^h in $t = 1$. If instead he reports θ^l then, by Lemma 7 and 6, he receives a rent in $t = 2$ of $2\Delta\theta\gamma$ if agent B happens to be efficient and $\Delta\theta\gamma$ if agent B happens to be inefficient. Hence, the contract Φ_1^{VS} is incentive compatibility to type θ^h if:

$$\omega_1^h - (1-\theta^h) + \mu\Delta\theta\gamma \geq \omega_1^l + \Delta\theta\gamma + \mu\Delta\theta\gamma.$$

Likewise, it follows from Lemma 6 and Lemma 7 that, irrespective of his report in $t = 1$, the inefficient agent A receives no rent. Hence, the contract Φ_1^{VS} is incentive compatibility to type θ^l if:

$$\omega_1^l + \mu\Delta\theta\gamma \geq \omega_1^h - (1-\theta^l).$$

Combining the previous two conditions shows that the contract Φ_1^{VS} is incentive compatible exactly when:

$$(1-\theta^h)+\Delta\theta\gamma \leq \omega_1^h - \omega_1^l \leq (1-\theta^l)+\mu\Delta\theta\gamma. \tag{15}$$

Implementability requires $1-\theta^h+\Delta\theta\gamma < 1-\theta^l+\mu\Delta\theta\gamma$ which holds exactly when $\gamma < 1/(1-\mu)$...

Case ii) $\mu \in (\hat{\mu}(\theta^l), \hat{\mu}(\theta^h)]$: From Lemma 6 and Lemma 7 it follows that the efficient agent A receives an expected information rent of $\mu\Delta\theta\gamma$ irrespective of his report in $t=1$. Hence, the contract Φ_1^{VS} is incentive compatibility to type θ^h if:

$$\omega_1^h - (1-\theta^h) + \mu\Delta\theta\gamma \geq \omega_1^l + \mu\Delta\theta\gamma.$$

Likewise, it follows from Lemma 6 that the inefficient agent A receives an expected rent $\mu\Delta\theta\gamma$ in $t=2$ from reporting θ^l. If he reports θ^h instead then, by Lemma 7, he receives no rent in $t=2$. Hence, the contract Φ_1^{VS} is incentive compatibility to type θ^l if:

$$\omega_1^l + \mu\Delta\theta\gamma \geq \omega_1^h - (1-\theta^l).$$

Combining the previous two conditions shows that the contract Φ_1^{VS} is incentive compatible exactly when:

$$(1-\theta^h) \leq \omega_1^h - \omega_1^l \leq (1-\theta^l). \tag{16}$$

Hence, implementability requires $1-\theta^h < 1-\theta^l$ which is always the case.

Case iii) $\mu > \hat{\mu}(\theta^h)$: From Lemma 6 and Lemma 7 it follows that the efficient agent A receives no rent in $t=2$ if he reports θ^h in $t=1$. If, instead, he reports θ^l then Lemma 7 and 6 imply that he receives a rent in $t=2$ of $\Delta\theta\gamma$ in $t=2$ if agent B happens to be efficient and no rent if agent B happens to be inefficient. Hence, the contract Φ_1^{VS} is incentive compatibility to type θ^h if:

$$\omega_1^h - (1-\theta^h) \geq \omega_1^l + \mu\Delta\theta\gamma...$$

Likewise, it follows from Lemma 6 that, irrespective of his report in $t=2$, the inefficient agent A receives no rent in $t=2$. Hence, the contract Φ_1^{VS} is incentive compatibility to type

θ^l if:

$$\omega_1^l \geq \omega_1^h - (1-\theta^l).$$

Combining the previous two conditions shows that the contract Φ_1^{VS} is incentive compatible exactly when:

$$(1-\theta^h) + \mu\Delta\theta\gamma \leq \omega_1^h - \omega_1^l \leq (1-\theta^l). \tag{17}$$

Implementability requires $1-\theta^h+\mu\Delta\theta\gamma < 1-\theta^l$ which holds exactly when $\gamma < 1/\mu$. ■

Proof of Proposition 3 The principal's optimal wage structure minimizes her expected wages:

$$E_\theta\left[\omega_1^a\right] = \mu\omega_1^h + (1-\mu)\omega_1^l,$$

and we are left to determine ω_1^h and ω_1^l. For $\mu \leq \hat{\mu}(\theta^l)$, the principal minimizes $E_\theta\left[\omega_1^a\right]$ under the incentive constraints (15) and the participation constraints,

$$\omega_1^h - (1-\theta^h) + \mu\Delta\theta\gamma \geq 0 \text{ and } \omega_1^l + \mu\Delta\theta\gamma \geq 0.$$

At the optimum the participation constraint of type θ^l and the incentive constraint of type θ^h are binding. This yields $\omega_1^h = 1-\theta^h+(1-\mu)\Delta\theta\gamma$ and $\omega_1^l = -\mu\Delta\theta\gamma$. The principal's payoff is $\Pi^V = \mu\theta^h + (\tilde{\theta}+\theta^l)\gamma$.

For $\hat{\mu}(\theta^l) \leq \mu \leq \hat{\mu}(\theta^h)$, the principal minimizes $E_\theta\left[\omega_1^a\right]$ under the incentive constraints (16) and the participation constraints,

$$\omega_1^h - (1-\theta^h) + \mu\Delta\theta\gamma \geq 0 \text{ and } \omega_1^l \geq 0.$$

At the optimum the participation constraint of type θ^l and θ^h are binding. This yields $\omega_1^h = 1-\theta^h-\mu\Delta\theta\gamma$ and $\omega_1^l = 0$. The principal's payoff is $\Pi^V = \mu(\theta^h+\mu\Delta\theta\gamma)+(1-(1-\mu)^2)(\theta^h+\theta^l)\gamma = \mu\theta^h + 2\mu(\theta^h+(1-\mu)\theta^l)\gamma$.

For $\mu > \hat{\mu}(\theta^h)$ we minimize $E_\theta\left[\omega_1^a\right]$ under the incentive constraints (17) and the participation constraints,

$$\omega_1^h - (1-\theta^h) \geq 0 \text{ and } \omega_1^l \geq 0.$$

At the optimum the participation constraint of type θ^l and the incentive constraint of type θ^h are binding. This yields $\omega_1^h = 1 - \theta^h + \mu\Delta\theta\gamma$ and $\omega_1^l = 0$. The principal's payoff is $\Pi^V = \mu(\theta^h - \mu\Delta\theta\gamma) + [2\mu^2\theta^h + (1-\mu)\mu(\theta^h + \theta^l)\gamma = \mu\theta^h + \mu(\theta^h + \theta^l)\gamma$. ■

Proof of Proposition 4 By Proposition 2, the horizontal hierarchy is optimal whenever first period separation in the vertical hierarchy is not implementable. We therefore only need to compare Π^H with Π^V. From Proposition 1 and 3, we compute $\Delta\Pi = \Pi^H - \Pi^V$ whenever first period separation in the vertical hierarchy is implementable. For $\mu \leq \hat{\mu}(\theta^l)$ and $\gamma < 1/(1-\mu)$ it follows $\Delta\Pi = \theta^l - \mu\theta^h$, which is positive exactly when $\mu < \theta^l/\theta^h = \hat{\mu}(0)$. Hence, if $\mu < \hat{\mu}(0)$ the horizontal hierarchy is optimal irrespective of whether γ. Instead, for $\hat{\mu}(0) < \mu \leq \hat{\mu}(\theta^l)$, the vertical hierarchy is optimal whenever it is implementable which is the case for $\gamma \geq 1/(1-\mu)$. For $\hat{\mu}(\theta^l) < \mu \leq \hat{\mu}(\theta^h)$, we have $\Delta\Pi = \theta^l - \mu\theta^h - \mu^2\Delta\theta < -\mu^2\Delta\theta < 0$, where the first inequality follows because $\mu \geq \hat{\mu}(\theta^l) > \hat{\mu}(0)$. Hence, the vertical hierarchy is optimal whenever implementable and implementability is always ensured. For $\mu > \hat{\mu}(\theta^h)$ and $\gamma < 1/\mu$ we have $\Delta\Pi = \theta^l - \mu\theta^h + \mu(\tilde{\theta} - \theta^l)\gamma = \theta^l - \mu\theta^h + \mu^2\Delta\theta\gamma$. This is negative exactly when $\gamma < \mu(\theta^h - \theta^l)/(\mu^2\Delta\theta)$. Because $\mu(\theta^h - \theta^l)/(\mu^2\Delta\theta) < \mu(\theta^h - \mu\theta^l)/(\mu^2\Delta\theta) < 1/\mu$. $\Delta\Pi < 0$ implies $\gamma < 1/\mu$ so that first period separation is implementable. ■

Proof of Lemma 9: With Ψ_{VH}, the sequential rational contract $\{q_2^{ab},\ W^{ab}\}^{b\in\{h,l\}}$ after a first period report θ^h maximizes Π_2^V subject to (13) and (14) with $\theta^a = \theta^h$. Lemma 6 provides the solution to this problem. With Ψ_{VH}, the sequential rational contract $\{q^{ab}, \omega_A^{ab}, \omega_B^{ab}\}^{b\in\{h,l\}}$ in $t = 2$ after agent A's report that $\theta^a = \theta^l$ in $t = 1$ follows directly from Lemma 3. ■

Proof of Proposition 5 From the first period contract Φ_1^Ψ with Ψ_{VH}, agent A of type θ^l expects the payoff:

$$\omega_1^l - (1-\theta^l)q_1^l + (1-\theta^l)\gamma - (1-\theta^l)\gamma = \omega_1^l - (1-\theta^l)q_1^l \tag{18}$$

from reporting type θ^l in $t = 1$. If this type, in contrast, reports θ^h, then it is optimal for him to reject the sequential rational contract offer Φ_2^Ψ of Lemma 9 in $t = 2$ for both parameter constellations $\mu \leq \hat{\mu}(\theta^h)$ and $\mu > \hat{\mu}(\theta^h)$. Hence, reporting θ^h in $t = 2$ yields type θ^l the payoff:

$$\omega_1^h - (1-\theta^l)q_1^h. \tag{19}$$

Type θ^l has an incentive to report truthfully in $t = 1$ exactly when (18) exceeds (19), i.e., exactly when

$$\omega_1^h - \omega_1^l \leq (1-\theta^l)(q_1^h - q_1^l). \tag{20}$$

For the parameter constellation $\mu \leq \hat{\mu}(\theta^h)$, it follows from Lemma 9 that truthfully reporting his type θ^h yields agent A:

$$\omega_1^h-(1-\theta^h)q_1^h+(2-\theta^h-\theta^l)\gamma-(1-\theta^h)\gamma-\mu(1-\theta^h)\gamma-(1-\mu)(1-\theta^l)\gamma = \omega_1^h-(1-\theta^h)q_1^h+\mu\Delta\theta\gamma. \tag{21}$$

For the parameter constellation $\mu > \hat{\mu}(\theta^h)$, it follows from Lemma 9 that truthfully reporting his type θ^h yields agent A:

$$\omega_1^h - (1-\theta^h)q_1^h + \mu[(2-\theta^h-\theta^h)\gamma - (1-\theta^h)\gamma - (1-\theta^h)\gamma] = \omega_1^h - (1-\theta^h)q_1^h. \tag{22}$$

Instead, reporting θ^l in $t = 1$ yields agent A of type θ^h:

$$w_1^l - (1-\theta^h)q_1^l + (1-\theta^l)\gamma - (1-\theta^h)\gamma = w_1^l - (1-\theta^h)q_1^l + \Delta\theta\gamma... \tag{23}$$

Hence, for $\mu \leq \hat{\mu}(\theta^h)$ agent A of type θ^h has an incentive to report honestly exactly when (21) exceeds (23), i.e., exactly when

$$\omega_1^h - \omega_1^l \geq (1-\theta^h)(q_1^h - q_1^l) + (1-\mu)\Delta\theta\gamma. \tag{24}$$

This is consistent with (20) only if $q_1^h = 1$, $q_1^l = 0$ and $(1-\mu)\gamma \leq 1$. Similarly, for $\mu > \hat{\mu}(\theta^h)$ agent A of type θ^h has an incentive to report honestly exactly when (22) exceeds (23), i.e., when

$$\omega_1^h - \omega_1^l \geq (1-\theta^h)(q_1^h - q_1^l) + \Delta\theta\gamma. \tag{25}$$

This is however always inconsistent with (20).

This proves the first claim of the proposition that only for $\mu \leq \hat{\mu}(\theta^h)$ and $(1-\mu)\gamma \leq 1$, the conditional hierarchy is implementable. Moreover, implementability requires $q_1^h = 1$ and $q_1^l = 0$ so that the principal's payoff is:

$$\mu[1-\omega_1^h+2\gamma-(2-\theta^h-\theta^l)\gamma]+(1-\mu)[-w_1^l+2\gamma-(1-\theta^l)\gamma-\mu(1-\theta^h)\gamma-(1-\mu)(1-\theta^l)\gamma] \quad (26)$$

The optimal contract maximizes (26) subject to $q_1^h = 1$, $q_1^l = 0$, the incentive constraints (20) and (24), and the participation constraints which require that (18) and (22) are non–negative. Straightforward calculations yield that this problem is maximized for $\omega_1^h = 1-\theta^h+(1-\mu)\Delta\theta\gamma$ and $\omega_1^l = 0$ with payoff:

$$\Pi^{VH} = \mu\theta^h + (\tilde{\theta} + \theta^l). \quad (27)$$

From Proposition 3 and 4 we can compare this expression with Π^V for the implementable range $\mu \leq \hat{\mu}(\theta^h)$ and $(1-\mu)\gamma \leq 1$. This yields $\Pi^{VH} = \Pi^V$ when $\mu \in (\hat{\mu}(0), \hat{\mu}(\theta^l)]$ and $\Pi^{VH} - \Pi^V = [(2-\mu)\theta^l - 2\mu(1-\mu)\theta^l - \mu\theta^h]\gamma < 0$ for $\mu \in (\hat{\mu}(\theta^l), \hat{\mu}(\theta^h)]$ and $\gamma < 1/(1-\mu)$. This proves the second claim of the proposition. ■

Proof of Lemma 10 With Ψ_{HV}, the sequential rational contract $\{q_2^{ab},\ W^{ab}\}^{b\in\{h,l\}}$ after a first period report θ^l maximizes Π_2^V subject to (13) and (14) with $\theta^a = \theta^l$. Lemma 6 provides the solution to this problem. With Ψ_{HV}, the sequential rational contract $\{q^{ab}, \omega_A^{ab}, \omega_B^{ab}\}^{b\in\{h,l\}}$ in $t=2$ after agent A's report that $\theta^a = \theta^h$ in $t = 1$ follows directly from Lemma 3. ■

Proof of Proposition 6 First, consider the parameter constellation $\mu \leq \hat{\mu}(\theta^l)$... From the first period contract Φ_1^Ψ with Ψ_{HV}, agent A of type θ^l expects the payoff:

$$\omega_1^l-(1-\theta^l)q_1^l+2(1-\theta^l)\gamma-(1-\theta^l)\gamma-\mu(1-\theta^h)\gamma-(1-\mu)(1-\theta^l)\gamma = \omega_1^l-(1-\theta^l)q_1^l+\mu\Delta\theta\gamma \quad (28)$$

from reporting type θ^l in $t = 1$. If this type, in contrast, reports θ^h, then it is optimal for him to reject the sequential rational contract offer Φ_2^Ψ of Lemma 9 in $t = 2$. Hence, reporting θ^h in $t = 2$ yields type θ^l the payoff:

$$\omega_1^h - (1-\theta^l)q_1^h. \quad (29)$$

Hence, type θ^l has an incentive to report truthfully in $t = 1$ exactly when (28) exceeds (29), i.e., exactly when

$$\omega_1^h - \omega_1^l \leq (1-\theta^l)(q_1^h - q_1^l) + \mu\Delta\theta\gamma. \tag{30}$$

From Lemma 10 it follows that truthfully reporting his type θ^h yields agent A:

$$\omega_1^h - (1-\theta^h)q_1^h + (1-\theta^h)\gamma - (1-\theta^h)\gamma = \omega_1^h - (1-\theta^h)q_1^h. \tag{31}$$

Instead, reporting θ^l in $t = 1$ yields agent A of type θ^h:

$$\omega_1^l - (1-\theta^h)q_1^l + (2-\theta^l-\theta^l)\gamma - (1-\theta^h)\gamma - \mu(1-\theta^h)\gamma - (1-\mu)(1-\theta^l)\gamma = w_1^l - (1-\theta^h)q_1^l + (1+\mu)\Delta\theta\gamma... \tag{32}$$

Type θ^h has an incentive to report truthfully when (31) exceeds (32), i.e., whenever

$$\omega_1^h - \omega_1^l \geq (1-\theta^h)(q_1^h - q_1^l) + (1+\mu)\Delta\theta\gamma. \tag{33}$$

Because $\gamma > 1 \geq q_1^h - q_1^l$, this incentive constraint is never consistent with (30).

Next, consider the parameter constellation $\mu > \hat{\mu}(\theta^l)$. In this case, truthfully reporting his type θ^l yields agent A the payoff:

$$\omega_1^l - (1-\theta^l)q_1^l + \mu[(2-\theta^l-\theta^h)\gamma - (1-\theta^l)\gamma - (1-\theta^h)\gamma = \omega_1^l - (1-\theta^l)q_1^l. \tag{34}$$

Also for $\mu > \hat{\mu}(\theta^l)$, this type rejects the sequential rational contract offer Φ_2^Ψ after reporting θ^h instead. Hence, reporting θ^h in $t = 2$ yields type θ^l the payoff (29). Hence, type θ^l has an incentive to report truthfully in $t = 1$ exactly when (34) exceeds (29), i.e., exactly when

$$\omega_1^h - \omega_1^l \leq (1-\theta^l)(q_1^h - q_1^l). \tag{35}$$

From Lemma 10 it follows that truthfully reporting his type θ^h yields agent A the payoff (31). Instead, reporting θ^l in $t = 1$ yields agent A of type θ^h the following payoff[12]:

$$w_1^l - (1-\theta^h)q_1^l + \mu[(2-\theta^l-\theta^h)\gamma - (1-\theta^h)\gamma - (1-\theta^h)\gamma] = w_1^l - (1-\theta^h)q_1^l + \mu\Delta\theta\gamma... \tag{36}$$

[12] As before, the sequential rational contract in Lemma 10 that is offered after a first period report θ^l is also incentive compatible for agent A of type θ^h.

Hence, type θ^h has an incentive to report truthfully when (31) exceeds (36), i.e., whenever

$$\omega_1^h - \omega_1^l \geq (1-\theta^h)(q_1^h - q_1^l) + \mu\Delta\theta\gamma. \tag{37}$$

This incentive constraint is only consistent with (35) if $q_1^h = 1$, $q_1^l = 0$, and $\mu\gamma \leq 1$. This proves the first statement of the proposition.

Hence, implementability requires $\mu > \hat{\mu}(\theta^l)$, $q_1^h = 1$, $q_1^l = 0$, and $\mu\gamma \leq 1$ so that the principal's payoff is:

$$\mu[1-\omega_1^h + 2\gamma - (1-\theta^h)\gamma - \mu(1-\theta^h)\gamma - (1-\mu)(1-\theta^l)\gamma] + (1-\mu)[-w_1^l + 2\gamma - (2-\theta^l-\theta^h)\gamma]$$

$$= \mu[1-\omega_1^h + (\theta^h + \tilde{\theta})\gamma] + (1-\mu)[-\omega_1^l + \mu(\theta^l+\theta^h)\gamma] \tag{38}$$

The optimal contract maximizes (38) subject to $q_1^h = 1$, $q_1^l = 0$, the incentive constraints (35) and (37), and the participation constraints which require that (31) and (34) are non–negative. Straightforward calculations yield that this problem is maximized for $\omega_1^l = 0$ and $\omega_1^h = 1-\theta^h + \mu\Delta\theta\gamma$ with the payoff:

$$\Pi^{HV} = \mu\theta^h + \mu(2-\mu)(\theta^h+\theta^l)\gamma... \tag{39}$$

Using Proposition 3 and 4, we can compare (39) with (i) Π^V for $\mu \in (\hat{\mu}(\theta^l), \hat{\mu}(\theta^h)]$ and $\gamma < 1/\mu$, and (ii) Π^V for $\mu > \hat{\mu}(\theta^h)$ and $\gamma < \frac{\mu\theta^h - \theta^l}{\mu^2\Delta\theta}$. From a direct comparison, $\Pi^{HV} - \Pi^V = -\mu^2\Delta\theta\gamma < 0$ in case (i) and $\Pi^{HV} - \Pi^V = \mu(1-\mu)\Delta\theta\gamma > 0$ in case (ii). ■

Appendix 2: Semi-Separating Strategies

We showed the optimality of vertical hierarchies by focusing on pooling and full separation contracts. It is well known however that, in general, optimal contract may also involve semi–separation. This is due to the limited commitment of the principal, which leads to a failure of the revelation principle. In this appendix we show that the superiority of vertical hierarchies is not due to our neglect of semi–separation. We, thereby, need to consider only semi separating strategies in the horizontal hierarchy, because we demonstrate that these outcomes lead to payoffs that are already lower than the principal's optimal contract in a full separating vertical hierarchy.

Bester and Strausz (2001) show that, despite imperfect commitment by the principal, direct mechanisms can implement any Pareto optimal outcome between the principal and the agent. These direct mechanisms induce the agent to report his type truthfully with a strict positive probability, but may also require the agent to misreport with a strictly positive probability. In order to represent such mechanisms in our context, let $\alpha^a \in (0,1]$ denote the probability that agent A of type θ^a reports his type truthfully. A combination (α^h, α^l) represents agent A's reporting strategy when facing a direct mechanism. Because we can always relabel messages, we may also restrict attention to reporting strategies with $\alpha^h \geq 1 - \alpha^l$. Pooling contracts then coincide with the reporting strategy $\alpha^h = \alpha^l = 1/2$ and (full) separation contracts coincide with the reporting strategy $\alpha^h = \alpha^l = 1$. Because we focus in this appendix on reporting strategies that imply neither full pooling or full separation, we consider only reporting strategies from the set:

$$\mathcal{A} \equiv \{(\alpha^h, \alpha^l) \in (0,1) \times (0,1) \mid \alpha^h \geq 1 - \alpha^l\} \backslash \{(1/2, 1/2), (1,1)\}.$$

Moreover, denote the principal's posterior belief that agent A is efficient after he reports θ_A^a in $t = 1$ by μ_2^a. For a given reporting strategy α^h and α^l, the principal's beliefs μ_2^h and μ_2^l, in equilibrium, must satisfy Bayes' rule and:

$$\mu_2^h(\alpha^h, \alpha^l) \equiv \frac{\mu\alpha^h}{\mu\alpha^h + (1-\mu)\alpha^l} \text{ and } \mu_2^l(\alpha^h, \alpha^l) \equiv \frac{\mu(1-\alpha^h)}{\mu(1-\alpha^h) + (1-\mu)\alpha^l}. \tag{40}$$

For any reporting strategy $(\alpha^h, \alpha^l) \in \mathcal{A}$ expression (40) implies $\mu_2^l < \mu < \mu_2^h$.

In an equilibrium, the principal's offer in $t = 2$ is sequential rational given agent A's reporting behavior. We can, therefore, view it as contingent on agent A's report θ^a and, by Lemma 1, also on agent B's type θ^b. The crucial question is whether the sequential rational contract itself is a pooling or a separation one. Therefore, let ρ^{ab} denote the probability that the principal offers a pooling contract in $t = 2$ given agent A's report θ^a and agent B's type

θ^b. By Lemma 2, sequentially rational behavior of the principal implies:

$$\rho^{ab} \in \begin{cases} \{1\} & \text{if } \mu_2^a < \hat{\mu}(\theta^b), \\ [0,1] & \text{if } \mu_2^a = \hat{\mu}(\theta^b), \\ \{0\} & \text{if } \mu_2^a > \hat{\mu}(\theta^b). \end{cases} \tag{41}$$

From $\mu_2^l < \mu < \mu_2^h$, it then follows that $\rho^{lb} \geq \rho^{hb}$. This means that, in equilibrium, it is more likely that the principal offers a pooling contract if agent A reports himself as inefficient. It follows, because the subsequent decrease in the principal's belief that agent A is efficient shifts the trade–off between a pooling and a separation contract more in favor of pooling contracts.

On the equilibrium path, any sequential rational contract in $t = 2$, leaves a rent to agent A only if the contract is a pooling one and, in this case, the rent is exactly $\Delta\theta\gamma$. Agent A of type θ^h therefore expects to receive this rent with probability $\mu\rho^{ah} + (1-\mu)\rho^{al}$ when he reports θ^a. Because, in equilibrium, agent A's reporting strategy (α^h, α^l) must be optimal, the reporting probability α^h satisfies:

$$\begin{aligned} \alpha^h \in \ & \arg\max_{\hat{\alpha}^h} \hat{\alpha}^h[\omega_1^h - (1-\theta^h) + (\mu\rho^{hh} + (1-\mu)\rho^{hl})\Delta\theta\gamma] \\ & + (1-\hat{\alpha}^h)[\omega_1^l + (\mu\rho^{lh} + (1-\mu)\rho^{ll})\Delta\theta\gamma]. \end{aligned} \tag{42}$$

Similarly, because agent A of type θ^l will never receive a rent in $t = 2$, the reporting strategy α^l, in equilibrium, satisfies:

$$\alpha^l \in \arg\max_{\hat{\alpha}^l} \hat{\alpha}^l\omega_1^l + (1-\hat{\alpha}^l)[\omega_1^h - (1-\theta^l)]. \tag{43}$$

With the help of α^h, α^l and ρ^{ab}, the principal's expected payoff in the beginning of $t = 1$ can be written as:

$$\begin{aligned} \Pi_{semi}^H = \ & \mu\{\alpha^h[1 - \omega_1^h + (\theta^h + \widetilde{\theta} - (\mu\rho^{hh} + (1-\mu)\rho^{hl})\Delta\theta)\gamma] \\ & +(1-\alpha^h)[-\omega_1^l + (\theta^h + \widetilde{\theta} - (\mu\rho^{lh} + (1-\mu)\rho^{ll})\Delta\theta)\gamma]\} \\ & +(1-\mu)\{\alpha^l[-\omega_1^l + (\mu\rho^{lh}(\theta^l + \theta^h) + (1-\mu)\rho^{ll}(\theta^l + \theta^l))\gamma] \\ & +(1-\alpha^l)[1 - \omega_1^h + (\mu\rho^{hh}(\theta^l + \theta^h) + (1-\mu)\rho^{hl}(\theta^l + \theta^l))\gamma]\}. \end{aligned} \tag{44}$$

Therefore, if semi-separation in the horizontal hierarchy is optimal, then there exists a combination $(\alpha^h, \alpha^l) \in \mathcal{A}$ that maximizes Π^H_{semi} subject to (40), (41), (42), and (43). In order to solve this problem, we first show that, with semi–separation, optimal reporting strategies are necessarily such that they induce a belief of the principal that makes her exactly indifferent about offering a pooling or separation contract in $t = 2$... The intuition behind this result is that these beliefs bring the principal's myopic second period behavior in line with the principal's ex ante perspective of $t = 1$. Because, in equilibrium, the principal's beliefs depend on the agent's reporting strategy, his reporting behavior is a key instrument for controlling the principal's myopic behavior in $t = 2$.

Lemma 11 *If reporting behaviors in $\mathcal{A}$ are optimal, then there exist an optimal α^h and α^l such that i) $\alpha^h = 1$ and $\mu^h_2(1, \alpha^l) = \hat{\mu}(\theta^h)$, or ii) $\alpha^l = 1$ and $\mu^l_2(\alpha^h, 1) = \hat{\mu}(\theta^l)$, or iii) $\mu^h_2(\alpha^h, \alpha^l) = \hat{\mu}(\theta^h)$ and $\mu^l_2(\alpha^h, \alpha^l) = \hat{\mu}(\theta^l)$.*

Proof. Suppose α^h and α^l are optimal and stochastic and $\mu^h_2(\alpha^h, \alpha^l) \neq \hat{\mu}(\theta^h)$ and $\mu^l_2(\alpha^h, \alpha^l) \neq \hat{\mu}(\theta^l)$. Because ρ^{ab} only changes at $\hat{\mu}(\theta^a)$, (44) is linear in α^h and α^l (for sufficiently small changes in α^h and α^l). However, the optimality of α^h and α^l implies that (44) can neither be increasing nor decreasing in α^h or α^l. Thus, (44) must be independent of α^h and α^l and we can increase or decrease α^h and α^l so that $\mu^h_2(\alpha^h, \alpha^l) = \hat{\mu}(\theta^h)$ or $\mu^l_2(\alpha^h, \alpha^l) = \hat{\mu}(\theta^l)$. Because the original ρ^{ab} is consistent with these adapted α^h and α^l, it yields the same expected payoff to the principal and is hence also optimal. Suppose now there exist no deterministic reporting strategy at the optimum. Then, $\alpha^h \in (0, 1)$ and $\alpha^l \in (0, 1)$ so that $\mu^h_2(\alpha^h, \alpha^l) = \hat{\mu}(\theta^h)$ or $\mu^l_2(\alpha^h, \alpha^l) = \hat{\mu}(\theta^l)$. If $\mu^h_2(\alpha^h, \alpha^l) = \hat{\mu}(\theta^h)$ and $\mu^l_2(\alpha^h, \alpha^l) \neq \hat{\mu}(\theta^l)$, then we can specify the implicit function $\widetilde{\alpha}^l(\alpha^h)$ by $\mu^h_2(\alpha^h, \widetilde{\alpha}^l(\alpha^h)) = \mu^h_2(\alpha^h, \alpha^l)$. Because $\widetilde{\alpha}^l$ is necessarily linear in α^h and (44) is linear in both α^h and α^l, replacing α_l with $\widetilde{\alpha}^l(\alpha^h)$ in (44) yields an expression for the principal's payoff $\widetilde{\Pi}^H_{semi}(\alpha^h)$ that is linear in α^h for a small change in α^h. Now, suppose $\mu^l_2(\alpha^h, \alpha^l) < \hat{\mu}(\theta^l)$, and consider the interval $[0, \bar{\alpha}^h]$, where $\bar{\alpha}^h$ is such that $\mu^l_2(\bar{\alpha}^h, \widetilde{\alpha}^l(\bar{\alpha}^h)) = \hat{\mu}(\theta^l)$. Note that if α^h and α^l maximize (44), then they must maximize $\widetilde{\Pi}^H_{semi}(\alpha^h)$ over $[0, \bar{\alpha}^h]$ at α^h, because ρ^{ab} remains unchanged for any combination $\{\alpha_h, \alpha^h, \widetilde{\alpha}^l(\alpha^h)\}$ with $\alpha^h \in [0, \bar{\alpha}^h]$. However, because $\widetilde{\Pi}^H_{semi}(\alpha^h)$ is linear in

α^h, it must be maximized at either 0 or $\bar{\alpha}^h$. It follows that (44) must also be optimized for $\{\alpha^h, \alpha^l\} = \{0, \widetilde{\alpha}^l(0)\}$ or $\{\alpha^h, \alpha^l\} = \{\bar{\alpha}^h, \widetilde{\alpha}^l(\bar{\alpha}^h)\}$. The first case, $\{\alpha^h, \alpha^l\} = \{0, \widetilde{\alpha}^l(0)\}$, contradicts the assumption that there only exist optimal outcome in purely mixed strategy, implying that $\{\bar{\alpha}^h, \widetilde{\alpha}^l(\bar{\alpha}^h)\}$ must be optimal. Yet, as we needed to show, this implies $\mu_2^l = \hat{\mu}(\theta^l)$. If $\mu_2^l(\alpha^h, \alpha^l) > \hat{\mu}(\theta^l)$ similar arguments show that there must also exist optimal α^h and α^l with $\mu_2^h(\alpha^h, \alpha^l) = \hat{\mu}(\theta^h)$ and $\mu_2^l(\alpha^h, \alpha^l) = \hat{\mu}(\theta^l)$. The case of $\mu_2^h(\alpha^h, \alpha^l) \neq \hat{\mu}(\theta^h)$ and $\mu_2^l(\alpha^h, \alpha^l) = \hat{\mu}(\theta^l)$ can be treated similarly to show that there must then also exist optimal α^h and α^l with $\mu_2^h(\alpha^h, \alpha^l) = \hat{\mu}(\theta^h)$ and $\mu_2^l(\alpha^h, \alpha^l) = \hat{\mu}(\theta^l)$. ■

With this characterization result of optimal reporting behaviors, we may compute the principal's maximum payoff for the three different cases and compute her maximum payoff, Π^H_{semi} among these cases.

Proposition 7 *The payoff Π^V exceeds the payoff from any contract in the horizontal hierarchy that induces a semi-separating whenever $\mu < \mu^m$ and, in particular, for any $\mu < \hat{\mu}(\theta^h)$, because:*

$$\hat{\mu}(\theta^h) < \mu^m \equiv \frac{2\theta^h\theta^l}{(\theta^h + \theta^l)\Delta\theta\gamma}.$$

Proof. We compute the payoffs for the three cases in Lemma 11. First, suppose $\alpha^h = 1$ and $\mu_2^h(1, \alpha^l) = \hat{\mu}(\theta^h)$ is optimal. Because $\mu_2^h > \mu$, this case exists only if $\hat{\mu}(\theta^h) > \mu$. Moreover, it holds $\mu_2^l = 0$ so that $\rho^{lh} = \rho^{ll} = 1$ and, due to $\mu_2^h(1, \alpha^l) = \hat{\mu}(\theta^h) > \hat{\mu}(\theta^l)$ it follows $\rho^{hl} = 0$ and the principal is indifferent between pooling and separation, and thus her expected payoff is independent of ρ^{hh}. Expression (44), therefore, simplifies to:

$$\begin{aligned}
&\mu[1 - \omega_1^h + (\theta^h + \tilde{\theta})\gamma] \\
&+(1-\mu)\{\alpha^l[-\omega_1^l + (\mu(\theta^l + \theta^h) + (1-\mu)(\theta^l + \theta^l))\gamma] + (1-\alpha^l)[1 - \omega_1^h]\},
\end{aligned}$$

and is to maximize under the incentive constraints,

$$\omega_1^h \geq (1 - \theta^h) + (1 - \mu\rho^{hh})\Delta\theta\gamma + \omega_1^l \quad \text{ and } \omega_1^l = \omega_1^h - (1 - \theta^l),$$

and the participation constraints,

$$\omega_1^l \geq 0 \quad \text{ and } \omega_1^h \geq 1 - \theta^h - \mu\rho^{hh}\Delta\theta\gamma.$$

Because ρ^{hh} does not influence the principal's expected payoff, but the constraints are unambiguously relaxed for larger ρ^{hh}, $\rho^{hh} = 1$ is optimal. As usual, the participation constraint for type θ^l together with the incentive constraint for type θ^h imply the participation constraint for type θ^h. It also follows that the two incentive constraints are consistent only if $(1-\mu)\gamma \leq 1$. In this case, $\omega_1^l = \omega_1^h - (1-\theta^l)$ implies that the incentive constraint for type θ^h is satisfied, because it follows:

$$(1-\theta^h) + (1-\mu)\Delta\theta\gamma + \omega_1^l = (1-\mu)\Delta\theta\gamma - \Delta\theta + \omega_1^h \leq \omega_1^h.$$

Hence, an optimal solution exists only if $(1-\mu)\gamma \leq 1$ and it exhibits:

$$\omega_1^h = 1 - \theta^l \text{ and } \omega_1^l = 0,$$

with the payoff:

$$\Pi_{semi}^H = \frac{2\mu\theta^h\theta^l + \left[(2\mu\theta^h - \mu^2\Delta\theta)\Delta\theta + 2\theta^l(\theta^h + \theta^l)\right]\gamma}{\theta^h + \theta^l},$$

which is smaller than Π^V for the entire implementable range $\mu < \hat{\mu}(\theta^h)$.

Suppose now $\alpha^l = 1$ and $\mu_2^l(\alpha^h, 1) = \hat{\mu}(\theta^l)$ is optimal. Because $\mu_2^l < \mu$, this case exists only if $\hat{\mu}(\theta^l) < \mu$. Moreover, it holds $\mu_2^h = 1$ so that $\rho^{hh} = \rho^{hl} = 0$ and, due to $\mu_2^l(\alpha^h, 1) = \hat{\mu}(\theta^l) < \hat{\mu}(\theta^h)$ it follows $\rho^{lh} = 1$ and the principal is indifferent between separation and pooling. Hence the principal's payoff in (44) is independent of ρ^{ll}, and simplifies to:

$$\begin{aligned}&\mu\{\alpha^h[1 - \omega_1^h + (\theta^h + \widetilde{\theta})\gamma] + (1-\alpha^h)[-\omega_1^l + (\theta^h + \widetilde{\theta} - \mu\Delta\theta)\gamma]\}\\ &+(1-\mu)[-\omega_1^l + \mu(\theta^l + \theta^h)\gamma],\end{aligned}$$

and is to maximize under the incentive constraints,

$$\omega_1^h = (1-\theta^h) + (\mu + (1-\mu)\rho^{ll})\Delta\theta\gamma + \omega_1^l \quad \text{ and } \omega_1^l \geq \omega_1^h - (1-\theta^l),$$

and the participation constraints,

$$\omega_1^l \geq 0 \quad \text{ and } \omega_1^h \geq 1 - \theta^h.$$

Again, the participation constraint of θ^h is implied by the incentive constraint of θ^h and the participation constraint of θ^l. Moreover, because the principal's expected payoff is decreasing in ω_1^h, and ω_1^h itself is decreasing in ρ^{ll}, a smaller ω_1^h relaxes the incentive constraint for type θ^l, and it follows that $\rho^{ll} = 0$ must be optimal. It then follows that the two incentive constraints are consistent only if $\mu\gamma \leq 1$. Only in this case, an optimal solution exists and exhibits:

$$\omega_1^h = (1-\theta^h) + \mu\Delta\theta\gamma \text{ and } \omega_1^l = 0,$$

with the payoff:

$$\Pi_{semi}^H = \frac{\mu{\theta^h}^2 - (2-\mu)\theta^h\theta^l + \mu(2-\mu)({\theta^h}^2 - {\theta^l}^2)\gamma}{\Delta\theta}.$$

This is smaller than Π^V for the range $\mu \in (\hat{\mu}(\theta^l), \hat{\mu}(\theta^h)]$. For $\mu > \hat{\mu}(\theta^h)$, straightforward calculations show that $\Pi_{semi}^H > \Pi^V$ when

$$\mu > \frac{2\theta^h\theta^l}{(\theta^h + \theta^l)\Delta\theta\gamma}.$$

Finally, suppose $\mu_2^h(\alpha^h, \alpha^l) = \hat{\mu}(\theta^h)$ and $\mu_2^l(\alpha^h, \alpha^l) = \hat{\mu}(\theta^l)$ is optimal. Because $\mu_2^l < \mu < \mu_2^h$ this case exists only if $\hat{\mu}(\theta^l) < \mu < \hat{\mu}(\theta^h)$. Due to $\mu_2^h(\alpha^h, \alpha^l) = \hat{\mu}(\theta^h) > \hat{\mu}(\theta^l)$, it follows $\rho^{hl} = 0$ and, due to $\mu_2^l(\alpha^h, \alpha^l) = \hat{\mu}(\theta^l) < \hat{\mu}(\theta^h)$, it follows $\rho^{lh} = 1$. Moreover, the principal is indifferent to ρ^{hh} and ρ^{ll}, and hence her expected payoff must be independent of both ρ^{hh} and ρ^{ll}. The expression in (44) simplifies to:

$$\begin{aligned}&\mu\{\alpha^h[1-\omega_1^h + (\theta^h + \widetilde{\theta})\gamma] + (1-\alpha^h)[-\omega_1^l + (\theta^h + \widetilde{\theta} - \mu\Delta\theta)\gamma]\}\\ &+(1-\mu)\{\alpha^l[-\omega_1^l + \mu(\theta^l + \theta^h)\gamma] + (1-\alpha^l)[1-\omega_1^h]\},\end{aligned} \tag{45}$$

and is to maximize under the incentive constraints,

$$\omega_1^h = (1-\theta^h) + (\mu(1-\rho^{hh}) + (1-\mu)\rho^{ll})\Delta\theta\gamma + \omega_1^l \quad \text{and } \omega_1^l = \omega_1^h - (1-\theta^l),$$

and the participation constraints,

$$\omega_1^l \geq 0 \quad \text{and } \omega_1^h \geq 1 - \theta^h + \mu\rho^{hh}\Delta\theta\gamma.$$

The incentive constraints are consistent only if $\mu(1-\rho^{hh})+(1-\mu)\rho^{ll}=1/\gamma$. This implies $\omega_1^h = 1-\theta^l+\omega_1^l$. Substituting out the wage ω_1^h in (45) demonstrates that the expression is decreasing in ω_1^l. The optimal wage is $\omega_1^l = 0$, which satisfies the participation constraint of type θ^h for $\rho^{hh}=0$. It yields the principal the payoff:

$$\Pi^H_{semi} = \frac{2\theta^h\theta^l(\mu(\theta^h+\theta^l)-2\theta^l)+\mu(2-\mu)(\theta^h+\theta^l)\Delta\theta^2\gamma}{\Delta\theta^2},$$

which is smaller than Π^V for the entire implementable range $\mu \in (\hat{\mu}(\theta^l), \hat{\mu}(\theta^h))$. ■

The proposition above states that, for $\mu \in (\hat{\mu}(0), \hat{\mu}(\theta^h))$, the vertical hierarchy is also optimal when considering semi–separating strategies. For $\mu \in (\hat{\mu}(\theta^h), 1)$, the region of vertical hierarchy in Figure 1 survives unless μ is very high. The intuition behind the result here is as follows. With the semi-separating strategy, the probability of zero output in both periods is strictly positive. However, there is a chance that separation takes places in $t=1$ with no rent provision in $t=2$. Thus, if it is highly likely that the agent is efficient, then the principal prefers to take a risk in $t=1$ by inducing the agent's randomization. By doing so the principal's cost to separate the types of agent A in $t=1$ becomes lower. If, however, the agent is not highly likely to be efficient, then the principal is better off by avoiding the semi-separating strategy.

References

[1] Aghion, P. and Tirole, J. (1997), "Formal and Real Authority in Organizations," *Journal of Political Economy*, 105, 1 – 29.

[2] Baron, D. and Besanko, D. (1992), "Information, Control, and Organizational Structure," *Journal of Economics and Management Strategy*, 1, 237 – 275.

[3] Bartolome, F. (1989), "Nobody Trusts the Boss Completely – Now What?," *Harvard Business Review*, March-April, 5 – 11.

[4] Bester, H. and Strausz, R. (2001), "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case," *Econometrica*, 69, 1077 – 1098.

[5] Bolton, P. and Dewatripont, M. (1994), "The Firm as a Communication Network," *Quarterly Journal of Economics*, 109, 809 – 839.

[6] Calvo, G. and Wellisz, S. (1979), "Hierarchy, Ability, and Income Distribution," *Journal of Political Economy*, 87, 991 – 1010.

[7] Dessein W. (2002), "Authority and Communication in Organizations," *Review of Economic Studies*, 69, 811 – 838.

[8] Friebel, G. and Raith, M. (2004), "Abuse of Authority and Hierarchical Communication," Rand Journal of Economics, 35, 224 – 244.

[9] Harris, M. and Raviv, A. (2002), "Organization Design," *Management Science*, 48, 852 – 865.

[10] Hart, O. and Moore, J. (2005), "On the Design of Hierarchies: Coordination versus Specialization," *Journal of Political Economy*, 113, 675 – 702.

[11] Laffont, J.-J. and Martimort, D. (1998), "Collusion and Delegation," *Rand Journal of Economics*, 29, 280 – 305.

[12] Laffont, J.-J. and Tirole, J. (1988), "The Dynamics of Incentive Contracts," *Econometrica*, 56, 1153 – 1175.

[13] Laffont, J.-J. and Tirole, J. (1993), *A Theory of Incentives in Procurement and Regulation*, MIT Press.

[14] Lazear, E. and Rosen, S. (1981), "Rank–Order Tournaments as Optimum Labor Contracts," *The Journal of Political Economy*, 89, 841–864

[15] Maskin, E. and Tirole, J. (1999), "Unforeseen Contingencies and Incomplete Contracts," *Review of Economic Studies*, 66, 83 – 114.

[16] McAfee, P. and McMillan, J. (1995), "Organizational Diseconomies of Scale," *Journal of Economics and Management Strategy*, 4, 399 – 426.

[17] Melumad, M., Mookherjee, D. and Reichelstein, S. (1992), "A Theory of Responsibility Centers," *Journal of Accounting and Economics*, 15, 445 – 484.

[18] Melumad, M., Mookherjee, D. and Reichelstein, S. (1995), "Hierarchical Decentralization of Incentive Contracts," Rand *Journal Economics*, 26, 654 – 672.

[19] Mintzberg, H. (1979), *The Structuring of Organizations*, Prentice-Hall.

[20] Moore, J. and Repullo, R. (1988), "Subgame Perfect Implementation," *Econometrica*, 56, 1191 – 1220.

[21] Qian, Y. (1994), "Incentives and Loss of Control in an Optimal Hierarchy," *Review of Economic Studies*, 61, 527 – 544.

[22] Radner, R. (1993), "The Organization of Decentralized Information Processing," *Econometrica*, 61, 1109 – 1146.

[23] Rajan, R. and Zingales, L. (2001), "The Firm as a Dedicated Hierarchy: A Theory of The Origins and Growth of Firms," *Quarterly Journal of Economics*, 116, 805 – 851.

[24] Rosen, S. (1982), "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics*, 13, 311 – 323.

[25] Williamson, O. (1967), "Hierarchical Control and Optimum Firm Size," *Journal of Political Economy*, 75, 123 – 138.