View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

Thünen-Series of Applied Economic Theory

Thünen-Reihe Angewandter Volkswirtschaftstheorie

Working Paper No. 70

Quality Monitoring, Collusion and Sub-contracting

by

Michael Kuhn and Qianhui Yan

Universität Rostock

Wirtschafts- und Sozialwissenschaftliche Fakultät
Institut für Volkswirtschaftslehre
2006

Quality Monitoring, Collusion and Sub-contracting

Michael Kuhn*, Qianhui Yan†

December 19, 2006

Abstract

We discuss the social welfare improvement under centralized and decentralized hierarchies and focus on supervisor's ability to monitor quality. Although the possibility of collusion against the principal is eliminated under decentralized hierarchy, the decentralization is dominating only if supervisory accuracy is large enough in the case of public information. Private information about the accuracy hurts the principal under both hierarchies. The optimal effort in hierarchy A is pooling one. The dominance of decentralization over centralization depend combination of accuracies of both the low and the high type supervisor.

1 Introduction

Health care organizations rely increasingly on performance monitoring as a means of guaranteeing quality care. Monitoring is usually carried out by an intermediary expert (supervisor). This is for instance the case in the secondary care sector within the English National Health Service (NHS), where the Healthcare Commission is responsible for auditing the performance of hospitals (Healthcare Commission 2004). While in many organizations the supervisor is only engaged in monitoring but not in the commissioning of care, there are institutional settings where the intermediary assumes a considerably greater role in being responsible for directly contracting with providers. For instance, Primary Care Trusts within the English NHS have been assigned the authority to design contracts with practices to induce quality enhancement and cost containment for the delivery of care (Department of Health 2000, 2001).

This paper examines, within a model of hierarchical agency, the conditions under which subcontracting by the intermediary improves social welfare. We compare two structures, as illustrated in Figures 1a and 1b.

*corresponding author: University of Rostock, Department of Economics and Social Sciences and MPIDR; Konrad-Zuse-Str. 1, 18057 Rostock, Germany; Tel +49 381 2081 105; e-mail: kuhn@demogr.mpg.de.

†DERS, University of York, UK.

[Insert Figures 1a and 1b]

Under hierarchy A (figure 1a), the principal writes separate contracts with the agent and the supervisor, where the latter collects information about quality. The supervisor receives a signal of the provider's quality (σ) and makes a report (r) of his observation. The transfers to the supervisor (t^S) and agent (t^A) are then paid out as a function of the report (r). Because his monitoring of quality is not perfect, both the supervisor and the agent may collude against the principal by hiding quality information.

Under hierarchy B (figure 1b), the principal delegates to the supervisor the provision of incentives to the agent for quality enhancement and cost reduction. She only designs a contract with the supervisor, specifying t^S as a function of r, whereby the supervisor then sub-contracts the agent.

Both hierarchies may involve private information not only about quality but also about supervisory accuracy. By 'accuracy' we mean the probability with which quality is observed. Imperfection of monitoring quality enables the supervisor to collude with the agent. Specifically, a low-quality-low-cost type may bribe the supervisor to be concealed in order to receive an information rent from the principal. For hierarchy A, a collusion-proof principle has been proven for the case that supervisory accuracy is common knowledge (Tirole 1986). Tirole considers a principal-supervisor-agent hierarchy where the supervisor is defined as a bridge linking the principal and the agent reporting verifiable (hard) information about the agent. With a probability, the supervisor can observe the agent's type. Under the three-tier organization, the supervisor has a dimension of discretion by which he can pretend he has observed nothing and conceal the verifiable information. This discretion opens a door of collusion between the supervisor and the agent.

In response to collusion, the principal designs an allocation mechanism that is characterized by the collusion-proof principle. The principle states that, under certain conditions, there exist mechanisms for which the supervisor and the agent have no incentive to hide information from the principal. The transfer payment to the supervisor must not be less than the stake of collusion discounted by transaction cost of collusion.

We will argue that, provided supervisory accuracy is common knowledge and information is hard, hierarchy B rules out the possibility of collusion between the supervisor and the agent. This is because delegation introduces a conflict between supervisor and agent. Specifically, it is optimal for the supervisor to design a contract so as to induce the agent to truthfully reveal quality information. But in such a contract the low-quality-low cost type receives an information rent. By construction, this rent cannot be transferred back to the supervisor as a bribe for collusion. Second, the supervisor lacks the discretion to hide information, i.e. he cannot report not to know the agent's type. It is common knowledge that the supervisor has designed a contract for the agent that elicits a truthful quality report. But then as the principal knows that the supervisor is informed (either through direct observation or through the agent's report), she can refuse a report of zero information. The supervisor must then reveal either high or low quality. Here, the assumption that information is verifiable rules out the possibility that the supervisor misreports signals.

While a greater accuracy in the supervisor's monitoring technology (i.e. a higher probability of detecting the agent's type) increases the social surplus under either organizational structure, we find that this is more pronounced under the decentralized organization B. Indeed, this structure dominates in welfare terms if and only if the accuracy is sufficiently high. When choosing an organizational form the principal needs to trade off two hierarchy specific inefficiencies. Under hierarchy A a rent has to be paid to the supervisor in order to avoid collusion, whereas under hierarchy B the subcontract between the supervisor and agent stipulates an inefficient level of cost-reducing effort. This latter inefficiency is well-known (Laffont and Tirole 1993: section B1.7) and arises from the supervisor maximising income rather than a broader welfare measure. As a higher probability of detecting the agent's true type increases the expected rent to the supervisor under hierarchy A and, at the same time, decreases the expected efficiency loss due to the distortion in hierarchy B, this implies dominance of hierarchy B in circumstances of high accuracy.

The information structure becomes more complex if the supervisor is privately informed about the accuracy of supervision. As the transfer to supervisor depends on the reported accuracy within both hierarchies, the supervisor has an incentive to misreport the accuracy of his monitoring technology. More specifically, the supervisor has an incentive to under-report accuracy in order to receive an information rent. Comparing again the hierarchies with regard to their welfare properties, it turns out that the decentralized hierarchy B dominates hierarchy A if and only if supervisory accuracy is sufficiently high *and* the spread between high and low accuracy is not too large. This additional requirement (as opposed to the case of publicly known supervisory ability) implies that the informational problem with respect to the supervisor's private information weighs in more heavily in the case of delegation. This is because in a decentralized hierarchy the principal has no scope to influence the agent's effort in order to extract the supervisor's rent.

A number of articles have studied the incentives within hierarchical organization. Tirole (1992) presents a comprehensive overview on collusion with hard information. Kofman and Lawarree (1993) propose that the problem of collusion between a manager and an internal auditor can be mitigated by randomly subjecting the organization to external audits. While external auditors operate with lower accuracy than internal ones, their independence insulates them against attempts of collusion. Another strand of literature addresses the information cost in different designs of organizations. Laffont and Martimort (1998) and Baron and Besanko (1992, 1994) focus on communication within an organization. With limits on communication, a decentralized structure dominates a centralized one. Analyses on collusion issues provide a better understanding of the determinants that should govern the choice between a central or decentral organization. Faure-Grimaud et al (2003) examine a model with collusion and communication in the presence of soft information. They show that delegation can be viewed as an implementation of the optimal collusion-proof contract within a centralised hierarchy. Thus they establish equivalence between organisational forms, with the decentralised organisation being weakly superior in that it allows to implement the efficient solution as a unique equilibrium.

All of these articles assume that supervisory accuracy is common knowledge within an organization and the supervisor has no private information about his ability. This is our main point of departure from the received literature, where we assume that in addition to quality supervisory accuracy is private information. With respect to supervisory accuracy, our study is related to Faure-Grimaud et al (1999). They assume that monitoring is costly. Taking the supervisory cost into account, they analyze the deadweight loss of delegated auditing in a three-tier hierarchy. In their model, supervisory accuracy is controlled by the principal, although they claim that this assumption can be relaxed, turning accuracy into a moral hazard variable (see their footnote 6).

The remainder of the paper is organized as follow. Section 2 presents the basic models. In section 3, we investigate the hierarchies when supervisory ability is public information. Section 4 provides the analysis for the case in which the supervisor's accuracy is private knowledge. Conclusions are presented in section 5.

2 The model

We follow Laffont and Tirole (1991) in modelling a three-tier hierarchy with a principal, supervisor or intermediary, and an agent, the latter caries out production and engaging in a cost-reducing effort.

The agent

The agent is a productive unit, say a physician practice, who provides services with quality q. To simplify analysis, we consider the case of one unit service. The cost of one unit of service

$$C = q - e, \tag{1}$$

increases in quality and decreases in cost reduction effort, e. We assume that the cost is observable to the principal (e.g. on the basis of accounting data). The agent bears a disutility of exerting cost reduction effort . The disutility is defined in the sense of non-monetary term and is represented by the term $\psi(e)$, where it is assumed that $\psi\prime(e) > 0$ and $\psi\prime\prime(e) > 0$.

The agent receives a financial transfer t^A and his utility can then be written as

$$U = t^A - \psi(e). \tag{2}$$

Quality is unobservable to the principal. As it is the aim of the paper to examine social welfare improvement in different hierarchies, we assume quality to be an adverse selection variable rather than a moral hazard one. We assume the service to be either of high quality $\overline{q}$,or of low quality $\underline{q}$ $\left(\Delta q = \overline{q} - \underline{q} > 0\right)$. The differences in quality levels may, for instance, relate to differences in a physician's (unobservable) skills. It is assumed that distribution of quality level is common knowledge among the players but its specifications are the agent's private information. With probability ξ, services are provided at quality $\overline{q}$, and with probability $(1-\xi)$, at quality $\underline{q}$.

The supervisor or subprincipal

The supervisor or subprincipal receives a signal σ of the agent's quality. With probability μ the supervisor/subcontractor observes the true quality ($\sigma = q$), and with probability $(1-\mu)$ he observes nothing $\sigma = \emptyset$. Following Tirole (1986, 1992) and Laffont and Tirole (1991) we assume that the signal is hard (verifiable) information. Thus if $\sigma = \emptyset$ this is the only thing the supervisor can report: $r = \emptyset$. However, if $\sigma = q$, the supervisor can opt to report the true quality $r = q$ or claim to have observed nothing, $r = \emptyset$.

The supervisor and subcontractor play different roles. The supervisor in hierarchy A just collects and reports information about quality. All of the contracting remains in the principal's hand. The supervisor is reimbursed for his efforts with a salary t^S. He may also receive a bribe from the agent b. Assuming he is only concerned about his private wealth, we can write the supervisor's utility as

$$V = t^S + b. \tag{3}$$

Finally, we assume that the supervisor is risk neutral but faces a liquidity constraint requiring $t^S \geq 0$.[1]

In hierarchy B the principal delegates to an intermediary the authority to enter into a contract with the agent. Thus acting as subprincipal, the intermediary receives a transfer t^S but has to bear the cost of services and the transfer to the agent so that

$$V = t^S - \left(t^A + C\right) = t^S - \left[(q-e) + t^A\right]. \tag{4}$$

Reflecting their broader remit, intermediaries will generally differ from supervisors. While the latter are in many instances employees or professionals operating at a scale that exposes them to liquidity constraints, intermediaries may engage in a variety of activities, enabling them to cross-subsidise. We therefore assume that in contrast to the supervisor the intermediary does not face a liquidity constraint. This certainly appears to reflect the context of UK primary care. Whereas supervisors are usually individuals who are employed by the Department of Health or another Government authority, the Primary Care Trusts acting as commissioners (subprincipals) of care usually face a large number of practices (agents). To this extent they are able to shift funds between practices, implying that they are not exposed to the same liquidity constraints as the supervisor.

The principal

The principal is concerned for social welfare, defined as net benefit from the health service plus the agents' and supervisors' utility. Here we consider the

[1] This would imply that the bribe b is in kind. Alternatively, one could specify $t^S + b \geq 0$. As it turns out this distinction is immaterial for our analysis.

case that publicly funded services, and patients do not need to pay directly for their consumption. Therefore, the shadow price λ of public funds is included in the social welfare function. Welfare can thus be written as

$$W = B(q) + U + V - (1+\lambda)\left(C + t^S + t^A\right)$$

under hierarchy A and

$$W = B(q) + U + V - (1+\lambda)\, t^S$$

under hierarchy B, where $B(q)$ denotes the patient's benefit from consumption of one unit of service with quality q. Substituting for t^S and t^A from (3) and (2) in case of hierarchy A and from (4) in case of hierarchy B, we obtain for both A and B

$$W = B(q) - (1+\lambda)\left[(q-e) + \psi(e)\right] - \lambda U - \lambda V$$

Under both hierarchies the principal faces a problem regarding the agent's private information about q. An additional complication arises (in section 4) when the principal faces a supervisor who is privately informed about the accuracy of monitoring μ. Under hierarchy A, the principal offers simultaneously a contract to the agent and a contract to the supervisor. Both contracts are implemented on the basis of the quality report r. Under these arrangements, the informational problems are decerebrated due to the scope for collusion between the supervisor and the agent which arises under imperfect quality monitoring.

Under hierarchy B, there is no direct communication between the principal and the agent. The principal only offers a contract to the intermediary who then sub-contracts the agent. Contrary to hierarchy A, here the principal has to move first and offers the contract to the intermediary, based on his report r. And then, for any given contract with the principal, the intermediary offers a (formal) contract to the agent, which is specified by the revelation principle. As we will argue in greater detail below the principal can design a set of transfers to the intermediary which rules out collusion. Specifically, this requires transfers to be contingent on the reported quality but not on the value of the signal σ. In this case, the intermediary becomes residual claimant of his efforts to reveal the agent's type which eliminates the scope for collusion. While subcontracting may fare better on these grounds, it also involves an inefficiency in effort due to the supervisor not internalising the agent's benefit.

Timing (Hierarchy A)

1. The agent learns privately quality q, and (in section 4) the supervisor learns the accuracy μ.
2. The principal offers a package of contracts to the supervisor $\left\{t^S, \mu, r\right\}$ and to the agent $\left\{t^A, C, r\right\}$, respectively.
3. Both determine whether or not to accept the contract. If the supervisor refuses, the game goes back to a two-tier hierarchy; if the agent refuses, all get reservation utility normalized to zero and the game ends. Supposing both parties participate the game continues as follows.

4. The supervisor self-selects μ, receives the signal σ and if nonbenevolent offers the agent a side contract that specifies a bribe b in exchange for hiding information when this may afford the agent a rent.

5. The agent exerts cost reduction effort, the supervisor reports r.

6. The transfers are implemented.

Timing (Hierarchy B)

1. The agent learn quality q, and (in section 4) the supervisor learns μ.

2. The principal offers a main contract $\left\{t^S, \mu, r\right\}$ to the intermediary. If he refuses the offer, the game ends and both the intermediary and the agent get reservation utility normalized at zero.

3. If he accepts the main contract, the intermediary reports μ, receives the signal σ and offers the agent a sub-contract $\left\{t^A, C, \sigma\right\}$.

4. If the agent refuses the sub-contract, the game ends and both the intermediary and the agent get reservation utility normalized to zero.

5. If the agent accepts the sub-contract, he exerts effort, and (where $\sigma = 0$) reports quality to the intermediary; the intermediary reports quality to the principal.

6. The main contract and the sub-contract are implemented.

3 Hierarchies with publicly known supervisory accuracy

To understand the sources of the information rent and the implementation of the contract, we first summarize the solution of a two-tier principal-agent model with asymmetry of information about quality. This problem is solved by the revelation principle (e.g., Baron and Myerson,1982).

An agent who provides low quality services $\underline{q}$ could lie about quality and hence needs to receive an information rent

$$\Psi\left(\overline{e}\right) = \psi\left(\overline{e}\right) - \psi\left(\overline{e} - \Delta q\right), \tag{5}$$

whilst the high quality type, $\overline{q}$, receives no rent. To extract the information rent, the principal imposes a second best effort $\overline{e}^{SB}$ on the high type, which is determined by

$$\psi\prime\left(\overline{e}\right) = 1 - \frac{\left(1-\xi\right)\lambda}{\xi\left(1+\lambda\right)}\Psi\prime\left(\overline{e}\right). \tag{6}$$

For the low type, the effort is imposed at the first best, i.e.

$$\psi\prime\left(\underline{e}\right) = 1. \tag{7}$$

Solving the equations above yields the optimal efforts levels

$$\underline{e} = e^* \quad \text{and} \quad \overline{e} = \overline{e}^{SB} < e^*.$$

In the following, we are going to examine how the principal can employ a supervisor in order to ameliorate the informational problem. As a frame of reference, note that after the audit there are four possible states of nature:

- with probability $\xi\mu$, the agent provides high quality $q = \overline{q}$ and the supervisor observes it, $\sigma = \overline{q}$,
- with probability $\xi(1-\mu)$, $q = \overline{q}$ but this remains unobserved, $\sigma = \emptyset$,
- with probability $(1-\xi)\mu$, the agent provides low quality $q = \underline{q}$ and the supervisor observes it $\sigma = \underline{q}$,
- with probability $(1-\xi)(1-\mu)$, $q = \underline{q}$ but no observation $\sigma = \emptyset$.

To describe the four states, we introduce the subscripts $i,\ j\ (i,\ j = 1,\ 2)$, where $i = 1$ stands for high quality, $i = 2$ for low quality, $j = 1$ for 'observed' and $j = 2$ for 'unobserved'. Thus, we define the probabilities of each state as:

$$\begin{aligned} p_{11} &\equiv \xi\mu & p_{12} &\equiv \xi(1-\mu) \\ p_{21} &\equiv (1-\xi)\mu & p_{22} &\equiv (1-\xi)(1-\mu). \end{aligned} \tag{8}$$

For ease of reference, we will frequently index variables according to the state (i, j).. In particular, we will use q_{ij}, where $q_{11} = q_{12} = \overline{q}$ and $q_{21} = q_{22} = \underline{q}$, $B_{ij} = B(q_{ij})$. We now proceed to analyses to the forms of hierarchy in separate, beginning with hierarchy A, where the supervisor only monitors the agent whereas the principal retains the control of the agent's contract.

3.1 Hierarchy A

Within this organizational structure, the supervisor will bridge the information gap but also has the scope to collude with the agent. Collusion arises under the following circumstances. In the presence of hard information the supervisor cannot falsify reports. Thus he can at best claim to have observed nothing ($r = \emptyset$) when really he has observed true quality $\sigma = q$. Further, collusion requires that there is some rent to be shared between the colluding parties. This rent can only be the information rent the agent would receive in a situation when nothing is observed. As we see from (5), this rent is obtained by a low quality type with an incentive to mimic high quality. Thus, collusion can occur only in state $(2, 1)$, where the supervisor has observed an agent producing low quality, $\sigma = \underline{q}$. As we also see from (5), the rent then depends on the effort that would be allocated to the high type ($i = 1$) in the absence of an observation ($j = 2$). Thus the stakes of collusion, equal to the information rent, are given by

$$\Psi(e_{12}) = \psi(e_{12}) - \psi(e_{12} - \Delta q). \tag{9}$$

We follow Tirole (1986, 1992) in modelling collusion with hard information, where the following points merit attention. First, information between the supervisor and the agent is perfect. Concretely, the agent knows whether or not the supervisor has observed quality. Second, the bargaining power within the coalition lies entirely with the supervisor, who can therefore extract the entire rent. Third, due to transaction costs only a share $k, k \in (0,1)$, of the rent can be consumed by the supervisor. That is, the supervisor only receives $k\Psi(e_{12})$ of the full bribe $\Psi(e_{12})$. Larger transaction costs correspond to smaller values of k. Fourth, we follow the literature by assuming that the bribe is of a monetary nature.

The prevention of collusion requires the satisfaction of a coalition incentive constraint (**CIC**).

$$t_{21}^S \geq t_{22}^S + k\Psi(e_{12}) \tag{CIC}$$

In state$(2,1)$, where collusion may occur the principal has to grant a transfer to the supervisor, which is not less than what he would receive were he to collude with the agent. Recalling that under collusion the supervisor reports $r = \emptyset$ and the agent chooses the low quality type's contract (including the rent), ex-post this corresponds to state $(2,2)$. Hence, under collusion the supervisor receives the transfer t_{22}^S from the principal and the bribe (discounted for transaction costs) $k\Psi(e_{12})$.

In addition to **CIC**, the principal faces the following constraints:

1. An individual rationality constraints for the agent (**AIR**) and a limited liability constraint for the supervisor (**SLL**)

   To guarantee participation, the agent's ex post utility at each state must not be less than his reservation utility, standardized to zero. That is

   $$U_{ij} = t_{ij}^A - \psi(e_{ij}) \geq 0 \quad (i,j = 1,2). \tag{AIR}$$

   Likewise, the supervisor's ex post utility at each state must not be less than his reservation utility, again standardized to zero. That is

   $$V_{ij} = t_{ij}^S \geq 0 \quad (i,j = 1,2). \tag{SLL}$$

2. The agent's incentive compatible constraint, **AIC**

   In state $(2,2)$, where the supervisor cannot observe low quality, the agent needs to be induced to self-select the own contract. [2] Thus,

   $$U_{22} \geq U_{12} + \Psi(e_{12}) \tag{AIC}$$

   with $\Psi(e_{12})$ as given by (9).

[2] Similarly, the high type agent ($i = 1$) is required telling truth. But, as is common, this constraint is not binding and can be ignored.

The principal's problem is to maximize social welfare with respect to the effort levels e_{ij} and supervisory accuracy, μ, subject to the constraints described above. That is

$$\max_{e_{ij}} EW$$
$$s.t. \;\; AIC, \;\; CIC, \;\; AIR, SLL,$$

where EW is equal to

$$EW = \sum_{i=1}^{2}\sum_{j=1}^{2} p_{ij}\left\{B_{ij} - (1+\lambda)\left[(q_{ij} - e_{ij}) + \psi(e_{ij})\right] - \lambda U_{ij} - \lambda V_{ij}\right\}$$

Because EW decreases in U_{ij}, V_{ij} $(i, j = 1, 2)$, the principal can arrange the transfers such that

$$U_{ij} = \begin{cases} \Psi(e_{12}) & if \;\; i = j = 2 \\ 0 & \text{Otherwise} \end{cases} \tag{10}$$

$$V_{ij} = \begin{cases} k\Psi(e_{12}) & if \;\; i = 2, j = 1 \\ 0 & \text{Otherwise} \end{cases} \tag{11}$$

Substituting into EW yields the equivalent maximization problem

$$\max_{e_{ij}} \left\{ \begin{array}{c} \sum_{i=1}^{2}\sum_{j=1}^{2} p_{ij}\left\{B_{ij} - (1+\lambda)\left[(q_{ij} - e_{ij}) + \psi(e_{ij})\right]\right\} \\ -p_{22}\lambda\Psi(e_{12}) - p_{21}\lambda k\Psi(e_{12}) \end{array} \right\}$$

Noting that $e_{12} = \widehat{e}_{12}$ is given by

$$\psi\prime(\widehat{e}_{12}) = 1 - \frac{\lambda}{1+\lambda}\frac{1-\xi}{\xi}\left(1 + \frac{\mu}{1-\mu}\right)\Psi\prime(\widehat{e}_{12}), \tag{12}$$

we can summarize the solution to this problem as follows.

Lemma 1 *Given that supervisory accuracy μ is public knowledge, the optimal contracts in hierarchy A can be characterized as follows.*

1. The optimal efforts are given by

$$\widehat{e}_{11} = \widehat{e}_{21} = \widehat{e}_{22} = e^*,$$
$$\widehat{e}_{12} < \overline{e}^{SB} < e^*.$$

2. The transfers to the agent and supervisor are given by

$$t_{ij}^{A} = \begin{cases} \psi(e^*) + \Psi(\widehat{e}_{12}) & \textit{if } i = j = 2 \\ \psi(\widehat{e}_{ij}) & \textit{otherwise} \end{cases}$$

and by

$$t_{ij}^{S} = \begin{cases} k\Psi(\widehat{e}_{12}) & \textit{if } i = 2, \; j = 1 \\ 0 & \textit{otherwise} \end{cases}.$$

3. The introduction of a supervisor improves social welfare.

4. Effort $\widehat{e}_{12}$ falls in the supervisory accuracy μ.

Proof. See Appendix. ■

It is well-known that the threat of collusion within a hierarchy with a supervior necessitates a further downward distortion of effort $\widehat{e}_{12}$ below the level that would be realised in a two-tier hierarchy. Here, the principal not only seeks to extract information rent from the low quality type [the second term on the RHS of (12)], but she also seeks to reduce the rent she needs to pay to the supervisor in order to prevent collusion [the third term on the RHS of (12)] . This additional distortion of effort in the presence of supervision is less costly in so far as the relevant state $(1, 2)$ becomes less likely.

Note that despite the threat of collusion it is always beneficial from a social welfare perspective to hire a supervisor. The reason is that due to the transaction costs of collusion, the introduction of a supervisor allows the principal to reduce the expected rental payments by an amount $(1-\xi)\,\mu\,(1-k)\,\Psi\,(e_{12})$ for any given effort e_{12}. While this holds for $\overline{e}^{SB}$, optimal choice of $\widehat{e}_{12} < \overline{e}^{SB}$ implies that welfare for the supervisor must be even greater. The savings on rent vanish in the absence of transaction costs ($k = 1$). In this case the principal will have to pay the rent for $i = 2$ irrespective of whether the agent's type is observed (in case of which the rent goes to the supervisor) or not (in case of which the rent goes to the agent). However, even in this case a second benefit justifies the introduction of supervision. The reason is that a positive probability of discovering the agent's quality (type) allows the principal to reduce the expected inefficiency associated with the state $(1, 2)$. Hence, even in the case of perfect collusion ($k = 1$) it pays the principal to employ the supervisor if only for the improvements in efficiency. However, as we note from the last part of the

Lemma, while welcome on welfare grounds an increase in supervisory accuracy also leads to a greater distortion in incentives for the agent. This is because precisely of the improvements in expected efficiency under better supervision. The principal trades off a part of this efficiency increase against a further reduction in rent.

3.2 Hierarchy B

This section considers a hierarchy in which the principal offers the main contract to an intermediary who in turn sub-contracts with the agent. The agent is the only productive unit. Inducing quality enhancement and optimal cost reduction effort becomes the intermediary's responsibility. Acting as subprincipal the intermediary plays two roles in the organization. One one hand, he has to collect information about quality. On the other hand, he has to offer the agent a contract that induces truthful reports of quality and optimal cost reduction effort.

We shall now argue under which circumstances this organizational structure eliminates collusion between the intermediary and the agent. We begin by noting that under our assumption of hard information, the intermediary will always report the true type to the principal. It is common knowledge that the intermediary has either observed the agent's type or otherwise has learned it by offering a self-selecting contract to the agent. The principal therefore knows that the intermediary is informed about the agent's type. She can then require that the intermediary to reveal true quality by refusing to reimburse the intermediary for a report $r = 0$. As before, the presence of hard information eliminates the possibility that the intermediary misreports quality information, and he will therefore reveal true quality, $r = q$.

However, while always learning the agent's type from the supervisor, the principal cannot tell from this whether the intermediary has received this information on the basis of an observation $\sigma = q$ or whether for $\sigma = \emptyset$ the intermediary had to elicit this information by paying a rent. It is easy to see that trying to elicit the information on the value of σ exposes the principal to the risk of collusion. For instance, in state $(2, 1)$, i.e. if $\sigma = q = \underline{q}$, the intermediary and agent have an incentive to collude against the principal and claim $\sigma = \emptyset$ as long as by so doing they induce the principal to increase the transfer to the intermediary in order to compensate him for the rental payment $\Psi(e_{12})$. Thus, if transfers are made contingent on the reported σ, the contract with the supervisor will have to be made collusion proof in a way similar to hierarchy A.

In the following we propose a different mechanism under which sub-contracting may eliminate the scope for collusion. If the intermediary is not liquidity contrained they can be exposed to losses in certain states. The only constraint that is relevant is an ex-ante participation constraint requiring that the intermediary's expected benefit from partaking in the hierarchy is non-negative. The principal can then specify a couple of transfers t_i^S that are only linked to the report $r = q_i$ but pool across the realisation of σ.The scope for collusion is then eliminated. In fact, as we will see shortly, as under hard information the intermediary has to make a truthful report, the transfer t^S can even be pooled across types. This arrangement is equal to one in which the principal provides a type adjusted budget to the intermediary and leaves him to sort out the sub-contracting. By making the supervisor residual claimant, the principal over-comes the moral hazard problem of collusion.

We solve the problem backwards. First, we solve for the optimal contract between the intermediary and the agent at stage 2. The solution is then used to determine the contract between the principal and the intermediary at stage 1. In stage 2, for any main transfer t^S, the supervisor has to offer a contract that induces truthful reports of quality and optimal effort. In stage 1, the main contract will specify a transfer to the supervisor, conditional on the quality report.

Stage 2

For any contract with the principal $\{t_i^S, r\}$, the intermediary will design a contract $\{t_{ij}^A, e_{ij}\}$to maximize his expected utility

$$EV = \sum_{i=1}^{2} \sum_{j=1}^{2} p_{ij} \left\{ t_i^S - \left[(q_{ij} - e_{ij}) + \psi(e_{ij}) \right] - U_{ij} \right\}$$

subject to (AIC) and (AIR). Solving the problem in the usual way yields that

1. For any t^S, the optimal effort is characterized by

$$\check{e}_{12} < \overline{e}^{SB} < \ \widehat{e}_{11} = \widehat{e}_{21} = \widehat{e}_{22} = e^*,$$

where $\check{e}_{12}$ is determined by

$$\psi\prime\left(\check{e}_{12}\right) = 1 - \frac{1-\xi}{\xi}\Psi\prime\left(\check{e}_{12}\right), \tag{13}$$

and where $\overline{e}^{SB}$ is determined by (6).

2. Transfers are equal to

$$t_{11}^A = t_{21}^A = \psi\left(e^*\right), \quad t_{12}^A = \psi\left(\check{e}_{12}\right), \quad t_{22}^A = \psi\left(e^*\right) + \Psi\left(\check{e}_{12}\right).$$

Stage 1

Let the intermediary's utility at each state be defined as

$$\check{V}_{11} = t_1^S - a_{11}$$
$$\check{V}_{12} = t_1^S - a_{12}$$
$$\check{V}_{21} = t_2^S - a_{21}$$
$$\check{V}_{22} = t_2^S - a_{22} - \Psi\left(\check{e}_{12}\right),$$

where

$$\begin{aligned} a_{11} &\equiv \left[\left(\overline{q} - e^*\right) + \psi\left(e^*\right)\right] \\ a_{12} &\equiv \left[\left(\overline{q} - \ \check{e}_{12}\right) + \psi\left(\ \check{e}_{12}\right)\right] \\ a_{21} = a_{22} &\equiv \left[\left(\underline{q} - \ e^*\right) + \psi\left(e^*\right)\right]. \end{aligned}$$

As there is no scope for collusion and as no misreporting can take place, the only constraint that needs to be satisfied is the intermediary's ex-ante participation constraint, where 'ex-ante' refers to the revelation of the signal σ :

$$\sum_{i=1}^{2}\sum_{j=1}^{2} p_{ij}\check{V}_{ij} \geq 0$$

As rents are costly for the principal, she chooses t_1^S and t_2^S in order to make the constraint bind, implying

$$\xi t_1^S + (1-\xi)\, t_2^S = \sum_{i=1}^{2}\sum_{j=1}^{2} p_{ij}a_{ij}.$$

With no further constraints a risk-neutral principal has a degree of freedom in choosing the t_i^S's. She may therefore also choose a unified budget

$$\begin{aligned} t^S &= \sum_{i=1}^{2}\sum_{j=1}^{2} p_{ij}a_{ij}. \\ &= \xi\left[a_{11} + (1-\mu)\, R_e\right] + (1-\xi)\left[a_{21} + (1-\mu)\,\Psi\left(\check{e}_{12}\right)\right], \end{aligned} \tag{14}$$

with

$$R_e :\equiv [e^* - \psi(e^*)] - [\check{e}_{12} - \psi(\check{e}_{12})] > 0$$

being the additional cost due to distortion of agent's effort. Note that ex-ante this budget does not leave a rent to the supervisor, whereas ex-post rents are possible in the favourable states $(1,1)$ and $(2,1)$.

We are now able to compare the two hierarchies for the case of a publicly known accuracy μ.

Proposition 2 *The contracts for the respective hierarchies A and B compare as follows.*

1. Optimal effort is characterized by

$$max\{\check{e}_{12}, \widehat{e}_{12}\} < \overline{e}^{SB}$$

and

$$\check{e}_{12} < \widehat{e}_{12} \Leftrightarrow \lambda k < \frac{1-\mu}{\mu}.$$

2. There exists a threshold of supervisory accuracy, $\mu_0 \in (0,1)$, *such that hierarchy B dominates hierarchy A if and only if* $\mu > \mu_0$.

Proof. See Appendix. ■

Although the contract under hierarchy B has eliminated collusion between the supervisor and the agent and therefore allows a reduction of the expected rent to the second-best level, one cannot conclude that this structure is always better than hierarchy A. This is because of a distortion in effort in the sub-contract, where the intermediary is overzealous to extract rent. As we see this private rent-shifting leads to an effort $\check{e}_{12}$ that is below the one set by the principal even under the threat of collusion, $\widehat{e}_{12}$ if the transaction costs of collusion are high (low k) if the shadow cost of funds, λ, is low, or if the accuracy of supervision, μ, is low.

To understand intuitively under which conditions hierarchy B dominates hierarchy A, consider the difference in social welfare

$$\Delta SW = \alpha + \beta + \gamma$$

where

$$\alpha = \xi(1+\lambda)(1-\mu)\left[(\check{e}_{12} - \psi(\check{e}_{12})) - (\widehat{e}_{12} - \psi(\widehat{e}_{12}))\right]$$

$$\beta = -(1-\xi)\lambda(1-\mu)\left[\Psi(\check{e}_{12}) - \Psi(\widehat{e}_{12})\right]$$

$$\gamma = (1-\xi)\lambda k\mu\Psi(\widehat{e}_{12}).$$

Here, the expression in α gives the net reduction in total cost of production (monetary + effort) that is realised under hierarchy B as opposed to A. The term in β gives the net reduction in rent paid to the agent when the organizational form is switched from A to B. The third term $\gamma > 0$ corresponds to the collusion rent which is saved under hierarchy B. It follows that for a high accuracy of supervision ($\mu \to 1$), hierarchy B will dominate. The distortion in efficiency relates to a state $(1,2)$ that is very unlikely, and almost certain welfare gains then arise from the avoidance of collusion. Noting that the expected cost $\xi(1+\lambda)(1-\mu)(\overline{q}-e_{12}+\psi(e_{12}))+(1-\xi)\lambda(1-\mu)\Psi(e_{12})$ is minimised by $\overline{e}^{SB}$, one can see that $\alpha+\beta\geq 0$ if and only if $\check{e}_{12}\geq\widehat{e}_{12}$. This again is true

as long as μ is sufficiently large. In this case, the effort realised in hierarchy B is more efficient, making this hierarchy dominate throughout. The situation changes as μ falls. Whereas hierarchy B still benefits from not exhibiting collusion, a distortion in effort $\check{e}_{12}<\widehat{e}_{12}$ will render this hierarchy less efficient than hierarchy A. It is then easy to see for low levels of μ and certainly for $\mu\rightarrow 0$, where collusion becomes irrelevant, hierarchy A will dominate. Note that a similar argument applies for low levels of k. As collusive rents do not play a big role here, hierarchy A is preferable on the grounds of its greater efficiency.

Under both structures, the transfer to the supervisor is a function of the accuracy μ. If this is private information, the supervisor may therefore manipulate reports about the accuracy in order to obtain a favourable transfer. As we will see in the following, this will change the conditions of domination.

4 Hierarchies with privately known supervisory accuracy

In the following, we assume the supervisor's, respectively intermediaries, accuracy in monitoring μ to be their private information. Let μ be drawn from a bivariate distribution, where $\mu\in\left\{\underline{\mu},\overline{\mu}\right\}$, $\Delta\mu=\overline{\mu}-\underline{\mu}>0$ and $\pi=prob\left(\mu=\overline{\mu}\right)$.As is common, we assume the distribution to be common knowledge.

Contracts are implemented on the basis of the reported accuracy. Under hierarchy A, the principal conditions the cost-reducing effort $\widehat{e}_{12}$ as well as some of the transfers on the reported μ^r; under hierarchy B the transfer t^S is a function of μ^r. We should therefore expect that for certain realisations of μ the supervisor has an incentive to misreport the accuracy. In the following we will show that in both hierarchies the supervisor has an incentive to under-report the accuracy.

Supervisor's behavior under hierarchy A

The supervisor is prone to under-report his ability. Indeed, from lemma 1, we know the transfer to the supervisor and, from this we are able to calculate his expected utility for the allocation under full information as

$$EV\left(\mu,\mu^r\right)=(1-\xi)\,k\mu\Psi\left[\widehat{e}_{12}\left(\mu^r\right)\right],$$

where μ and μ^r denote the true and reported accuracy, respectively. The principal imposes effort $\widehat{e}_{12}$ based on reported accuracy. That is to say, $\widehat{e}_{12}=\widehat{e}_{12}\left(\mu^r\right)$, with $\frac{\partial\widehat{e}_{12}}{\partial\mu^r}<0$ as shown in lemma 1. It is now readily verified that

$$\frac{dEV}{d\mu^r}=(1-\xi)\,k\mu\Psi\left(\cdot\right)\frac{\partial\widehat{e}_{12}}{\partial\mu^r}<0$$

Here we use the facts that, from the assumption of $\psi\prime(\cdot) > 0$, we have $\Psi\prime(\cdot) > 0$. Consequently, $EV\left(\mu, \underline{\mu}\right) > EV\left(\mu, \overline{\mu}\right)$ for either μ. The supervisor has an incentive to under-report μ.

Intermediary's behavior under hierarchy B

Under hierarchy B the expected transfer to the supervisor directly depends on the principal's belief μ^r, where we obtain from (14)

$$EV\left(\mu, \mu^r\right) = t^S\left(\mu^r\right) - \sum_{i=1}^{2}\sum_{j=1}^{2} p_{ij} a_{ij} = \left(\mu - \mu^r\right)\left[\xi R_e + (1-\xi)\,\Psi\left(\check{e}_{12}\right)\right], \quad (15)$$

with

$$R_e \equiv \left[e^* - \psi\left(e^*\right)\right] - \left[\check{e}_{12} - \psi\left(\check{e}_{12}\right)\right] > 0.$$

Note that $\check{e}_{12}$, as determined by (13) is unrelated to μ and μ^r. From (15) it is thus obvious that $\frac{dEV}{d\mu^r} < 0$, implying that the intermediary, too, has an incentive to under-report accuracy.

Principal's strategy

We know establish for both hierarchies the optimal contract for the principal in the presence of this additional informational constraint before combining them in terms of expected welfare. Under hierarchy A, the principal has two regulatory instruments: the transfer payments and the choice of effort, which she can adjust in order to minimize the total information rent.

4.1 Hierarchy A

The principal's objective can now be written as $\pi\overline{EW}^A + (1-\pi)\,\underline{EW}^A$, where

$$\overline{EW}^A = \sum_{i=1}^{2}\sum_{j=1}^{2} \overline{p}_{ij}\left\{B\left(q_{ij}\right) - (1+\lambda)\left[\left(q_{ij} - \overline{e}_{ij}\right) + \psi\left(\overline{e}_{ij}\right)\right] - \lambda\overline{U}_{ij} - \lambda\overline{V}_{ij}\right\}$$

with $\overline{(\cdot)}$ denoting the variables corresponding to a high accuracy, $\overline{\mu}$, and where

$$\underline{EW}^A = \sum_{i=1}^{2}\sum_{j=1}^{2} \underline{p}_{ij}\left\{B\left(q_{ij}\right) - (1+\lambda)\left[\left(q_{ij} - \underline{e}_{ij}\right) + \psi\left(\underline{e}_{ij}\right)\right] - \lambda\underline{U}_{ij} - \lambda\underline{V}_{ij}\right\}$$

with $\underline{(\cdot)}$ denoting the variables corresponding to a low accuracy, $\underline{\mu}$.

In order to induce truthful reporting of the supervisor's type, the principal has to design the contract in a way that the supervisor's incentive compatibility constraint is satisfied. Generally, this can be expressed as

$$EV\left(\mu, \mu\right) \geq EV\left(\mu, \mu^r\right), \quad \mu, \mu^r \in \left\{\underline{\mu}, \overline{\mu}\right\}, \quad \text{(SIC)}$$

where $EV\left(\mu, \mu^r\right)$ it type μ's expected utility when μ^r is the reported type. More specifically, we can write

$$\begin{aligned} EV\left(\mu, \mu^r\right) &= \left[\begin{array}{c} \xi\mu t_{11}^S\left(\mu^r\right) + \xi\left(1-\mu\right) t_{12}^S\left(\mu^r\right) \\ +\left(1-\xi\right)\mu t_{21}^S\left(\mu^r\right) + \left(1-\xi\right)\left(1-\mu\right) t_{22}^S\left(\mu^r\right) \end{array}\right]. \\ &= y\left(\mu^r\right) + \mu z\left(\mu^r\right), \qquad (16) \end{aligned}$$

with

$$
\begin{aligned}
y(\mu^r) &:= \xi t_{12}^S(\mu^r) + (1-\xi)\, t_{22}^S(\mu^r) \\
z(\mu^r) &:= \xi\left[t_{11}^S(\mu^r) - t_{12}^S(\mu^r)\right] + (1-\xi)\left[t_{21}^S(\mu^r) - t_{22}^S(\mu^r)\right]. \qquad (17)
\end{aligned}
$$

The incentive compatibility constraint for the high type can then be written as

$$
\begin{aligned}
EV(\overline{\mu},\overline{\mu}) &\geq EV(\overline{\mu},\underline{\mu}) \\
&\Leftrightarrow \overline{\mu}\left[z(\overline{\mu}) - z(\underline{\mu})\right] \geq y(\underline{\mu}) - y(\overline{\mu}) \qquad (\overline{SIC})
\end{aligned}
$$

and

$$
\begin{aligned}
EV(\underline{\mu},\underline{\mu}) &\geq EV(\underline{\mu},\overline{\mu}) \\
&\Leftrightarrow \geq y(\underline{\mu}) - y(\overline{\mu}) \geq \underline{\mu}\left[z(\overline{\mu}) - z(\underline{\mu})\right]. \qquad (\underline{SIC})
\end{aligned}
$$

Note that satisfaction of both constraints ($\overline{SIC}$) and ($\underline{SIC}$) implies

$$
\overline{\mu}\left[z(\overline{\mu}) - z(\underline{\mu})\right] \geq y(\underline{\mu}) - y(\overline{\mu}) \geq \underline{\mu}\left[z(\overline{\mu}) - z(\underline{\mu})\right]. \qquad (18)
$$

and, thus, the monotonicity condition $\left[z(\overline{\mu}) - z(\underline{\mu})\right]\Delta\mu > 0$. It is then immediate that the implementability condition

$$
\left[z(\overline{\mu}) - z(\underline{\mu})\right] \geq y(\underline{\mu}) - y(\overline{\mu}) \geq 0 \qquad (19)
$$

has to be satisfied.

We also note that there is nothing to suggest a priori that the principal cannot structure the transfers in a way that guarantees satisfaction of ($\overline{SIC}$) and ($\underline{SIC}$).

The remaining constraints relate to the supervisor's limited liability

$$
t_{ij}^S(\mu) \geq 0, \qquad \text{(SLL)}
$$

coalition incentive compatibility

$$
t_{21}^S(\mu) \geq t_{22}^S(\mu) + k\Psi\left[e_{12}(\mu)\right], \quad \mu = \underline{\mu}, \overline{\mu}, \qquad \text{(CIC)}
$$

agent individual rationality

$$
U_{ij}(\mu) = t_{ij}^A(\mu) - \psi\left[e_{ij}(\mu)\right] \geq 0, \quad i,j = 1,2, \quad \mu = \underline{\mu}, \overline{\mu}\ , \qquad \text{(AIR)}
$$

and agent incentive compatibility

$$
U_{22}(\mu) \geq U_{12}(\mu) + \Psi\left[e_{12}(\mu)\right], \quad \mu = \underline{\mu}, \overline{\mu}. \qquad \text{(AIC)}
$$

The principal's maximization problem is then given by

$$
\begin{aligned}
&\max_{\overline{e}_{ij}, \underline{e}_{ij}, t_{ij}^S(\overline{\mu}), t_{ij}^S(\underline{\mu})} \pi \overline{EW}^A + (1-\pi)\, \underline{EW}^A \qquad \text{(P)} \\
&s.t.\ \overline{SIC}\ ,\ \underline{SIC},\ SLL, CIC,\ AIR,\ AIC
\end{aligned}
$$

The solution of this problem is summarized in the following Proposition.

Proposition 3 *In the presence of asymmetric information on μ, the contract in hierarchy A is a pooling contract with the following properties.*

1.
$$\bar{t}^S_{ij} = \underline{t}^S_{ij} = \begin{cases} k\Psi\left(e^P_{12}\right) & \text{if } i=2,\ j=1 \\ 0 & \text{otherwise} \end{cases} .$$

2.
$$\bar{e}_{ij} = \underline{e}_{ij} = \begin{cases} e^P_{12} < e^* & \text{if } i=1,\ j=2 \\ e^* & \text{otherwise} \end{cases}$$

where
$$\breve{e}_{12} < e^P_{12} \Leftrightarrow \lambda k < \frac{(1-E\mu)}{E\mu},$$

with
$$E\mu := \pi\overline{\mu} + (1-\pi)\underline{\mu}.$$

Proof. See appendix. ■

Under asymmetric information on the supervisor's accuracy μ, the principal is unable to separate out the types and implements a pooling contract, where the distorted effort e^P_{12} in state $(1,2)$ is determined by

$$\psi\prime\left(e^P_{12}\right) = 1 - \frac{\lambda}{1+\lambda}\frac{1-\xi}{\xi}\left(1 + \frac{E\mu}{1-E\mu}k\right)\Psi\prime\left(e^P_{12}\right). \tag{20}$$

Although the principal has a sufficient number of instruments to separate the types, he cannot properly employ them. The reason here is an instant of non-responsiveness (see e.g. Laffont and Martimort 2002: section 2.10.2). Regarding the transfers as such this is not immediate from the implementability condition (19). In prinicple, there is a set of transfers that would satisfy this condition and consequently also satisfy (18) . However, as it turns out it always pays the principal to reduce the total pay-out to the minimum. This implies that all transfers are set to zero apart from $t^S_{21}(\mu) = k\Psi[e_{12}(\mu)]$, as required by (CIC). However, (SIC) no longer allows a differentiation according to the supervisor's type of effort e_{12}.

Considering the problem as one of choosing optimal effort e_{12}, we see that a conflict arises between the structure of effort that should be implemented under full information as opposed to asymmetric information. Under full information, the extraction of the *collusive* rent requires $\underline{e}_{12} > \bar{e}_{12}$ (see Lemma 1). However, this is over-turned under asymmetric information about μ, where the extraction of the *information* rent requires $\underline{e}_{12} < \bar{e}_{12}$. This conflict cannot be resolved without any further instruments and the principal has no choice but to pool the types.

Remark 4 *One may argue that pooling arises due to a lack of transfers $t(\mu^r)$ that depend only on reported types but not on the state ij. Consider for, instance, the following*

$$\begin{aligned} y(\mu^r) &: = t^S(\mu^r) \\ z(\mu^r) &: = (1-\xi)\,t^S_{21}(\mu^r) = k\Psi[e_{12}(\mu^r)]. \end{aligned}$$

In this case, condition (18) reads

$$\overline{\mu}k\left\{\Psi\left[e_{12}\left(\overline{\mu}\right)\right]-\Psi\left[e_{12}\left(\underline{\mu}\right)\right]\right\}\geq t^{S}\left(\underline{\mu}\right)-t^{S}\left(\overline{\mu}\right)\geq\underline{\mu}k\left\{\Psi\left[e_{12}\left(\overline{\mu}\right)\right]-\Psi\left[e_{12}\left(\underline{\mu}\right)\right]\right\}$$

which implies $\Delta\mu k\left\{\Psi\left[e_{12}\left(\overline{\mu}\right)\right]-\Psi\left[e_{12}\left(\underline{\mu}\right)\right]\right\}>0$ *and, since* $\Delta\mu>0$, *we must have* $\Psi\left[e_{12}\left(\overline{\mu}\right)\right]\geq\Psi\left[e_{12}\left(\underline{\mu}\right)\right]$. *However, this requires* $\overline{e}_{12}>\underline{e}_{12}$ *which, again, contradicts the structure of effort under full information,* $\overline{e}_{12}<\underline{e}_{12}$. *Thus, non-responsiveness leads to pooling under these circumstances, too.*

4.2 Hierarchy B

Hierarchy B has eliminated the possibility of collusion between the agent and the intermediary. However, this organization limits the principal's regulatory power to only one instrument—reimbursing the supervisor. In particular, the principal lacks the power of manipulating effort in order to curtail the intermediary's rent when accuracy is private information. Furthermore, as we have seen in the previous section, the principal's device to avoid collusion within this hierarchy was to pool across the intermediary's signal. Note that according to the intermediary's participation constraint (SIR), the principal could differentiate the transfers according to the report on the *agent's* type. However, as (SIR) implies that the expected transfer (across agent's type) decreases in accuracy, a $\overline{\mu}$-intermediary would still always opt for the $\underline{\mu}$ allocation irrespective of the structure of the transfers. Indeed, pooling the intermediary's types is the only option available to the principal. As we have seen the high type $\overline{\mu}$ has an incentive to underreport accuracy. Therefore, assuming that shut-down of low-types does not occur, all types will be reimbursed according to $\underline{\mu}$ and the high type receives a rent equal to

$$EV\left(\overline{\mu},\underline{\mu}\right)=\Delta\mu\left[\xi R_{e}+\left(1-\xi\right)k\Psi\left(\check{e}_{12}\right)\right], \tag{21}$$

with

$$R_{e}\equiv\left[e^{*}-\psi\left(e^{*}\right)\right]-\left[\check{e}_{12}-\psi\left(\check{e}_{12}\right)\right]>0.$$

Comparison of Social Welfare

We now proceed to examine how the two organizational structures compare in social welfare terms when there is asymmetric information about the supervisor. Recall for hierarchy A, that only at state $(2,2)$, the agent's utility is larger than zero, $U_{22}=\lambda\Psi\left(\widehat{e}_{12}\right)$, and only at state $(2,1)$, the supervisor's utility is larger than zero, $V_{21}=\lambda k\Psi\left(e_{12}^{P}\right)$. Taking into account that the optimal contract involves pooling, we can write social welfare for a high realization of accuracy, $\overline{\mu}$,

$$\begin{aligned}\overline{EW}^{A} &= \sum_{i=1}^{2}\sum_{j=1}^{2}\overline{p}_{ij}\left\{B\left(q_{ij}\right)-\left(1+\lambda\right)\left[\left(q_{ij}-e_{ij}^{P}\right)+\psi\left(e_{ij}^{P}\right)\right]\right\}\\ &\quad -\lambda\overline{p}_{22}\Psi\left(e_{12}^{P}\right)-\lambda\overline{p}_{21}k\Psi\left(e_{12}^{P}\right)\end{aligned}$$

Likewise, for a low realization of accuracy, $\underline{\mu}$,

$$\begin{aligned}\underline{EW}^A &= \sum_{i=1}^{2}\sum_{j=1}^{2}\underline{p}_{ij}\left\{B\left(q_{ij}\right)-(1+\lambda)\left[\left(q_{ij}-e_{ij}^P\right)+\psi\left(e_{ij}^P\right)\right]\right\}\\ &\quad -\lambda\underline{p}_{22}\Psi\left(e_{ij}^P\right)-\lambda\underline{p}_{21}k\Psi\left(e_{ij}^P\right).\end{aligned}$$

Under hierarchy B, as defined in the last section, the intermediary's expected utility is equal to

$$\overline{EV}^B = EV\left(\overline{\mu},\underline{\mu}\right) = \Delta\mu\left[\xi R_e + (1-\xi)\,k\Psi\left(\ \breve{e}_{12}\right)\right]$$

as from (21), and $\underline{EV}^B = 0$, as from (SIR). Therefore, social welfare is given by

$$\begin{aligned}\overline{EW}^B &= \sum_{i=1}^{2}\sum_{j=1}^{2}\overline{p}_{ij}\left\{B\left(q_{ij}\right)-(1+\lambda)\left[\left(q_{ij}-\breve{e}_{ij}\right)+\psi\left(\breve{e}_{ij}\right)\right]\right\}\\ &\quad -(1-\xi)\left(1-\overline{\mu}\right)\lambda\Psi\left(\breve{e}_{12}\right)-\lambda\Delta\mu\left[\xi R_e+(1-\xi)\,k\Psi\left(\breve{e}_{12}\right)\right]\end{aligned}$$

for a high realization of accuracy, $\overline{\mu}$, and by

$$\begin{aligned}\underline{EW}^B &= \sum_{i=1}^{2}\sum_{j=1}^{2}\underline{p}_{ij}\left\{B\left(q_{ij}\right)-(1+\lambda)\left[\left(q_{ij}-\breve{e}_{ij}\right)+\psi\left(\breve{e}_{ij}\right)\right]\right\}\\ &\quad -(1-\xi)\left(1-\underline{\mu}\right)\lambda\Psi\left(\breve{e}_{12}\right)\end{aligned}$$

for a low realization of accuracy, $\underline{\mu}$. The difference of expected social welfare between hierarchy A and hierarchy B is then defined as

$$\Delta SW \equiv \left[\pi\overline{EW}^B+(1-\pi)\,\underline{EW}^B\right]-\left[\pi\overline{EW}^A+(1-\pi)\,\underline{EW}^A\right].$$

We can now establish the following.

Proposition 5 *If accuracy is private information, the following is true.*

1. If asymmetric information on μ is unimportant, i.e. if $\pi\Delta\mu \to 0$, then hierarchy B dominates if and only if $\underline{\mu} > \underline{\mu}_0$.

2. If asymmetric information on μ is important i.e. if $\pi\Delta\mu \geq 0$, then for the special case that $\underline{\mu} \to 0$, hierarchy B is dominated by hierarchy A for either high or low values of $\pi\Delta\mu$. However, conditions can be found such that there exist intermediate values of $\pi\Delta\mu$ for which hierarchy B dominates.

Proof. See appendix ■

It is less than straightforward to characterise the impact of asymmetric information on the social preferrability of either hierarchy A or B. Obviously, if asymmetric information is not very important, either because $\pi \to 0$ is low, so that the allocation is governed by $\underline{\mu}$ or if $\Delta\mu \to 0$ then the hierarchies can be ranked under similar criteria as in the case of publicly known accuracy (Proposition 2). If asymmetric information matters $\pi\Delta\mu > 0$ the ranking becomes less than straightforward. Writing

$$\begin{aligned}\Delta SW &= \left\{ \begin{array}{c} \xi\left(1+\lambda\right)\left(1-E\mu\right)\left[\left(\breve{e}_{12}-\psi\left(\breve{e}_{12}\right)\right)-\left(e_{12}^{P}-\psi\left(e_{12}^{P}\right)\right)\right] \\ -\left(1-\xi\right)\lambda\left(1-E\mu\right)\left[\Psi\left(\breve{e}_{12}\right)-\Psi\left(e_{12}^{P}\right)\right] \end{array} \right\} \\ &\quad +\left(1-\xi\right)\lambda k E\mu\Psi\left(e_{12}^{P}\right) \\ &\quad -\lambda\pi\Delta\mu\left[\xi R_e+\left(1-\xi\right)k\Psi\left(\breve{e}_{12}\right)\right].\end{aligned}$$

we see that the difference in welfare is again determined by 'allocative efficiency' (the first line) which can favour either hierarchy depending on the ordering between $\breve{e}_{12}$ and e_{12}^{P}, the collusion rent to be paid under hierarchy A (the second line) and the information rent for hierarchy B (the third line). Increases in $\pi\Delta\mu$ have two effects. On the one hand, they increase $E\mu$, where $E\mu = \underline{\mu}+\pi\Delta\mu$. As we have argued earlier, this tends to favour hierarchy B. On the other hand, the information rent in hierarchy A grows. Obviously, if $\pi\Delta\mu$ grows large, hierarchy B is dominated since it implies large information rents, whereas for a very high $\pi\Delta\mu$ the collusion rent can be made very small through a strong reduction in e_{12}^{P}. However, for intermediate levels of $\pi\Delta\mu$ hierarchy B may dominate.

5 Conclusion and Discussion

We have compared two organizational structures, one with a supervisor, which may be plagued by collusion, one involving delegation to an intermediary. This latter structure eliminates collusion but it is plagued by inefficiency due to a misallocation of effort and possible information rents in the presence of asymmetric information about the supervisor's accuracy in monitoring the agent. We show that if accuracy is publicly known greater levels of accuracy tend to favour delegation. This result is weakened under asymmetric information about accuracy. Here, hierarchy A dominates for high levels of accuracy as they would imply a large informational rents to he intermediary.

6 APPENDIX

6.1 Proof of Lemma 1

1. Follows immediately from the first-order conditions for e_{ij} $(i, j = 1, 2)$ and comparing (6) with (12).

2. Follows from (10) in connection with (2) and (11) in connection with (3).

3. Welfare is given by

$$\begin{aligned}\widehat{EW} &= \sum_{i=1}^{2}\sum_{j=1}^{2} p_{ij}\left\{B_{ij} - (1+\lambda)\left[(q_{ij} - \widehat{e}_{ij}) + \psi(\widehat{e}_{ij})\right]\right\} \\ &\quad -\lambda(p_{22} + p_{21}k)\,\Psi(\widehat{e}_{12})\end{aligned}$$

for hierarchy A, with $\widehat{e}_{11} = \widehat{e}_{21} = \widehat{e}_{22} = e^*$, and by

$$EW^{SB} = \sum_{i=1}^{2} p_{i2}\left\{B_{i2} - (1+\lambda)\left[\left(q_{i2} - e_{i2}^{SB}\right) + \psi\left(e_{i2}^{SB}\right)\right]\right\} - \lambda p_{22}\Psi\left(e_{12}^{SB}\right),$$

for the two-tier structure, where $e_{12}^{SB} = \overline{e}^{SB}$ and $e_{22}^{SB} = e^*$. After straightforward calculations we find

$$\begin{aligned}\Delta SW &= \widehat{EW} - EW^{SB} \\ &= \xi(1+\lambda)\mu\left\{[e^* - \psi(e^*)] - \left[\overline{e}^{SB} - \psi\left(\overline{e}^{SB}\right)\right]\right\} \\ &\quad + \left\{\begin{array}{c}\xi(1+\lambda)(1-\mu)\left[(\widehat{e}_{12} - \psi(\widehat{e}_{12})) - \left(\overline{e}^{SB} - \psi\left(\overline{e}^{SB}\right)\right)\right] \\ -(1-\xi)\lambda\left[1-(1-k)\mu\right]\left[\Psi(\widehat{e}_{12}) - \Psi\left(\overline{e}^{SB}\right)\right]\end{array}\right\} \\ &\quad + (1-\xi)\lambda(1-k)\mu\Psi\left(\overline{e}^{SB}\right)\end{aligned}$$

The first and second terms are positive because

$$e^* = \arg\max\left[e - \psi(e)\right]$$

$$\widehat{e}_{12} = \arg\max\left\{\begin{array}{c}\xi(1+\lambda)(1-\mu)\left[e_{12} - \psi(e_{12})\right] \\ -(1-\xi)\lambda\left[1-\mu(1-k)\right]\Psi(e_{12})\end{array}\right\}.$$

Therefore, $\Delta SW > 0$.

4. Consider the partial-total differential of (12)

$$\psi''(\widehat{e}_{12})\frac{\partial \widehat{e}_{12}}{\partial \mu} = -\frac{\lambda}{1+\lambda}\frac{1-\xi}{\xi}\left\{\Psi''(\widehat{e}_{12})\frac{\partial \widehat{e}_{12}}{\partial \mu} + \frac{k\Psi'(\widehat{e}_{12})}{(1-\mu)^2} + \frac{k\mu\Psi''(\widehat{e}_{12})}{1-\mu}\frac{\partial \widehat{e}_{12}}{\partial \mu}\right\}$$

$$\frac{\partial \widehat{e}_{12}}{\partial \mu}\left[\psi''(\widehat{e}_{12})\frac{1+\lambda}{\lambda}\frac{\xi}{1-\xi} + \Psi''(\widehat{e}_{12})\frac{1-\mu(1-k)}{1-\mu}\right] = -\frac{k}{(1-\mu)^2}\Psi'(\widehat{e}_{12})$$

Given $\psi(\cdot)$ is convex it follows that $\frac{\partial \widehat{e}_{12}}{\partial \mu} < 0$.
QED

6.2 Proof of Proposition 2

1. The condition follows immediately from comparison of (12) and (13).
2. Write social welfare under hierarchies A and B as

$$\begin{aligned}EW^{A} &= \sum_{i=1}^{2}\sum_{j=1}^{2} p_{ij}\left\{B(q_{ij}) - (1+\lambda)\left[(q_{ij} - \widehat{e}_{ij}) + \psi(\widehat{e}_{ij})\right]\right\} \\ &\quad -(1-\xi)(1-\mu)\lambda\Psi(\widehat{e}_{12}) - (1-\xi)\mu\lambda k\Psi(\widehat{e}_{12})\end{aligned}$$

and

$$
\begin{aligned}
EW^B &= \sum_{i=1}^{2}\sum_{j=1}^{2} p_{ij}\left\{B\left(q_{ij}\right)-(1+\lambda)\left[\left(q_{ij}-\breve{e}_{ij}\right)+\psi\left(\breve{e}_{ij}\right)\right]\right\} \\
&\quad -(1-\xi)(1-\mu)\lambda\Psi\left(\breve{e}_{12}\right)
\end{aligned}
$$

respectively. The difference in welfare is thus given by

$$
\begin{aligned}
\Delta SW &= EW^B-EW^A \\
&= \xi(1+\lambda)(1-\mu)\left[\left(\breve{e}_{12}-\psi\left(\breve{e}_{12}\right)\right)-\left(\widehat{e}_{12}-\psi\left(\widehat{e}_{12}\right)\right)\right] \\
&\quad -(1-\xi)\lambda(1-\mu)\left[\Psi\left(\breve{e}_{12}\right)-\Psi\left(\widehat{e}_{12}\right)\right] \\
&\quad +(1-\xi)\lambda k\mu\Psi\left(\widehat{e}_{12}\right).
\end{aligned}
$$

The proof proceeds as follows. We first show that $\Delta SW|_{\mu=0} < 0$ and $\Delta SW|_{\mu=1} > 0$ and then go on to show that $\frac{d\Delta SW}{d\mu}|_{\Delta SW=0} > 0$. Combined this proves the existence of a unique boundary value $\mu_0 \in (0,1)$ such that $\Delta SW > 0 \Longleftrightarrow \mu > \mu_0$.

As a first step, rewrite

$$
\Delta SW = Y(\mu)+(1-\xi)\lambda k\mu\Psi\left(\breve{e}_{12}\right),
$$

with

$$
Y(\mu)=\left\{\begin{array}{c}\xi(1+\lambda)(1-\mu)\left[\left(\breve{e}_{12}-\psi\left(\breve{e}_{12}\right)\right)-\left(\widehat{e}_{12}-\psi\left(\widehat{e}_{12}\right)\right)\right] \\ -(1-\xi)\lambda\left[(1-\mu)+k\mu\right]\left[\Psi\left(\breve{e}_{12}\right)-\Psi\left(\widehat{e}_{12}\right)\right]\end{array}\right\} \leq 0, \tag{22}
$$

where the inequality follows from the fact that

$$
\widehat{e}_{12}=\arg\max\left\{\xi(1+\lambda)(1-\mu)\left(e_{12}-\psi\left(e_{12}\right)\right)-(1-\xi)\lambda\left[(1-\mu)+k\mu\right]\Psi\left(e_{12}\right)\right\}.
$$

Thus, $\Delta SW|_{\mu=0} = Y(0) < 0$

since $Y(\mu) \leq 0$ for all μ and $\widehat{e}_{12} < \breve{e}_{12}$ for $\mu = 0$.Next, consider

$$
\begin{aligned}
\Delta SW|_{\mu=1} &= Y(1)+(1-\xi)\lambda k\Psi\left(\breve{e}_{12}\right) \\
&= (1-\xi)\lambda k\Psi\left(\widehat{e}_{12}\right) > 0.
\end{aligned}
$$

Finally, consider

$$
\begin{aligned}
\frac{d\Delta SW}{d\mu} &= \frac{\partial\Delta SW}{\partial\mu}+\frac{\partial\Delta SW}{\partial\widehat{e}_{12}}\frac{d\widehat{e}_{12}}{d\mu}=\frac{\partial\Delta SW}{\partial\mu} \\
&= X+(1-\xi)\lambda k\Psi\left(\widehat{e}_{12}\right)
\end{aligned} \tag{23}
$$

where $\frac{\partial\Delta SW}{\partial\widehat{e}_{12}} = 0$ from the envelope theorem and where

$$
X:=\left\{\begin{array}{c}\xi(1+\lambda)\left[\left(\widehat{e}_{12}-\psi\left(\widehat{e}_{12}\right)\right)-\left(\breve{e}_{12}-\psi\left(\breve{e}_{12}\right)\right)\right] \\ -(1-\xi)\lambda\left[\Psi\left(\widehat{e}_{12}\right)-\Psi\left(\breve{e}_{12}\right)\right]\end{array}\right\}. \tag{24}
$$

Note that the sign of X can be positive or negative so that the sign of $\frac{d\Delta SW}{d\mu}$ is ambiguous. Observing that we can write

$$Y(\mu) := -(1-\mu)X - (1-\xi)\lambda k\mu\left[\Psi(\breve{e}_{12}) - \Psi(\widehat{e}_{12})\right]$$

it follows that

$$\begin{aligned}\Delta SW &= -(1-\mu)X - (1-\xi)\lambda k\mu\left[\Psi(\breve{e}_{12}) - \Psi(\widehat{e}_{12})\right] + \mu(1-\xi)k\lambda\Psi(\breve{e}_{12}) \\ &= -(1-\mu)X + \mu(1-\xi)k\lambda\Psi(\widehat{e}_{12}) = 0 \\ \iff & X > 0.\end{aligned}$$

But then, it follows from (23) that $\frac{d\Delta SW}{d\mu}|_{\Delta SW=0} > 0$, which completes the proof. *QED*.

6.3 Proof of Proposition 3

1. We first show that the optimal contract always involves a pooling of transfers, i.e. $\bar{t}_{ij}^S = \underline{t}_{ij}^S = t_{ij}^P$. In order to demonstrate this, we consider a simplified problem. Note that (CIC) implies $t_{21}^S(\mu^r) - t_{22}^S(\mu^r) = k\Psi[e_{12}(\mu^r)]$ .Using this in (17) gives

$$\begin{aligned} y(\mu^r) &:= \xi t_{12}^S(\mu^r) + (1-\xi)t_{22}^S(\mu^r) \\ z(\mu^r) &:= \xi\left[t_{11}^S(\mu^r) - t_{12}^S(\mu^r)\right] + (1-\xi)k\Psi[e_{12}(\mu^r)]. \end{aligned}$$

implying that for any $e_{12}(\mu^r)$ we can construct any $y(\mu^r) \geq y_{\min}$ and $z(\mu^r) \geq z_{\min}$, with $y_{\min}, z_{\min} \in [0, k\Psi[e_{12}(\mu^r)]]$ and $z_{\min}+y_{\min} = k\Psi[e_{12}(\mu^r)]$, by finding an appropriate combination of $t_{11}^S(\mu^r), t_{12}^S(\mu^r)$ and $t_{22}^S(\mu^r)$. But then the original problem (P) implies the following problem

$$\max_{\overline{y},\underline{y},\overline{z},\underline{z}} = -\left[\pi(\overline{y} + \overline{\mu z}) + (1-\pi)\left(\underline{y} + \underline{\mu z}\right)\right]$$

subject to

$$\Delta y \geq \underline{\mu}\Delta z, \qquad (\underline{SIC})$$

$$\overline{\mu}\Delta z \geq \Delta y, \qquad (\overline{SIC})$$

$$\Delta z \geq 0, \qquad \text{(MC1)}$$

$$\Delta y \geq 0, \qquad \text{(MC2)}$$

$$\overline{y} - y_{\min} \geq 0, \qquad \text{(LL1)}$$

$$\underline{z} - z_{\min} \geq 0, \qquad \text{(LL2)}$$

where $\overline{y} := y(\overline{\mu}), \underline{y} := y\left(\underline{\mu}\right), \overline{z} := z(\overline{\mu})$, etc., and where $\Delta y := \underline{y} - \overline{y}$ and $\Delta z := \overline{z} - \underline{z}$. The Lagrangian function of the problem is

$$L = \left\{ \begin{array}{c} -\left[\pi(\overline{y} + \overline{\mu z}) + (1-\pi)\left(\underline{y} + \underline{\mu z}\right)\right] \\ +\Omega_1\left(\Delta y - \underline{\mu}\Delta z\right) + \Omega_2(\overline{\mu}\Delta z - \Delta y) + \Omega_3\Delta z + \Omega_4\Delta y + \Omega_5(\overline{y} - y_{\min}) + \Omega_6\left[\underline{z} - z_{\min}\right] \end{array} \right\}.$$

The first-order conditions for $\overline{y}, \underline{y}, \overline{z}$ and $\underline{z}$ are given by

$$-\pi\overline{\mu} - \Omega_1\underline{\mu} + \Omega_2\overline{\mu} + \Omega_3 = 0, \tag{25}$$

$$-(1-\pi)\,\underline{\mu} + \Omega_1\underline{\mu} - \Omega_2\overline{\mu} - \Omega_3 + \Omega_6 = 0, \tag{26}$$

$$-\pi - \Omega_1 + \Omega_2 - \Omega_4 + \Omega_5 = 0, \tag{27}$$

$$-(1-\pi) + \Omega_1 - \Omega_2 + \Omega_4 = 0. \tag{28}$$

We obtain from (27) with (28)

$$\Omega_5 = 1 > 0 \quad \Longrightarrow \overline{y} = y_{\min}$$

and from (25) with (26)

$$\Omega_6 = \pi\overline{\mu} + (1-\pi)\,\underline{\mu} = E\mu > 0 \quad \Longrightarrow \underline{z} = z_{\min}.$$

Furthermore, we notice that the four combinations of (Ω_1, Ω_2), i.e. $(\Omega_1 = 0, \Omega_2 = 0)$, $(\Omega_1 = 0, \Omega_2 > 0)$, $(\Omega_1 >$ and $(\Omega_1 > 0, \Omega_2 > 0)$, are all feasible a priori. However, we can now show that all of the four imply pooling. Indeed

1. $(\Omega_1 = 0, \Omega_2 = 0) \Longrightarrow$ from (25) and (28)

   $$\Omega_3,\ \Omega_4 > 0 \quad \Longrightarrow \Delta y = \Delta z = 0.$$

2. $(\Omega_1 = 0, \Omega_2 > 0) \Longrightarrow$ from (28)

   $$\Omega_4 > 0 \quad \Longrightarrow \Delta y = 0$$

   and from $(\underline{SIC})$

   $$\Delta z = 0.$$

3. $(\Omega_1 > 0, \Omega_2 = 0) \Longrightarrow$ from (25)

   $$\Omega_3 > 0 \quad \Longrightarrow \Delta z = 0$$

   and from $(\overline{SIC})$

   $$\Delta y = 0.$$

4. $(\Omega_1 > 0, \Omega_2 > 0) \Longrightarrow$ Both $(\overline{SIC})$and $(\underline{SIC})$ are binding. From (18)

   $$\overline{\mu}\Delta z = \Delta y = \underline{\mu}\Delta z \quad \Longrightarrow \Delta y = \Delta z = 0.$$

Because $\overline{y} = y_{\min}$ it then follows $\underline{y} = y_{\min}$. Thus the principal can make the arrangement such that $\bar{t}_{12}^S = \underline{t}_{12}^S = t_{12}^P = 0$, and $\bar{t}_{22}^S = \underline{t}_{22}^S = t_{22}^P = 0$.Likewise, $\underline{z} = \overline{z} = z_{\min}$. Here, for $\bar{t}_{12}^S = \underline{t}_{12}^S = t_{12}^P = 0$ we have $\bar{t}_{11}^S = \underline{t}_{11}^S = t_{11}^P = 0$. Finally, from $\overline{CIC}$ and $\underline{CIC}$ it follows that $\bar{t}_{21}^S = k\Psi\left[e_{12}\left(\overline{\mu}\right)\right] = \underline{t}_{21}^S = k\Psi\left[e_{12}\left(\underline{\mu}\right)\right] = t_{21}^P$. But then it follows that the principal has to impose pooling effort $\overline{e}_{12} = \underline{e}_{12} = e_{12}^P$.

2. We can now consider the maximization problem

$$\max_{e_{ij}^P} \pi \overline{EW}^A + (1-\pi)\, \underline{EW}^A$$

where

$$\overline{EW}^A = \left\{ \begin{array}{c} \sum_{i=1}^2 \sum_{j=1}^2 \overline{p}_{ij} \left\{ B\left(q_{ij}\right) - (1+\lambda)\left[\left(q_{ij} - \ e_{ij}^P\right) + \psi\left(\ e_{ij}^P\right)\right]\right\} \\ -\lambda \overline{p}_{22} \Psi\left(\ e_{12}^P\right) - \lambda \overline{p}_{21} k \Psi\left(\ e_{12}^P\right) \end{array} \right\}$$

$$\underline{EW}^A = \left\{ \begin{array}{c} \sum_{i=1}^2 \sum_{j=1}^2 \underline{p}_{ij} \left\{ B\left(q_{ij}\right) - (1+\lambda)\left[\left(q_{ij} - \ e_{ij}^P\right) + \psi\left(\ e_{ij}^P\right)\right]\right\} \\ -\lambda \underline{p}_{22} \Psi\left(\ e_{12}^P\right) - \lambda \underline{p}_{21} k \Psi\left(\ e_{12}^P\right) \end{array} \right\}.$$

Note that the agent's (AIR) and (AIC) constraints have been taken into account as usual. From the first-order conidtions it is readily verified that

$$\psi\prime\left(e_{12}^P\right) = 1 - \frac{\lambda}{1+\lambda} \frac{1-\xi}{\xi} \left(1 + \frac{E\mu}{1-E\mu} k\right) \Psi\prime\left(e_{12}^P\right),$$

where $E\mu := \pi\overline{\mu} + (1-\pi)\underline{\mu}$. Finally, comparison with (13) gives the condition reported. *QED*.

6.4 Proof of Proposition 4

The difference in welfare can be expressed by

$$\begin{aligned} \Delta SW \quad = \quad & \left\{ \begin{array}{c} \xi(1+\lambda)(1-E\mu)\left[\left(\breve{e}_{12} - \psi\left(\breve{e}_{12}\right)\right) - \left(e_{12}^P - \psi\left(e_{12}^P\right)\right)\right] \\ -(1-\xi)\lambda(1-E\mu)\left[\Psi\left(\breve{e}_{12}\right) - \Psi\left(e_{12}^P\right)\right] \end{array} \right\} \\ & + (1-\xi)\lambda k E\mu \Psi\left(e_{12}^P\right) \\ & - \lambda\pi\Delta\mu\left[\xi R_e + (1-\xi) k \Psi\left(\breve{e}_{12}\right)\right]. \end{aligned}$$

Note that $E\mu = \underline{\mu} + \pi\Delta\mu$. We can, thus, establish as a benchmark, the case $\pi\Delta\mu = 0$, i.e. a situation in which asymmetric information is irrelevant. For this case, the results in Proposition 2 apply.

Let us now consider, the impact of asymmetric information, where $\pi\Delta\mu \geq 0$. In order to make analysis tractable, we focus on the special case, where $\underline{\mu} = 0$ and $E\mu = \pi\Delta\mu$. Again, in analogy to Proposition 2, we find

$$\Delta SW\left|_{\pi\Delta\mu=0}\right. = Y(0) < 0,$$

where $Y(\pi\Delta\mu) \leq 0$, as defined in (22), for all $\pi\Delta\mu$ and $\widehat{e}_{12} < \breve{e}_{12}$ for $\pi\Delta\mu = 0$.Next, consider

$$\begin{aligned}\Delta SW\left|_{\pi\Delta\mu=1}\right. &= Y(1) + (1-\xi)\lambda k\Psi(\breve{e}_{12}) - \lambda\left[\xi R_e + (1-\xi)k\Psi(\breve{e}_{12})\right] \\ &= (1-\xi)\lambda k\left[\Psi\left(e_{12}^P\right) - \Psi(\breve{e}_{12})\right] - \lambda\xi R_e < 0\end{aligned}$$

since $e_{12}^P\left|_{\pi\Delta\mu=1}\right. << \breve{e}_{12}$. Finally, consider

$$\frac{d}{d(\pi\Delta\mu)}\Delta SW = X + (1-\xi)\lambda k\Psi\left(e_{12}^P\right) - \lambda\left[\xi R_e + (1-\xi)k\Psi(\breve{e}_{12})\right]$$

where X as defined in (24). Furthermore, it is readily established that

$$\frac{d^2}{d(\pi\Delta\mu)^2}\Delta SW = \left[\xi(1+\lambda)\left(1-\psi'\left(e_{12}^P\right)\right) - (1-\xi)\lambda(1+k)\Psi'\left(e_{12}^P\right)\right]\frac{de_{12}^P}{d(\pi\Delta\mu)}$$

Noting that $\frac{de_{12}^P}{d(\pi\Delta\mu)} = \frac{\partial e_{12}^P}{\partial E\mu} < 0$, it is easily checked with reference to (20) that $\frac{d^2}{d(\pi\Delta\mu)^2}\Delta SW \geq 0 \Leftrightarrow E\mu = \pi\Delta\mu \leq \frac{1}{2}$.

Suppose now there exists a $\widetilde{\pi\Delta\mu} \geq \frac{1}{2}$ for which $\frac{d}{d(\pi\Delta\mu)}\Delta SW = 0$ with $\frac{d^2}{d(\pi\Delta\mu)^2}\Delta SW < 0$ is true. This implies a maximum of ΔSW. Noting that

$$\begin{aligned}\frac{d}{d(\pi\Delta\mu)}\Delta SW &= 0 \Leftrightarrow \\ X &= \lambda\left[\xi R_e + (1-\xi)k\Psi(\breve{e}_{12})\right] - (1-\xi)\lambda k\Psi\left(e_{12}^P\right) - \lambda\left[\xi R_e + (1-\xi)k\Psi(\breve{e}_{12})\right]\end{aligned}$$

and writing, in analogy to Proposition 2,

$$\Delta SW = -(1-\pi\Delta\mu)X + \pi\Delta\mu(1-\xi)k\lambda\Psi\left(e_{12}^P\right) - \lambda\pi\Delta\mu\left[\xi R_e + (1-\xi)k\Psi(\breve{e}_{12})\right],$$

it follows that

$$\max\Delta SW = \Delta SW\left|_{\frac{d}{d(\pi\Delta\mu)}\Delta SW=0}\right. = (1-\xi)k\lambda\Psi\left(e_{12}^P\right) - \lambda\left[\xi R_e + (1-\xi)k\Psi(\breve{e}_{12})\right].$$

Obviously, $\max\Delta SW > 0$ if and only if $(1-\xi)k\Psi\left(e_{12}^P\left|_{\widetilde{\pi\Delta\mu}}\right.\right) - \lambda\left[\xi R_e + (1-\xi)k\Psi(\breve{e}_{12})\right] > 0$. This condition holds if ξ is sufficiently small and if $e_{12}^P > \breve{e}_{12}$ at the maximum.
Noting that

$$\frac{d}{d(\pi\Delta\mu)}\Delta SW\left|_{\Delta SW=0}\right. = \frac{\lambda\pi\Delta\mu}{1-\pi\Delta\mu}\left\{(1-\xi)k\left[\Psi\left(e_{12}^P\right) - \Psi(\breve{e}_{12})\right] - \xi R_e\right\}$$

it is readily verified that there exist two boundary levels $\underline{\pi\Delta\mu}$ and $\overline{\pi\Delta\mu}$ such that

$\Delta SW > 0$ if and only if the following conditions are satisfied:

1. there exist $\widetilde{\pi\Delta\mu} := \arg\max \Delta SW$ and $(1-\xi)\, k\Psi \left(e_{12}^{P} \Big|_{\widetilde{\pi\Delta\mu}} \right) - \lambda \left[\xi R_e + (1-\xi)\, k\Psi \left(\breve{e}_{12} \right) \right] > 0$
2. $\pi\Delta\mu \in \left[\underline{\pi\Delta\mu}, \overline{\pi\Delta\mu} \right]$.

Otherwise $\Delta SW < 0$ for all $\pi\Delta\mu$. *QED*.

Acknowledgement 6 *We are grateful to Hugh Gravelle and Luigi Siciliani for helpful comments. All errors are ours. Financial contributions from DERS towards attending the conference is acknowledged by the second author.*

References

[1] Baron, D. P. and R. Myerson (1982). "Regulating a Monopoly with Unknown Costs." Econometrica 50: 911-930.

[2] Baron, D. P. and D. Besanko (1992). "Information, Control, and Organizational Structure." Journal of Economics & Management Strategy 1(2): 237-75.

[3] Baron, D. P. and D. Besanko (1994). Information Hierarchies, Self-Remedying Hidden Gaming, and Organizational Neutrality, Stanford University, Research Paper No. 1315: 1-26.

[4] Faure-Grimaud, A., J. J. Laffont, et al. (1999). "The endogenous transaction costs of delegated auditing - American Economic Review." European Economic Review 43(4-6): 1039-1048.

[5] Faure-Grimaud, A., J. J. Laffont, et al. (2003). "Collusion, Delegation and Supervision with Soft Information." Review of Economic Studies 70(2): 253-280.

[6] Department of Health (2000), The NHS Plan. A Plan for Investment. A Plan for Reform, London: Stationery Office.

[7] Department of Health (2001), Primary Care, General Practice and the NHS Plan; Information for GPs, Nurses, Other Health Professionals and Staff Working in Primary Care in England, London: Stationery Office.

[8] Gravelle, H. (1999). "Capitation contracts: access and quality." Journal of Health Economics 18(3): 315-340.

[9] Gravelle, H. and R. Rees (2004). Microeconomics. New York, Pearson Education.

[10] Healthcare Commission (2004), Corporate Plan 2004/2008, London: Healthcare Commission (www.healthcarecommission.org.uk).

[11] Kofman, F. and J. Lawarree (1993). "Collusion in Hierarchical Agency." Econometrica 61(3): 629.

[12] Laffont, J.-J. and J. Tirole (1991). "The Politics of Government Decision-making: a Theory of Regulatory Capture", Quarterly Journal of Economics 106, 1089-1127.

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

[14] Laffont, J. and D. Martimort (2002). The Theory of Incentives: The Principal-Agent Model. Princeton and Oxford, Princeton University Press.

[15] Laffont, J. J. and D. Martimort (1998). "Collusion and Delegation." Rand Journal of Economics 29(2): 280-305.

[16] Maskin, E. S. and J. Tirole (1990). "The Principal-Agent Relationship with an Informed Principal: The Case of Private Values." Econometrica 58(2): 379-409.

[17] Maskin, E. S. and J. Tirole (1992). "The Principal-Agent Relationship with an Informed Principal, II: Common Values." Econometrica 60(1): 1-42.

[18] Tirole, J. (1986). "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations." Journal of Law, Economics and Organization 2(2): 181-214.

[19] Tirole, J. (1992). Collusion and the theory of organizations. Advances in Economic Theory: Sixth World Congress Vol. II. J.-J. Laffont. Cambridge, UK, Cambridge University Press.

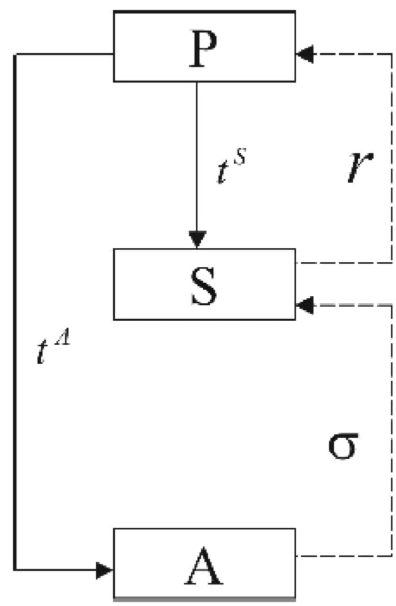

Figure 1a: Hierarchy A. Principal P contracts with agent A *and* supervisor S.

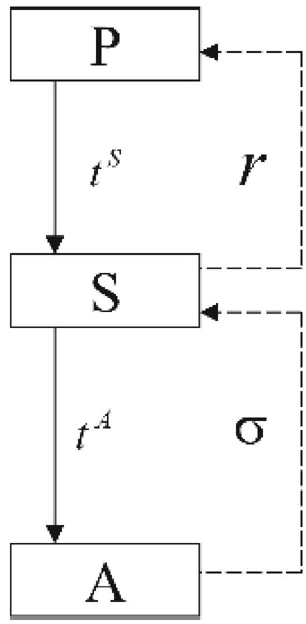

Figure 1b: Hierarchy B. Principal P contracts with intermediary S; intermediary contracts with agent A.