TSE WP working paper series

12-343

Research Group: Public Economics

October 2012

"Moral Hazard in Hierarchies and Soft Information"

Charles Angelucci and Antonio Russo

Toulouse School of Economics (TSE) - Manufacture des Tabacs
Aile Jean-Jacques Laffont - 21, allée de Brienne - 31015 TOULOUSE Cedex 6
Tél : +33 (0) 5 61 12 85 89 - Fax : +33 (0) 5 61 12 86 37 - **www.tse-fr.eu** - contact@tse-fr.eu

Moral Hazard in Hierarchies and Soft Information*

Charles Angelucci
Department of Economics
Harvard University
and Toulouse School of Economics
Email: cangelucci@fas.harvard.edu

Antonio Russo
Robert Schuman Center for Advanced Studies, EUI
and Toulouse School of Economics
Email: antonio.russo@eui.eu

October 10, 2012

Abstract

We investigate the scope for supervisory activities in organizations in which information is non-verifiable and opportunism severe. A principal-supervisor-agent hierarchy is considered. Side-contracts between supervisor and agent may be reached both before and after the agent has chosen his hidden action. We find that the supervisor is useful if and only if appointed before the agent has chosen his action. We also show that delegation of payroll authority is suboptimal. Finally, some insights concerning the optimal design of verification activities are provided: when information is non-verifiable, the supervisor should be employed as a monitor rather than as an auditor.

JEL Classification: D82, D86, L23

Keywords: collusion, extortion, delegation, mechanism design

*We thank Philippe Aghion, Andrea Attar, Julia Cagé, Jacques Crémer, Emmanuel Farhi, Renato Gomes, Olga Gorelkina, Martijn Han, Oliver Hart, Holger Herz, Jacques Lawarrée, Yassine Lefouili, Patrick Legros, Ching-to Albert Ma, Inés Macho-Stadler, David Martimort, Volker Nocke, Simone Meraglia, Humberto Moreira, Tommaso Reggiani, Patrick Rey, François Salanié, Wilfried Sand-Zantman, Jean Tirole, Kouroche Vafaï and seminar participants in Alicante (ASSET 2010), Crest-Lei, Ecole Polytechnique, Paris-10, ECARES, Cargese (ESNIE 2011) and Oslo (EEA 2011) for helpful discussions and comments. Opinions and errors are ours.

1 Introduction

Informal agreements—agreements that are not enforced by courts but sustainable because of trust, repeated relationships, violence, etc—play a prominent role in the economy. In particular, and as emphasized by sociologists, hidden contracts between members of an organization are ubiquitous (see Roethilsberger and Dickson [1947] and Dalton [1959]). These may aim at cooperative ends, but also opportunistic ones. The scope for such opportunism is much greater when the information the organization relies upon is manipulable. Yet, the reliance on subjective evaluation of employees by supervisors is widespread in firms (see Gibbons [2005]).[1] In this paper we investigate the optimal organizational response to informal agreements between supervisors and supervisees in an environment in which information is entirely manipulable.

Supervision is here taken as consisting in both the provision of incentives to a subordinate and the gathering of information concerning the latter's performance. Both tasks, we show, are tightly intertwined. A supervisor is said to be cooperative if she treats her subordinate fairly, despite other attitudes being more profitable (e.g. being tough). Such *cooperation*, not surprisingly, is beneficial to the organization. *Opportunism*, which may either be *collective* or *individual*, is instead detrimental. Consider collective opportunism: coalitions of individuals may emerge so as to enforce their own objectives. Such *collusion* may range from the simple exchange of favors to outright fraud, involving the hidden transfer of money or goods. As a leading example of this type of opportunism, consider the case of payroll fraud:

> *"(one) way to obtain approval of a fraudulent time card is to collude with a supervisor who authorizes timekeeping information. In these schemes, the supervisor knowingly signs false time cards and usually takes a portion of the fraudulent wage. In some cases, the supervisor may take the entire amount of the overpayment. In an example, a supervisor assigned employees to better work areas or better jobs, but in return demanded payment. (...) The employees were compensated for fictitious overtime, which was kicked back to the supervisor."*
>
> Joseph T. Wells, chairman of the Association of Certified Fraud Examiners.

Individual opportunism operates differently: it benefits only the instigator. A supervisor may be tempted to report falsely the performance of a subordinate if lucrative to do so or simply because acting in a negligent way. In the extreme, a supervisor may even be tempted to

[1] As a further example, Gibbs, Merchant, Van der Stede, and Vargus [2004] report that incentive payments for 23 percent of managers in car dealerships are tied to a subjective appraisal of their performance.

engage in *extortion*: threaten to make a false report affecting negatively her subordinate so as to extract bribes or favors. In this vein, one may take sexual harassment or bullying at the workplace as examples of such behavior.[2]

Economics, since Tirole [1986], has addressed these issues, investigating the possibility for agents to reach binding, though hidden, agreements.[3] The relevance of such "side contracting" depends on the nature of the information gathered by supervisors. This, in turn, determines the usefulness of the supervisor. Rather intuitively, from existing literature, it emerges that the less verifiable information is, the larger the scope for opportunism. In a model where reports on an agent's hidden action are fully verifiable (i.e. hard), Kessler [2000], for instance, shows that opportunism is harmless and that, consequently, the supervisor's information is made full use of. Khalil, *et al* [2010] instead argue that if information can be collectively manipulated (but not individually), there exists a tension between preventing both individual and collective opportunism. Because of this tension, the organization is unable to make full use of the supervisor's information. In the extreme, Tirole [1986] argues that, when the supervisor's reports are entirely non-verifiable, opportunism is so pervasive that no use can be made of her reports.

Intuitively, non-verifiability of information facilitates several forms of opportunism. It exposes members of an organization to supervisors' individual opportunism. Falsely accusing subordinates of errors not actually made or unfair evaluation of performances are recurrent instances of workplace bullying by managers or negligence.[4] It may also increase the scope for collective opportunism (i.e., supervisor-agent collusion), as recognized by Dalton [1959].[5] In practice, however, non-verifiable supervisory reports are pervasive in organizations, thereby contradicting theoretical predictions. While research has been conducted in adverse selection environments[6] , a further investigation in the context of moral hazard is warranted to reconcile

[2]The two forms of opportunism impact differently the organization. As pointed out by Khalil, Lawarrée and Yun [2010] (hereafter KLY) and Vafaï [2002, 2010], while collusion reduces the agent's gains from shirking (since a bribe needs to be paid not to be caught) extortion instead punishes the hard working agent (since a bribe needs to be paid not to be unfairly punished).

[3]One may invoke honor, friendship, and repeated relationships to justify their enforceability. As Itoh [1993] puts it: "*this assumption is clearly extreme. However, it also appears extreme to assume that no promise can be honored.*"

[4]See the survey by Fevre *et al.* [2011].

[5]In his study of the "Milo Fractionating Plant", he reports cases of costs being exaggerated by the creation of fictitious personnel, overstatement of costs of equipment, and even the manipulation of scientific experiments and data.

[6]In an adverse selection set-up, Faure-Grimaud, Laffont and Martimort [2003] show that the supervisor is valuable with soft information if collusion happens under asymmetric information and if she is risk averse. We instead consider a risk neutral supervisor. Also in an adverse selection set-up, Baliga [1999] shows that there exists an equilibrium in which hard and soft information are equivalent in Tirole [1992]'s model. Individual

the theory with actual practice.

To study the issue at hand we adopt a principal-supervisor-agent set-up, with a moral hazard problem at the bottom. Information is entirely non-verifiable and supervisor and agent can stipulate agreements that may lead to either *cooperation, collusion,* or *extortion.* Importantly, these agreements (or side-contracts) can be made both before (*ex ante*) and after (*ex post*) the agent has chosen his action. The possibility for *ex ante* side contracting differentiates us from existing literature on three-tier hierarchies.[7] We believe this to be a particularly reasonable description of relationships within organizations such as firms: these tend to foster close ties.[8] The supervisor is assumed particularly opportunistic in the sense that she may be willing to inflict harm upon her subordinate and the principal even when other payoff equivalent options exist. She may wish to do so, for instance, to retaliate in case the agent refused to collude. Further, all incentives are pecuniary, although these can be given broader interpretations, but hidden transfers involve some transaction costs. Such a model allows us to investigate (i) the extent to which the organization is able to make use of the supervisor's information and (ii) whether delegating payroll authority to the supervisor is optimal. Some insights for the optimal design of verification activities are also provided.

Pervasiveness of opportunism. We begin by considering a situation in which side-contracts can be signed only once the agent has chosen how much effort to exert. We refer to these as *ex post* side-contracts. In such instances, the supervisor's opportunism is so pervasive that her information is useless to the organization. Since her behavior does not affect the agent's (already chosen) action, she designs the most profitable informal agreement (either involving collusion or extortion) which involves herself and the agent making reports that are independent of the supervisor's private information, and thus useless to the principal.[9]

Things differ radically when *ex ante* side-contracts can also be signed. These are side-contracts signed before the agent chooses his action. As long as she can commit to such agreements, the supervisor's information becomes valuable to the organization. While increasing the number of possible collusive strategies, *ex ante* side contracting also forces

opportunism, however, is not considered.

[7]See for instance Khalil, *et al* and Vafaï.

[8]Other types of corruption, such as undisciplined drivers trying to bribe policemen, are probably better explained by *ex post* side contracting. However, our set-up can also apply to prolonged monitoring of firms by public officials. An example can be the monitoring of environmental impacts of infrastructure projects by public agencies.

[9]Formally, we focus on Markov Perfect Equilibria (see Maskin and Tirole [2001]) in which players do not make their strategies contingent on payoff irrelevant information. Such equilibria indeed exacerbate the problem caused by non-verifiability of information and opportunism.

the supervisor to internalize the consequences of her opportunism on the agent's effort. Such a thing is not possible when agreements between the two can be signed only *ex post*. The principal thus designs an incentive scheme rewarding *both* players when the agent's task is successful. It then becomes in the supervisor's best interest to act *cooperatively*. This, we show, necessarily involves a promise not to engage in extortion *ex post*, precisely to induce the agent into exerting high effort. This result establishes that supervisors can be useful even when information is entirely non-verifiable, as commonly observed in practice, and opportunism severe, as documented by sociologists.

Payroll authority. We also investigate whether the optimal organizational response to opportunism involves delegating payoff authority to the supervisor. The reason why a situation in which side contracting occurs *ex ante*, rather than *ex post*, dominates is that it leads to the supervisor internalizing the consequences of opportunism on the agent's incentives to exert high effort. Taking this logic to its extreme, relying entirely on the supervisor to provide the right incentives to the agent could be a powerful organizational arrangement to alleviate opportunism. We find, however, that such a structure is in fact suboptimal. If the principal makes transfers only to the supervisor, information manipulation by the latter does not impact the agent's incentives to exert high effort directly. The supervisor thus simply makes the report with the highest associated transfer, and her information becomes useless to the principal. This is in contrast to other moral hazard set-ups in which side contracting, unlike us, occurs among agents performing almost identical tasks (see Holmstrom and Milgrom [1990], Itoh [1993] and Baliga and Sjöström [1998]).

Optimal design of verification activities. The above findings have direct implications for the optimal design of verification activities. In the spirit of Strausz [2005], we compare two alternative forms of verification: one in which the supervisor functions as a *monitor* (starting her work before the agent chooses his effort level) and one in which she functions rather as an *auditor* (being called for only once the agent's task is concluded). We find that the two organizational structures are equivalent when supervisory information is verifiable. However, monitoring dominates auditing in the presence of non-verifiable information. These findings complement those of Strausz, but in an environment where verification is delegated by the principal to a third player and taking into account the scope for opportunism.[10]

[10] Strausz's main focus is on the principal's inability to commit to a costly verification effort level. Indeed, there is no supervisor in his model.

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 solves the game first allowing supervisor and agent to side-contract only once the latter has chosen his action, and then introducing the possibility of side contracting also before that. It ends by deriving implications for the delegation of payroll authority to the supervisor and for the optimal organization of internal audits. Section 4 concludes.

2 Set-up of the Model

2.1 Information and Players

A principal P (the organization) contracts both with a supervisor S (she), and an agent A (he).[11] All players are assumed risk neutral. A and S have zero reservation utilities. None of them has private wealth and both are protected by limited liability.

Information. There is a publicly available signal π and a signal σ observed only by S and A. Both are correlated with A's action in a way that will be shown below. Finally, we have the (publicly observable) pair of reports $m = (m_S, m_A)$ which S and A respectively make to P. We define a state as the combined realization of π, σ and m.

Agent. A unobservably chooses a binary action $e \in \{\underline{e}, \bar{e}\}$. This action may be given various interpretations, but, to fix ideas, we interpret it as the choice between two levels of effort on a certain task assigned by P: low $\underline{e}$ and high $\bar{e}$. Effort e implies disutility $\psi(e)$ for A, where $\psi(\bar{e}) = \psi > \psi(\underline{e}) = 0$. Signal π takes one of two values: high $\pi = \bar{\pi}$ or low $\pi = \underline{\pi}$ and is correlated with e. It may be interpreted, for instance, as the profitability of the project A works on. Its conditional distribution is as follows:

	$e = \bar{e}$	$e = \underline{e}$
$\bar{\pi}$	ρ_π	$1 - \rho_\pi$
$\underline{\pi}$	$1 - \rho_\pi$	ρ_π

[11] Note that our hierarchy may be given interpretations other than the one developed in the introduction. Still focusing on a corporate environment S may for instance be interpreted as a "compliance officer" recruited by S to ensure that A abides by the law. If one sticks to such an interpretation this paper can then be said to compute the agency costs associated with the implementation of a compliance program (see Angelucci and Han (2012) for a treatment of corporate crime). Another possible interpretation is that of a regulatory context in which transfers are permitted between P, the "political principal" and both S, the regulator, and A, the regulated firm.

where $\rho_\pi > \frac{1}{2}$; i.e. $\bar{\pi}$ ($\underline{\pi}$) is more likely when high (low) effort is made. As we show in the analysis, the public signal is a key tool at the disposal of P as it leads to S internalizing the consequences of her opportunism. A receives a state-contingent transfer $t_{\pi m} \geq 0$ from P and makes/receives side-transfer $y_{\pi\sigma m}$ from/to S. The latter is assumed to be positive when going from S to A and negative otherwise. A's utility function is $U^A = t_{\pi m} + y_{\pi\sigma m} - \psi(e)$.

Supervisor. The supervisor possesses a verification technology allowing her, with probability ρ_σ, to obtain signal σ that is perfectly informative of A's effort e.[12] The signal is instead empty with probability $1-\rho_\sigma$. The conditional distribution of σ is thus:

	$e=\bar{e}$	$e=\underline{e}$
$\sigma = N$	$1-\rho_\sigma$	$1-\rho_\sigma$
$\sigma = G$	ρ_σ	0
$\sigma = B$	0	ρ_σ

We refer to G as "good news" concerning A's effort, B as "bad news" and N as "no news". In the following, we study the implications, for the optimal organizational design, of the nature of the supervisor's information. Two polar cases—that have been the main focus of the analysis—are considered: *verifiable* (or "hard") and *non-verifiable* (or "soft") information. If σ is verifiable, anticipating on the relevant message spaces, then it can be hidden, but not forged. More precisely, given σ, $m_S \in [\sigma, N]$ if information is hard. Instead, if σ is *soft* information, m_S can take any value on the support of σ, regardless of its actual realization.[13][14]

In the case of soft information, we focus on Markov Perfect equilibria whereby strategically equivalent subgames have the same equilibrium, i.e., players do not make their strategies contingent on payoff irrelevant information (see Maskin and Tirole [2001] for a formal definition of the "Markov Principle"). This assumption allows us to (i) focus on the situation in which opportunism is potentially the most costly to the organization—since S and A's behavior may not vary with σ—and (ii) isolate well the value to the organization of appointing S early so as to exploit her commitment abilities. In addition, we assume that S's threats to A, when

[12] S is not part of the production process. This differentiates our setting from other papers also focusing on moral hazard environments, in which S has a role in production (see Holmstöm and Milgrom [1990], Itoh [1993] and Baliga and Sjoström [1998]). As emphasized by Radner [1992] it is very commonly observed that management is a very distinct activity from production.

[13] A's information, in contrast, is always taken to be non-verifiable.

[14] KLY [2010] proposed a hybrid information structure in which information is soft under collusion and otherwise hard. We discuss below what our results would be in such a scenario. These would be qualitatively the same as in the case of fully soft information.

coherent with equilibrium behavior, are credible. This implies that when S and A bargain over a side-contract, A's outside option is determined by the worst equilibrium (from his perspective) in the relevant continuation game. This, once again, captures well the spirit of opportunism.

S's utility function is $U^S = s_{\pi m} - K(y_{\pi\sigma m}) \cdot y_{\pi\sigma m}$, where $s_{\pi m} \geq 0$ and $y_{\pi\sigma m}$ denote respectively the state-contingent salary received from P and the side-transfer made to/received from A. $K(y_{\pi\sigma m})$ captures the transaction cost of organizing side-transfers for S. We have $K(y_{\pi\sigma m}) = k$ if $y_{\pi\sigma m} > 0$ and $K(y_{\pi\sigma m}) = \frac{1}{k}$ otherwise. If S wants to send 1 dollar to A, she has to pay $k > 1$ dollars, while if A sends her 1 dollar, S receives only $\frac{1}{k} < 1$ dollars.

Principal. Total (state contingent) payments by P are denoted $z_{\pi m} = t_{\pi m} + s_{\pi m}$. Its utility function is $U^P = u(\pi) - z_{\pi m}$. We assume that P gets some utility $u(.) > 0$ when the project A works on is successful, i.e., $\pi = \bar{\pi}$, and none otherwise. Such utility is high enough that P always prefers to induce A into choosing $\bar{e}$. This is why P's objective is to induce high effort at the lowest possible cost. Such cost is measured by $E(z) = \sum_\pi \sum_\sigma p^{\bar{e}}_{\pi\sigma} (t_{\pi m} + s_{\pi m})$, where $p^e_{\pi\sigma}$ is the probability, conditional on e, of state (π, σ, m) taking place.[15] It is useful to explicitly write the joint distribution of probabilities $p^e_{\pi\sigma}$

	$e = \bar{e}$	$e = \underline{e}$		$e = \bar{e}$	$e = \underline{e}$
$p^e_{\bar{\pi}G}$	$\rho_\sigma \rho_\pi$	0	$p^e_{\underline{\pi}G}$	$\rho_\sigma (1 - \rho_\pi)$	0
$p^e_{\bar{\pi}N}$	$(1 - \rho_\sigma) \rho_\pi$	$(1 - \rho_\sigma)(1 - \rho_\pi)$	$p^e_{\underline{\pi}N}$	$(1 - \rho_\sigma)(1 - \rho_\pi)$	$(1 - \rho_\sigma) \rho_\pi$
$p^e_{\bar{\pi}B}$	0	$\rho_\sigma (1 - \rho_\pi)$	$p^e_{\underline{\pi}B}$	0	$\rho_\sigma \rho_\pi$

We also denote the expected transfers to A and S, as well as side-transfers, conditional on the effort e, on the actual realization of σ and on m, respectively as

$$E^e_\sigma t_m \equiv \sum_\pi p^e_{\pi\sigma} t_{\pi m} \quad E^e_\sigma s_m \equiv \sum_\pi p^e_{\pi\sigma} s_{\pi m} \quad E^e_\sigma y_{m(\sigma)} \equiv \sum_\pi p^e_{\pi\sigma} y_{\pi\sigma m} \quad \forall e, \sigma, m$$

2.2 Contracts

Grand-Contract. We give full bargaining power to P. We denote M_S and M_A the set of messages that, respectively, S and A can send to P. A grand-contract GC is a collection of the two message spaces, M_S and M_A, and two functions defined as:

[15] Since m is not stochastic, $p^e_{\pi\sigma m}$ is invariant with respect to it. This is why we drop the index.

1. The transfer to A, t: $\pi \times M_S \times M_A \rightarrow \mathbb{R}_+$,

2. The transfer to S, s: $\pi \times M_S \times M_A \rightarrow \mathbb{R}_+$.

We denote m_A and m_S generic elements belonging to, respectively, M_A and M_S and $m = (m_S, m_A)$. Throughout we denote $t_{\pi m}$ and $s_{\pi m}$ the transfers made to A and S for a given realization of the profits π and a given pair of messages m.

Side-Contract. We assume S has full bargaining power in designing side-contracts (due to her superior position in the hierarchy). A side-contract is a mapping from the profits π, the realization of the signal σ, the two message spaces M_S and M_A specified in GC and the internal message space $\tilde{M}_A$ to three functions defined on the product of these spaces:

1. The side-transfer, y: $\pi \times \sigma \times M_S \times M_A \times \tilde{M}_A \rightarrow \mathbb{R}$,

2. S's reporting strategy, m_S: $\pi \times \sigma \times M_S \times M_A \times \tilde{M}_A \rightarrow M_S$,

3. A's reporting strategy, m_A: $\pi \times \sigma \times M_S \times M_A \times \tilde{M}_A \rightarrow M_A$.

The *Revelation Principle* holds at the side contracting stage. Our focus is on the equilibria in which S opts for a direct (side) mechanism. Throughout we denote $y_{\pi\sigma m}$ the side-transfer for a given realization of profits π, private signal σ, and a pair of messages m.[16] In each state, money can flow from or to A, but him and S may not exchange more than what is specified in GC, i.e., $-t_{\pi m} \leq y_{\pi\sigma m} \leq s_{\pi m}$ for $\forall \pi, \sigma, m$. The commitment on m_i in the side-contract, where $i = S, A$, is credible if and only if the report is interim rational given transfers.[17]

Different cases are considered in turn. We first assume, in line with existing literature, that S can design her side-contract only *ex post* (we denote the ensuing side-contract by $\bar{SC}$), i.e., only at the time she observes the realization of signal σ, after A has chosen $e \in \{\underline{e}, \bar{e}\}$. In a following section, we let S design her side-contract also *ex ante*, that is, before A has chosen his effort. We denote the *ex ante* side-contract by $\underline{SC}$.[18]

[16] To save on notation we do not write side-transfers as being explicitly contingent on the side message $\tilde{m}_A$ as this plays a (very) limited role in the analysis.

[17] S can commit only to side-transfers, which, as specified above, are contingent both on m and σ. It is therefore by committing to a given set of side-transfers that S determines the interim rationality of sending a given report for a given σ. This assumption is in line with existing literature. If S and A could directly commit to any m results would be almost identical.

[18] Renegotiation is ruled out by assumption here. It is however fairly straightforward to show that if we allowed for $\underline{SC}$ to be collectively renegotiated at stage 4 results would be identical.

2.3 Timing

The sequence of events is as follows. Note that communication between the players may occur at any time during the game; we thus do not write it explicitly in the timing.

1. P offers a grand-contract GC to S and A.
2. S offers side-contract $\underline{SC}$ to A. We refer to this as the "*ex ante*" side contracting stage.
3. A chooses e.
4. π and σ are realized. If no $\underline{SC}$ was agreed upon, S offers side-contract $\bar{SC}$ to A. We refer to this as the "*ex post*" side contracting stage.
5. Communication ends, i.e., reports $m = (m_S, m_A)$ are complete.
6. GC and either $\underline{SC}$ or $\bar{SC}$ are executed.

As mentioned above we allow for two distinct side contracting stages. The first takes place before A has chosen his action (Stage 2). We refer to its result as an *Ex Ante* Side-Contract ($\underline{SC}$). The second takes place after A has chosen his action, the supervisor has gathered information on A's behavior and σ is realized (Stage 4). We refer to its result as an *Ex Post* Side-Contract ($\bar{SC}$). $\bar{SC}$ is relevant only if $\underline{SC}$ has not been agreed upon. Both A and S can also decide not to participate to any of the two side contracting rounds, in which case they play GC non cooperatively.[19][20]

3 Solving the Model

Our main focus is on the study of the optimal organization of supervisory activities under different assumptions on the timing of side contracting and on the nature of the supervisor's information. First, we provide three benchmark GCs that will be useful references for the analysis to follow. Then, we consider the optimal GC if side contracting is allowed only at Stage 4. Finally, we allow for side contracting to occur both at Stage 2 and Stage 4.

[19] Note also that here S's signal σ realizes at the same time as public signal π does. Reports concerning σ thus cannot physically occur before knowing π. An equally plausible set-up could have π realize after σ. We looked at this alternative scenario in an earlier version of the paper and the main results were qualitatively unchanged.

[20] Renegotiation is ruled out by assumption here. It is however fairly straightforward to show that if we allowed for $\underline{SC}$ to be collectively renegotiated at stage 4 results would be identical. We do not include the proof to save on space.

3.1 Benchmarks

In this subsection, when relevant, we restrict communication to be exclusively with S, and concerning only σ. This assumption is without loss of generality and made for expositional clarity only.

No supervision. Suppose P does not hire S. In that case, $\sigma = N$ always. It is then straightforward to show that the best P can do is to offer a contract specifying a positive transfer to A, equal to $\frac{\psi}{(2\rho_\pi - 1)}$, if and only if $\pi = \bar{\pi}$. Total expected payments are then $E(z)^{SB} = \frac{\rho_\pi \psi}{(2\rho_\pi - 1)}$. We refer to this benchmark as the second-best (SB) contract. Obviously, in no circumstances will P hire S and design a GC such that $E(z) \geq E(z)^{SB}$.

Opportunism-free supervision. Suppose now that S is *benevolent*, i.e., always reports information truthfully. Both side contracting stages are void, regardless of whether σ is verifiable or not. The incentive compatibility constraint at A's level then takes the form

$$\sum_{\sigma} E_{\sigma}^{\bar{e}} t_{\sigma} - \psi \geq \sum_{\sigma} E_{\sigma}^{\underline{e}} t_{\sigma} \tag{1}$$

An optimal way of inducing A into exerting high effort involves setting $t_{\bar{\pi}G} = \frac{\psi}{\rho_\sigma \rho_\pi}$ and all other transfers to A and S to zero. We refer to the contract just described as *Grand Contract* GC^I. Not surprisingly, since a perfectly informative signal on e is available at no cost, GC^I allows P to reach the first-best level of expected payments. We have $E(z)^I = E(z)^{FB} = \psi$.

Verifiable supervisor information. Suppose S is no longer benevolent but *opportunistic*: i.e., she can engage in side contracting with A and make the report m_S that maximizes her payoff. Suppose also that her information σ is *verifiable* (i.e. hard). Then, following Kessler [2000], we have

Lemma 1. *When S is opportunistic and σ is verifiable (i.e. hard) information, the opportunism-free GC^I is optimal. The supervisor's opportunism is of no consequence to the organization.*

Proof. See Kessler [2000]. □

When information is hard, the only way in which S can exploit her discretionary power is by hiding evidence of good or bad behavior by A. P can optimally neutralize this threat by promising payments to the latter only when reporting good news. Then, since G cannot be forged, there is no scope for collusion. This also eradicates any possibility for S to threaten A with unfavourable reports when the latter is working hard (and ask for extorsive payments).[21] Importantly, this result holds regardless of the timing of collusion, that is, regardless of whether the side-contract is struck before or after the action e.

3.2 Non-verifiable information and opportunistic supervisor with only *ex post* side contracting

We now consider a situation in which S is opportunistic *and* her information is entirely *non-verifiable* (i.e. soft). Assume, for the moment, that side contracting occurs only *ex post* (i.e., Stage 2 is void). It can easily be shown that, in this scenario, making use of S's information never allows P to reduce expected payments below $E(z)^{SB}$.

Lemma 2. *If the supervisor's information σ is non-verifiable and she can sign side-contracts with the agent only after the latter has chosen his level of effort e (i.e., Stage 2 is void), the optimal contract is the Second-Best contract.*

Proof. Suppose Stage 2 to be void: S and A cannot agree on $\underline{SC}$. Suppose further to be at Stage 4. If A rejects $\bar{SC}$, messages are sent to P non cooperatively. Denote l the pair of messages $m_A \in M_A$ and $m_S \in M_S$ sent in the equilibrium of such a non-cooperative reporting game (the equilibrium guaranteeing the lowest payoff to A). By the "Markov Principle" to which we adhere (see Section 2), the equilibrium of this subgame is independent of the actual realization of σ (since the realization of π has already been observed).[22]
When designing $\bar{SC}$, for a given l and a given π, S solves

$$\max_{\{m, y_{\pi m}\}} \{s_{\pi m} - K(y_{\pi m}) \cdot y_{\pi m}\} \quad \text{s.t.}$$

$$s_{\pi m} \geq y_{\pi m} \geq t_{\pi m}, \tag{2}$$

$$t_{\pi m} + y_{\pi m} \geq t_{\pi l}. \tag{3}$$

[21]Any possible extorsive threat can be undone by promising S an infinitely small extra payment when reporting G compared to when reporting N (the optimal payment being zero in that case).

[22]Assuming instead that S and A coordinate on different equilibria depending on payoff irrelevant information is in fact tantamount to assuming partially verifiable information, which is not in the spirit of our analysis.

(2) ensures that $\bar{SC}$ is feasible and (3) ensures participation by A. The chosen pair of messages m and side-transfers $y_{\pi m}$ are invariant with respect to the true state of the world (note that S could be indifferent between several such pairs) and, therefore, so are payoffs. Consequently, it is useless for P to make transfers contingent on her communication with S and A. The best it can achieve is the second-best, by exploiting the public signal π. □

As previously conjectured by Tirole [1986] and, more recently, KLY [2010], non-verifiability of information may make the supervisor's presence useless. The intuition is as follows. First, since S wishes to extract as much as possible from her relationship with A, she credibly commits, in case A rejects $\bar{SC}$, to the reporting strategy that leads to the equilibrium with the lowest possible payoff to A, so as to relax (3). Such a reporting strategy is invariant with private information σ since information is soft. Second, when S and A side contract, they choose their reporting strategies in a way that is independent of private information σ because, once again, information is fully manipulable and because A's outside option is also invariant with σ. It follows, then, that in such a situation P cannot hope to make any use of S's presence since she systematically receives the same messages. Note, finally, that this result is independent of the actual value of the inefficiency of side contracting k.

As a result S is simply not hired and left with her reservation utility. She is, in a sense, a victim of her own opportunistic behavior. If she were able to commit otherwise, the principal may find a way to usefully integrate her into the organization. However, this conclusion is obtained under the implicit assumption that the supervisor does not internalize the consequences of her opportunism. This follows from the fact that side contracting can happen only once the agent has chosen his action. At that point, the supervisor has, trivially, no influence on such a choice. As we will show below, this possibility instead exists if A and S can side contract before the former has chosen his action. This may radically change the optimal incentive scheme.

3.3 Non-verifiable information and opportunistic supervisor with both *ex ante* and *ex post* side contracting

We now study the situation in which S can side contract with A *both before and after* the latter has decided his effort level e. That is, side contracting can happen both at Stage 2 and 4. In particular, if $\underline{SC}$ is not agreed upon—either because A rejected it or because S didn't offer it, $\bar{SC}$ may still be agreed upon at Stage 4.

We formally show in the appendix that it is without loss of generality for P (i) to communicate only with S and (ii) exclusively concerning σ, that is, $M_S = \{G, N, B\}$ and $M_A = \emptyset$. When commenting on the results in this section, we take this as given.

We solve this game by backward induction. First, we compute S's optimal design of a side-contract $\underline{SC}$, given a grand-contract GC. Secondly, we solve for the optimal GC given S's best reaction.

Supervisor's problem. For a given GC, S reacts by designing an $\underline{SC}$. Any given $\underline{SC}$ induces a choice by A of either $\bar{e}$ or $\underline{e}$. There is thus no loss of generality in focusing only on the $\underline{SC}$ that, for a given action, guarantees S the highest expected payoff, given transfers from P and side-transfers that may be exchanged with A. We denote the side-contract inducing high effort by $\underline{SC}_{\bar{e}}$, and the one inducing low effort by $\underline{SC}_{\underline{e}}$.

Suppose S wishes to induce A into choosing action $e = e$ instead of e', with $e, e' \in \{\underline{e}, \bar{e}\}$. She then chooses a schedule of side-transfers $\boldsymbol{y}$ and of messages $\boldsymbol{m}$ to solve the following program:

$$\max_{\{\boldsymbol{m},\boldsymbol{y}\}} \sum_{\sigma} \left\{ E_\sigma^e s_{m(\pi,\sigma)} - E_\sigma^e K\left(y_{m(\pi,\sigma)}\right) y_{\sigma m(\pi,\sigma)} \right\} \quad \text{s.t.}$$

$$\sum_{\sigma} E_\sigma^e \left(t_{m(\pi,\sigma)} + y_{m(\pi,\sigma)}\right) - \psi(e) \geq \sum_{\sigma} E_\sigma^{e'} \left(t_{m(\pi,\sigma)} + y_{\sigma m(\pi,\sigma)}\right) - \psi(e'), \tag{SIC}$$

$$\sum_{\sigma} E_\sigma^e \left(t_{m(\pi,\sigma)} + y_{m(\pi,\sigma)}\right) - \psi(e) \geq \underline{U}, \tag{SPC}$$

$$s_{\pi m(\pi,\sigma)} \geq y_{\pi\sigma m(\pi,\sigma)} \geq -t_{\pi m(\pi,\sigma)} \qquad \forall \{\pi, \sigma\}. \tag{LL}$$

We denote $\underline{U}$ the outside option of A when rejecting $\underline{SC}$, i.e., the expected payoff (net of side-transfers paid to S) in the continuation game in which side contracting takes place only at Stage 4.[23] S must ensure that A prefers the intended effort level e (constraint (SIC)). She also has to make sure that participating in the *ex ante* side-contract makes A better off (SPC) than simply postponing side contracting to Stage 4. Finally, by limited liability, side-transfers cannot be larger than the payments promised by P. We denote by $E^e(s - y)$ the expected payoff enjoyed by S when offering the most profitable $\underline{SC}_e$.[24]

[23] We do not write it explicitly here as it varies non-continuously with the payments designed in the GC.

[24] For expositional clarity we have also omitted constraints ensuring that the reporting function $m(\pi, \sigma)$ be interim rational. It is however trivial that S is capable of finding a schedule of out-of-equilibrium transfers ensuring interim rationality of any such reporting function. Note, in addition, that even if equilibrium side-transfers are zero, the $\underline{SC}$ is not necessarily null as out-of-equilibrium payments could be playing an important

Principal's problem. P's problem is that of designing a grand-contract GC such that S proposes A an $\underline{SC}$ inducing high effort $\bar{e}$, while minimizing expected payments

$$\min_{\{\boldsymbol{t},\boldsymbol{s}\}} E(z) = \sum_{\sigma} E_{\sigma}^{\bar{e}} \left(s_{m(\pi,\sigma)} + t_{m(\pi,\sigma)} \right) \quad \text{s.t.}$$

$$E^{\bar{e}} \left(s - y \right) \geq E^{\underline{e}} \left(s - y \right). \tag{4}$$

That S designs $\underline{SC}_{\bar{e}}$ rather than $\underline{SC}_{\underline{e}}$ is ensured by (4). When looking at S's problem, it is immediate that (4) makes incentive compatibility and participation constraints at A's level redundant. As long as it holds, S will, *if necessary*, top up the transfers promised to A by P in order to have him choose $\bar{e}$, and it is can then only be optimal for P to give enough resources to S to do so, for otherwise GC fails to induce A into choosing $e = \bar{e}$. In such instances S is said to have payroll authority. Naturally, there could be several ways to optimally organize this hierarchy, with varying degrees of payroll authority delegated to S. We show below that *the optimal organizational form is a centralized one.*

The following Proposition describes the optimal GC, solution the maximization problem presented above. In the following, we have that $k^1 = \frac{\rho_\sigma(1-\rho_\pi)+(2\rho_\pi-1)}{\rho_\sigma(1-\rho_\pi)} < k^2 = \frac{\rho_\sigma(1-\rho_\pi)+2(2\rho_\pi-1)}{\rho_\sigma(1-\rho_\pi)}$.

Proposition 1. *When the supervisory signal σ is non-verifiable and S can sign side-contracts with A before the latter has chosen his effort level (i.e. at Stage 2), making use of the supervisor's information is strictly optimal for the organization. In particular, the optimal grand-contract is GC^{II} is such that*

(i) if either "$k^1 < k < k^2$ and $\rho_\sigma < \frac{(2\rho_\pi-1)}{\rho_\pi}$" or "$k \geq k^2$" then

$$t_{\bar{\pi}G} = \frac{\psi}{\rho_\sigma\rho_\pi}, \qquad s_{\bar{\pi}G} = \max\left(\left(\frac{\rho_\sigma}{(2\rho_\pi - 1)} - \frac{1}{\rho_\pi} \right) \frac{\psi}{\rho_\sigma k}, 0 \right) < s_{\bar{\pi}N} = s_{\bar{\pi}G} + \frac{t_{\bar{\pi}G}}{k}, \qquad s_{\bar{\pi}B} \in [0, s_{\bar{\pi}N}],$$

and all other transfers are set to zero, and

(ii) otherwise then

$$\begin{aligned} t_{\bar{\pi}G} =& t_{\bar{\pi}N} = \frac{\psi}{\rho_\pi - (1-\rho_\sigma)(1-\rho_\pi)}, \quad t_{\bar{\pi}B} = (1-\rho_\sigma)\, t_{\bar{\pi}G}, \\ s_{\bar{\pi}G} =& s_{\bar{\pi}N} = s_{\bar{\pi}B} = \frac{1}{k}\left(\frac{\psi}{2\rho_\pi - 1} - \frac{\psi}{\rho_\pi - (1-\rho_\sigma)(1-\rho_\pi)} \right), \end{aligned}$$

role. They could, for example, allow S to credibly commit to reporting G, should it turn up, without extorting A.

and all other transfers are set to zero.

The total expected payment is such that $E(z)^{II} < E(z)^{SB}$ *for* $\forall k > 1$.

Proof. See Appendix. □

As stated in Lemma 2, when S is not benevolent and in the absence of *ex ante* side contracting, P has no choice but not to make use of her non-verifiable information. However, when the possibility of *ex ante* side contracting is considered, the landscape for the optimal incentive scheme radically changes: S becomes useful. In a nutshell, this is because P is now able to exploit S's commitment ability by interesting her in the outcome of the project.

Intuition. Let us comment on the optimal GC. To convey the intuition behind Proposition 1, it is simpler for now to take it as given that the optimal way to organize the hierarchy is such that P does not delegate any payroll authority to S (i.e., that P does not rely on S making side-transfers to A in equilibrium). We provide the intuition for this second set of results below.

Note that GC^* may be of two types—one more high powered than the other—depending on the value of the inefficiencies of side contracting k and the precision of S's information ρ_σ. We comment on both types in turn. In either grand-contract, however, all transfers are set to zero when profits are low ($\pi = \underline{\pi}$). Indeed, P finds it optimal to do so as (i) low profits are indicative of A not having exerted high effort and (ii) S *is to be made interested in the outcome of the project.*

Consider the high powered grand-contract first. In this contract, A's incentive scheme is the same as in the "benevolent supervisor" benchmark. P is able to do so because it can both exploit S's commitment ability and the inefficiencies of side-transfers. To see this, note that because $t_{\bar{\pi}G} > t_{\bar{\pi}N} = t_{\bar{\pi}B} = 0$ the $S - A$ coalition has incentives to systematically report good news. If it was to do so (i.e., if $t_{\bar{\pi}G} = \frac{\psi}{\rho_\sigma \rho_\pi}$ was paid by P whenever profits are high) S would let A pocket $t_{\bar{\pi}G}$ only when $\sigma = G$ and otherwise pocket it (and thus receive $\frac{\psi}{k\rho_\sigma \rho_\pi}$). However, it is immediate that P can do better by setting $s_{\overline{\pi}N}$ at least equal to $s_{\overline{\pi}G} + \frac{\psi}{k\rho_\sigma \rho_\pi}$ and reduce expected equilibrium payments by $\frac{\psi}{\rho_\sigma \rho_\pi}\left(\frac{k-1}{k}\right)$. Of course by setting $s_{\overline{\pi}N} > s_{\overline{\pi}G}$, S may then be tempted to systematically report N (and engage in a behavior akin to extortion). This however is not profitable to S, as she would then have to pay A out of her own pocket (since $t_{\bar{\pi}N} = 0$) to have him work hard. Thus, her desire to have A work hard removes any incentive S may have in suppressing good news, i.e., *in engaging in extortion.*

Finally, to have S prefer having A work hard, attention must also be paid to her payoff when letting shirk happen. It is useful to have in mind that in these instances, money then flows from the latter to the former.[25] Thus, when engaging in collective opportunism, S finds it optimal to send message $m_S \in \{G, N, B\}$ with the highest associated term $s_{\overline{\pi} m_S} + \frac{t_{\overline{\pi} m_S}}{k}$, which here means sending message N. When ρ_σ is low (but not too low, see below), the payment $s_{\overline{\pi} N}$ is made often in equilibrium and is sufficient in providing the right incentives to S. Contrarily, when ρ_σ is high, and payment $s_{\overline{\pi} N}$ is made rarely, P has no choice but to set $s_{\overline{\pi} G} > 0$ also. To conclude, in the high powered grand-contract, A is left with no rent and S is left with a rent which decreases with the transaction costs k and increases with the precision of her information ρ_σ.

When k is low and/or ρ_σ is high, the payment $s_{\overline{\pi} N}$ is made often and thus offering a very high powered scheme to A becomes very costly to P. Consequently, P prefers offering a flatter incentive scheme; this is the second shape that GC can take. When opportunism is very costly to deter, P prefers to rely less on S's information so as to lower the incentives the $S - A$ coalition may have in manipulating information: P does not distinguish between good and no news. Note, however, that S's information is still of some use. Because it is in S's best interest to have A work hard, P can essentially count on S to punish A whenever she observes $\sigma = B$ by acting as a “bounty hunter”.

Recognizing the strong impact of the S's opportunism on A's incentives to work hard, P optimally stimulates her interest in the successful outcome of the project she is supervising. With only *ex post* side contracting *S's* role is simply to reduce the information gap between the top and the bottom of the organization. Instead, when *ex ante* side contracting is considered, she is given, *a priori*, a chance to intervene directly (through side-agreements) on A's incentives. As we will discuss more in detail in a following section, it is actually never the case that P lets her make transfers to A in equilibrium. However, the mere fact of giving S such a role allows to discipline her opportunism, inducing her to cooperate with A in the best interest of the organization. With *ex ante* side contracting S becomes, in a sense, a “second principal.” It is in her own interest to commit not to behave opportunistically vis-à-vis A. However, since P cannot contract directly on the side-contract that S may propose to A, she is left with a rent *in equilibrium.*

Note, finally, that our results would be qualitatively the same if we had adopted the information structure considered by KLY [2010]. They consider supervisory information that

[25] This must be true for otherwise S could replicate the same reporting strategies as the ones in an equilibrium in which money flows in the opposite direction, set all side transfers to zero, and be strictly better off.

is soft only under collusion, and otherwise hard. In that case, the S's opportunism is less incisive than with fully soft information. Indeed, using such an information structure, when side contracting takes place only in Stage 4 (which is the scenario studied by KLY), the organization optimally uses the S's information, but only to a partial extent. In contrast, when *ex ante* side-contracts can be signed, then the optimal contract is the same as the one we described in Proposition 1.

Payroll Authority. We now comment as to why, as anticipated above, it is strictly suboptimal to rely, in equilibrium, on S making payments to A. As the results presented in Proposition 1 suggest, an effective way of coping with opportunism is to make S internalize the adverse consequences of her misbehavior. This, we have shown, is possible only if she is is appointed *ex ante*. Building on this intuition, the delegation of payroll authority—P delegates to S the task of contracting (partially or fully) with A—constitutes perhaps the strongest manner by which to make S an essential input in the production process of the organization. The success of the project indeed then entirely rests on the her shoulders. Limited liability, however, slows down the extent to which S can be made to internalize her behavior. Non-verifiability of information, finally, allows S to inflate payroll expenses. In this subsection, we investigate which of these forces dominates. To keep the comparison between both forms of contracting on an equal foot we continue assuming that side-transfers involve a transaction cost k. This is relaxed below.

Corollary 1. *It is strictly suboptimal to delegate payroll authority to S.*

Proof. See Appendix, Lemma 7. □

Intuition. Several factors lead to the fact that it is strictly optimal for P to retain all payroll authority. First, and ignoring the scope for information manipulation, transaction costs k are such that P is better off paying directly any amount of money to A rather than paying k times this amount to S to have her pay the amount to A. Finally, because of both information manipulation and because S may find it profitable to let A shirk, a rent must be given up to S. Consequently, to have A receive a dollar through S, the latter must receive a dollar inflated by the importance of the information rent. In other words, a double marginalization of rents occurs and is such that, independently of transaction costs k, P is better off making payments directly to A.

We comment on related literature in the following subsection.

3.4 Arm's length relationships

We now briefly investigate whether an arm's length relationship—P contracts with S who in turn contracts with A—may be an optimal way of organizing the hierarchy *when side-transfers do not involve transaction costs*. The same factors cited above in favor of delegating payroll authority are still present, to which need to be added new ones. First, the scope for collusion aiming at raising the agent's rent disappears as the latter's transfers are, by definition, set to zero. Second, the principal may hope to benefit better from the supervisor's superior information: under delegated payroll authority the agent is offered by the supervisor the most efficient employment contract and, consequently, the amount needed to reimburse the latter's payroll expenses is (potentially) low.

The factors going against the delegation payroll authority cited above are also still present (limited liability and non-verifiability of information). However, the pervasiveness of the information manipulation increases. Because transfers to the agent by the principal are set to zero, the consequences on the former's payoff of the supervisor's opportunism are less important than in a centralized hierarchy. This, in turn, tends to attenuate the internalization by the supervisor of the detrimental consequences of information manipulation. In this subsection we investigate which of these forces dominate.[26]

Proposition 2. *An arm's length relationship is strictly suboptimal.*

Proof. Consider a given GC specifying transfers $s_{\pi m}$ to S. S reacts by designing $\underline{SC}$, specifying (i) a reporting strategy m^* and (ii) a schedule of side-transfers $\boldsymbol{y}^*$. In such an arrangement it is optimal for S to send the message with the highest associated payment, regardless of the actual value of σ and regardless of the action e induced by $\underline{SC}$. Indeed, all transfers to A being set to zero, this represents the Pareto optimal strategy. It follows that it is useless for P to elicit the value of σ and consequently transfers to S are contingent only on the profits π.

Since P wishes S to induce $e = \bar{e}$ it is optimal to reward S with a positive transfer if and only if $\pi = \bar{\pi}$. To continue, because S's cost of inducing high effort need to be reimbursed, where these equal at least disutility ψ, it follows immediately that P's expected salary cost under delegation is at least $\frac{\rho_\pi \psi}{2\rho_\pi - 1}$, i.e., weakly worse than a situation without the presence of the supervisor, and consequently weakly worse than a situation in which S is appointed *ex ante* but P retains payroll authority. $\square$

[26]We here shut down the communication channel from P to A for notational purposes. If this assumption was relaxed, results would be qualitatively identical.

Intuition. Because, by definition, transfers to A from P are set to zero, individual opportunism by S does not impact directly A's payoff. Regardless of the report made by S, A receives a zero transfer from P. What matters to A are the side-transfers promised by A. Bearing this in mind, A finds it optimal to make the report with the highest associated (expected) payment. Her information then however becomes useless and transfers are contingent only the profits π. The best P can guarantee itself is the second best payoff, which is weakly lower than than under *ex ante* side contracting but with a centralized grand-contract. This result highlights the fact that for S to internalize the effect of opportunism on A's incentives, it must necessarily be the case that the grand-contract specifies positive transfers to the latter.

Related literature. In moral hazard set ups with side contracting between (risk averse) agents performing almost identical productive tasks Holmstrom and Milgrom [1990] and Itoh [1993] show that some degree of delegation may be optimal, so as to foster cooperation. In the former case this is due to better risk sharing between agents, while in the latter mutual monitoring constitutes the main reason. In a similar set-up to theirs, but under risk neutrality, Baliga and Sjostrom [1998] show that the optimal centralized contract can be implemented through decentralization.[27] Although in our model the coalition also enjoys information not shared with the principal, the scope for opportunism is such that centralization is strictly optimal. To continue, our result is also in contrast to principal-supervisor-agent set ups under *adverse selection*. In FGLM [2003], for instance, delegation is one way to implement the optimal (collusion proof) centralized mechanism. The difference lies in P's ability to exploit the inefficiencies between S and A due to information asymmetries concerning the latter's type. In a similar environment, but under a different structure of types, Celik [2009] shows that delegation is strictly suboptimal. As in our model, he finds that delegation is so detrimental that P would in fact be better off by not hiring the supervisor.

3.5 Optimal design of supervisory activities: monitoring versus auditing

Our model can also be used to study the optimal design of verification procedures. The objective of this section is to compare two alternative organizations: borrowing the terminology

[27] Macho-Stadler and Perez-Castrillo [1998] investigate issues of delegation when the principal suffers from commitment issues and provide conditions under which a decentralized structure is equivalent to a centralized one with collusion. Also, Che [1995] and Alger and Ma [2003] study models in which tolerating some collusion is best.

used by Strausz [2005], we will compare an organization in which S is a *monitor* to one in which she is an *auditor*. They differ in the timing of verification procedures.

Monitoring: S is hired to check on A just before the latter begins her task. The timing of the game is as described in Section 2.3.

Auditing: S is hired (and her report produced) only once A's work is concluded. The timing of the game is as described in Section 2.3, except that Stage 2 is void.[28]

The comparison between the two forms of verification turns out to depend crucially on the nature of information. We have

Proposition 3. *If the supervisor's information is verifiable, the principal is indifferent between having her perform monitoring or auditing functions. Instead, if information is non-verifiable, the supervisor should perform monitoring functions.*

Proof. In the case of verifiable information, treated in Lemma 1, the optimal GC does not depend on the timing of side contracting. Indeed, the opportunism-free GC^{I} is always optimal. In the case of non-verifiable information, Lemma 2 shows that, if auditing is adopted, supervisory information is useless. When adopting, instead, monitoring as a verification procedure, P can implement GC^{II} and make use of supervisory information in an effective way (see Proposition 1). As a consequence, monitoring is the preferable procedure. $\square$

Our results suggest that, in a setup in which the principal has to delegate the verification activity to a non-benevolent third party, the choice between monitoring and auditing depends on the nature of supervisory information. When it is verifiable, the two forms of verification are equivalent, since S's opportunism is never relevant. When it is not, it turns out that monitoring is always optimal. This follows from the findings of Section 3.3. When auditing is adopted, P waits until A has chosen her action before asking S to verify it. This, as we have seen above, implies that S cannot internalize the consequences of her opportunism for the agent's incentives. When monitoring is chosen, instead, S verifies A's actions while he is still working on the task assigned. This might increase the chances of collusion between the two, but also implies that S has to take into account the consequences of her behavior.

[28] In our set-up, the supervisor‘s intervention is optimally requested independently of the realization of π. This is because monitoring is not costly and π can take both high and low values irrespectively A's effort. As a consequence, the optimal contract with auditing would not change if the principal waited until the agent‘s action has generated publicly available results before hiring the supervisor.

In a different set-up, Strausz [2005] points out that the organization of verification activities (concerning an agent's unobservable action) depends on the P's ability to commit to costly effort to gather information. In his model, however, P carries out the verification itself, so there is no supervisor involved. We study a complementary situation in which supervision has to be delegated by P to a third player. To S, verification is effortless, but she has to be provided incentives to report truthfully. Our findings are, indeed, quite different from those of Strausz.

The results of this section run somewhat against existing literature that, instead, typically advocates appointing supervisors for short periods of time.[29] This is thought to make collusion more difficult, chiefly because supervisor and supervisee do not have time to develop the kind of relationships able to sustain sophisticated informal agreements. What our result suggests is that by reducing the factors that are thought to facilitate collusion one necessarily also reduces those facilitating cooperation. In a similar set-up to ours, Hiriart, Martimort and Pouyet [2011] let supervision/collusion occur both before and after a possible (say environmental) accident and show that it is preferable to appoint two distinct supervisors so as to reduce the pervasiveness of the first round of side contracting. The drastic difference between our results lies essentially in that, contrarily to us, both their side contracting stages occur once the agent has chosen his hidden action.[30]

4 Conclusion

We have investigated the impact of informal agreements on the functioning of organizations, using a model based on a three-tier-hierarchy framework (in the spirit of Tirole's [1986] seminal paper), with a moral hazard problem at the bottom and soft supervisory information. We have highlighted the importance of the timing of appointment of supervisors to contrast the pervasiveness of opportunistic behavior. Our results suggest that allowing supervisor and supervisee(s) to side-contract before the latter chooses his action can be beneficial to the organization. The optimal incentive scheme provides group-based incentives, thus contrasting collective opportunism, while also eradicating individual opportunism. This is because if the supervisor is interested in the positive outcome of the agent's task, then she has no incentives to act opportunistically vis-à-vis the latter. This has implications for the optimal design of verification activities: in particular, monitoring is found to be, in the presence of soft

[29] A large part of this literature also advocates splitting information gathering tasks amongst several supervisors (see for instance Kofman and Lawarée [1993] and Laffont and Martimort [1998, 1999]).

[30] Also, in their model, reports are verifiable and extortion is ruled out.

supervisory information, always superior to auditing. Finally, we studied delegation of payroll authority to the supervisor. Our results suggest that, with soft information, this is not optimal.

References

[1] Alger, I. and Albert Ma, Ching-to. "Moral hazard, insurance, and some collusion." *Journal of Economic Behavior & Organization*, Elsevier, vol. 50(2), pages 225-247, February 2003.

[2] Angelucci, C. and Han, M.. “Private and Public Control of Management.” Mimeo, Amsterdam, 2012.

[3] Baliga, S. and Sjostrom, T.. "Decentralization and Collusion." *Journal of Economic Theory*, Elsevier, vol. 83(2), pages 196-232, December 1998.

[4] Baliga, S.. "Monitoring and Collusion with "Soft" Information." *Journal of Law, Economics and Organization*, Oxford University Press, vol. 15(2), pages 434-40, July 1999.

[5] Celik, G.. "Mechanism design with collusive supervision." *Journal of Economic Theory*, Elsevier, vol. 144(1), pages 69-95, January 2009.

[6] Chandler, A., 1962, Strategy and Structure, MIT Press, Cambridge.

[7] Che, Y.-K.. “Revolving doors and the optimal tolerance for agency collusion.” *RAND Journal of Economics*, Vol. 26, No. 3, Autumn 1995 pp. 378-393.

[8] Dalton, M., 1959, Men who Manage, Wiley and Sons, New York.

[9] Faure-Grimaud, A., Laffont J.-J. and Martimort, D.. "The endogenous transaction costs of delegated auditing." *European Economic Review*, Elsevier, vol. 43(4-6), pages 1039-1048, April 1999.

[10] Faure-Grimaud, A., Laffont, J.-J. and Martimort, D.. "Collusion, Delegation and Supervision with Soft Information." *Review of Economic Studies*, Blackwell Publishing, vol. 70(2), pages 253-279, 04, 2003.

[11] Fevre, R., Lewis, D., Robinson, A. and Jones, T.. “Insight into Ill-Treatment in the Workplace: Patterns, Causes and Solutions.” *ESRC report*, 2011.

[12] Gibbons, R.. “Incentives between firms (and within).” *Management Science* 51(1), 2–17, 2005.

[13] Gibbs, M., Merchant, K. A., Van der Stede, W. A., and Vargus, M. E.. "Determinants and effects of subjectivity in incentives." *Accounting Review* 79(2), 409–436, 2004.

[14] Hiriart, Y., Martimort, D., and Pouyet, J.. "The public management of risk: Separating ex ante and ex post monitors." *Journal of Public Economics*, Elsevier, vol. 94(11-12), pages 1008-1019, December 2011.

[15] Holmström, B. and Milgrom, P.. "Regulating Trade Among Agents." *Journal of Institutional and Theoretical Economics*. Vol. 146, pp. 85-105, 1990.

[16] Itoh, H.. "Coalitions, Incentives, and Risk Sharing." *Journal of Economic Theory*, Elsevier, vol. 60(2), pages 410-427, August 1993.

[17] Kessler, A. S.. "On Monitoring and Collusion in Hierarchies." *Journal of Economic Theory*, Elsevier, vol. 91(2), pages 280-291, April 2000.

[18] Khalil F., Lawarrée J. and Yun S.. "Bribery versus extortion: allowing the lesser of two evils." *RAND Journal of Economics*, RAND Corporation, vol. 41(1), pages 179-198, 2010.

[19] Kofman, F. and Lawarrée, J.. "Collusion in Hierarchical Agency." *Econometrica*, Econometric Society, vol. 61(3), pages 629-56, 1993.

[20] Laffont, J.-J. and Tirole, J.. "Cost padding, auditing and collusion." *Annales d'économie et de Statistique*. n.25/26, pp. 205-226, 1992.

[21] Laffont, J.-J. and Martimort, D.. "Collusion and Delegation." *RAND Journal of Economics*, The RAND Corporation, vol. 29(2), pages 280-305, Summer 1998.

[22] Laffont, J.-J. and Martimort, D. "Separation of Regulators Against Collusive Behavior," *RAND Journal of Economics*, The RAND Corporation, vol. 30(2), pages 232-262, Summer 1999.

[23] Laffont, J.-J. and Martimort, D.. "Mechanism Design with Collusion and Correlation." *Econometrica*, Econometric Society, vol. 68(2), pages 309-342, March 2000.

[24] Laffont, J.-J. and Martimort, D.. "The Theory of Incentives: the Principal-Agent model." Princeton, N.J. ; Oxford : Princeton University Press, 2002.

[25] Macho-Stadler, I. and Perez-Castrillo, J. D.. "Centralized and Decentralized Contracts in a Moral Hazard Environment." *Journal of Industrial Economics*, Wiley Blackwell, vol. 46(4), pages 489-510, December 1998.

[26] Maskin, E. and Tirole, J.. "Markov Perfect Equilibrium: I. Observable Actions," *Journal of Economic Theory*, Elsevier, vol. 100(2), pages 191-219, October 2001.

[27] Radner, R.. "Hierarchy: The Economics of Managing." *Journal of Economic Literature*, vol. 30 (Sept. 1992), 1382-1415

[28] Roethlisberger, F. and Dickson, W., 1947. "Management and the Worker." Harvard University Press, Cambridge.

[29] Strausz, R.. "Timing of verification procedures: Monitoring versus auditing." *Journal of Economic Behavior & Organization*, Elsevier, vol. 59(1), pages 89-107, January 2005.

[30] Tirole, J.. "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations." *Journal of Law, Economics and Organization*, Oxford University Press, vol. 2(2), pages 181-214, Fall 1986.

[31] Tirole, J.. "Collusion and the Theory of Organizations." in J. J. Laffont (ed.), Advances in Economic Theory, Vol. 2 (Cambridge: Cambridge University Press), 151-206, 1992.

[32] Vafaï, K.. "Preventing Abuse of Authority in Hierarchies." *International Journal of Industrial Organization,* 20, 1143-1166, 2002.

[33] Vafaï, K.. "Opportunism in Organizations." *Journal of Law, Economics, & Organization,* 26, 158-181, 2012.

[34] Wells, J. T.. "Corporate Fraud Handbook: Prevention and Detection". Second Edition. ACFE, 2011.

Appendix

A Proof of Proposition 1

4.0.1 Preliminary Information and Structure of the Proof

Since S can always offer an empty $\underline{SC}$, we w.l.o.g. restrict our attention to GCs that are such that a $\underline{SC}$ is signed by A.

We begin by stating S's optimization problems. If S designs $\underline{SC_e}$, her problem then is

$$\max_{\{\boldsymbol{m},\boldsymbol{y}\}} \left\{ \sum_{\sigma} E_{\sigma}^{e} \left(s_{m(\sigma)} - K \left(y_{\sigma m(\pi,\sigma)} \right) y_{\sigma m(\pi,\sigma)} \right) \right\} \qquad \text{s.t.}$$

$$\sum_{\sigma} E_{\sigma}^{e} \left(t_{m(\pi,\sigma)} + y_{\sigma m(\pi,\sigma)} \right) \geq \sum_{\sigma} E_{\sigma}^{\bar{e}} \left(t_{m(\pi,\sigma)} + y_{\sigma m(\pi,\sigma)} \right) - \psi, \tag{5}$$

$$\sum_{\sigma} E_{\sigma}^{e} \left(t_{m(\pi,\sigma)} + y_{\sigma m(\pi,\sigma)} \right) \geq \underline{U}, \tag{6}$$

$$s_{\pi m(\pi,\sigma)} \geq y_{\pi\sigma m(\pi,\sigma)} \geq -t_{\pi m(\pi,\sigma)} \qquad \forall \{\pi, \sigma\}. \tag{7}$$

Inequality (5) is the side incentive compatibility constraint, while (6) is the side participation constraint. Finally, (7) ensures that the side-contract is feasible. We comment on $\underline{U}$ below. If S instead designs $\underline{SC_{\bar{e}}}$, her problem then is

$$\max_{\{\boldsymbol{m},\boldsymbol{y}\}} \left\{ \sum_{\sigma} E_{\sigma}^{\bar{e}} \left(s_{m(\pi,\sigma)} - K \left(y_{\sigma m(\pi,\sigma)} \right) y_{\sigma m(\pi,\sigma)} \right) \right\} \qquad \text{s.t.}$$

$$\sum_{\sigma} E_{\sigma}^{\bar{e}} \left(t_{m(\pi,\sigma)} + y_{\sigma m(\pi,\sigma)} \right) - \psi \geq \sum_{\sigma} E_{\sigma}^{e} \left(t_{m(\pi,\sigma)} + y_{\sigma m(\pi,\sigma)} \right), \tag{8}$$

$$\sum_{\sigma} E_{\sigma}^{\bar{e}} \left(t_{m(\pi,\sigma)} + y_{\sigma m(\pi,\sigma)} \right) - \psi \geq \underline{U}, \tag{9}$$

$$s_{\pi m(\pi,\sigma)} \geq y_{\pi\sigma m(\pi,\sigma)} \geq -t_{\pi m(\pi,\sigma)} \qquad \forall \{\pi, \sigma\}. \tag{10}$$

Inequality (8) is the side incentive compatibility constraint, while (9) is the side participation constraint. Finally, (7) ensures that the side-contract is feasible. In either problems, $\underline{U}$ denotes the expected payoff to A when rejecting $\underline{SC}$. It is equal to the expected payoff accruing from $\bar{SC}$ and thus depends on schedules of transfers $\boldsymbol{t}$ and $\boldsymbol{s}$ as well as message spaces M_A and M_S.

Modified Set-up. The rest of the proof is structured as follows. We restrict our attention to GCs that are such that $\underline{SC}$ is agreed upon *but* we solve for a modified optimization problem in which P has one more "degree of freedom" compared to the actual game. In particular, *we assume that* $\underline{U}$ *is a separate choice variable at the disposal of* P. Said differently, we assume that $\underline{U}$ (i) no longer depends on $\boldsymbol{t}$, $\boldsymbol{s}$, M_A and M_S and (ii) is chosen by P. By a simple replication argument, in this modified optimization problem, P can only be better-off.

We then solve for the optimal GC in such a modified set-up and show that the same outcome can be implemented in the actual game in which $\underline{U}$ is *not* a distinct choice variable. This modification is made primarily to simplify the exposition of the proof.[31]

The proof is structured as follows. In Section 4.0.2 we provide lemmas that show that we may w.l.o.g. restrict the scope for communication in this game. Section 4.0.3 instead contains lemmas allowing us to write down P's problem, which we solve in Section 4.0.4. Finally, we show the implementation result in Section 4.0.5.

4.0.2 Scope for Communication

Lemma 3. *It is without loss of generality for* S *to restrict the side message space* $\hat{M}_A$ *to be empty.*

Proof. Once P has offered and committed to GC, specifying message spaces M_A and M_S and schedules of transfers $\boldsymbol{t}$ and $\boldsymbol{s}$, S's problem is identical to a standard principal-agent moral hazard problem in which S acts as a principal inducing actions e and m_A. The Revelation Principle then holds and, as is by now well understood, in this class of games with entirely soft information, there is no scope for communication (see for instance Laffont-Martimort [2002], chapter 4). $\square$

We proceed by assuming that S designs $\underline{SC}$ such that indeed there is no side-communication. Consequently, and this will be of importance later, when designing $\underline{SC}^*$, S makes the reporting strategies m_S^* and m_A^* contingent *at most only on the realizations of* π *and* σ.

In the following, denote $M_S^*(i,\pi) \subseteq M_S$ and $M_A^*(i,\pi) \subseteq M_A$ the message spaces (chosen by S) from which S and A select their messages at communication round i for a given realization of π (if relevant) under $\underline{SC}_{\overline{e}}^*$, and $m_S^*(i,\pi)$ and $m_A^*(i,\pi)$ their generic elements. Similarly, denote $\tilde{M}_S^*(i,\pi) \subseteq M_S$ and $\tilde{M}_A^*(i,\pi) \subseteq M_A$ the message spaces (chosen by S) from which S

[31] Formally, this modification implies that we need not worry about changes in the payoffs associated to the continuation game in which A rejects $\underline{SC}$; these are difficult to keep track of.

and A select their messages at communication round i or a given realization of π (if relevant) under $\underline{SC}_{e}^{*}$, and $\tilde{m}_{S}^{*}(i,\pi)$ and $\tilde{m}_{A}^{*}(i,\pi)$ their generic elements.

Lemma 4. *Suppose P designs a given GC specifying M_S and M_A and such that communication occurs in N rounds. Then it is necessarily the case that $\tilde{M}_{S}^{*}(i,\pi) \subseteq M_{S}^{*}(i,\pi)$ and $\tilde{M}_{A}^{*}(i,\pi) \subseteq M_{A}^{*}(i,\pi)$ for $\forall i, \pi$, where $i = 1, ..., N$ and $\pi \in \{\underline{\pi}, \overline{\pi}\}$.*

Proof. We prove this by contradiction. Suppose that $\tilde{m}_{S}^{*}(i,\pi)$ and/or $\tilde{m}_{A}^{*}(i,\pi)$ do not belong to, respectively, $M_{S}^{*}(i,\pi)$ and $M_{A}^{*}(i,\pi)$ for some i. GC must then be such that in equilibrium the associated (to these messages) pair of transfers is equal to zero so as to optimally relax (SIC), *for P is able to perfectly infer that the coalition deviated.* This, in turn, however implies that the coalition would have been (weakly) better-off opting for reporting strategies such that $\tilde{m}_{S}^{*}(i,\pi) \in M_{S}^{*}(i,\pi)$ and $\tilde{m}_{A}^{*}(i,\pi) \in M_{A}^{*}(i,\pi)$ in the first place, a contradiction. □

Lemma (4) is helpful because it allows us to pin down the consequences on a deviating $S-A$ coalition of changes in message spaces M_S and M_A. The next lemma states that the scope for communication between P on the one hand, and S and A on the other, is rather limited in equilibrium.

Lemma 5. *It is without loss of generality for P to restrict message spaces to be $M_S = \{G, N, B\}$ and $M_A = \emptyset$.*

Proof. Throughout the proof we set $\underline{U}$ fixed.
We first prove that it is optimal for P not to communicate with A. Take a given GC with message spaces M_S and M_A and schedules of transfers $\boldsymbol{t}$ and $\boldsymbol{s}$. S reacts by designing $\underline{SC}^{*}$, specifying schedule of side-transfers $\boldsymbol{y}$ and reporting strategies m_{S}^{*} and m_{A}^{*}. Payoffs are then determined by $GC \circ \underline{SC}^{*}$.

Suppose now that P offers an alternative mechanism $\tilde{GC}$ with message spaces $\tilde{M}_S = M_S \times M_A$ and $\tilde{M}_A = \emptyset$ and schedules of transfers $\tilde{\boldsymbol{t}}$ and $\tilde{\boldsymbol{s}}$. $\tilde{GC}$ is such that, for $\forall (m_A, m_S)$ and $\forall \pi$, $t_{\pi m_S m_A} = \tilde{t}_{\pi m_S m_A}$ and $s_{\pi m_S m_A} = \tilde{s}_{\pi m_S m_A}$, where by construction the collection of message spaces is identical under either grand-contract.

It must now be optimal for S to design an alternative side-contract $\underline{\tilde{SC}}^{*}$ such that $\tilde{m}_{S}^{*} = m_{S}^{*} \times m_{A}^{*}$ and $\tilde{m}_{A}^{*} = \emptyset$. Indeed, if it weren't so, it would contradict the optimality of m_{S}^{*} and m_{A}^{*} as a response to GC. Note that this is true regardless of the action e induced by $\underline{SC}$, that is, inequality (SIC) is left unchanged *on both sides*. An optimal way of solving for GC is thus to shut down communication with A. This concludes the first part of the proof.

We now prove that it is optimal for P to have communication with S concern only σ. Suppose P offers a given GC with message space M_S. Recall that, by Lemma 3, when designing its side-contract $\underline{SC}^*$, S makes its reporting strategy contingent only on the realizations of π and σ (there are six possible joint realizations.) P thus anticipates six different messages (or vectors of messages in case there are several communication rounds) consistent with the equilibrium path. In addition, recall from Lemma 4 that a deviating S (that is, one inducing $e = \underline{e}$) has no choice but to also choose its messages from this subset of six messages. Consequently, there are at most six relevant pairs of transfers both on and off the equilibrium path.

It follows that by requesting reports exclusively concerning $\sigma \in \{G, N, B\}$ and using $\pi \in \{\underline{\pi}, \overline{\pi}\}$, P has enough degrees of freedom to design an alternative grand-contract $\tilde{GC}$ that leaves the coalition's payoffs unchanged both on and off the equilibrium path (if not it would contradict the optimality of $\underline{SC}^*_{\bar{e}}$ and $\underline{SC}^*_{\underline{e}}$ as responses to GC), thereby leaving unaffected inequality (*SIC*) and replicating its payoff under $GC \circ \underline{SC}^*_{\bar{e}}$.

This concludes the proof. $\square$

From now on we denote $m_S = m$ and no longer state transfers as depending on m_A.

4.0.3 Delegation and Collusion Proofness

We now proceed with two lemmas that will enable us to write down P's optimization problem.

Lemma 6. *It is without loss of generality for P to design GC such that $m^*(\pi, \sigma) = \sigma$ on the equilibrium path.*

Proof. Throughout the proof we set $\underline{U}$ fixed.

Suppose P designs a given GC, with schedules of transfers $\boldsymbol{t}$ and $\boldsymbol{s}$ and $M_S = \{G, N, B\}$, inducing S into designing $\underline{SC}^*_{\bar{e}}$. Suppose further that $\underline{SC}^*_{\bar{e}}$ specifies reporting strategy $m^*(\pi, \sigma)$ where $m^*(\pi, \sigma) \neq \sigma$ for some (possibly all) $\sigma \in \{G, N, B\}$ and, in particular, such that (i) for a subset $\tilde{\Omega} \subseteq \{G, N, B\} \cup \emptyset$ we have that $m^*(\sigma) = \sigma$ for $\forall \sigma \in \tilde{\Omega}$ but (ii) for a subset $\tilde{\tilde{\Omega}} \subseteq \{G, N, B\} \cup \emptyset$ we have that $m^*(\sigma) \neq \sigma$ for $\forall \sigma \in \tilde{\tilde{\Omega}}$ where $m^* \in \{G, N, B\}$ and where $\{G, N, B\} \equiv \tilde{\Omega} \cup \tilde{\tilde{\Omega}}$. Denote $M_S^* \subseteq \{G, N, B\}$ the set of relevant messages in the equilibrium induced by GC, that is, the set of messages specified in $\underline{SC}^*_{\bar{e}}$.

Suppose now that P designs an alternative $\tilde{GC}$ with associated schedules of transfers $\tilde{\boldsymbol{t}}$ and $\tilde{\boldsymbol{s}}$, where $\tilde{t}_{\pi\sigma} = t_{\pi m^*(\pi,\sigma)}$ and $\tilde{s}_{\pi\sigma} = s_{\pi m^*(\pi,\sigma)}$ for $\forall \sigma \in \Omega$ and for $\forall \pi \in \{\underline{\pi}, \bar{\pi}\}$. S's best response now involves setting $\tilde{m}^*(\pi, \sigma) = \sigma$ for $\forall \sigma \in \{G, N, B\}$, and the payoffs accruing to all players under $GC \circ \underline{SC}^*_{\tilde{e}}$ are replicated (if not, it would be in contradiction with the optimality of $\underline{SC}^*_{\tilde{e}}$ as a best response to GC).

Finally, the coalition's payoff in case S designs $\underline{SC}^*_{\underline{e}}$is also unchanged since (i) by Lemma 4 we know that in the equilibrium induced by GC, S has to choose messages from M^*_S and (ii) $\tilde{GC}$ simply amounts to a relabeling of the relevant transfers associated to M^*_S. Inequality (SIC) is thus unchanged.

This concludes the proof. ☐

Lemma (6) allows us to seek for the optimal GC within the set of mechanisms for which S truthfully reveals the realization of σ *in equilibrium.*

Lemma 7. *It is optimal for P to design GC such that there are no side-transfers on the equilibrium path.*

Proof. Throughout the proof we set $\underline{U}$ fixed.
From Lemma 6 we know that there is no loss of generality for P in designing GC such that $m(\pi, \sigma) = \sigma$ in equilibrium. Suppose we focus on this class of grand-contracts. To each message $m \in \{G, N, B\}$ is associated, respectively, a pair of (schedules of) transfers $(\boldsymbol{t_m}, \boldsymbol{s_m})$. S reacts by designing $\underline{SC}^{\overline{e}}$ specifying, in addition to reporting strategies $m(\pi, \sigma) = \sigma$, a schedule of side-transfers $\boldsymbol{y}$. *Ex post* payoffs (for a given realization of π) are then $(t_{\pi\sigma} + y_{\pi\sigma\sigma}, s_{\pi\sigma} - K(y_{\pi\sigma\sigma})\, y_{\pi\sigma\sigma})$ for $\forall \sigma, \pi$, where the first component is A's *ex post* payoff and the second component is S's *ex post* payoff.

P can always design an alternative grand-contract $\tilde{GC}$ such that the two pairs of transfers become $\left(\tilde{t}_{\pi\sigma} = t_{\pi\sigma} + y_{\pi\sigma\sigma}, \tilde{s}_{\pi\sigma} = s_{\pi\sigma} - K(y_{\pi\sigma\sigma})\, y_{\pi\sigma\sigma}\right)$. One can prove by contradiction that the optimal reaction by S involves setting no side-transfers. Furthermore, it is strictly optimal for P to do so as $\tilde{t}_{\pi\sigma} + \tilde{s}_{\pi\sigma} = t_{\pi\sigma} + y_{\pi\sigma\sigma} + s_{\pi\sigma} - K(y_{\pi\sigma\sigma})\, y_{\pi\sigma\sigma} < t_{\pi\sigma} + s_{\pi\sigma}$. To see suppose first that $y_{\pi\sigma\sigma} > 0$, and then we have indeed that $\tilde{t}_{\pi\sigma} + \tilde{s}_{\pi\sigma} = t_{\pi\sigma} + y_{\pi\sigma\sigma} + s_{\pi\sigma} - k y_{\pi\sigma\sigma} < t_{\pi\sigma} + s_{\pi\sigma}$ since $k > 1$. Second, if $y_{\pi\sigma\sigma} < 0$ instead, this is also true since, $\tilde{t}_{\pi\sigma} + \tilde{s}_{\pi\sigma} = t_{\pi\sigma} + y_{\pi\sigma\sigma} + s_{\pi\sigma} - \frac{1}{k} y_{\pi\sigma\sigma} = t_{\pi\sigma} + s_{\pi\sigma} + \left(1 - \frac{1}{k}\right) y_{\pi\sigma\sigma} < t_{\pi\sigma} + s_{\pi\sigma}$.

Note that we have disregarded $\underline{SC}^{\underline{e}}$. However since the sum of transfers are now smaller $\forall \sigma, m$, where $m \in \{G, N, B\}$, it can only be the case that S is worse-off under $\tilde{GC}$ than under GC, thereby optimally relaxing inequality (SIC).

This concludes the proof. ☐

This lemma implies that in equilibrium we shall not be concerned by inequalities (10), that is, we shall not be concerned by the feasibility of the equilibrium side-contract. Lemma 7 also proves that delegation, either partial or full, is strictly suboptimal.

The following lemma allows us to characterize the optimal value for $\underline{U}$. In terms of notation, $m^*(\pi,\sigma)$ and $y^*_{\pi\sigma m^*(\pi,\sigma)}$ are the messages and side-transfers specified in $\underline{SC}^*_{\overline{e}}$ for given realizations of π and σ, while $\tilde{m}^*(\pi,\sigma)$ and $\tilde{y}_{\pi\sigma\tilde{m}^*(\pi,\sigma)}$ are the reporting strategies and schedules of side-transfers specified in $\underline{SC}^*_{\underline{e}}$.

Lemma 8. *It is weakly optimal to set*

$$\underline{U} \geq \sum_\sigma E^{\underline{e}}_{\overline{\sigma}}\left(t_{m^*(\pi,\sigma)} + y^*_{\sigma m^*(\pi,\sigma)}\right).$$

In addition, the optimal GC must be such that (6) and (9) bind. Consequently we have that

$$\begin{aligned}\underline{U} &= \sum_\sigma E^{\underline{e}}_{\overline{\sigma}}\left(t_{\tilde{m}^*(\pi,\sigma)} + \tilde{y}_{\sigma\tilde{m}^*(\pi,\sigma)}\right),\\ &= \sum_\sigma E^{\overline{e}}_{\sigma}\left(t_{m^*(\pi,\sigma)}\right) - \psi.\end{aligned}$$

Proof. Consider first $\underline{SC}^{\underline{e}}$. Suppose S binds (5). It is then optimal for S to set $\tilde{y}^*_{\pi\sigma\tilde{m}^*(\pi,\sigma)} = -t_{\pi\tilde{m}^*(\pi,\sigma)}$, $\forall\sigma$, thereby still leaving (5) slack. It is immediate however that such a $\underline{SC}^{\underline{e}}$ then violates (6). We can thus conclude that (6) necessarily binds, i.e., $\underline{U} = \sum_\sigma E^{\underline{e}}_{\overline{\sigma}}\left(t_{\tilde{m}^*(\pi,\sigma)} + \tilde{y}_{\sigma\tilde{m}^*(\pi,\sigma)}\right)$.

Consider now $\underline{SC}^{\overline{e}}$. Suppose (8) binds. It is then optimal to increase $\underline{U}$ at least up to the point at which (8) and (9) coincide, i.e., $\underline{U} \geq \sum_\sigma E^{\underline{e}}_{\overline{\sigma}}\left(t_{m^*(\pi,\sigma)} + y^*_{\sigma m^*(\pi,\sigma)}\right)$, as this leaves $\underline{SC}^{\overline{e}}$ unchanged, but weakly decreases S's payoff when designing $\underline{SC}^{\underline{e}}$, as can be seen from (6). We thus also have that (9) binds, i.e. $\underline{U} = \sum_\sigma E^{\overline{e}}_{\sigma}\left(t_{m^*(\pi,\sigma)}\right) - \psi$. ∎

We are now in a position to write down P's modified optimization problem.

4.0.4 Principal's Problem

Before stating the problem we note that it must necessarily be the case that $\underline{SC}^*_{\underline{e}}$ is such that the expected side-transfer $\sum_\sigma E^{\underline{e}}_{\overline{\sigma}}\left(\tilde{y}_{\sigma\tilde{m}^*(\pi,\sigma)}\right) \leq 0$, for otherwise S could be better-off, for instance, simply by not offering any side-contract and replication reporting strategies.

Using Lemmas 3-8, P's problem takes the form

$$\min_{\{\boldsymbol{t},\boldsymbol{s},\underline{U}\}} \quad \left\{ \sum_{\pi} p^{\bar{e}}_{\pi G} \left(t_{\pi G} + s_{\pi G}\right) + \sum_{\pi} p^{\bar{e}}_{\pi N} \left(t_{\pi N} + s_{\pi N}\right) \right\} \quad \text{s.t.}$$

$$\sum_{\sigma}\sum_{\pi} p^{\bar{e}}_{\pi\sigma} s_{\pi\sigma} \geq \sum_{\sigma}\sum_{\pi} p^{\underline{e}}_{\pi\sigma} s_{\pi\tilde{m}^*(\pi,\sigma)} - \frac{1}{k}\sum_{\sigma}\sum_{\pi} \min\left(\underbrace{\left(p^{\bar{e}}_{\pi\sigma} t_{\pi\sigma} - \psi\right) - p^{\underline{e}}_{\pi\sigma} t_{\pi\tilde{m}^*(\pi,\sigma)}}_{\sum_{\sigma} E^{\underline{e}}_{\sigma}\left(\tilde{y}_{\sigma\tilde{m}^*(\pi,\sigma)}\right)}, 0 \right), \tag{11}$$

$$\underline{U} \geq E^{\underline{e}}_{N}\left(t_N\right), \tag{12}$$

$$\sum_{\pi}\sum_{\sigma} p^{\bar{e}}_{\pi\sigma} t_{\pi\sigma} - \psi = \underline{U}, \tag{13}$$

$$s_{\pi\sigma} \geq s_{\pi\sigma'} + K(t_{\pi\sigma} - t_{\pi\sigma'}) \cdot (t_{\pi\sigma} - t_{\pi\sigma'}) \;\; \forall \pi, \sigma, \sigma', \tag{14}$$

where $\sigma \neq \sigma'$ and $\sigma, \sigma' \in \{G, N, B\}$.

Inequality (11) is S's incentive compatibility constraint: it ensures that S prefers offering $\underline{SC}_{\bar{e}}$ to either offering $\underline{SC}_{\underline{e}}$ or not offering any $\underline{SC}$ at all. Indeed, recall that we seek GC^* within the space of mechanisms in which S offers $\underline{SC}$, even in this modified set-up. When writing (11), we have made use of Lemmas 6-8: (i) S is reporting information truthfully in equilibrium (Lemma 6), not exchanging any side-transfers in equilibrium (Lemma 7) and (ii) if designing $\underline{SC}_{\underline{e}}$ out-of-equilibrium, $\sum_{\sigma} E^{\underline{e}}_{\sigma}\left(\tilde{y}_{\sigma\tilde{m}^*(\pi,\sigma)}\right)$ is necessarily found by binding (6). The "min" condition captures the fact that $\sum_{\sigma} E^{\underline{e}}_{\sigma}\left(\tilde{y}_{\sigma\tilde{m}^*(\pi,\sigma)}\right) \leq 0$ necessarily (see above). We dispense with it in the remainder of the proof, and show that it indeed holds *ex post*.

Further, we introduce inequality (12) into P's problem as we have shown in Lemma 12 that there is no loss of generality in seeking a GC that respects it. Importantly, we have substituted in the fact that it is optimal, in equilibrium, for the supervisor to set $y^*_{\pi B m^*(\pi,B)} = -t_{\pi m^*(\pi,B)}$, for $\forall \pi$, so as to relax her own problem (this is not in contradiction with Lemma 7 as we show below that it is weakly optimal to set $t_{\pi B} = 0$ for $\forall \pi$).

From Lemma 7 we know that it is optimal to seek GC^* within the space of mechanisms in which there are no side-transfers in equilibrium. It thus follows, also making use of Lemma 8 that (13) must hold. Finally, (14) ensures that $\underline{SC}^*_{\bar{e}}$ is such that $m^*(\pi,\sigma) = \sigma$ for $\forall \sigma$, that is, (14) ensures truthful information revelation on the equilibrium path. Note, to conclude, that $\underline{U}$ is indeed treated as a distinct choice variable in this optimization problem.

Observe that we entirely disregard feasibility considerations regarding the side-contracts. This is so for two reasons. First, Lemma 7 tells us that no side-transfers need to be exchanged in equilibrium. This implies that the *equilibrium* side-contract trivially satisfies conditions (10). Second, and as we discuss below in greater detail, $\underline{SC}^*_{\underline{e}}$ is always such that side-feasibility is inherently satisfied and thus cannot be exploited by P.

The problem can be simplified by plugging the value for $\underline{U}$ pinned down by (13) into (11) and (12). It then becomes

$$\min_{\{\boldsymbol{t},\boldsymbol{s}\}} \quad \left\{ \sum_{\pi} p^{\bar{e}}_{\pi G} \left(t_{\pi G} + s_{\pi G}\right) + \sum_{\pi} p^{\bar{e}}_{\pi N} \left(t_{\pi N} + s_{\pi N}\right) \right\} \text{ s.t.}$$

$$\sum_{\sigma}\sum_{\pi} p^{\bar{e}}_{\pi\sigma} s_{\pi\sigma} \geq \sum_{\sigma}\sum_{\pi} p^{\underline{e}}_{\pi\sigma} s_{\pi \tilde{m}^*(\pi,\sigma)} + \frac{1}{k}\left(\underbrace{\sum_{\sigma} E^{\underline{e}}_{\sigma}\left(t_{\tilde{m}^*(\pi,\sigma)}\right) - \sum_{\sigma}\sum_{\pi} p^{\bar{e}}_{\pi\sigma} t_{\pi\sigma} + \psi}_{\sum_{\sigma} E^{\underline{e}}_{\sigma}\left(\tilde{y}_{\sigma \tilde{m}^*(\pi,\sigma)}\right)} \right), \tag{15}$$

$$\sum_{\pi}\sum_{\sigma} p^{\bar{e}}_{\pi\sigma} t_{\pi\sigma} - \psi \geq E^{\underline{e}}_{N}\left(t_{N}\right) \tag{16}$$

$$s_{\pi\sigma} \geq s_{\pi\sigma'} + K(t_{\pi\sigma'} - t_{\pi\sigma}) \cdot (t_{\pi\sigma'} - t_{\pi\sigma}) \ \forall \pi, \sigma, \sigma', \tag{17}$$

where $\sigma \neq \sigma'$ and $\sigma, \sigma' \in \{G, N, B\}$.

Observe, from looking at (15) and recalling the nature of $\tilde{m}^*$ (see S's problem above), that it must necessarily be the case that $\tilde{m}^*\left(\overline{\pi}, \sigma\right)$ and $\tilde{m}^*\left(\underline{\pi}, \sigma\right)$ are determined by respectively

$$s_{\overline{\pi}\tilde{m}^*(\overline{\pi},\sigma)} + \frac{1}{k} t_{\overline{\pi}\tilde{m}^*(\overline{\pi},\sigma)} = \max\left(s_{\overline{\pi}G} + \frac{1}{k} t_{\overline{\pi}G}, s_{\overline{\pi}N} + \frac{1}{k} t_{\overline{\pi}N}, s_{\overline{\pi}B} + \frac{1}{k} t_{\overline{\pi}B} \right) \quad \forall \sigma, \tag{18}$$

and

$$s_{\underline{\pi}\tilde{m}^*(\underline{\pi},\sigma)} + \frac{1}{k} t_{\underline{\pi}\tilde{m}^*(\underline{\pi},\sigma)} = \max\left(s_{\underline{\pi}G} + \frac{1}{k} t_{\underline{\pi}G}, s_{\underline{\pi}N} + \frac{1}{k} t_{\underline{\pi}N}, s_{\underline{\pi}B} + \frac{1}{k} t_{\underline{\pi}B} \right) \quad \forall \sigma. \tag{19}$$

This formally proves that side-feasibility can be ignored also when it comes to $\underline{SC}_{\underline{e}}$.

We proceed by first solving the problem in case (15) is slack (Case 1), then when it is instead binding (Case 2). For the sake of crispness, we anticipate throughout that $t_{\overline{\pi}G} \geq t_{\overline{\pi}N}$. We show that this anticipation is w.l.o.g. below.

Case 1. Suppose first that the solution is such that (15) is slack. If (16) is also slack, then it is optimal to set all transfers to zero as it minimizes the objective function and satisfies inequality (17). This however cannot be a solution as it fails to induce $e = \bar{e}$. Thus, (16) is binding, and rearranging it yields

$$\rho_\pi \rho_\sigma t_{\overline{\pi}G} + \rho_\pi (1-\rho_\sigma) t_{\overline{\pi}N} + (1-\rho_\pi) \rho_\sigma t_{\underline{\pi}G} + (1-\rho_\pi)(1-\rho_\sigma) t_{\overline{\pi}N} - \psi = \rho_\pi (1-\rho_\sigma) t_{\underline{\pi}N} + (1-\rho_\pi)(1-\rho_\sigma) t_{\overline{\pi}N}. \tag{20}$$

Ignoring for now (17), simply by inspecting this last equation and the objective function it is immediate that it is optimal to set all transfers to zero when $\pi = \underline{\pi}$; in particular it is (i) strictly optimal to set $t_{\underline{\pi}N} = 0$ and (ii) weakly optimal to set $t_{\underline{\pi}B} = t_{\underline{\pi}G} = 0$. This however also leads to (17) costlessly holding whenever $\pi = \underline{\pi}$ since, if transfers are zero, there is no scope for information manipulation. We thus set all transfers to zero when $\pi = \underline{\pi}$.

We rearrange (20) as

$$t_{\overline{\pi}G} = \frac{\psi - (1-\rho_\sigma)(2\rho_\pi - 1) t_{\overline{\pi}N}}{\rho_\sigma \rho_\pi}. \tag{21}$$

We now anticipate that (17) binds if and only if $\pi = \overline{\pi}$, $\sigma = N$ and $\sigma' = G$, and check *ex post* that this is indeed the case. Said differently, we anticipate that the only tension in terms of information revelation occurs when S has observed $\sigma = N$ (it is trivial to show that if (17) is completely disregarded then GC fails to induce $e = \overline{e}$). We thus have that $s_{\overline{\pi}N} = s_{\overline{\pi}G} + \frac{1}{k}(t_{\overline{\pi}G} - t_{\overline{\pi}N})$ necessarily. Since $t_{\overline{\pi}B}$ then appears nowhere in the problem, it is weakly optimal to set it to zero.

Substituting in $s_{\overline{\pi}N} = s_{\overline{\pi}G} + \frac{1}{k}(t_{\overline{\pi}G} - t_{\overline{\pi}N})$ and the value of $t_{\overline{\pi}G}$ given by (21), the objective function becomes

$$\rho_\pi s_{\overline{\pi}G} + \psi - (1-\rho_\sigma)(2\rho_\pi - 1) t_{\overline{\pi}N} + \rho_\pi (1-\rho_\sigma)\left(t_{\overline{\pi}N} + \frac{1}{k}\left(\frac{\psi - (1-\rho_\sigma)(2\rho_\pi - 1) t_{\overline{\pi}N}}{\rho_\sigma \rho_\pi} - t_{\overline{\pi}N}\right)\right).$$

This function is increasing in $t_{\overline{\pi}N}$ whenever $k > \frac{\rho_\sigma(1-\rho_\pi)+(2\rho_\pi-1)}{\rho_\sigma(1-\rho_\pi)} > 1$. Suppose this last condition holds, then, from (21), we know that it is optimal to set $t_{\overline{\pi}G} = \frac{\psi}{\rho_\sigma \rho_\pi}$ and, from (17), $s_{\overline{\pi}N} = \frac{1}{k}\frac{\psi}{\rho_\sigma \rho_\pi}$, and all other transfers to zero. We must verify that these schedules of transfers are such that (15) is indeed slack. Substituting in the solution we find that it is if and only if $\rho_\sigma < \frac{(2\rho_\pi - 1)}{\rho_\pi}$. In case $k \leq \frac{\rho_\sigma(1-\rho_\pi)+(2\rho_\pi-1)}{\rho_\sigma(1-\rho_\pi)}$, one can show that the ensuing solution necessarily violates (15). Therefore, the solution derived here in case (15) is slack prevails if and only if both $\rho_\sigma > \frac{(2\rho_\pi - 1)}{(k-1)(1-\rho_\pi)}$ and $\rho_\sigma < \frac{(2\rho_\pi - 1)}{\rho_\pi}$.

To conclude *Case 1* we now verify that the anticipations made earlier hold. Let us look into (17) first. Since $s_{\pi B} = t_{\pi B} = 0$, it is clear that S would never find it optimal to untruthfully report B. Also, it is optimal for S to design a schedule of (out-of-equilibrium) side-transfers that ensure truthful reporting of $\sigma = B$, as this constitutes an optimal way of relaxing (8). Thus, $\sigma = B$ is also reported truthfully.[32] It remains to check that S does not prefer reporting N instead of G. This is ensured by the fact that $s_{\overline{\pi} G} = 0 \geq \left(\frac{1}{k} - k\right) \frac{\psi}{\rho_\sigma \rho_\pi}$, since $k > 1$. Finally, substituting in the solution, we see that $\sum_\sigma E^e_\sigma \left(\tilde{y}_{\sigma \tilde{m}^*(\pi,\sigma)}\right) = \psi - \rho_\sigma \rho_\pi \frac{\psi}{\rho_\sigma \rho_\pi} = 0$, i.e., the "max" condition discussed earlier holds.

Case 2. Suppose now (15) to be binding. Writing it explicitly, (15) becomes

$$\begin{aligned} \rho_\pi \rho_\sigma s_{\overline{\pi} G} + \rho_\pi (1-\rho_\sigma) s_{\overline{\pi} N} + (1-\rho_\pi) \rho_\sigma s_{\underline{\pi} G} + (1-\rho_\pi)(1-\rho_\sigma) s_{\underline{\pi} N} = \\ (1-\rho_\pi) s_{\overline{\pi} m^*(\overline{\pi},\sigma)} + \rho_\pi s_{\underline{\pi} \tilde{m}^*(\underline{\pi},\sigma)} + \frac{1}{k}\left((1-\rho_\pi) t_{\overline{\pi} \tilde{m}^*(\overline{\pi},\sigma)} + \rho_\pi t_{\underline{\pi} \tilde{m}^*(\underline{\pi},\sigma)} + \psi\right) - \\ \frac{1}{k}\left(\rho_\pi \left(\rho_\sigma t_{\overline{\pi} G} + (1-\rho_\sigma) t_{\overline{\pi} N}\right) + (1-\rho_\pi)\left(\rho_\sigma t_{\underline{\pi} G} + (1-\rho_\sigma) t_{\underline{\pi} N}\right)\right). \end{aligned} \tag{22}$$

Regardless of the value $\tilde{m}^*(\underline{\pi}, \sigma)$ takes (that is, be it G, N or B), an optimal way of relaxing (15) is by setting all transfers to zero when $\pi = \underline{\pi}$. Exactly as before, doing so also optimally relaxes (16), (17), and lowers the objective function.

As it is unclear whether (16) also binds, we rewrite it as

$$\bar{t}_G = \frac{\psi + \triangle - (1-\rho_\sigma)(2\rho_\pi - 1)\bar{t}_N}{\rho_\sigma \rho_\pi} \tag{23}$$

where $\triangle \geq 0$. It becomes clear below that setting $\triangle = 0$ is optimal, i.e., binding (16) is optimal. We once again anticipate that (17) binds if and only if $\pi = \overline{\pi}$, $\sigma = N$ and $\sigma' = G$, and check *ex post* that this is indeed the case. We thus have that $s_{\overline{\pi} N} = s_{\overline{\pi} G} + \frac{1}{k}\left(t_{\overline{\pi} G} - t_{\overline{\pi} N}\right)$; which determines also the solution to (18). Since $t_{\overline{\pi} B}$ then appears nowhere, it is weakly optimal to set it to zero. Substituting in $s_{\overline{\pi} N}$, and rearranging, (15) becomes

[32] Even if $t_{\pi B} > 0$ the supervisor finds it optimal to report truthfully. She however sets (out-of-equilibrium) side-transfers such that the agent ends up with a payoff equal to zero.

$$k\left(2\rho_\pi-1\right)s_{\bar{\pi}G}+\rho_\pi\left(1-\rho_\sigma\right)\left(t_{\bar{\pi}G}-t_{\bar{\pi}N}\right)=$$
$$\left(1-\rho_\pi\right)\left(t_{\bar{\pi}G}-t_{\bar{\pi}N}\right)+\left(\left(1-\rho_\pi\right)t_{\bar{\pi}N}+\psi-\rho_\pi\rho_\sigma t_{\bar{\pi}G}-\rho_\pi\left(1-\rho_\sigma\right)t_{\bar{\pi}N}\right),$$

which simplifies to

$$k\left(2\rho_\pi-1\right)s_{\bar{\pi}G}=\psi-\left(2\rho_\pi-1\right)t_{\bar{\pi}G}. \tag{24}$$

Substituting in $s_{\bar{\pi}G}$ as defined by (24) and $t_{\bar{\pi}G}$ as defined by (23), the objective becomes

$$\frac{\rho_\pi}{k}\left(\frac{\psi}{2\rho_\pi-1}+\frac{\psi+\triangle-\left(1-\rho_\sigma\right)\left(2\rho_\pi-1\right)t_{\bar{\pi}N}}{\rho_\sigma\rho_\pi}\right)+\psi+$$
$$\triangle+\left(1-\rho_\sigma\right)\left(1-\rho_\pi\right)t_{\bar{\pi}N}+$$
$$\rho_\pi\left(1-\rho_\sigma\right)\frac{1}{k}\left(\frac{\psi+\triangle-\rho_\pi t_{\bar{\pi}N}+\left(1-\rho_\sigma\right)\left(1-\rho_\pi\right)t_{\bar{\pi}N}}{\rho_\sigma\rho_\pi}\right).$$

It is optimal to set $\triangle=0$ as it enters everywhere positively, i.e., it is optimal to bind (16). This function is decreasing in $\bar{t}_N$ if and only if

$$k>\left(\frac{2\left(2\rho_\pi-1\right)+\rho_\sigma\left(1-\rho_\pi\right)}{\rho_\sigma\left(1-\rho_\pi\right)}\right). \tag{25}$$

If (25) holds, then it is optimal to set $t_{\bar{\pi}G}=\frac{\psi}{\rho_\sigma\rho_\pi}$, $s_{\bar{\pi}G}=\frac{1}{k}\left(\frac{\psi}{2\rho_\pi-1}-\frac{\psi}{\rho_\sigma\rho_\pi}\right)$ and $s_{\bar{\pi}N}=\frac{1}{k}\left(\frac{\psi}{2\rho_\pi-1}-\frac{\psi}{\rho_\sigma\rho_\pi}\right)+\frac{\psi}{k\rho_\sigma\rho_\pi}$, *while setting all other transfers to zero.*

We now verify that the anticipations made earlier hold. Take (17) first. S is never tempted to report untruthfully B as both her and A's payments are then equal to zero. Further, as before, S finds it weakly optimal to choose a schedule of out-of-equilibrium side-transfers that leads her to report truthfully $\sigma=B$, as this optimally relaxes (8) in her problem. It thus remains to check that S prefers to report $\sigma=G$ truthfully rather than N; this is so because

$$s_{\bar{\pi}G}=\frac{1}{k}\left(\frac{\psi}{2\rho_\pi-1}-\frac{\psi}{\rho_\sigma\rho_\pi}\right)\geq\frac{1}{k}\left(\frac{\psi}{2\rho_\pi-1}-\frac{\psi}{\rho_\sigma\rho_\pi}\right)+\frac{\psi}{k\rho_\sigma\rho_\pi}-k\frac{\psi}{\rho_\sigma\rho_\pi}<0,$$

which holds because $k>1$.

Finally, substituting in the solution, we see that $\sum_\sigma E^e_\sigma\left(\tilde{y}_{\sigma\tilde{m}^*(\pi,\sigma)}\right)=\psi-\rho_\sigma\rho_\pi\frac{\psi}{\rho_\sigma\rho_\pi}=0$, i.e., the "max" condition discussed earlier holds.

If instead (25) does not hold, then it is instead optimal, by equation (23), to set

$t_{\overline{\pi}G} = t_{\overline{\pi}N} = \frac{\psi}{\rho_\pi-(1-\rho_\sigma)(1-\rho_\pi)}$ and $s_{\overline{\pi}G} = s_{\overline{\pi}N} = \frac{1}{k}\left(\frac{\psi}{2\rho_\pi-1} - \frac{\psi}{\rho_\pi-(1-\rho_\sigma)(1-\rho_\pi)}\right)$. It can easily be shown that all anticipations made earlier hold.

Solution to the Modified Optimization Problem

Summarizing, we have found that the solution to this optimization problem is as follows.

I. If $k < \frac{\rho_\sigma(1-\rho_\pi)+(2\rho_\pi-1)}{\rho_\sigma(1-\rho_\pi)}$, then it is optimal to set

$$
\begin{aligned}
t_{\overline{\pi}G} = t_{\overline{\pi}N} &= \frac{\psi}{\rho_\pi - (1-\rho_\sigma)(1-\rho_\pi)}, \\
s_{\overline{\pi}G} = s_{\overline{\pi}N} &= \frac{1}{k}\left(\frac{\psi}{2\rho_\pi - 1} - \frac{\psi}{\rho_\pi - (1-\rho_\sigma)(1-\rho_\pi)}\right),
\end{aligned}
$$

and all other transfers to zero.

II. If $\frac{2(2\rho_\pi-1)+\rho_\sigma(1-\rho_\pi)}{\rho_\sigma(1-\rho_\pi)} > k > \frac{\rho_\sigma(1-\rho_\pi)+(2\rho_\pi-1)}{\rho_\sigma(1-\rho_\pi)}$ and

(i) if $\rho_\sigma < \frac{(2\rho_\pi-1)}{\rho_\pi}$, then it is optimal to set

$$
\begin{aligned}
t_{\overline{\pi}G} &= \frac{\psi}{\rho_\sigma\rho_\pi}, \\
s_{\overline{\pi}N} &= \frac{1}{k}\frac{\psi}{\rho_\sigma\rho_\pi},
\end{aligned}
$$

and all other transfers to zero.

(ii) if $\rho_\sigma > \frac{(2\rho_\pi-1)}{\rho_\pi}$, then it is optimal to set

$$
\begin{aligned}
t_{\overline{\pi}G} = t_{\overline{\pi}N} &= \frac{\psi}{\rho_\pi - (1-\rho_\sigma)(1-\rho_\pi)}, \\
s_{\overline{\pi}G} = s_{\overline{\pi}N} &= \frac{1}{k}\left(\frac{\psi}{2\rho_\pi - 1} - \frac{\psi}{\rho_\pi - (1-\rho_\sigma)(1-\rho_\pi)}\right),
\end{aligned}
$$

and all other transfers to zero.

III. If $k > \frac{2(2\rho_\pi-1)+\rho_\sigma(1-\rho_\pi)}{\rho_\sigma(1-\rho_\pi)}$ then it is optimal to set

$$
\begin{aligned}
t_{\overline{\pi}G} &= \frac{\psi}{\rho_\sigma \rho_\pi}, \\
s_{\overline{\pi}G} &= \frac{1}{k}\max\left(\frac{\psi}{2\rho_\pi - 1} - \frac{\psi}{\rho_\sigma\rho_\pi}, 0\right), \\
s_{\overline{\pi}N} &= \frac{1}{k}\left(\frac{\psi}{2\rho_\pi - 1} - \frac{\psi}{\rho_\sigma\rho_\pi}\right) + \frac{\psi}{k\rho_\sigma\rho_\pi},
\end{aligned}
$$

and all other transfers to zero.

Proof that $t_{\overline{\pi}G} < t_{\overline{\pi}N}$ is Suboptimal

We now prove that the anticipation made earlier is without consequences.

Case 1. Suppose that S's incentive compatibility constraint is not binding. It can easily be argued that the A's incentive compatibility constraint must nevertheless bind (see above for an identical argument.) Assume now that $t_{\overline{\pi}G} < t_{\overline{\pi}N}$ and write $t_{\overline{\pi}G} + x = t_{\overline{\pi}N}$, with $x > 0$. Using the fact that (12) is binding, we can write

$$
x = \frac{\psi - (2\rho_\pi - 1 + \rho_\sigma(1-\rho_\pi))\,\bar{t}_G}{(1-\rho_\sigma)(2\rho_\pi - 1)},
$$

anticipating that all transfers when $\pi = \underline{\pi}$ are optimally set to zero (this can be proven following exactly the same steps as in the proof above.) Note that since $x > 0$, we need

$$
t_{\overline{\pi}G} < \frac{\psi}{2\rho_\pi - 1 + \rho_\sigma(1-\rho_\pi)}
$$

to hold. Now, using the value of x so obtained, we have

$$
t_{\overline{\pi}N} = t_{\overline{\pi}G} + x = \frac{\psi - \rho_\sigma\rho_\pi \bar{t}_G}{(1-\rho_\sigma)(2\rho_\pi - 1)},
$$

and, since $t_{\overline{\pi}N} \geq 0$,

$$
t_{\overline{\pi}G} \leq \frac{\psi}{\rho_\sigma\rho_\pi}
$$

must also hold (for otherwise A's incentive compatibility constraint is slack and P could be made better-off). Hence, we have $t_{\overline{\pi}G} < \min\left(\frac{\psi}{\rho_\sigma\rho_\pi}, \frac{\psi}{2\rho_\pi - 1 + \rho_\sigma(1-\rho_\pi)}\right) = \frac{\psi}{2\rho_\pi - 1 + \rho_\sigma(1-\rho_\pi)}$.

Suppose now that the objective function (after all the necessary replacements) is increasing in $t_{\overline{\pi}G}$. It follows that it is optimal to set it to zero. Hence,

$$
t_{\overline{\pi}N} = \frac{\psi}{(1-\rho_\sigma)(2\rho_\pi - 1)},
$$

which, as can be easily verified, implies a $E(z)$ higher than in the Second Best contract.

Suppose instead that the objective function (after all the necessary replacements) is decreasing in $t_{\overline{\pi}G}$. Then it is optimal to set such a transfer as high as possible, i.e., to have

$$t_{\overline{\pi}G} = \frac{\psi}{2\rho_\pi - 1 + \rho_\sigma (1 - \rho_\pi)} - \varepsilon,$$

with $\varepsilon > 0$ and small. Replacing in $t_{\overline{\pi}N}$, we have

$$t_{\overline{\pi}N} = \frac{\psi}{2\rho_\pi - 1 + \rho_\sigma (1 - \rho_\pi)} + \varepsilon,$$

where ε is (optimally) set infinitesimally small and henceforth ignored. Furthermore, notice that, in this case, $s_{\overline{\pi}G} = s_{\overline{\pi}N} + \frac{1}{k}(t_{\overline{\pi}N} - t_{\overline{\pi}G})$ (since now the risk is that G is not reported truthfully.) It follows that the solution in Case 1 is either the same as or dominated by those that were found when setting $t_{\overline{\pi}G} \geq t_{\overline{\pi}N}$. Consequently, this solution is weakly inferior.

Case 2. Suppose now that constraint (11) is binding. To begin, let us establish that (12) binds as well. In this case, $s_{\overline{\pi}G} = s_{\overline{\pi}N} + \frac{1}{k}(t_{\overline{\pi}N} - t_{\overline{\pi}G})$ (since now the risk is that G is not reported truthfully.) Since we cannot be sure a priori that (12) binds, we write

$$t_{\overline{\pi}N} = \frac{\psi + \triangle - \rho_\sigma \rho_\pi \bar{t}_G}{(1 - \rho_\sigma)(2\rho_\pi - 1)}.$$

Using similar reasonings as in the previous proof we obtain

$$k(2\rho_\pi - 1) s_{\overline{\pi}N} \quad = \psi - (2\rho_\pi - 1) t_{\overline{\pi}N} \quad .$$

Replacing for $s_{\overline{\pi}N}$, $t_{\overline{\pi}N}$ and $s_{\overline{\pi}G} = s_{\overline{\pi}N} + \frac{1}{k}(t_{\overline{\pi}N} - t_{\overline{\pi}G})$ in the objective function, one can show that it is increasing in Δ. Hence, it is optimally set to zero, meaning that (12) binds. Having established this, one can follow the proof as in Case 1 above to establish that, since

$$t_{\overline{\pi}N} = t_{\overline{\pi}G} + x = \frac{\psi - \rho_\sigma \rho_\pi t_{\overline{\pi}G}}{(1 - \rho_\sigma)(2\rho_\pi - 1)},$$

two possibilities exist. First, the objective function can be increasing in $t_{\overline{\pi}G}$, which means, as established for Case 1, that having $t_{\overline{\pi}G} < t_{\overline{\pi}N}$ is necessarily suboptimal. Second, the objective function can be decreasing in $t_{\overline{\pi}G}$. Then, as in Case 1, we have that it is optimal to set $t_{\overline{\pi}G} = \frac{\psi}{2\rho_\pi - 1 + \rho_\sigma(1-\rho_\pi)} - \varepsilon$ and $t_{\overline{\pi}N} = \frac{\psi}{2\rho_\pi - 1 + \rho_\sigma(1-\rho_\pi)} + \varepsilon$. Since constraint (11) is binding by assumption, the solution is

$$t_{\overline{\pi}G} = \frac{\psi}{2\rho_\pi - 1 + \rho_\sigma (1 - \rho_\pi)} - \varepsilon \quad t_{\overline{\pi}N} = \frac{\psi}{2\rho_\pi - 1 + \rho_\sigma (1 - \rho_\pi)} + \varepsilon,$$

$$s_{\overline{\pi}N} = s_{\overline{\pi}G} = \frac{1}{k}\left(\frac{\psi}{2\rho_\pi - 1} - \frac{\psi}{2\rho_\pi - 1 + \rho_\sigma\left(1-\rho_\pi\right)} - \varepsilon\right).$$

Recall that ε is optimally set infinitesimally small and can, therefore, be ignored. Once again, the solution is either the same (equivalent) as that obtained assuming $t_{\overline{\pi}G} \geq t_{\overline{\pi}N}$, or it involves strictly higher expected payments by P. It is therefore weakly inferior.

This concludes the proof that setting $t_{\overline{\pi}G} < t_{\overline{\pi}N}$ is necessarily suboptimal.

4.0.5 Implementability of the Solution in the Actual Hierarchy

We now prove that the solution to the modified optimization problem stated above is also the solution to the actual optimization problem faced by P. Recall that in the modified optimization problem, P was given an extra degree of freedom since it could treat $\underline{U}$ as a distinct choice variable. This implied that P could only be better-off compared to the actual problem it faced.

We now show that P's payoff in the modified set-up under GC^* (computed above) can also be reached in the actual game. In the actual game—that is, the one described in Section 2—$\underline{U}$ is not longer a distinct choice variable, but instead determined by the schedules of transfers $\boldsymbol{t}$ and $\boldsymbol{s}$. In particular, $\underline{U}$ is now equal to A's expected payoff accruing from $GC \circ \bar{SC}$ and is found by solving S's problem as explained when proving Lemma 2 above (see (2) and (3)).

Case 1. Suppose first that either "$k < \frac{\rho_\sigma(1-\rho_\pi)+(2\rho_\pi-1)}{\rho_\sigma(1-\rho_\pi)}$" or "$\frac{2(2\rho_\pi-1)+\rho_\sigma(1-\rho_\pi)}{\rho_\sigma(1-\rho_\pi)} > k > \frac{\rho_\sigma(1-\rho_\pi)+(2\rho_\pi-1)}{\rho_\sigma(1-\rho_\pi)}$ and $\rho_\sigma > \frac{(2\rho_\pi-1)}{\rho_\pi}$". Suppose further that we set all the transfers exactly as in in "**I.**" or "**II.** *(ii)*" above, except for $s_{\overline{\pi}B}$, which we set equal to $s_{\overline{\pi}N} + \epsilon$, and $t_{\overline{\pi}B}$, which we set equal to$(1-\rho_\sigma)\, t_{\overline{\pi}G} = (1-\rho_\sigma)\, \frac{\psi}{\rho_\pi-(1-\rho_\sigma)(1-\rho_\pi)}$.

First, as we know from S's problem at the *ex post* side contracting stage, we have that $\underline{U} = (1-\rho_\pi)\,(1-\rho_\sigma)\,\frac{\psi}{\rho_\pi-(1-\rho_\sigma)(1-\rho_\pi)}$, which is identical to the optimal value of $\underline{U}$ chosen by P in the modified optimization problem, as can be seen substituting in the solution stated in **I.** or **II.** *(ii)* in (13) and rearranging. In addition, it is still the case that information is truthfully reported in equilibrium since (i) S does not untruthfully report B as the associated pair of transfers is lower to that when reporting either G or N and (ii) it is still optimal for S to report B truthfully (and setting $y_{\overline{\pi}BB} = -t_{\overline{\pi}B}$) as it optimally relaxes (8) in her own problem. The other deviations continue to be of no concern since the relevant transfers are unchanged compared to above.

It follows therefore that the solutions to $\underline{SC_{\overline{e}}}$ and $\underline{SC_{\underline{e}}}$ are unchanged and thus that P's problem is identical to that in the modified set-up. The optimal GC is thus identical to that described above, except that $s_{\overline{\pi}B} = s_{\overline{\pi}N} + \epsilon$ and $t_{\overline{\pi}B} = (1-\rho_\sigma)\, t_{\overline{\pi}G} = (1-\rho_\sigma)\,\frac{\psi}{\rho_\pi - (1-\rho_\sigma)(1-\rho_\pi)}$.

Case 2. An almost identical reasoning prevails when instead either "$\frac{2(2\rho_\pi - 1)+\rho_\sigma(1-\rho_\pi)}{\rho_\sigma(1-\rho_\pi)} > k > \frac{\rho_\sigma(1-\rho_\pi)+(2\rho_\pi-1)}{\rho_\sigma(1-\rho_\pi)}$ and $\rho_\sigma < \frac{(2\rho_\pi - 1)}{\rho_\pi}$" or "$\frac{2(2\rho_\pi - 1)+\rho_\sigma(1-\rho_\pi)}{\rho_\sigma(1-\rho_\pi)} < k$". In these cases, it is (weakly) optimal to replicate exactly the schedules of transfers stated above for the relevant cases. In particular, because $t_{\overline{\pi}B} = s_{\overline{\pi}B} = 0$, we have that $\underline{U} = 0$, exactly as in the modified optimization problem. The actual problem then is identical in every respect to the modified one and the GC computed above for these cases is optimal.

This concludes the proof.