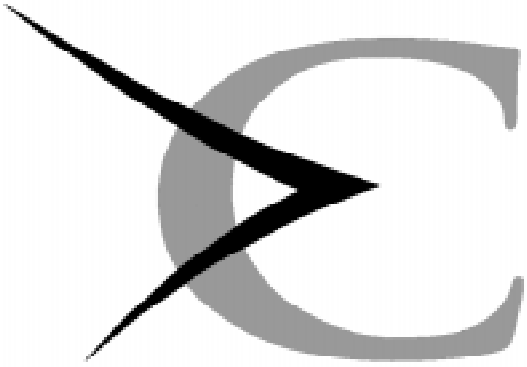


CIRANO

Centre interuniversitaire de recherche
en analyse des organisations

Série Scientifique
Scientific Series

99s-07

A Theory of Abuse of Authority in Hierarchies

Kouroche Vafaï

Montréal
Février 1999

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du ministère de l'Industrie, du Commerce, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche. La *Série Scientifique* est la réalisation d'une des missions que s'est données le CIRANO, soit de développer l'analyse scientifique des organisations et des comportements stratégiques.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de l'Industrie, du Commerce, de la Science et de la Technologie, and grants and research mandates obtained by its research teams. The Scientific Series *fulfils one of the missions of CIRANO: to develop the scientific analysis of organizations and strategic behaviour.*

Les organisations-partenaires / The Partner Organizations

- École des Hautes Études Commerciales
- École Polytechnique
- Université Concordia
- Université de Montréal
- Université du Québec à Montréal
- Université Laval
- Université McGill
- MEQ
- MICST
- Alcan Aluminium Ltée
- Banque Nationale du Canada
- Bell Canada
- Développement des ressources humaines Canada (DRHC)
- Egis
- Fédération des caisses populaires Desjardins de Montréal et de l'Ouest-du-Québec
- Hydro-Québec
- Imasco
- Industrie Canada
- Microcell Labs inc.
- Raymond Chabot Grant Thornton
- Téléglobe Canada
- Ville de Montréal

ISSN 1198-8177

A Theory of Abuse of Authority in Hierarchies*

Kouroche Vafaï†

Résumé / Abstract

Cet article étudie l'abus d'autorité dans une hiérarchie principal-superviseur-agent avec risque moral. Dans une telle structure, le superviseur dont le rôle consiste à faire un rapport sur le niveau d'effort de l'agent, peut abuser de son autorité en exerçant un chantage sur son subordonné. Lorsque le superviseur observe que l'agent a fourni le niveau d'effort désiré par le principal, il peut menacer l'agent d'un faux rapport dans lequel il signale n'avoir rien observé. Afin d'empêcher une telle menace, l'agent peut avoir intérêt à accepter le chantage du superviseur en lui versant un tribut. On étudie les contrats optimaux dans une telle structure hiérarchique. On montre qu'il est optimal pour le principal d'offrir des contrats qui empêchent l'abus d'autorité. Ces contrats détruisent l'enjeu du chantage en augmentant l'espérance de gain de l'agent. Ce résultat est surprenant car, paradoxalement, la possibilité d'abus d'autorité dans l'organisation profite finalement à la victime (l'agent) et non à l'instigateur (le superviseur).

This paper analyzes abuse of authority in a principal-supervisor-agent hierarchy under moral hazard. We characterize the optimal contracts when the supervisor takes advantage of his authority by blackmailing the agent. We show that the optimal policy for the principal is to deter abuse of authority. We find that, paradoxically, the existence of abuse of authority in hierarchies benefits in fine *the victim (the agent) and not the instigator (the supervisor). Our analysis also reveals that the existence of abuse of authority in organizations expands the range of contractual incompleteness.*

Mots Clés : Contrats, risque moral, abus d'autorité, chantage, hiérarchie, contrats incomplets

Keywords : Agency, moral hazard, abuse of authority, blackmail, hierarchy, incomplete contracts

JEL : D82, L20, M12

---

* Corresponding Author: Kouroche Vafaï, CIRANO, 2020 University Street, 25th floor, Montréal, Qc, Canada H3A 2A5 Tel: (514) 985-4000 Fax: (514) 985-4039 e-mail: vafaik@cirano.umontreal.ca
I would like to thank Marcel Boyer, Jacques Robert and Bernard Sinclair-Desgagné for their comments. I gratefully acknowledge financial support from CIRANO.

† CIRANO

1 Introduction

1.1 Overview

In agency models contracts are conditioned by unofficial activities of the agents. Recently, the basic agency model has been extended to account for more complex unofficial activities. Among these activities are collusion (Tirole (1986, 1992), Laffont and Tirole (1993)), hidden games (Laffont (1988, 1990)), influence activities (Milgrom (1988)), sabotage (Lazear (1989)), employee crime (Dickens and al. (1989)), conformism (Scharfstein and Stein (1990)), conservatism (Zwiebel (1995)) and favoritism (Prendergast and Topel (1996)). While some of these unofficial activities may exist in a principal-agent relationship, others are specific to a multi-level principal-supervisor-agent hierarchy. When a principal employs a supervisor to obtain and report information about an agent, she may be exposed to two main forms of unofficial activities. The supervisor and the agent may play against the principal by colluding. The supervisor may also blackmail the agent. Since Tirole's (1986) influential paper, collusion has become the focus of a growing literature. Collusion refers to situations when the supervisor and the agent agree to form a coalition which is detrimental to the principal. The agent then pays the supervisor to conceal unfavorable information. Thus, collusion is a mutually advantageous agreement between the supervisor and the agent. In contrast to collusion, the abuse of authority in organizations remains, theoretically, almost unexplored. The aim of this paper is to take a step towards analyzing this issue. We characterize the design of optimal contracts when the supervisor takes advantage of his authority by blackmailing the agent. In our hierarchical model, the agent (she) has the choice between working or shirking. The principal (she) must therefore give an incentive to the agent to induce her to work. Given that the outcome of this hierarchical relationship is assumed to be revealed in the long run, contracts are contingent only on the supervisor's report on the agent's effort level. The information transmitted by the supervisor (he) to the principal is considered to be hard evidence. Therefore, the only way to manipulate information is to conceal it. As we assume that the supervision technology is imperfect, there is a probability that the supervisor will not observe the agent's effort level. The supervisor has then discretion to conceal information by pretending that he has observed nothing. In this context, we refer to blackmail as situations when the supervisor observes that the agent works but threatens the agent with concealing this information. Accordingly, in the case of blackmail, the supervisor threatens the agent with concealing favorable

information, that is, information that the agent would like the supervisor to accurately report. Unlike collusion, blackmail is not a collective (supervisor-agent coalition) form of unofficial activity and, therefore it is not mutually profitable. When blackmailing the agent, the supervisor manipulates information to the detriment of the entire organization. The major result of this paper is related to the highly imperfect supervision technology case. That is, when there is a low probability that the supervisor observes the agent's effort level. In this case, we show that when the supervisor is truthful, the agent receives three different wages contingent on the supervisor's report on her effort. The principal uses all of the supervisor's information to remunerate the agent. As the supervisor's report becomes more favorable, the agent's wage increases. The agent receives the highest wage when the supervisor reports that she has worked. Next, we consider the case of a venal supervisor who takes advantage of his authority by blackmailing the agent. We show that the optimal policy for the principal is to deter blackmail. However, unlike in the case of collusion, to deter blackmail, the principal has not the choice between designing incentive payments for the supervisor and reducing his discretion. The only available policy against blackmail is to reduce the supervisor's authority over the agent. This means that blackmail can only be deterred by destroying its stake. This is done by ignoring available information provided by the supervisor. The principal then uses aggregate information to pay the agent. Formally, the supervisor's authority is reduced through the destruction of the stake of blackmail, by raising the agent's contingent wage connected with an uninformative report. The agent is then offered the same wage as the one which was contingent on the most favorable report in the truthful supervisor case whether the supervisor's report shows that the agent has worked or whether the report is uninformative. Therefore, when blackmail occurs, the agent is the one who gets the informational rent connected with blackmail. This means that the existence of blackmail in hierarchies benefits *in fine* the victim (the agent) and not the instigator (the supervisor). We refer to this phenomenon as the "blackmail paradox" in hierarchies. Despite the fact that the supervisor may always successfully blackmail the agent in multi-level hierarchy models, extortion and blackmail have not received much attention from agency theorists (e.g., Tirole, 1986; Kofman and Lawarrée, 1993; Laffont and Tirole, 1993, among others). Two exceptions are Laffont's analysis of hidden games and Hindriks, Keen and Muthoo's tax evasion model. Laffont (1988, 1990) analyzes hidden games in a principal-supervisor-two-agents hierarchy under moral hazard. In his model the supervision technology is perfect, that is, the supervisor always observes the agents'

production levels. Laffont assumes that the supervisor observes individual production levels while the principal observes only total production. Accordingly, in his model, information about total production is hard whereas the supervisor's information about individual production levels is partially soft. That is, the supervisor can not only conceal information but also partially lie about individual production levels. In this context, Laffont considers hidden games organized by the supervisor. The supervisor can collude with one of the agents at the expense of the other. The supervisor can also face the two agents with a prisoner-dilemma game. A prisoner-dilemma game is a collective extortion organized by the supervisor who threatens the two agents with permuting their mutual production levels in his report unless they pay him a bribe. When organizing this type of hidden game, the supervisor extracts benefits from each agent by threatening to favor the other agent. Laffont shows that hidden games are deterred only when information is "highly"soft. This contrasts with our result since we show that in a three-level hierarchy with hard information blackmail must always be deterred. Moreover, in Laffont's model, hidden games are not deterred by raising the expected utility of the agents but rather by using anonymous contracts. That is, contracts which are not tied to the agent's individual production levels. Therefore, unlike in our model, the hidden games organized by the supervisor do not *in fine* benefit the agents.[1] Accordingly, there is no "hidden games paradox" in Laffont's model. This is due to the fact that in Laffont's model information is soft. Therefore, it is impossible to destroy the stake of the supervisor's hidden games by raising the expected utility of the agents since this policy cannot prevent the supervisor from misreporting the agent's individual performances. Aside from Laffont's hidden games, Hindriks, Keen and Muthoo (1996) consider also extortion in a three-level hierarchy. Their model is one of tax evasion in an adverse selection hierarchy composed of a government (the principal), a tax inspector (the supervisor) and a tax paying citizen (the agent). They study optimal mechanism design in a setting where the tax inspector may collude with the agent or extort her. In contrast to our paper, the agent in their model is a tax paying citizen rather than an employee of the organization and, is therefore not paid by the principal. Accordingly, in their analysis extortion takes place between a member of the organization (the tax inspector) and an outsider (the tax paying citizen) rather than between two members of the same organization. Therefore, their model concerns external extortion while ours focuses on internal

[1] In Laffont (1988), the two agents are compensated for the supervisor's extortion. Their expected utilities are therefore not raised.

extortion.[2]

In order to overcome the credibility issue of the supervisor's blackmail threat in our model, we must refer to the same exogenous mechanisms postulated in the literature on collusion and in Laffont's analysis of hidden games. In the literature on collusion, the agent's promise to pay a bribe to the supervisor after the supervisor has concealed information lacks credibility. Similarly, in Laffont's hidden games model, the supervisor cannot credibly commit to the payoffs defining the hidden games. As expected, if exogenous mechanisms are not invoked, collusion and hidden games cannot be studied in a one-shot hierarchical relationship. Following Tirole (1986), the growing literature[3] which studies collusion in static three-level hierarchies is based on the assumption that exogenous mechanisms exist which make unofficial promises credible. In order to restore the credibility of promises in one-shot hierarchies, three exogenous mechanisms are postulated by Tirole (1986, 1992) and Laffont and Tirole (1993), namely "word-of-honor", reputation and psychology. In his analysis of hidden games, Laffont also invokes reputation to overcome the supervisor's commitment problem. Since we consider a static hierarchical relationship, exogenous mechanisms have to be called upon to study blackmail. In section 4.1, we discuss how the same exogenous mechanisms as in the cases of collusion and hidden games can be invoked to overcome commitment issues in our three-level hierarchy.

1.2 Examples

Blackmail in hierarchies may be observed in the case of sexual harassment. Legal approaches differentiate two forms of sexual harassment at work. The first form is a "demand by a supervisor directed to a subordinate that the subordinate grant the supervisor sexual favors in order to obtain or keep certain job benefits, be it a wage increase, a promotion, a transfer, or the job itself" (Husbands, 1992, p. 541). This type of sexual harassment has been labelled "quid pro quo" sexual harassment or job-related sexual blackmail. The second form, called "hostile environment" sexual harassment, refers to "unwelcome sexual advances, requests for sexual favors or other verbal, non-verbal or physical conduct of sexual nature which has the purpose or effect of unreasonably interfering with an individual's work performance or creating an intimidating, hostile, abusive or offensive working environment" (Husbands, 1992, p. 541). Hostile environment sexual harassment is different from "quid-pro-quo"

[2] Konrad and Skaperdas (1997) also consider external extortion. Their model is one of criminal extortion by a gang.

[3] For a survey of this literature, see Laffont and Rochet (1997).

harassment in that "the complainant does not have to show a tangible economic loss through being dismissed or forfeiting a promotion or a wage increase"(Husbands,1992, p.541). Therefore, our paper is concerned with "quid pro quo" sexual harassment rather than with hostile environment sexual harassment. A comparative analysis by Husbands (1992) of sexual harassment law in 23 industrialized countries shows that "quid pro quo" sexual harassment is the most widespread form of sexual harassment. Husband's analysis also shows that in many countries (the United States and Spain for example), the employer (the principal) is the one who is liable for damage to the victim. These definitions and findings support the analysis of sexual harassment in a three-level hierarchical agency. Moreover, our assumption that exogenous devices may sustain the credibility of the supervisor's blackmail threat is also corroborated by case studies. Although in most situations the supervisor harasser receives no monetary benefits for carrying out his threat if his subordinate refuses to give up, threats are nevertheless carried out in many cases. For example, the Australian annual reports of the Human Rights and Equal Opportunity Commission (HREOC) which contain numerous case studies of sexual harassment, report cases in which a supervisor carries out his threat by denying work to a person who has refused to give him the demanded tribute.[4] In the battle against sexual harassment, many policies have been proposed and used. The most common type of policy is monetary compensation which is given to the victim by the employer. Another solution is to impose sanctions on the harasser such as transfer, demotion or dismissal. However, these solutions can be very costly[5] not only in actual monetary terms but also because they tarnish the organization's reputation. In this paper, we propose a contractual or organizational solution to the issue of sexual harassment at work and more generally to the issue of abuse of authority in organizations. This solution is to reduce the supervisor's authority, that is, to destroy the sexual harassment stake. This policy has the advantage of deterring sexual harassment and thus avoids having to go to court or having to dismiss valuable employees. Moreover, this policy is advantageous because it prevents trauma and harm to employees by preventing abuse of authority. In cases such as those of sexual harassment, deterring blackmail and preventing trauma and harm is more ethical than the compensation of a victim after-the-fact. In this sense, the "blackmail paradox" in hierarchies can be interpreted as an ethical principle. Other opportunities of blackmail in hierarchies may take the

[4]Human Rights and Equal Opportunity Commission: *Annual Report, 1989-90*, pp. 84-85.

[5]Especially when the employer is liable.

form of any type of extortion of favors (not exclusively sexual) from a subordinate by her (or his) superior.

1.3 Related literature

1.3.1 Bureaucracy and incomplete contracts

Aside from lying within the scope of the literature which considers the impact of varying forms of agents' unofficial activities on organizational design, this paper is also related to the growing literature on the benefits of bureaucratization and rules. During the last decade, agency theorists have emphasized the importance of bureaucratization in reducing the discretion of agents responsable for supervisory or managerial tasks (Milgrom (1988), Shleifer and Vishny (1989), Tirole (1986, 1992), Laffont (1990), Kofman and Lawarée (1993), Prendergast and Topel (1996), among others). Bureaucratization has been shown to be an optimal device to fight employees' unofficial activities. It may require to ignore available information such as part of that provided by the supervisor. Bypassing available information through bureaucratization in order to reduce employees' unofficial activities may also shed new light on the foundations of incomplete contracts. If we define an incomplete contract as a contract which deliberately does not use all the available information,[6] or uses aggregate information, the need for rules may also be a part of a theory of incomplete contracts. The usual arguments to justify the use of incomplete contracts are transaction costs, such as those associated with bounded rationality or with writing complete contracts. The literature on the benefits of bureaucracy suggests that although the principal can write complicated contracts, simple contracts which deliberately ignore certain information may be more efficient to curb employees' unofficial activities.

1.3.2 Blackmail

This paper may also shed light on the differences between blackmail within organizations and blackmail outside organizations. Blackmail within organizations is different from blackmail which occurs outside organizational contexts such as when one person demands money from another and promises not to reveal some information in exchange (e.g.

[6] This definition of contractual incompleteness is used by Spier (1992) in a principal-agent framework. It captures the essence of contractual simplicity. This definition is also one among many reviewed by Bernheim and Whinston (1998).

one party is cheating on her (or his) spouse).[7] One difference lies in the fact that in non-organizational contexts, blackmail may benefit both the instigator and the victim. The victim pays because his utility increases when the information is concealed and the blackmailer's utility is obviously increased when the victim pays. The second difference lies in the type of information manipulation. In non-organizational situations, the blackmailer threatens to reveal the truth. That is, he threatens to disclose some information. In contrast, in an organizational context, the blackmailer threatens to conceal the truth. That is, he threatens to manipulate information. Finally, in an organizational context, the blackmailer threatens to hurt the victim by concealing information from a third party (the principal) who employs and pays both of them, whereas in a non-organizational setting, the blackmailer threatens to hurt the victim by disclosing information to an outside third party who is involved in a relationship with only one of them. The major consequence of these differences is that non-organizational blackmail is deterred by paying the blackmailer while we show that organizational blackmail is deterred by paying the victim.

1.3.3 Exploitation in organizations

Finally, understanding blackmail and extortion in organizations can be considered a step towards building a general theory of exploitation and abuse of authority in organizations. Pioneering works in this domain are Laffont's analysis of hidden games and Prendergast and Topel's study of favoritism in organizations.

The remainder of this paper is organized as follows. Section 2 outlines the model. Sections 3 and 4 characterize optimal contracts in a hierarchy with a truthful supervisor and in a hierarchy with a venal supervisor who blackmails the agent, respectively. Section 5 concludes the paper.

2 The model

We consider a principal-supervisor-agent hierarchy under moral hazard. The top of the hierarchy (the principal) is the residual claimant of profits generated by the whole structure. The bottom of the hierarchy (the agent) is in charge of production. The intermediate layer (the supervisor) is in charge of collecting information on the agent's effort level. The

[7] For a complete economic analysis of blackmail outside organizational contexts, see the May 1993 issue on blackmail of the University of Pennsylvania Law Review.

principal and the supervisor are risk neutral whereas the agent is risk averse. The agent may or may not exert productive effort, $e \in \{0,1\}$. If she exerts no effort, that is, if she shirks, she produces nothing. At the one-effort level she produces x. Her utility function is $U(w,e) = u(w) - \varphi e$ with $u' > 0$, $u" < 0$, where w is the monetary transfer she receives from the principal and $\varphi > 0$ is her disutility of effort. Without loss of generality, we assume that $u(0) = 0$ and $u'(0) \neq +\infty$.[8] Since the principal cannot observe the agent's effort level, the role of the supervisor is to provide a report r on this effort level. The supervisor's utility function is $V(s,a) = s - \zeta a$, where s is his monetary transfer and $\zeta > 0$ is his disutility of supervising effort. The supervisor has the choice between two supervising effort levels, $a \in \{0,1\}$. At the zero supervising effort level he observes nothing. We assume imperfect supervision technology, that is, there is only a probability p that the supervisor does actually observe the agent's effort level when he chooses $a = 1$. The probability p then represents the degree of the perfection of the supervision technology. Consequently, the supervisor's report can take three values $r \in \{0,1,\emptyset\}$, where $r = \emptyset$ means that the supervisor has not observed the agent's effort. We assume that the supervisor's information is hard, that is, verifiable by the principal. Thus, the only way to manipulate information is to conceal it. The supervisor may conceal information by reporting $r = \emptyset$ when he observes $e = 0$. In this case, he colludes with the agent. He can also threaten to conceal information when he observes $e = 1$. We refer to this form of information manipulation as blackmail and focus on this case throughout the paper.[9] The agent's and supervisor's reservation utilities are assumed to be $\underline{u}$ and $\underline{v}$, respectively. According to these assumptions, the principal's problem is to elicit the production effort level $e = 1$ and the supervision effort level $a = 1$.[10] Assuming that the outcome x of the hierarchy is revealed in the long run, the agent's and the supervisor's contracts depend only on the supervisor's report of the agent's effort level.[11] Accordingly, the principal offers a contract

[8] Throughout the paper we will assume that the agent is protected by limited liability. That is, her wage must exceed some lower limit. To simplify the analysis, we set this limit at zero. For this reason, we must assume that $u'(0) \neq +\infty$.

[9] A complete analysis should also envision other coalitions including the one between the supervisor and the agent. However, in order to focus on blackmail, we neglect these coalitions. Moreover, it can be shown that in this setting the supervisor-agent and principal-supervisor coalitions (collusion) are irrelevant (Vafaï (1998)). Consequently, the blackmail-free outcome is immune from coalitions between the supervisor and the agent and between the principal and the supervisor.

[10] The outcome x is assumed to be sufficiently large that it is in the principal's interest to engage in production.

[11] This assumption is identical to assuming that whereas the supervisor observes the outcome x which is revealed at the end of the period, the principal does not. The

$(\underline{w}, \widehat{w}, \overline{w})$ to the agent where $\underline{w}$ is the wage she receives when $r = 0$ and $\widehat{w}$ (resp. $\overline{w}$) is the wage she receives when $r = \emptyset$ (resp. $r = 1$). Similarly, the principal offers a contract $(\underline{s}, \widehat{s}, \overline{s})$ to the supervisor.

Definition. The principal uses aggregate information to remunerate the agent if she chooses to reduce the value set of the supervisor's report by considering only the two reports $r = 0$ and $r \neq 0$.[12] Similarly, a contract is complete (incomplete) if it specifies different (identical) wages for each value of the supervisor's report (two different values of the supervisor's report).

This definition captures the essence of contractual simplicity. Spier (1992) uses this definition of contractual incompleteness in a principal-agent model. She shows that asymmetric information expands the range of contractual incompleteness. In her model, the principal offers an incomplete contract to the agent in order to signal her type through incompleteness. In this paper, we suggest another explanation for the expansion of contractual incompleteness. We show that it is the existence of abuse of authority which expands the range of contractual incompleteness.

It is important to note that, in the case of collusion, it is not relevant to know whether it is the supervisor who threatens the agent with revealing the truth or the agent who asks the supervisor to conceal the truth. This is due to the fact that when the agent shirks and her effort level is observed by the supervisor, forming a coalition to conceal the truth is mutually advantageous. Collusion is therefore an agreement between a dishonest agent (a shirker) and a dishonest supervisor (a supervisor who accepts bribes). Accordingly, since collusion is mutually beneficial, the supervisor does not take advantage of his authority when colluding with the agent. In contrast, blackmail involves an honest agent (an agent who works and who wants the supervisor to accurately report her effort level) and a dishonest supervisor (a supervisor who demands a tribute to reveal the truth). Therefore, blackmail is an abuse of authority.

The timing of the game is as follows

Insert Figure

At the first stage, the principal offers a contract $(\underline{w}, \widehat{w}, \overline{w})$ to the agent and a contract $(\underline{s}, \widehat{s}, \overline{s})$ to the supervisor. The agent and the supervisor then makes a report on the outcome.

[12] Laffont (1990) uses the term aggregate information in a different meaning. In his three-level pyramidal hierarchy with two agents, the supervisor can provide information on total production level as well as on individual production levels. Aggregate information then refers to information on total production level.

pervisor decide whether to accept or refuse the contract. If one of them refuses, the game ends and they both receive their reservation utility. At stage two, the agent and the supervisor play a simultaneous move game. The agent chooses whether to work or to shirk. The supervisor chooses whether to supervise or not. If the supervisor decides to supervise, the imperfect supervision technology reveals the agent's effort level with probability p. At the third stage, if the supervision technology reveals that the agent works, the supervisor takes advantage of his authority by blackmailing the agent. He threatens the agent with an uninformative report ($r = \emptyset$) unless the agent pays him a tribute. The agent accepts or refuses to pay the demanded tribute. At the fourth stage, the supervisor produces a report for the principal. At stage 5, the parties exchange transfers according to the contractual agreements. At the last stage, the agent and the supervisor exchange a side-transfer (tribute) according to their side-agreement. We look for a Subgame Perfect Equilibrium of this game, and we restrict attention on pure strategies.

Throughout this paper we will consider the collusion model and its main results as a comparison benchmark for our analysis. For this purpose, we adopt the major assumptions of the canonical collusion model. That is, we assume that blackmail takes place after the supervisor has observed the agent's effort. Furthermore, we assume that the risk-neutral supervisor has all the bargaining power in the unofficial subgame with the agent and that he is protected by limited liability. Therefore, the principal must give the supervisor a rent for his information. Finally, we assume that a mechanism exists which makes unofficial promises between the supervisor and the agent credible. In our setting, this last assumption means that there is a mechanism which makes the agent's promise to pay a tribute to the supervisor after he has reported the truth credible. In section 4.1 we show that this last assumption must be extended in order to account for the supervisor's blackmail threat.

We will first turn to the characterization of contracts when the supervisor reports truthfully.

3 A truthful supervisor

In this section we analyze the benchmark case in which the supervisor is truthful and does not take advantage of his authority by blackmailing the agent. Therefore, the third stage of the game is removed. Given the imperfect supervision technology, the principal wishes to elicit the production effort level $e = 1$. Thus, the agent's participation and no-

shirking constraints are respectively,

$$\begin{aligned} pu(\overline{w}) + (1-p)u(\widehat{w}) - \varphi &\geq \underline{u} && (1)\\ pu(\overline{w}) + (1-p)u(\widehat{w}) - \varphi &\geq pu(\underline{w}) + (1-p)u(\widehat{w}) \quad \text{that is,} \\ p(u(\overline{w}) - u(\underline{w})) - \varphi &\geq 0 && (2) \end{aligned}$$

Similarly, the principal wishes to elicit the supervision effort level $a = 1$. Accordingly, the supervisor's participation and no-shirking constraints are respectively,

$$\begin{aligned} p\overline{s} + (1-p)\widehat{s} - \zeta &\geq \underline{v} && (3)\\ p\overline{s} + (1-p)\widehat{s} - \zeta &\geq \widehat{s} \text{ that is, } p(\overline{s} - \widehat{s}) - \zeta \geq 0 && (4) \end{aligned}$$

Equations (1) and (3) state that the agent and the supervisor must obtain at least their reservation utility. Equations (2) and (4) are the agent's and supervisor's *no-shirking* constraints, making them weakly prefer to exert effort in equilibrium.[13]

It is important to note that in our setting, $\underline{w}$ and $\underline{s}$ are never paid since the principal wishes to elicit the production effort level $e = 1$ and the supervision effort level $a = 1$. However, $\underline{w}$ and $\underline{s}$ are used to elicit the equilibrium production effort level $e = 1$. Formally, to elicit $e = 1$, the principal must also provide incentives to the supervisor to report the truth when he observes $e = 0$. That is, we also have to consider the supervisor's participation and no-shirking constraints for $e = 0$,

$$\begin{aligned} p\underline{s} + (1-p)\widehat{s} - \zeta &\geq \underline{v} && (5)\\ p\underline{s} + (1-p)\widehat{s} - \zeta &\geq \widehat{s} \text{ that is, } p(\underline{s} - \widehat{s}) - \zeta \geq 0 && (6) \end{aligned}$$

We assume that the agent and the supervisor are protected by limited liability. This requires that their wage exceeds a lower limit. Without loss of generality, we set this limit at zero. Therefore,

$$\underline{w} \geq 0,\ \widehat{w} \geq 0,\ \overline{w} \geq 0,\ \underline{s} \geq 0,\ \widehat{s} \geq 0,\ \overline{s} \geq 0 \qquad (7)$$

Given our assumptions the principal's program reduces to a cost minimization subjected to the supervisor and the agent participating and supplying full effort, that is,

[13] We assume that when the agent is indifferent between working and shirking (resp. when the supervisor is indifferent between supervising or not), she (he) behaves in the way the principal wants her (him) to. That is, she chooses to work (resp. he chooses to supervise).

$$[P_0] \qquad \min_{(\underline{w},\widehat{w},\overline{w}),(\underline{s},\widehat{s},\overline{s})} \; p(\overline{w}+\overline{s})+(1-p)(\widehat{w}+\widehat{s})$$
$$\text{s.t. } (1),(2),(3),(4),\ (5),(6) \text{ and } (7)$$

The optimal solution to this program is summarized in the following proposition:

Proposition 1.

(i) There exists $\widetilde{p}$, such that the principal uses (offers) aggregate information (an incomplete contract to the agent) iff $p \geq \widetilde{p}$.

(ii) For $p \geq \widetilde{p}$, the agent's contract is characterized by $0 \leq \underline{w} \leq u^{-1}(\varphi + \underline{u} - \frac{\varphi}{p})$ and $\widehat{w} = \overline{w} = u^{-1}(\underline{u} + \varphi)$.

(iii) For $p < \widetilde{p}$, the agent's contract is characterized by $\underline{w} = 0$, $\widehat{w} = u^{-1}(\frac{\underline{u}}{1-p})$ and $\overline{w} = u^{-1}(\frac{\varphi}{p})$.

(iv) The principal keeps the supervisor at his reservation utility level.

Proof. see Appendix

When the supervision technology is weakly imperfect, that is, when the agent's effort may be observed with a high probability, the principal can elicit full production effort by giving a flat wage schedule to the agent whether $r = 1$ or $r = \emptyset$. The principal then uses aggregate information, that is, she does not use all the information available from the supervisor's report to pay the agent. Since $\underline{w}$ is never paid at equilibrium, the agent is "fully" insured in this case. In contrast, when the supervision technology is highly imperfect, that is, when the probability that the supervisor observes the agent's effort level is under some treshold value, the principal uses all the information available from the supervisor's report. The intuition behind this is that, since the probability of observing the agent's effort level is low, the incentive problem of the agent is exacerbated. She has less incentive to work and more incentive to shirk (since working and shirking have a small probability to be detected). In order to give incentives to the agent to motivate her to work, the optimal policy for the principal is then to offer a riskier wage schedule to the agent. This is done by creating a strong gap between contingent wages. Indexing the contingent wages connected with the $p < \widetilde{p}$ case by $*$ and those connected with the $p \geq \widetilde{p}$ case by o, one has $\underline{w}^* \leq \underline{w}^o < \widehat{w}^* < \widehat{w}^o = \overline{w}^o < \overline{w}^*$ for $p < \widetilde{p}$.

As for the supervisor, his participation constraint is binding. That is, he is set to his individual rationality level. His contract is such that $\overline{s} > \widehat{s}$, otherwise he will not exert any supervisory effort and, consequently, an agency relationship will not be possible. Similarly, we have $\underline{s} > \widehat{s}$,

otherwise the supervisor will not report truthfully when he observes $e = 0$.

We now turn to the case of a venal supervisor who takes advantage of his authority by blackmailing the agent.

4 A venal supervisor

4.1 Unofficial threats and promises in hierarchies

Among the strong assumptions on which the literature based on Tirole (1986) has been erected, the most crucial one is that side-agreements between the supervisor and the agent are binding in the collusion subgame. This means that it is credibly possible to commit to a promise[14] in an unofficial one-shot relationship. Formally, once the supervisor has reported in accordance with the side-agreement, the agent no longer has an incentive to pay the promised bribe, but she nevertheless stands by her promise and pays the supervisor.[15] In order to overcome this credibility issue, the literature on collusion appeals to exogenous mechanisms. Tirole (1992) and Laffont and Tirole (1993) invoke three exogenous mechanisms which may sustain promises in a static collusive side-agreement. The first mechanism is "word-of-honor". The parties to the side-contract pledge their word and loathe to cheat on promises with other parties. The second mechanism is reputation. It is assumed that the relationship is an ongoing one and therefore, the parties are concerned about their reputation of being able to abide to their promise in order to increase their payoffs. That is, the agent and the supervisor are concerned about their reputation of being trustworthy and fair. According to this argument, the binding promise framework is a short-cut to a dynamic theory relying on reputation considerations.[16] The third mechanism is psychology or emotions. Tirole (1992, pp. 155-156) suggests that one party will keep her promise worrying "that the other party would be upset by the breach of agreements and would seek revenge". Similarly, in Laffont's static analysis of hidden games, the supervisor's commitment to the payoffs defining the hidden games may lack credibility. Laffont also invokes reputation to overcome the supervisor's commitment problem. In other words, collusion and hidden games models implicitly assume that a pro-

[14] For a discussion of definitions of threat and promise, see Klein and O'Flaherty (1993).

[15] This approach is termed enforceability approach in contrast to the self-enforceability approach. For a complete analysis of these issues and for examples, see Tirole (1992).

[16] See Tirole (1992) for the first formal treatment of self-enforcing side-agreements.

hibitive cost exists (which is exogenous in the models) due to the loss of one's reputation or honor. This assumption makes the agent's strategy not to honor her promise in collusion models and the supervisor's strategy not to abide to the payoffs defining the hidden games in Laffont's analysis dominated strategies. As in the cases of collusion and hidden games, in the analysis of blackmail we must consider the credibility of the agent's promise to pay the supervisor after he has reported $r = 1$ as well as the credibility of the supervisor's threat to report $r = \emptyset$ unless the agent pays him a tribute. In our model, when the supervisor observes that the agent works, he threatens to conceal this information if the agent does not pay him a tribute. This threat, as well as the agent's promise of paying a tribute to the supervisor after he has reported the truth, lacks credibility. When the supervisor threatens the agent with an uninformative report and the agent refuses to comply, the supervisor gains $\overline{s} - \zeta$ if he does not carry out his threat and $\widehat{s} - \zeta$ otherwise. Since the principal must elicit the supervision effort level $a = 1$, we have $\overline{s} > \widehat{s}$. Therefore, it is not in the supervisor's interest to carry out his threat. However, in the case of sexual harassment, previous reports[17] have emphasized that the supervisor harasser often carries out his threat and harms his "non-cooperative" subordinate. The supervisor may do this even if it is not in his immediate monetary interest to do so. Therefore, there are long term monetary gains and costs as well as non-monetary gains and costs related to carrying out one's threat. In our setting, the supervisor's threat becomes credible if we rely on the same exogenous mechanisms as those suggested in collusion and hidden games models. These models implicitly assume that the agent incurs a prohibitive exogenous cost when she does not keep her promise toward the supervisor (collusion models) and that the supervisor incurs a prohibitive cost when he does not abide to the payoffs defining the hidden games (hidden games models). In order to analyze blackmail in a static hierarchy, we can similarly call upon reputation to overcome the commitment problem of the supervisor. We can assume that the hierarchical relationship is ongoing and, therefore, the supervisor's opportunity to benefit from blackmail in the future depends on his reputation for carrying out his threats. As in the case of collusion, our assumption that the supervisor can credibly commit to carrying out his threat is then a short-cut to a dynamic theory relying on reputation considerations. However, in contrast to the collusion and hidden games cases, in the blackmail case the supervisor is the only one who is concerned about his reputation. Moreover, he is

[17] For instance the Australian Human Rights and Equal Opportunity Commission: *Annual Report, 1989-90*, pp. 84-85.

not concerned about his reputation of being trustworthy but instead of being vindictive and agressive. That is, the supervisor wants the agent to know that he is capable of punishing her if she does not comply or if she does not keep her promise. Aside from reputational concerns, emotions or psychology can also overcome the credibility issue of commitments. In our setting, this can be done by assuming that the supervisor makes a belief-independent or belief-dependent emotional response to the agent's strategy to refuse to pay a tribute. Belief-independent emotional response refers to exogenous emotional response in games, that is, an emotional response which is always the same for a given outcome. In contrast, belief-dependent emotional response refers to an emotional response which differs depending on the parties' beliefs at different times.[18] Considering belief-independent emotional response in our setting means that there is a prohibitive psychological or emotional cost for the supervisor to not carry out his threat. This cost may be due to frustration and anger caused by not seeking revenge, that is, by letting the agent goes unpunished for her act of bravado.[19] Thus, the agent will accept to pay a tribute at the blackmail stage of the game (stage 3) because she knows that the supervisor will otherwise seek revenge by reporting $r = \emptyset$. Similarly, the agent will make the promised side-transfer after the supervisor has made his report (stage 6) because she knows that the supervisor will otherwise take his revenge by harming her. Thus, threats and promises are now credible due to the commonly known aggressive psychological profile of the supervisor. It is important to note that the psychological or emotional paradigm can be viewed as an extreme case of a reputation model in which the prior probability of the supervisor being aggressive is equal to one. Emotions and psychology are appealing exogenous mechanisms which sustain threats and promises, especially in cases such as those of sexual harassment or more generally those involving extortion of favors. This is the case since these situations involve strong feelings

[18]Games considering belief-dependent emotional responses are termed psychological games. For belief-independant emotional responses in games see Frank (1988) or Huang and Wu (1992), and for a complete definition and description of psychological games see Geanakoplos, Pearce and Stacchetti (1989).

[19]Elster (1996, p. 1392) writes:"...we may observe that emotions contribute to welfare in several ways. First, in any given encounter with the world there may arise occurrent emotions that are immediate sources of happiness and unhappiness. Secondly, emotional dispositions may shape the outcome of any such encounter. If people know that I am subject to fits of destructive anger, I will usually get my way when I deal with them. Thirdly, the dispositions tend to shape the stream of such encounters. If people know that I am irascible, they will avoid dealing with me. *Fourth, if I control my anger to prevent such effects, I incur psychological costs that may be quite severe. Suppression of spontaneous emotional experiences and action tendencies may have a large negative impact on soma and psyche".*(our emphasis)

(such as those related, for instance, to sexual attraction) as well as social, moral and professional stakes. Aside from Tirole (1992), authors such as Hirshleifer (1987), Frank (1987, 1988) and Huang and Wu (1992) as well as others have considered severe emotional gains and costs as devices which make commitments credible in one-shot interactions. These authors show how non-material gains and costs such as those related to joy, anger or frustration may offset material incentives so as to make some behaviors credible. For example, Huang and Wu (1992) consider both belief-independent and belief-dependent emotional responses in litigation. They first show that without emotional responses, the threat to go trial by a defendant is not credible. Then they show that both types of emotional responses make this threat credible. To summarize this section, reputation as well as emotions are devices which make threats and promises credible in one-shot unofficial relationships in organizations.

We will now turn to the analysis of the blackmail subgame.

4.2 Blackmail

When the agent decides to work and her effort level is observed by the supervisor, she is blackmailed by him. The supervisor then threatens to report $r = \emptyset$. It may be in the agent's interest to pay the tribute demanded by the supervisor to prevent information concealment. In accordance with Tirole (1992), we assume that side-transfers between the agent and the supervisor are made at a rate $k \in (0,1)$ which is proportional to the size of the tribute. That is, if the agent transfers t, the supervisor receives kt. This assumption means that the transfer technology between the agent and the supervisor is not totally efficient. This may be, for example, because it is costly to organize unofficial activities in organizations. Since we assume that an exogenous mechanism exists which makes threats and promises credible in unofficial subgames, $\overline{s} + kt^{\min} - \zeta$ and $\overline{s} - \zeta$ represent the supervisor's payoffs corresponding respectively to blackmailing or not the agent. The minimum monetary tribute $t^{\min}$ claimed by the supervisor to report the truth is then solution to $\overline{s} + kt^{\min} - \zeta \geq \overline{s} - \zeta$, that is, $t^{\min} \geq 0$. Thus, it is in the supervisor's interest to blackmail the agent as long as $t^{\min} \geq 0$. The agent accepts to pay the demanded tribute if $u(\overline{w} - t^{\max}) - \varphi \geq u(\widehat{w}) - \varphi$, that is, if $t^{\max} \leq \overline{w} - \widehat{w}$ where $t^{\max}$ is the maximum monetary tribute that the agent who does not shirk is ready to pay to the supervisor so that he will report the truth. If we consider the amount of the blackmail tribute transferred by the agent to the supervisor as the solution of a Nash bargaining game between the supervisor and the agent, this amount will be determined by their respective bargaining powers. Following Tirole

(1986), we assume that the supervisor has all the bargaining power. The supervisor then proposes a take-it-or-leave-it offer t to the agent. Since the agent is ready to pay the maximum amount of $t^{\max} = \overline{w} - \widehat{w}$ and the supervisor is always better off blackmailing the agent than not, the blackmail tribute related to the case in which the agent does not shirk and the supervisor observes her effort is $t = \overline{w} - \widehat{w}$.

Note that the fact that it is in the supervisor's interest to blackmail the agent as long as $t^{\min} \geq 0$ is independent of the respective bargaining powers. Even if the supervisor does not not have all the bargaining power, it is still beneficial for him to blackmail the agent as long as $t^{\min} \geq 0$. Therefore, we show that it is impossible to deter blackmail through the supervisor's contract by giving him incentives. Blackmail deterrence can only be carried out through the contract offered to the agent.

We now turn to the determination of the expressions of the participation and no-shirking constraints.

The agent's participation and no-shirking constraints are respectively,

$$\begin{aligned} pu(\overline{w} - t) + (1-p)u(\widehat{w}) - \varphi &\geq \underline{u} \quad \text{that is,} \\ u(\widehat{w}) - \varphi - \underline{u} &\geq 0 \qquad (8) \\ pu(\overline{w} - t) + (1-p)u(\widehat{w}) - \varphi &\geq pu(\underline{w}) + (1-p)u(\widehat{w}) \quad \text{that is,} \\ p(u(\widehat{w}) - u(\underline{w})) - \varphi &\geq 0 \qquad (9) \end{aligned}$$

The supervisor's participation and no-shirking constraints are respectively,

$$\begin{aligned} p(\overline{s} + k(\overline{w} - \widehat{w})) + (1-p)\widehat{s} - \zeta &\geq \underline{v} \text{ that is,} \\ p(\overline{s} - \widehat{s}) + \widehat{s} + pk(\overline{w} - \widehat{w}) - \zeta - \underline{v} &\geq 0 \qquad (10) \\ p(\overline{s} + k(\overline{w} - \widehat{w})) + (1-p)\widehat{s} - \zeta &\geq \widehat{s} \text{ that is,} \\ p(\overline{s} - \widehat{s}) + pk(\overline{w} - \widehat{w}) - \zeta &\geq 0 \qquad (11) \end{aligned}$$

We also have to consider the supervisor's participation and no-shirking constraints for $e = 0$ as well as the limited liability constraints. Therefore, we add constraints (5), (6) and (7) to the previous constraints.

The supervisor is better off blackmailing the agent as long as $t = \overline{w} - \widehat{w} \geq 0$. Accordingly, if $\overline{w} < \widehat{w}$, since the tribute t is then negative, the supervisor will not blackmail the agent. However, as we showed, the resulting program (program $[P_0]$ with $\overline{w} < \widehat{w}$) does not have any solution. We must therefore necessarily have $\overline{w} \geq \widehat{w}$. Consequently, when the supervisor is venal and takes advantage of his authority by

blackmailing the agent, the principal faces a new constraint,

$$\overline{w} - \widehat{w} \geq 0 \tag{12}$$

Under our assumptions the principal's program can be written as

$$[P_1] \qquad \min_{(\underline{w}, \widehat{w}, \overline{w}), (\underline{s}, \widehat{s}, \overline{s})} \; p(\overline{w} + \overline{s}) + (1-p)(\widehat{w} + \widehat{s})$$
$$\text{s.t. } (5), (6), (7), (8), (9), (10), (11) \text{ and } (12)$$

The solution to this program is summarized in the following proposition:

Proposition 2.

(i) The optimal policy for the principal is to deter abuse of authority.

(ii) The principal always uses (offers) aggregate information (an incomplete contract to the agent).

(iii) When the supervision technology is weakly imperfect (i.e. $p \geq \widetilde{p}$):

a. the agent's contract is characterized by $0 \leq \underline{w} \leq u^{-1}(\varphi + \underline{u} - \frac{\varphi}{p})$ and $\widehat{w} = \overline{w} = u^{-1}(\underline{u} + \varphi)$.

b. the existence of abuse of authority does not affect the efficiency of the organization.

(iv) When the supervision technology is highly imperfect (i.e. $p < \widetilde{p}$):

a. the agent's contract is characterized by $\underline{w} = 0$ and $\widehat{w} = \overline{w} = u^{-1}(\frac{\varphi}{p})$.

b. the existence of abuse of authority decreases the efficiency of the organization.

(v) The principal keeps the supervisor at his reservation utility level.

Proof. see Appendix

We therefore have

Corollary.

The existence of abuse of authority in hierarchies has the following consequences:

1. It expands the range of contractual incompleteness.

2. It benefits in fine the victim (the agent) and not the instigator (the supervisor).

As in the truthful supervisor case, the supervisor's participation constraint is saturated. Since $\widehat{w} = \overline{w}$ the supervisor's contract is unchanged. The principal offers the supervisor the same contract whether he is assumed to be truthful or venal and engaging in blackmail. As established in the truthful supervisor case, when the supervision technology is weakly imperfect, the principal uses aggregate information to remunerate the agent. In this case, blackmail cannot occur because there is no stake for it. Accordingly, blackmail is deterred at no cost and has no impact on contracts. In contrast, the agent's contract is no longer the same when the supervision technology is highly imperfect and blackmail occurs. In this case, giving incentives to the agent to make her work has a higher cost than when the supervisor is truthful. That is, the agent's no-shirking constraint becomes more costly to fulfill. This is due to the fact that if the agent decides to work she has a low probability that her effort level will be observed and, in case of control, the supervisor blackmails her. Allowing abuse of authority has therefore a higher cost than deterring it. Consequently, when the supervision technology is highly imperfect and the supervisor is venal and blackmails the agent, the principal deters abuse of authority. He does so by increasing the agent's contingent wage $\widehat{w}$ so that $\widehat{w} = \overline{w}$. The principal then gives a flat wage w to the agent whether $r = \emptyset$ or $r = 1$. Accordingly, the existence of abuse of authority in hierarchies expands the range of contractual incompleteness. Since the supervisor can therefore no longer blackmail the agent, using (offering) aggregate information (an incomplete contract to the agent) deters abuse of authority. This policy implies that the agent's expected utility increases when the supervisor is venal and that she is "fully" insured even when $p < \widetilde{p}$. The agent is then the one who benefits from the supervisor's blackmail activity. Thus, although blackmail is originally organized to the detriment of the agent and indirectly to the detriment of the principal, it is the agent who finally captures the informational rent attached to the supervisor's blackmail by enjoying a uniform wage associated with both $r = 1$ and $r \neq \emptyset$ reports. That is, the agent receives a flat wage when $r \neq 0$. When the supervision technology is highly imperfect and the supervisor is truthful (resp. venal), the cost of the three-level hierarchy is $C^t = \zeta + \underline{v} + pu^{-1}(\frac{\varphi}{p}) + (1-p)u^{-1}(\frac{u}{1-p})$ (resp. $C^v = \zeta + \underline{v} + u^{-1}(\frac{\varphi}{p})$). Therefore, when $p < \widetilde{p}$, the effect of the supervisor's abuse of authority on the principal's welfare is given by $C^v - C^t = (1-p)(u^{-1}(\frac{\varphi}{p}) - u^{-1}(\frac{u}{1-p})) > 0$. That is, when the supervision technology is highly imperfect, the existence of abuse of authority decreases the efficiency of the organization. Although the model developed here is a moral hazard one, it is interesting to compare the optimal policy

which deters blackmail in this setting with the one which deters collusion in adverse selection models.[20] In the adverse selection literature on collusion, the principal has the choice between two policies to deter collusion. The first policy is to create incentive payments for the supervisor and the second is to destroy the agent's stake in collusion. In Tirole's (1992) terminology, these policies are termed incentive policy and bureaucratic policy, respectively. The principal's choice is then conditioned by the unofficial transfer rate $k \in (0,1)$. When k is smaller than a threshold value, incentive policy becomes optimal.[21] Incentive policy operates by inducing the supervisor to report truthfully. That is, if we refer to the stake of collusion as the informational rent which is kept by the agent when the supervisor hides the truth, the payment to the supervisor must exceed this stake discounted by the transaction cost of collusion. If we assume that the supervisor has all the bargaining power in the collusion subgame, the collusion rent is then captured by the supervisor. On the other hand, bureaucratic policy works by eliminating the supervisor's discretion or, similarly, by destroying the agent's stake in any collusion with the supervisor. This policy implies lower levels of rent (than in the case without collusion) for the agent in order to reduce the possibility that he will corrupt the supervisor. That is, collusion threat may finally reduce the agent's expected utility. Thus, whenever collusion threat is beneficial to someone, it is usually to the supervisor, that is, to one of the collusion instigators. This conclusion is intuitive since collusion in hierarchies is a mutually advantageous coalition between two employees. In contrast to collusion, blackmail in hierarchies is an abuse of authority and thus is not mutually advantageous. Therefore, the main result of our analysis, that the possibility of blackmail in hierarchies only benefits the victim and not the instigator, is striking. We refer to this phenomenon as the "blackmail paradox" in organizations. It can be explained in the following way. Unlike collusion, blackmail cannot be deterred by making a payment to the supervisor which exceeds the stake of blackmail. That is, blackmail cannot be deterred by raising the supervisor's contingent wage, since blackmail is always beneficial to the supervisor. Thus, the only way to deter blackmail is through the agent's contract. The principal must then use aggregate information. She must pay the agent the same wage whether the supervisor reports $r \neq \emptyset$ or $r = 1$. That is, she offers a flat wage to the agent when $r \neq 0$. This policy destroys the supervisor's stake in any blackmail activity. When she offers the contract $\widehat{w} = \overline{w} = w = u^{-1}(\frac{\varphi}{p})$, the principal deters blackmail *ex ante* since the

[20] Collusion has been considered mainly in adverse selection environments. Exceptions are Macho-Stadler and Pérez-Castrillo (1991, 1998).

[21] For a survey and complete discussion of these issues, see Tirole (1992).

supervisor can no longer blackmail the agent due to the fact that reporting $r \neq \emptyset$ or $r = 1$ results in the same wage being given to the agent. Thus, in the case of blackmail, the incentive policy is ineffective. To deter blackmail, the principal must reduce the supervisor's discretion. Accordingly, in this setting and for any $k \in (0, 1)$, a bureaucratic policy is the only one that deters blackmail. This conclusion may shed new light on the issue of sexual harassment at work. When blackmail takes the form of sexual harassment, it is ethically reprehensible. Thus, preventing abuse of authority by introducing more rules in organizations may also have an ethical dimension. Bureaucracy may therefore have ethical virtues at times.

5 Conclusion

We analyzed a principal-supervisor-agent hierarchy under moral hazard where contracts are contingent only on the supervisor's report on the agent's effort level. The imperfection of the supervision technology allows for information manipulation by the supervisor. We focused on one form of information manipulation, namely blackmail, which has been ignored until very recently. In contrast to the existing literature on extortion and blackmail, we focused on internal blackmail. Unlike collusion, blackmail in hierarchies occurs when the supervisor's information is favorable to the agent. The supervisor then threatens the agent with concealing the truth. We showed that the optimal policy for the principal is to deter blackmail. We found that, paradoxically, blackmail in hierarchies benefits *in fine* the victim (the agent) rather than the supervisor. This is due to the fact that, in contrast to collusion, deterring blackmail can only be done through destroying its stake. This is done by raising the agent's contingent wage in the case of an uninformative report and paying her a flat wage whether the supervisor's report is favorable to the agent or whether it contains no information. Our analysis also revealed that the expansion of the range of contractual incompleteness may be the consequence of the existence of abuse of authority in hierarchies.

Appendix

Proof of proposition 1.

The program $[P_0]$ has the Lagrangian:

$L = p(\overline{w} + \overline{s}) + (1-p)(\widehat{w} + \widehat{s}) - \lambda_1[pu(\overline{w}) + (1-p)u(\widehat{w}) - \varphi - \underline{u}] - \lambda_2[p(u(\overline{w}) - u(\underline{w})) - \varphi] - \lambda_3[p\overline{s} + (1-p)\widehat{s} - \zeta - \underline{v}] - \lambda_4[p(\overline{s} - \widehat{s}) - \zeta] - \lambda_5[p\underline{s} + (1-p)\widehat{s} - \zeta - \underline{v}] - \lambda_6[p(\underline{s} - \widehat{s}) - \zeta] - \lambda_7\underline{w} - \lambda_8\widehat{w} - \lambda_9\overline{w} - \lambda_{10}\underline{s} - \lambda_{11}\widehat{s} - \lambda_{12}\overline{s}$

where λ_i, $i = 1, .., 12$ are non-negative multipliers.

The Kuhn-Tucker conditions for this concave minimization program are:

$$\delta L/\delta \underline{w} = \lambda_2 p u'(\underline{w}) - \lambda_7 = 0 \quad (13)$$

$$\delta L/\delta \widehat{w} = (1-p)(1 - \lambda_1 u'(\widehat{w})) - \lambda_8 = 0 \quad (14)$$

$$\delta L/\delta \overline{w} = p(1 - (\lambda_1 + \lambda_2)u'(\overline{w})) - \lambda_9 = 0 \quad (15)$$

$$\delta L/\delta \underline{s} = -p(\lambda_5 + \lambda_6) - \lambda_{10} = 0 \quad (16)$$

$$\delta L/\delta \widehat{s} = (1-p)(1 - \lambda_3 - \lambda_5) + p(\lambda_4 + \lambda_6) - \lambda_{11} = 0 \quad (17)$$

$$\delta L/\delta \overline{s} = p(1 - \lambda_3 - \lambda_4) - \lambda_{12} = 0 \quad (18)$$

plus the constraints and their complementary slackness conditions.

From the supervisor's no-shirking constraint $p(\overline{s} - \widehat{s}) - \zeta \geq 0$, one has $\overline{s} > 0$ and accordingly, $\lambda_{12} = 0$. One then has $\lambda_3 = 1 - \lambda_4$ from Eq. (18). From Eq. (16), one has $\lambda_5 = \lambda_6 = \lambda_{10} = 0$. Therefore, substituting $\lambda_3 = 1 - \lambda_4$ into Eq. (17) leads to $\lambda_4 = \lambda_{11}$. One possibility is then that $\lambda_4 = \lambda_{11} = 0$ and $\widehat{s} = 0$. It follows that $\lambda_3 = 1$, that is, the supervisor's participation constraint $p\overline{s} + (1-p)\widehat{s} - \zeta - \underline{v} \geq 0$ is binding and therefore $\overline{s} = \frac{\zeta + \underline{v}}{p}$. Similarly, from $\lambda_5 = \lambda_6 = \lambda_{10} = 0$ and $\widehat{s} = 0$, one has $\underline{s} = \frac{\zeta + \underline{v}}{p}$. From the agent's no-shirking constraint,

on has $\overline{w} > 0$ and therefore $\lambda_9 = 0$. From Eq. (13), there are two possibilities to be distinguished: (i) $\lambda_7 > 0$, that is, $\lambda_2 > 0$; (ii) $\lambda_7 = 0$, that is, $\lambda_2 = 0$. If (i) $\lambda_7 > 0$ and thus $\lambda_2 > 0$ and $\underline{w} = 0$, one has $\lambda_1 + \lambda_2 = \frac{1}{u'(\overline{w})}$ and $\lambda_8 = (1-p)(1 - \lambda_1 u'(\widehat{w})) \geq 0$ from Eqs. (14) and (15). Consider 3 possible cases: (a) $\lambda_1 = 0$; (b) $\lambda_1 \in (0, \frac{1}{u'(\widehat{w})})$; (c) $\lambda_1 = \frac{1}{u'(\widehat{w})}$. Cases (a) and (b) imply $\lambda_8 > 0$ and therefore $\widehat{w} = 0$. Case (a) yields a contradiction since one then has $u(\overline{w}) \geq \frac{\varphi + \underline{u}}{p}$ and $u(\overline{w}) = \frac{\varphi}{p}$. Similarly, case (b) yields a contradiction since one has $u(\overline{w}) = \frac{\varphi + \underline{u}}{p}$ and $u(\overline{w}) = \frac{\varphi}{p}$. Case (c) implies $\lambda_8 = 0$ and therefore $\widehat{w} \geq 0$. In this case $\lambda_2 = \left[\frac{1}{u'(\overline{w})} - \frac{1}{u'(\widehat{w})}\right] > 0$. That is, $\overline{w} > \widehat{w}$. $\widehat{w} = 0$ yields a contradiction since

one then has $u(\overline{w}) = \frac{\varphi+\underline{u}}{p}$ and $u(\overline{w}) = \frac{\varphi}{p}$. Accordingly, in case (c), one necessarily has $\widehat{w} > 0$. Furthermore, in this case, the agent's no-shirking and participation constraints are binding and one has $\widehat{w} = u^{-1}(\frac{\underline{u}}{1-p})$ and $\overline{w} = u^{-1}(\frac{\varphi}{p})$. Since one must have $\overline{w} > \widehat{w}$, this contract is only feasible if $p < \widetilde{p} = \frac{\varphi}{\varphi+\underline{u}}$. If (ii) $\lambda_7 = 0$ and thus $\lambda_2 = 0$ and $\underline{w} \geq 0$, one has $\lambda_1 = \frac{1}{pu'(\overline{w})}$ and $\lambda_8 = (1-p)(1-\lambda_1 u'(\widehat{w})) \geq 0$ from Eqs. (14) and (15). Consider 2 possible cases: (a) $\lambda_1 \in (0, \frac{1}{u'(\widehat{w})})$; (b) $\lambda_1 = \frac{1}{u'(\widehat{w})}$. Case (a) implies $\lambda_8 > 0$ and therefore $\widehat{w} = 0$. This case yields a contradiction since one then has $\lambda_1 = \frac{1}{pu'(\overline{w})} < \frac{1}{pu'(0)}$. Case (b) implies $\lambda_8 = 0$ and accordingly, $\widehat{w} \geq 0$. $\widehat{w} = 0$ yields a contradiction since one then has $\lambda_1 = \frac{1}{pu'(\overline{w})} = \frac{1}{pu'(0)}$. In this case one necessarily has $\widehat{w} > 0$ and therefore $\lambda_1 = \frac{1}{pu'(\overline{w})} = \frac{1}{pu'(\widehat{w})}$. It follows that $\widehat{w} = \overline{w}$. Then, from the agent's binding participation constraint one has $w = \widehat{w} = \overline{w} = u^{-1}(\varphi + \underline{u})$. Since the agent's no-shirking constraint requires $0 \leq u(\underline{w}) \leq u(w) - \frac{\varphi}{p} = \varphi + \underline{u} - \frac{\varphi}{p}$, one must also have $\varphi + \underline{u} - \frac{\varphi}{p} \geq 0$. This contract is then only feasible if $p \geq \widetilde{p}$.

Proof of proposition 2.

The program $[P_1]$ has the Lagrangian:

$L = p(\overline{w}+\overline{s})+(1-p)(\widehat{w}+\widehat{s})-\lambda_1[u(\widehat{w})-\varphi-\underline{u}]-\lambda_2[p(u(\widehat{w})-u(\underline{w}))-\varphi]-$
$\lambda_3[p(\overline{s}-\widehat{s})+\widehat{s}+pk(\overline{w}-\widehat{w})-\zeta-\underline{v}]-\lambda_4[p(\overline{s}-\widehat{s})+pk(\overline{w}-\widehat{w})-\zeta]-$
$\lambda_5[p\underline{s}+(1-p)\widehat{s}-\zeta-\underline{v}]-\lambda_6[p(\underline{s}-\widehat{s})-\zeta]-\lambda_7\underline{w}-\lambda_8\widehat{w}-\lambda_9\overline{w}-\lambda_{10}\underline{s}-$
$\lambda_{11}\widehat{s}-\lambda_{12}\overline{s}-\lambda_{13}[\overline{w}-\widehat{w}]$

where λ_i, $i = 1,..,13$ are non-negative multipliers.

The Kuhn-Tucker conditions for this concave minimization program are:

$$\begin{aligned}
\delta L/\delta\underline{w} &= \lambda_2 pu'(\underline{w}) - \lambda_7 = 0 && (19)\\
\delta L/\delta\widehat{w} &= (1-p) - \lambda_1 u'(\widehat{w}) - p[\lambda_2 u'(\widehat{w}) - k(\lambda_3+\lambda_4)] \\
-\lambda_8 + \lambda_{13} &= 0 && (20)\\
\delta L/\delta\overline{w} &= p(1-k(\lambda_3+\lambda_4)) - \lambda_9 - \lambda_{13} = 0 && (21)\\
\delta L/\delta\underline{s} &= -p(\lambda_5+\lambda_6) - \lambda_{10} = 0 && (22)\\
\delta L/\delta\widehat{s} &= (1-p)(1-\lambda_3-\lambda_5) + p(\lambda_4+\lambda_6) - \lambda_{11} = 0 && (23)\\
\delta L/\delta\overline{s} &= p(1-\lambda_3-\lambda_4) - \lambda_{12} = 0 && (24)
\end{aligned}$$

plus the constraints and their complementary slackness conditions.

From Eq. (22), one has $\lambda_5 = \lambda_6 = \lambda_{10} = 0$. One also has $\widehat{w} > 0$ and $\overline{w} > 0$ from the agent's participation and no-shirking constraints and from the constraint $\overline{w} \geq \widehat{w}$. Therefore, $\lambda_8 = \lambda_9 = 0$ and Eq. (21) can be rewritten as $\lambda_{13} = p(1 - k(\lambda_3 + \lambda_4))$. Since $\lambda_3 + \lambda_4 \leq 1$ from Eq. (24) and $k < 1$, one has $\lambda_{13} > 0$, that is, $\widehat{w} = \overline{w} = w$. From the supervisor's no-shirking constraint $p(\overline{s} - \widehat{s}) - \zeta \geq 0$ one has $\overline{s} > 0$ that is, $\lambda_{12} = 0$. One then has $\lambda_3 = 1 - \lambda_4$ from Eq. (24). Substituting $\lambda_3 = 1 - \lambda_4$ into Eq. (23) leads to $\lambda_4 = \lambda_{11}$. One possibility is then that $\lambda_4 = \lambda_{11} = 0$ and $\widehat{s} = 0$. It follows that $\lambda_3 = 1$, that is, the supervisor's participation constraint is binding and $\overline{s} = \frac{\zeta + v}{p}$. Similarly, from $\lambda_5 = \lambda_6 = \lambda_{10} = 0$ and $\widehat{s} = 0$, one has $\underline{s} = \frac{\zeta + v}{p}$. From Eq. (19), two cases are to be distinguished: (i) $\lambda_7 > 0$, that is, $\lambda_2 > 0$; (ii) $\lambda_7 = 0$, that is, $\lambda_2 = 0$. If (i) $\lambda_7 > 0$ and thus $\lambda_2 > 0$ and $\underline{w} = 0$, one has $\lambda_1 + p\lambda_2 = \frac{1}{u'(w)}$ from Eqs. (20) and (21). From the agent's binding no-shirking constraint, one has $w = u^{-1}(\frac{\varphi}{p})$. As one must have $\lambda_1 \geq 0$, the agent's participation constraint requires $u(w) \geq \varphi + \underline{u}$, that is, $\frac{\varphi}{p} - \varphi + \underline{u} \geq 0$. This contract is then only feasible if $p \leq \widetilde{p}$. If (ii) $\lambda_7 = 0$ and thus $\lambda_2 = 0$ and $\underline{w} \geq 0$, one has $\lambda_1 = \frac{1}{u'(w)}$ from Eqs. (20) and (21). From the agent's binding participation constraint, one has $w = u^{-1}(\varphi + \underline{u})$. Since the agent's no-shirking constraint requires $0 \leq u(\underline{w}) \leq u(w) - \frac{\varphi}{p} = \varphi + \underline{u} - \frac{\varphi}{p}$, this contract is then only feasible if $p \geq \widetilde{p}$.

References

Bernheim, D., Winsthon, M., 1998. Incomplete contracts and strategic ambiguity. American Economic Review 88, 902-932.

Dickens, William T., Katz, Lawrence F., Lang, Kevin and Lawrence H. Summers, 1989. Employee crime and the monitoring puzzle. Journal of Labor Economics 7, 331-347.

Elster, J., 1996. Rationality and the Emotions. Economic Journal 106, 1386-1397.

Frank, R., 1987. If Homo Economicus could choose his own utility function, would he want one with a conscience ? American Economic Review 77, 593-604.

Frank, R., 1988, Passions within reason: the strategic role of the emotions, (New York: W.W. Norton and company).

Geanakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological games and sequential rationality. Journal of Games and Economic Behavior 1, 60-79.

Hindriks, J., Keen, M., Muthoo, A., 1996. Corruption, extortion and evasion. Cahiers de Recherche, Facultés universitaires Notre-Dame de la Paix, Namur.

Hirshleifer, J., 1987, On the emotions as guarantors of threats and promises, In: John Dupré (Ed.), The latest on the best. MA: MIT Press, Cambridge, pp. 307-326.

Huang, P.H., Wu, H.-M., 1992. Emotional responses in litigation. International Review of Law and Economics 12, 31-44.

Human Rights and Equal Opportunity Commission: *Annual Report, 1989-90,* 1990*,* Australian Government Publishing Service, Canberra.

Husbands, R., 1992. Sexual harassment law in employment: An international perspective. International Labour Review 131, 535-559.

Klein, D.B., O'Flaherty, B., 1993. A game-theoretic rendering of promises and threats. Journal of Economic Behavior and Organization 21, 295-314.

Kofman, F., Lawarrée, J., 1993. Collusion in hierarchical agency. Econometrica 61, 629-656.

Konrad, K.A., Skaperdas, S., 1997. Credible threats in extortion. Journal of Economic Behavior and Organization 33, 23-39.

Laffont, J.J., 1988. Hidden gaming in hierarchies: facts and models. Economic Record 10, 295-306.

Laffont, J.J., 1990. Analysis of hidden gaming in a three-level hierarchy. Journal of Law, Economics and Organization 4, 301-324.

Laffont, J.J., Rochet, J.C., 1997. Collusion in organizations. Scandinavian Journal of Economics 99, 485-495.

Laffont, J.J., Tirole, J., 1993, A Theory of incentives in procurement and regulation, MIT Press, Cambridge, MA.

Lazear, E.P., 1989. Pay equality and industrial politics. Journal of Political Economy 97, 561-580.

Macho-Stadler, I., Pérez-Castrillo, D., 1991. Double risque moral et délégation. Recherches Economiques de Louvain 57, 277-296.

Macho-Stadler, I., Pérez-Castrillo, D., 1998. Centralized and decentralized contracts in a moral hazard environment. Journal of Industrial Economics XLVI, 489-510.

Milgrom, P., 1988. Employment contracts, Influence activities, and efficient organization design. Journal of Political Economy 96, 42-60.

Prendergast, C., Topel, R.H., 1996. Favoritism in organizations. Journal of Political Economy 104, 958-978.

Scharfstein, D., Stein, J., 1990. Herd behavior and investment. American Economic Review 80 , 465-479.

Shleifer, A., Vishny, R.W., 1989. Management entrenchment: The case of manager-specific investments. Journal of Financial Economics 25, 123-139.

Spier, K.E., 1992. Incomplete contracts and Signalling. Rand Journal of Economics 23, 432-443.

Symposium BLACKMAIL, 1993, University of Pennsylvania Law Review 141.

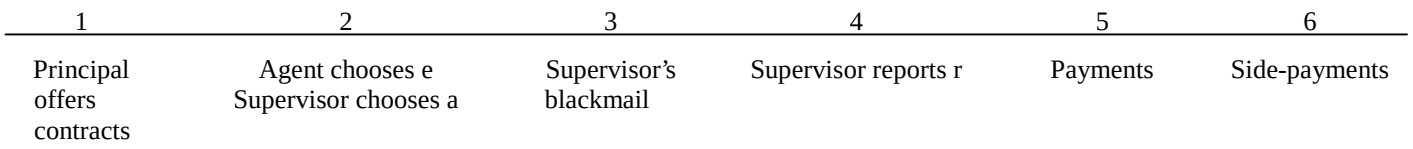
Tirole, J., 1986. Hierarchies and bureaucracies: On the role of collusion in organizations. Journal of Law, Economics, and Organization 2, 181-214.

Tirole, J., 1992, Collusion and the theory of organizations, in: Laffont, J.J. (ed.), Advances in Economic Theory, Vol. 2, Cambridge University Press, 151-206.

Vafaï, K., 1998. Collusion and blackmail in hierarchies. Mimeo, CIRANO, Montréal.

Zwiebel, J., 1995. Corporate conservatism and relative compensation. Journal of Political Economy 103, 1-25.

Figure


1
2
3
4
5
6
Principal offers contracts
Agent chooses e Supervisor chooses a
Supervisor's blackmail
Supervisor reports r
Payments
Side-payments

Liste des publications au CIRANO *

Cahiers CIRANO / *CIRANO Papers* (ISSN 1198-8169)

96c-1 Peut-on créer des emplois en réglementant le temps de travail ? / Robert Lacroix

95c-2 Anomalies de marché et sélection des titres au Canada / Richard Guay, Jean-François L'Her et Jean-Marc Suret

95c-1 La réglementation incitative / Marcel Boyer

94c-3 L'importance relative des gouvernements : causes, conséquences et organisations alternative / Claude Montmarquette

94c-2 Commercial Bankruptcy and Financial Reorganization in Canada / Jocelyn Martel

94c-1 Faire ou faire faire : La perspective de l'économie des organisations / Michel Patry

Série Scientifique / *Scientific Series* (ISSN 1198-8177)

99s-07 A Theory of Abuse of Authority in Hierarchies / Kouroche Vafaï

99s-06 Specific Investment, Absence of Commitment and Observability / Patrick González

99s-05 Seasonal Nonstationarity and Near-Stationarity / Eric Ghysels, Denise R. Osborn et Paulo M. M. Rodrigues

99s-04 Emerging Markets and Trading Costs / Eric Ghysels et Mouna Cherkaoui

99s-03 Sector-Specific Training and Mobility in Germany / Lars Vilhuber

99s-02 Women's Wages in Women's Work: A US/Canada Comparison of the Roles of Unions and 'Public Goods' Sector Jobs / Michael Baker et Nicole M. Fortin

99s-01 Globalization and Worker Welfare in Late Nineteenth Century Europe / Michael Huberman et Wayne Lewchuk

98s-41 Dynamic Equilibrium with Liquidity Constraints / Jérôme Detemple et Angel Serrat

98s-40 Monetary Policy Rules with Model and Data Uncertainty / Eric Ghysels, Norman R. Swanson et Myles Callan

98s-39 Environmental Risks: Should Banks Be Liable? / Karine Gobert et Michel Poitevin

98s-38 Better Observability Promotes the Adoption of More Flexible Technologies / Marcel Boyer, Armel Jacques et Michel Moreaux

98s-37 Job Characteristics, Wages and the Employment Contract / W. Bentley MacLeod et Daniel Parent

98s-36 Observability, Commitment and Flexibility / Marcel Boyer, Armel Jaques et Michel Moreaux

98s-35 Pricing ang Hedging Derivative Securities with Neural Networks and a Homogeneity Hint / René Garcia et Ramazan Gençay

98s-34 Gender Composition and Wages: Why Is Canada Different from the United States? / Michael Baker et Nicole M. Fortin

98s-33 Effects of the Fiscal Treatment of Tax Losses on the Efficiency of Markets and the Incidence of Mergers / Michel Poitevin

---

* Vous pouvez consulter la liste complète des publications du CIRANO et les publications elles-mêmes sur notre site World Wide Web à l'adresse suivante :
http://www.cirano.umontreal.ca/publication/page1.html