"Money Or Reputation" - A Dynamic Model of Blackmail[1]

Paper under revision

Mikko Leppämäki[2]

Helsinki School of Economics

June 5, 2015

[1] I am grateful for Patrick Bolton for his comments and all the help. I also wish to thank Mathias Dewatripont, Leonardo Felli, Guido Friebel, Joe Harrington, Ian Jewitt, Ariane Lambert-Mogiliansky, Espen Moen, Carolyn Pitchik, Michael Raith, Martin Ruckes and Hannu Salonen and the seminar audiencies in Bergen, Copenhagen, ECARES, Humbold (EEA), Leuven(EARIE), Montreal(NASM), SNF-Oslo, Stockholm School of Economics, and University of Helsinki for their comments. This paper is a revised version of Chapter 4 of my Ph.D. dissertation at the London School of Economics. Any remanining erros are my own.

[2] **Address:** Graduate School of Finance, Helsinki School of Economics, P.O. Box 1210, FI-00101 Helsinki, Finland. e-mail: mikko.leppamaki@hse.fi

Abstract

This paper develops a dynamic model of blackmail, where a piece of information an agent prefers to keep private may facilitate blackmail when another agent, namely a blackmailer, threatens to reveal that information. The crucial feature of the blackmail game is the commitment problem from the blackmailer's side. The blackmailer can not commit not to come back in future to demand more despite the payments received in the past. The paper outlines conditions under which successful extortion may arise, and shows that there is a blackmail equilibrium, which gives a precise prediction how much money the blackmailer is able to extort from the victim. It is also shown that the blackmailer receives a blackmail premium that compensates the blackmailer for not taking money from the victim and revealing information anyway.

JEL Classification: D82, D23, C72

Keywords: Blackmail, extortion, collusion, organisations, game theory

1 Introduction

"Even if you pay the blackmail, that is no guarantee that he won't demand more later." (The Oxford Thesaurus. An A-Z Dictionary of Synonyms.)

"The blackmailed person to Madson: "How can I trust the boy that he won't come back to ask more money?" Madson: Don't worry, he is just a kid, he is not a professional blackmailer." (Madson, BBC 1, 1996.)

What is the economics behind the phenomenon of blackmail? Should a victim pay a blackmailer or not? How much, if anything, should the victim pay the blackmailer? When is the blackmailer's threat to carry out his action credible? What if the blackmailer comes back to ask for more money? These are among the questions the present paper tackles.

In recent years relatively much has been written about collusion (bribery) in organisations.[3] Very little thought has been given to blackmail, which, in general, forms a complementary part of corruption. The concept of blackmail in legal and sociological literature has originally been used to refer to payments to avoid physical harm; today it primarily refers to payments to avoid revelation of discreditable information.[4] This is precisely how blackmail is modelled in the present paper. Even though collusion and blackmail are closely related, there is, however, an important difference. Under collusion two parties (e.g., a manager and an auditor) act together and collude against a principal (e.g., the shareholders), for example, by agreeing on information manipulation. They enter the collusive relationship voluntarily, and after successful collusion they are both better off than by not colluding. In the case of blackmail, a blackmailer operating alone, is able to hurt a victim. The blackmailer extorts the victim by threatening to reveal a piece of information which the victim prefers to keep private. The relationship between the blackmailer and the victim is involuntary and takes the form of pure extortion. Furthermore, after successful blackmail the blackmailer is better off, but the victim worse off compared to the case of no blackmail.

[3] See, for instance, Tirole(1992) and Laffont (1997) on collusion in organizations.

[4] See Hepworth (1980), who discusses several aspects of the phenomenon of blackmail from a sociological point of view.

This paper develops a dynamic model of blackmail. It takes into consideration the blackmail game, where a piece of information which an agent prefers to keep private may facilitate blackmail when the other agent, namely the blackmailer, is able to reveal that information. The crucial feature of the blackmail game is the commitment problem from the blackmailer's side. The blackmailer cannot commit not to come back in future to demand more despite the payments received in the past.[5]

The infinite horizon blackmail game has a very simple structure. Two players move sequentially in every period. First the victim decides how much to hand over as a blackmail payment, and then the blackmailer decides whether to reveal or suppress a piece of information about the victim. The equilibrium concept we use is a Markov Perfect equilibrium. That is, we consider strategies that are conditioned only on the payoff-relevant variables, and not on the entire history of the game.

We analyse first a situation where there exist no rewards for information revelation. This includes the cases where, for example, the tabloid press do not pay rewards for scandal stories, an organisation designer does not pay rewards to organisation members who turn in fellow workers, and so on. In this case it is shown that there is a Markov Perfect equilibrium where blackmail does not arise, and the victim pays nothing, and the blackmailer suppresses information.

Next we introduce rewards for information revelation, and consequently then the blackmailer has two potential buyers for a piece of information: the victim or the tabloid press. When a piece of information is revealed to the tabloid press, the game ends. On the contrary, if the information is suppressed, the game continues, and the blackmailer comes back in the next period to demand more money. In this case we show that in an equilibrium blackmail is an issue. The victim pays the blackmailer and information is not revealed. This Markov Perfect equilibrium gives us a precise prediction how much money the blackmailer will get by extorting the victim.

Our results remain the same in case where the blackmailer announces a

[5] Imagine, for example, a case where the blackmailer's evidence is a videotape. The victim can not be sure when buying the tape that the blackmailer does not have a copy of the tape. If the blackmailer has a copy, he can come back to ask for more money.

blackmail demand in a beginning of each period. What is important for the outcome is that the blackmailer always moves after the potential blackmail payment, and the blackmailer has an option to end the game by revealing the information. This, by the way, works partially against the blackmailer, who is unable to fully exploit the victim. Interestingly, the blackmail payment does not depend on the victim's valuation of the information, but only on the external net reward.

Due to the blackmailer's commitment problem, blackmail appears in the form of an infinite stream of small blackmail payments, rather than in the form of a large one-off payment in the beginning of the relationship. Since there is nothing which forces the blackmailer to suppress the information, even though he has received money from the victim, the optimal blackmail payment turns out to be a combination of the offered external net reward and a blackmail premium. The premium compensates the blackmailer for not taking money from the victim and revealing information anyway. The case when potential rewards are the blackmailer's private information is also considered. It is shown that information about the victim will be revealed with positive probability.

Recently some authors have also considered collusion in a dynamic framework. Acemoglu (1995) develops a dynamic model of implicit collusion between an auditor and a manager. Martimort (1996) in turn has proposed a model of self-enforcing collusion by modelling a static adverse selection problem as an infinitely repeated game. However, none of those papers examines the phenomenon of blackmail and the question of how the benefits of information suppression will be shared. update.

The outline for the rest of this paper is as follows. Section 2 describes the model. The analysis is carried out and the main results are provided in sections 3.1 and 3.2. Some interpretations and potential extensions are discussed in section 3.3, while section 4 concludes.

2 The Model

2.1 The players

Consider a model with two agents: a blackmailer (B) and a blackmailed person (V); we refer to V also as the victim. V has on his possession a piece of information (I), which he prefers to keep private: i.e., not to share with others. The monetary value of V's privacy (or, say, reputation) is equal to $v(I)$ and $v(I) = 0$ if information is revealed. In particular, we assume that the value of privacy is a per-period benefit. Later, we use v as a short-hand notation for $v(I)$. The time horizon we consider is infinite. We also assume that information revelation causes the victim a permanent damage. That is, the victim never get back his privacy.

The blackmailer, B also has access to a piece of information I, and is able to reveal it to a third party. The cost of information revelation for B is a fixed cost, $c > 0$. We assume, for the sake of simplicity, that B also learns v.[6] The blackmailer does not derive any direct utility himself from releasing discreditable information about V, and B hopes that the victim will pay him to suppress the information. In his article Hepworth (1990) puts this nicely: "At the heart of reputational blackmail lies the willingness of the blackmailer to exploit the victim's desire to prevent others sharing a secret".

The model we consider has many potential interpretations. Within an organisation, V may be a civil servant who has taken bribes from a contractor, and B is another civil servant who has observed this and blackmails V. Alternatively, V may be a politician who has had an "affair", and B is a person who blackmails the politician by threatening to reveal that information to the tabloid press. This latter example is more closely related to the very idea of reputational blackmail ("newspaper blackmail") on which we want to concentrate here.

So far the model has a simple feature where one agent prefers his privacy, and the other agent is able to provide that by remaining silent. The blackmail game has a very intuitive and familiar interpretation, namely that of a seller

[6] This is not a restricting assumption, since it is shown later on that the optimal blackmail payment does not depend on the victim's valuation at all.

and a buyer. Here the blackmailer (the seller) is able to provide the victim (the buyer) a service by remaining silent. The question is how much the buyer has to pay for this service. Or, to put it otherwise, we can even state that trade has already occurred, because the victim has on his possession an infinite surplus stream from privacy: $v(I)/(1-\delta)$, where δ is a common discount factor. Now the question becomes, how will the benefits from privacy be shared?

Note that in contrast to a normal seller-buyer model, the blackmailer is not needed here to create a surplus, since it already exists. However, the blackmailer is able to destroy the surplus permanently simply by revealing the information. Note that if the blackmailer reveals the information, he basically then also destroys his only asset, and certainly will not get anything out of the relationship in future.

The trade here is not a one-off event, since the victim's payment to the blackmailer does not end the game. In every period, given that no information has been revealed in the past, a new surplus arrives and the blackmailer is around demanding a share of it. Before describing the blackmail game, we make the following assumptions.

Assumption 1: No contracts can be conditioned on whether the blackmailer comes back in future to demand more money.

Assumption 1 is a crucial one, and states the fundamental commitment problem we believe captures the very essential feature of blackmail. The blackmailer can not commit not to come back in future to demand more money. That is, the blackmailer is unable to commit not to exercise a profitable action. For example, this is the case of unfortunate and well-documented practices among small businesses who have to pay "protection money" to criminals just to be able to run their businesses. In those cases blackmail exists exactly in the form of "small" blackmail payments criminals collect from small businesses, say every week, rather than in a form of a large one-off payment. The model presented here describes therefore also racketeering.[7]

Assumption 2: No contracts can be conditioned on information suppres-

[7] Interestingly, Konrad and Skaperdas (1997) do not consider this feature in their extortion analysis.

sion.

According to assumption 2 such contracts are not enforceable. The only reason why the blackmailer may suppress information is that it is individually rational to do so. That is, even if the victim hands over money, it is no guarantee that the blackmailer will suppress the information. This last point further emphasizes the fact that the relationship between the victim and the blackmailer is a game and not an enforceable contract.

How much money the blackmailer is able to extort from the victim is what we call blackmail, and a blackmail payment is labelled as m. There are presumably many alternative ways to model blackmail, and we have chosen to stick with that of dynamic games.[8]

2.2 Timing, Strategies, and the Solution concept

As stated above, the time horizon we consider is infinite ($T = \infty$). Within every period of the blackmail game, and given that information has not been revealed, there are two sequential stages (see figure 1). First, the victim decides how much to hand over as a blackmail payment (m). In stage 2, the blackmailer decides whether to reveal (r) or suppress (s) the information.[9] The blackmailer either accepts V's offer and suppresses the information, or rejects it and reveals the information. Since the blackmailer pockets m in any case, we can further simplify the blackmailer's action in each period. He either reveals or suppresses information. If B suppresses information, the game continues and in a next period stages 1 and 2 are repeated, and so on. Information revelation ends the blackmail game immediately in that very period, because after that there is no valuable information left anymore. It is in the blackmailer's hands whether the game continues or not.

The introduction of an external reward (R) has a crucial effect on the black-

[8] The Rubinstein-Ståhl model could provide another way to proceed. However, there an acceptance by one party ends the game, which is not valid here. There a contract between players is an enforceable contract; here the relationship is a non-cooperative game.

[9] Later, it is shown that this order of moves is preferred by both the victim and the blackmailer. The blackmailer does not want to move first, but he prefers to move after the vicitm has handed over the blackmail payment.

mail game, since then the blackmailer is able to sell a piece of information also to the third party. We do not incorporate the third party explicitly in the model, and it is assumed that when offered, the exogenously given R is available in every period to the blackmailer, who exposes discreditable information about the victim.

Rather than finding all the perfect equilibria of the blackmail game, we restrict ourselves to the Markov Perfect Equilibrium (MPE). That is, we consider only Markov strategies, where player i's strategy in period t does not depend on the whole history of the game, but only on the variables that affect its present period's payoffs.[10] Once we have constrained the set of strategies to Markov, the structure of the game becomes stationary, and on the condition that information has not been revealed earlier, the blackmail game looks similar in every period.

Consider next the players' strategies. The victim moves first by handing out the blackmail payment m_t, which is chosen from interval $[0, v]$. The blackmailer's strategy space is discrete: he either suppresses (s) information or reveals (r) it:$b_t \in \{s, r\}$. Define the available net reward for the blackmailer in the beginning of period t as $\theta_t = (R_t - c)$. We say that θ_t describes "a state of the system" in period t.

Define the history of the game in period t as $h_t = \{(m_1, b_1, \theta_1), (m_2, b_2, \theta_2), ...(m_{t-1}, b_{t-1}, \theta_{t-1})\}$.Note that the game has a history in period t only when no information has been revealed earlier. In period t the only aspect of the history that directly affects the victim's action in the present period is θ_t. Interestingly, since the external net reward is exogenously given, θ_t depends only on what B did in period $t-1$. Therefore if $b_{t-1} = r$, then the game would have ended, and if $b_{t-1} = s$, then $\theta_t = (R_t - c)$. Thus, the victim's strategy depends only on θ_t: $m_t(\theta_t)$. Due to the sequential timing of moves in each period, the aspect of the history that directly affects the blackmailer's payoff is θ_t, but the blackmailer's payoff relevant history also includes m_t, the blackmail payment handed over by the victim in that very period: $b_t(m_t, \theta_t)$. Moreover, the reactions by the victim and the blackmailer do not depend on the calendar time, but only on the state variable and the state of the system. The victim's

[10] See Maskin and Tirole (1997) for the Markov Perfect Equilibrium.

strategy is then a reaction function: $m_t = R(\theta_t)$, and the blackmailer's strategy is: $b_t = R(m_t, \theta_t)$.

Both the victim's and the blackmailer's objective is to maximize the present discounted value of their payoffs:

$$\sum_{t=1}^{t=\infty} \delta^{t-1} \pi_t^V(m_t, b_t) \tag{1}$$

$$\sum_{t=1}^{t=\infty} \delta^{t-1} \pi_t^B(m_t, b_t) \tag{2}$$

Define V^V and V^B respectively as net present values of being the victim and the blackmailer at the beginning of a period. Now we can use dynamic programming and write both the victim's and the blackmailer's payoffs (intertemporal profits from period t onwards) as valuation functions:

$$V_t^V = Max_{m_t} \pi_t^V(m_t, b_t) + \delta V_{t+1}^V. \tag{3}$$

$$V_t^B = Max_{b_t} \pi_t^B(m_t, b_t) + \delta V_{t+1}^B. \tag{4}$$

Now we are in a position to write down the payoffs in terms of the strategies. The victim's payoffs in the case of information suppression and revelation are respectively:

$$V_t^V = \pi_t^V(m_t, s) + \delta V_{t+1}^V = (v - m_t) + \delta V_{t+1}^V. \tag{5}$$

$$V_t^V = \pi_t^V(m_t, r) = -m_t. \tag{6}$$

If the blackmailer suppresses the information in period t, the victim's payoff is the net benefit $(v - m_t)$ plus the continuation value. In the case of information revelation, the victim loses his reputation immediately as well as the payment m_t, which the victim handed over. Note that this means that the victim pays a blackmail payment out of his pocket, and that v is realised at the end of each period only in the case of information suppression. The blackmailer's payoffs are:

$$V_t^B = \pi_t^B(r, m_t) = (m_t - c) + R_t. \tag{7}$$

$$V_t^B = \pi_t^B(m_t, s) + \delta V_{t+1}^B = m_t + \delta V_{t+1}^B. \quad (8)$$

In the case of information suppression, the blackmailer pockets m_t immediately, and he gets the continuation value as well. If the blackmailer decides to reveal the information he bears a cost of revelation, and receives $(m_t - c) + R_t$ and the blackmail game ends immediately in that very period.

3 The Analysis

3.1 No Rewards for Information Revelation

This section considers a case where no rewards exist for an agent who reveals some discreditable information in public. In practice this includes cases where the tabloid press pays no rewards for scandal stories, an organisation designer does not reward workers who disclose information about fellow workers' activities, and so on. The case of "no rewards" further clarifies the idea that the blackmailer does not enjoy any direct utility from releasing discreditable information about the victim.

The blackmail game proceeds as follows. In every period t, given that information has not been released earlier, the victim hands over mt to the blackmailer. The blackmailer's strategies are then simply: to suppress (s) or reveal (r) the information. That strategy corresponds to the decision of ending the game or letting it to continue.

In the following we argue and prove that in the case of "no rewards", blackmail will not arise. The victim pays nothing and the blackmailer suppresses the information. In order to derive an optimal blackmail payment, we have to find out the blackmailer's optimal responses to an arbitrary blackmail payment. It is shown that B's best response to any blackmail payment is to suppress the information. And, of course, then the victim optimally pays nothing for information suppression.

Consider period t and the blackmailer to whom the victim has handed over m_t. If the blackmailer suppresses the information, he receives m_t and the game continues. If B reveals the information, the game ends immediately in period t, and the blackmailer receives $(m_t - c)$. Note that once the information is revealed the blackmailer can not come back in future. In fact, a piece of information is like an asset for the blackmailer, and revelation of the information destroys this asset. Thus we have

Lemma 1 *Given that the victim has handed over a blackmail payment, the blackmailer's best response is to suppress information.*

Proof. Suppose not, and assume that B reveals the information, thus ending the game immediately. In that case B gets $(m_t - c) < m_t + \delta V^B$, which he would get by suppressing information. Clearly, it is optimal for the blackmailer to suppress the information, and we have a contradiction, and information suppression is the blackmailer's optimal response. ■

Consider next in turn the victim's problem of choosing m_t, when information has not been released earlier. Note that if the information had been released in period $t-1$, the game would have ended and V's action would be irrelevant. In the following we show that an optimal offer $m^* = 0$, which makes the blackmailer suppress the information and the game continues. Therefore we have:

Lemma 2 *Given that the blackmailer suppresses information after any arbitrary blackmail payment, the victim's optimal response is $m^* = 0$.*

Proof. Suppose not, and assume that the victim hands over $m > 0$, given that the blackmailer always suppress the information. In this case the victim receives $(v - m) + \delta V^V < v + \delta V^V$,which the victim receives when he pays $m = 0$. Thus, we have a contradiction, and the victim's optimal and unique best response is $m* = 0$. ■

The main result in the case when there are no rewards is:

Proposition 1 *In the case of "no rewards" there is a unique MPE of no blackmail. The victim pays nothing and the information is suppressed; $m^* = 0, V^V = v/(1-\delta), V^B = 0$.*

Proof. We want to show that a pair of strategies, $\{m* = 0 ands\}$ forms an equilibrium. We refer here to "one-period deviation - principle". By playing his equilibrium strategy, $m* = 0$, the victim gets: $v/(1-\delta)$, which is greater than $(v - \epsilon) + v/(1-\delta)$, which is what he would get by deviating and handing over $m = \epsilon$, and then conforming to the equilibrium strategy. By playing the equilibrium strategy, the blackmailer receives: $0 + V^B > 0 - c$, which he would get by deviating and revealing the information. In sum, neither the victim nor the blackmailer prefers to deviate from the equilibrium path. ■

The economics behind Proposition 3 is strikingly simple and yet it confirms the intuition Hepworth (1980) provides: "If the blackmailer is unable to persuade the victim he has access to a receptive audience he is powerless to commercialise or gain any other kind of advantage from the information which has fallen in his hands." In short, when a piece of information is not valuable to anyone apart from the victim, the blackmailer is unable to extort money from the victim.

Note that it is not important for the above result that there is a positive cost of information revelation. Even if c went to zero, information suppression would remain the blackmailer's optimal response, since by suppressing information he would get $0 + V^B \geq 0$. More precisely, information revelation is then weakly dominated by information suppression. And after the elimination of dominated strategies, information suppression remains a weakly dominant strategy.The victim would choose $m^* = 0$, the blackmailer would suppress information, and the "no-blackmail equilibrium" would survive.

So far we have assumed that the victim moves first, and the blackmailer is passive, and only reacts to the blackmail payment. Would the blackmailer ever prefer to move first? It is clear that this is never the case, since by moving first the blackmailer can not receive anything which he does not get by waiting for the victim's blackmail payment. The blackmailer does not want to move first, but waits until the victim has handed over the payment, and then moves. But here also the blackmailer is a "passive player". Since we consider blackmail or extortion in terms of how much money the blackmailer is able to extort from the victim, a far more interesting and relevant case for us is a situation where the blackmailer is able to announce a blackmail demand, dt in the beginning of period t. See figure 2 in Appendix. This is actually a very interesting case, since it is a kind of robustness test of Proposition 3.

Suppose now that B moves first, and announces a blackmail demand, d_t. How does the victim react to the blackmailer's demand? He can either accept it and pay dt or he can reject it. Even if he rejects the blackmailer's demand, the victim still has to decide how much to hand over as a blackmail payment. But now the set-up is exactly identical to the case where the victim moved first, and of course it is optimal to pay the same amount as before, $m^* = 0$. And for the blackmailer it is optimal to suppress the information. Consequently, no matter what the blackmailer's demand is, the victim pays only $m^* = 0$.

The logic behind the result remains the same even though the blackmailer moves first. What is crucial for the result is that he also moves last. He moves after the blackmail payment has been handed over. As the last mover the blackmailer has an option to end the game by revealing information. Here the last mover has a disadvantage, since he will always prefer that the game continues, and this destroys the credibility of his threat of terminating the game by revealing information if his blackmail demand is not matched. Consequently, we have Corollary 4:

Corollary 1 *If the blackmailer moves first by making a blackmail demand $d_t >$ $m^*$the result of no blackmail holds. The victim pays nothing, and the information will be suppressed.*

Corollary 2 *If the blackmailer moves first by making a blackmail demand ($d_t >$ m*), the result of no blackmail holds. The victim pays nothing, and the information will be suppressed.*

Proof: The logic is exactly similar to that of the proof of Proposition 3. Assume that B has announced a blackmail demand $\mathrm{d}_t >$ m*. V can pay this, in which case he gets (v- d_t) + $\delta \mathrm{V}^V$. Alternatively, he can reject B's demand and pay m* = 0, which is the optimal payment he would choose when moving first. From earlier we know that B's optimal response to that is to suppress the information, which gives V: v + $\delta \mathrm{V}^V$ >(v- d_t) + $\delta \mathrm{V}^V$. Therefore, no matter what B asks, the victim hands over m*, and B suppresses the information. □

This an important and interesting result. Indeed, the "no-blackmail" result will hold even if the blackmailer moves first by announcing a blackmail demand. The reason behind this is that the blackmailer is unable to make use of his threat of information revelation, since he is the agent who moves last. In particular, he moves after the victim has handed over m*. In effect, this resembles a situation where the victim makes a "take it or leave it" offer, and where the last mover has a disadvantage, since it is optimal for him to accept the payment and suppress the information.

3.2 Rewards for Information Revelation

In this section we consider whether the "no-blackmail" result will hold if there is a reward for information revelation. In practice, for example, the tabloid press do pay rewards for "scandal stories"; similarly an organisation designer may reward a worker who exposes wrongdoing in the workplace, and so on. Here we analyse whether the victim now can decline to pay anything, and how much money the blackmailer is able to extort from the victim.

Note that now the blackmailer can sell a piece of information also to a third party, which values information as well. However, there is an important difference whether a piece of information is sold to the tabloid press or to the victim. In the former case, revelation of the information will end the blackmail game, since the blackmailer no longer has valuable information on his possession. That is, the value of his asset has disappeared. In the latter case, there is nothing to keep the blackmailer from returning back to demand more money in future despite the payments he received earlier. From the victims's point of view, it is not optimal to hand over a blackmail payment which matches the reward offered by the third party, since the blackmailer would be back asking for more in the very next period.

Recall that the relationship between the victim and the blackmailer is a game, and not an enforceable contract. Therefore, there is nothing to keep the blackmailer from taking money from the victim and revealing the information anyway. The reward for information revelation represents here the blackmailer's opportunity cost of not releasing information.[11] Before proceeding any further, we make the following assumption:

[11]The role of an external reward is quite similar to that of the outside option in the Rubinstein-Ståhl Bargaining model. here, however, the external reward is available to the blackmailer in every period, even after the victim's payment.

Assumption 3*:* An external reward, R is small compared to the damage of information revelation; $(R - c) < v/(1 - \delta)$.

If Assumption 3 were violated, and (R-c) was greater than the damage, the blackmailer would not bother to blackmail, but would end the blackmail game by revealing the information and collecting the net reward in the very first period.[12]

We start again by considering B's best responses to V's offer. Now the victim has to take into consideration that if he does not pay the blackmailer, the blackmailer may reveal the information, in which case the victim loses his reputation immediately. In the next lemma we derive an optimal payment m* which makes the blackmailer indifferent between releasing information and suppressing it.

Lemma 3 *There exists m*=(R-c)(1-δ)/δ such that if V offers m < m*, B's optimal response is to reveal the information. Alternatively if V offers m $\geq$m*, B's optimal response is to suppress the information.*

Proof. First we derive an optimal payment m*. Assume that V offers m, and the blackmailer suppresses the information. In this case B gets: m + δV^B. If B reveals the information, he keeps m that V handed over, but he also receives a net reward (R-c). The payoff for information revelation becomes: m + (R-c). The victim's problem is to choose m he hands over to the blackmailer in every period such that B is just indifferent between suppressing and revealing the information: m + δV^B = m + (R-c). That is, m + δm + δ^2m + ... = m + (R-c); and after some manipulation, this gives m* = (R-c)(1-δ)/δ. The claim is that if V's offer m< m*, then B's best response is to reveal the information. Assume that V hands over (m*-ϵ)<m*. If B suppresses the information, it yields a payoff : (m*-ϵ) + δV^B = (m*- ϵ)/(1- δ). Alternatively, if B reveals the information, the payoff is: (m*-ϵ) + (R-c). Now it is clear that B will reveal the information, since (m*-ϵ) + (R-c) > (m*-ϵ) /(1- δ), where m*= (R-c)(1-δ

[12] Interestingly, in this case it might also happen that the victim sells his own story, since that would be the profit-maximizing strategy. Presumably, this is why we see in practice that occasionally celebrities sell their own (scandal) stories to the tabloid press, and indeed this is rational behaviour.

$)/\delta$. Alternatively, if the victim offers m$\geq$ m* - say, for example, m = m* - the blackmailer's best response is to suppress the information. ■

An important economic insight follows immediately from Lemma 5: the blackmailer will remain silent as long as the victim hands over $m^* = (R-c)(1-\delta)/\delta$. If the victim deviates from this by handing less than m*, the blackmailer will reveal the information; and this ensures him a payoff equal to that of the continuation value. In an equilibrium path, V is willing to pay B for information suppression, and the blackmailer is willing to suppress the information. We have a case of successful extortion.

The victim's problem of choosing an optimal offer m* becomes now quite straightforward, since he knows that the blackmailer will reveal the information and end the game if m m*. Thus, we have Lemma 6:

Lemma 4 *Given that the blackmailer reveals the information if $m < m^*$ and suppresses it if $m \geq m^*$, then the victim's optimal response is to hand over m^*.*

The proof of Lemma 6 is included in the proof of Proposition 7 below, and is thus omitted here.

The main result in the case when there is an external reward for information revelation is stated as Proposition 7:

Proposition 2 *When there is a reward for information revelation, there is a MPE of blackmail. The victim pays the blackmailer m^* in every period, who suppresses the information; $m^* = (R-c)(1-\delta)/\delta, V^V = (v-m^*)/(1-\delta), V^B = m^*/(1-\delta)$.*

Proof. We want to show that a pair of strategies {(m*-ϵ) , r} and {m*, s} forms an equilibrium. Once again we refer to the "one deviation only-principle". Given that V and B play equilibrium strategies, we show that neither player prefers to deviate. The victim's equilibrium payoff is: (v-m*) + δ(v-m*) + δ^2(v-m*) + ... = (v-m*)/(1-δ), where m* = (R-c)(1-δ)/δ. Suppose that he deviates from his equilibrium strategy by handing over (m*-ϵ). Then, for a moment, the victim potentially has on his possession v-(m*-ϵ) from that period. However, we know that in this case B will reveal the information immediately. The victim loses his reputation, and the blackmail game ends; and the victim's payoff is:

- (m*-ϵ). The victim does not want to deviate. The blackmailer's equilibrium payoff is: (m*) + δ(m*) + δ^2(m*) + ... = (m*)/(1-δ). If B deviates and reveals the information, he pockets m* + (R-c) and the game ends. However, by playing his equilibrium strategy, he receives: m* + $V^B \geq$ m* + (R-c), and $V^B \geq$ (R-c)/δ. The blackmailer has no incentive to deviate from equilibrium. ■

Note that a blackmail payment $m^* = (R-c)(1-\delta)/\delta$ does not depend on the victim's valuation (v) at all. It only depends on the external net reward and a discount factor . Now we can see that m* increases in (R-c) and decreases in δ, the latter meaning that when the blackmailer becomes less patient, the optimal blackmail payment increases. Interestingly, the blackmailer is able to get more money by extorting the victim than by selling the information directly to a third party. The intuition behind this is that the victim has to compensate the blackmailer for not taking money (m*) and revealing the information anyway, and thus we have;

Corollary 3 *In an equilibrium path the blackmailer gets more by extorting the victim than the potential net reward (R-c) he would get by selling a piece of information directly to the third party.*

Proof. We know that $m* = (R-c)(1-\delta)/\delta$. The flow payment that would match the external net reward is $(R-c)(1-\delta)$, which we label as M. Now $m*/M = 1/\delta$. And $1/\delta > 1$, since $\delta < 1$. ■

We see from Proposition 8 that the more impatient the blackmailer is, the bigger is the premium which the victim has to pay over the external net reward.

As in the earlier section, we see immediately that the blackmailer prefers for the victim to move first. By moving after the victim, the blackmailer is able to get $(R-c)/\delta$. By moving first the blackmailer is able to get at most (R-c). Then it is clear that the blackmailer, as well as the victim, prefers that the victim moves first and the blackmailer last.[13]

[13] Notice that there is of course another equilibrium, where the victim pays nothing and the blackmailer always reveals. That is, the victim pays nothing in the first period and the blackmailer reveals the information, and therefore the game ends. However this equilibrium is Pareto dominated by the blackmail equilibrium, and theferfore we concentrate only on that equilbrium.

What if the blackmailer is able to announce a blackmail demand before the victim moves? Intuitively, now the blackmailer should be in a stronger position, since he can make profits by revealing information to the third party if the victim declines to pay. One would expect that the blackmailer would get more out of the victim than in the case considered so far. Interestingly, and against intuition, this is not the case. In arguments which are very similar to those in section 3, it can be shown that the equilibrium payoffs do not change. Irrespective of what the blackmailer demands, the victim hands over $m^* = (R - c)(1 - \delta)/\delta$, and the blackmailer will suppress the information. He does not get any more out of the victim even though he moves first by announcing a blackmail demand. And thus we have:

Corollary 4 *When there exists a reward and the blackmailer demands,* $d_t > m^*$*, then in an equilibrium path the result of Proposition 7 holds: the victim pays* m^* *in every period and the information is suppressed.*

Proof. Assume that B moves first by announcing a blackmail demand $d_t > m^*$. The vicitm can either pay it, and then his payoff is: $(v - d_t) + \delta V^V$, since B suppresses the information. Alternatively, the vicitm can reject the demand, and hand over less than demanded d_t. Suppose that the victim hands over $(d_t - \epsilon) > m^*$. How will the blackmailer react? If B reveals the information, he gets: $(d_t - \epsilon) + (R - c)$. However, if he suppresses the information he receives:$(d_t - \epsilon) + V^B$. Now, we know that since $m^* = (R - c)(1 - \delta)/\delta$, we can write $(R - c) = m^* \delta/(1 - \delta)$. Then, by revealing the information, B gets: $(d_t - \epsilon) + m^* \delta/(1 - \delta)$. And now since the victim is ready to pay $(d_t - \epsilon)$ in every period for information suppression, the blackmailer gets $(d_t - \epsilon) + \delta(d_t - \epsilon)/(1 - \delta)$ by suppressing the information. Clearly, it is a dominant strategy for the blackmailer to suppress the information, since $(d_t - \epsilon) + \delta(d_t - \epsilon)/(1 - \delta) > (d_t - \epsilon) + m^* \delta/(1 - \delta)$. But now it is evident that V can lower his payment all the way down to m* - i.e., the optimal blackmail payment which the victim chose when moving first. And from earlier we know that the blackmailer will suppress the information. This concludes the proof.□ ■

So far we have seen that in an equilibrium, information is not revealed, and the blackmailer is able to extort money from the victim only in the case when there exists a third party which values the information as well. In the next section we consider whether the fact that information about the potential reward is the blackmailer's private information will change this result, and, in particular, whether a blackmailer who does not have access to an external reward is able to extort money from the victim.

4 Discussion: Some Applications

4.1 Escalation of Corruption in Organisations

It is a well-known fact that corruptive activities in organisations often escalate from one level to another.[14] Here we want to point out how bribery in the first place may lead blackmail. In the following we assume that bribe-taking is against the rules in the civil service, and that the penalty for being caught is that a civil servant is fired.

Imagine a simple organisation, say a government agency with three members and an outside contractor. The members of the organisation are two civil servants (A and B) and their superior civil servant (P), whose preferences are the same as the government's. A's task is to choose an outside contractor who will supply material to the government agency and to approve the quality of delivered materials. Suppose that the contractor bribes A in one way or another in return for accepting low-quality material. That is, the government agency pays the contractor according to the standard price, the contractor's profit margins are higher due to the less costly low-quality materials it supplies, and A is bribed by the contractor. On the whole, the government is losing money, which is going into the contractor's and A's pockets. Note that bribery here is not just a redistribution of wealth, but it has serious economic consequences as well,

[14] See, for instance, Basu et al (1992), who consider the controlling of corruption when corruption may escalate. See also Carrillo(1995), who also considers corruption in a case where there exist potentially dishonest agents at several levels of a hierarchy.

since the government is worse off due to the low-quality materials. Assume now that B, the second civil servant, observes the bribery with probability one. And after observing it, B may start to blackmail A. Will he reveal information about the bribery to their superior or nor? We keep the assumption of a small cost, c due to information revelation.

We consider first the case where the organisation designer has not offered a reward for a member of the organisation who exposes any wrongdoing. We know from earlier that in this case B will not reveal the information, since the net reward is negative. A pays nothing to B, who, however, suppresses the information about the bribery. The bribery between the contractor and A takes place, but here blackmail is not an issue. The contractor and A end up being better off. The economic consequences are severe, since the government loses money and receives low-quality materials.

Assume now that there is a reward $R > c$ for a whistle-blower who exposes bribery. Now it is clear from our earlier analysis that blackmail may arise, since B is able to reveal information about bribery, which will also bring him a positive net reward. Hence, bribery in the first place facilitates blackmail, and nobody blows the whistle. In short, the corrupted civil servant buys silence from the initially honest civil servant, who becomes a blackmailer. The contractor, A and now also B end up being better off, and again the government is worse off. Compared to the earlier case of no reward, here the second civil servant, B also gets his share of bribes. At the aggregate level the economic consequences of bribery are the same as above.

Note that the possibility for blackmail would arise even if the first civil servant had been corrupted only once. What is required for the successful blackmail is that A has been corrupted at least once, and that there is a positive net reward for the civil servant who blows the whistle. In that case blackmail arises, redistributing the wealth between A and B. Note that if the organisation designer wants to prevent corruptive activities altogether, he has to destroy the roots of corruption in the first place.

4.2 Implications for the Organisation Design

From the theoretical literature we have learned a reasonable number of measures organisations may adopt in preventing collusion. Among them is rotation of workers. The idea is that rotation prevents collusion, since it blocks long-term relationships that are essential for collusion to be feasible.

Interestingly, rotation alone does not necessarily prevent blackmail; rather it helps the blackmailer, since now the blackmailer can credibly commit not to come back to ask for more in the very next period. Assume that a worker is able to blackmail only when being an employee in a position to expose a wrongdoer. In other words, a reward for whistle-blowing is available only in the period when corruption has taken place.

We introduce here rotation by assuming that B, the second civil servant, is rotated in every period. Therefore, A meets a different civil servant in each period. In effect, rotation changes the relationship between A and B to a one-shot game.

What happens when B is rotated and there is no reward for whistle-blowing? The only change is that A and B play a one-shot game. Since information revelation is costly, we know that B will not reveal his information, and thus A pays nothing to B. Therefore, A will be bribed by the contractor, and blackmail is not an issue. The economic consequences are the same as in the case of no rotation. That is, the government loses money and receives low-quality materials, and the contractor and A are better off.

Consider now the case where the organisation designer rewards a whistle-blower. In this case rotation has an effect, and corruptive activities will be prevented altogether. To see this, note that if A does not pay B, B will certainly blows the whistle. And if A pays B, the blackmailer will reveal the information in any case. In short, B can not commit not to blow the whistle after any bribe paid by A. Now B's threat to reveal information about the bribery is credible. Therefore, A knows that B will expose him in any case, and thus A won't get involved in bribery in the first place.

Here the possibility of blackmail is beneficial from the organisation's point of view, since it prevents all corruptive activities. Rotation and a positive net reward are a powerful combination in preventing corruption. The economic consequences are as follows. A does not get involved in bribery, the government receives high-quality materials, and blackmail is not an issue. The organisation designer does not have to pay rewards, since nobody blows the whistle. The simple economics behind this result is the very powerful last period effect due to rotation. B can commit not to come back, but he can not commit not to reveal even after A has paid him.

4.3 A Creditor and a Firm

Consider a firm that has borrowed amount D from a creditor to realise a research and development project. When raising the debt, the firm has to disclose and share valuable information about the project with the creditor. The debt contract defines the repayment schedule, where the firm agrees to make fixed payments p in each period. If the firm does not make the payment, the creditor is able to end the project by liquidating the assets. Liquidation of a means the same as selling or revealing information to an interested third party - for example, a competitor. That is, the creditor is able to destroy the firm's potential profit stream in the case of default.[15]

Now two interesting questions arise. First, how much will the creditor get back from the firm? Secondly, what is the smallest amount the firm has to pay the creditor so that the creditor does not end the relationship by liquidating the assets? Assume that in period t the firm, for one reason or another, makes a payment $m_t < p$. What does the creditor do? In principle, he has two alternatives. The creditor can either end the game by liquidating the firm with a small cost c. In this case, the creditor gets $m_t + (L - c)$ immediately and nothing in future. Alternatively, the creditor can accept the smaller payment, and let the project go ahead. In this case the creditor gets: $m_t + V^C$, where V^C is the creditor's continuation value. Clearly, what is optimal for the creditor depends on $(L - c)$.

Suppose first that the liquidation value is low. Then, of course, the firm's

[15] We do not consider the possibility that the firm is excluded from the credit market in future, and we laso rule out enforceable credit contracts.

position is now stronger and it has to pay only the project's termination value, and the creditor lets the project go ahead. Assume an extreme case, $L = 0$, where no third party values the intangible assets of the s option works against him, since he will always prefer to continue the game and wait one more period for the payments. In this extreme case, the firm pays nothing, and the creditorcan not do anything but let the project go ahed.

Here the last mover, the creditor, has an option to end the game by revealing information to third party - i.e. , by liquidating the R&D project. This option works against him, since he will always prefer to continue the game and wait one more period for the payments. For the creditor it is better to accept lower payments than to terminate the relationship by liquidation, since the continuation value of the project is always greater or equal to the termination value. The model above seems to belong to a class of models that have in common a so-called last-mover disadvantage.

Gromb (1994) considers repeated lending between a creditor and a borrower. In his analysis a creditor, the last mover whose decision is whether or not to refinance a project, gets zero surplus, which is a return for the termination of a project. This is due to the fact that a creditor cannot fully commit to terminate the relationship if the borrower does not meet his repayment, since it is mutually beneficial for them to write a new contract under which both are better off.

In the present model the creditor gets a termination value as well, which is (L-c). That is the value of the R&D project to the interested third party minus the cost of revelation. But, in addition to that, here the last mover (the creditor) receives a premium, which compensates him for not taking money from the firm and selling information about the project to the third party in spite of the payment. To see this consider next a case where the liquidation value is higher. As earlier, the amount the firm has to pay the creditor depends on (L-c). Interestingly, it can be shown that here the firm may have to pay more than the agreed fixed payment p. Recall that no enforceable debt contracts are in place, and the firm has to hand over the payment that makes the creditor indifferent between liquidating the firm (revealing information) and letting the project go ahead (suppressing information).[16] That is, it has to be the case

[16] Alternatively, if we assume that the creditor is able to liquidate only if p<d, then p is the

that: $m_t + \delta V^C = m_t + (L - c)$,and thus $m^* = (L - c)(1 - \delta)/\delta$. Therefore, it may well happen that the flow payment $m^* > p$. The firm has to pay more than the agreed fixed payment p.

In this latter case, the creditor is able to blackmail the firm, since the creditor is able to sell information about the Of course, the possibility of blackmail here is quite an extreme case, but presumably not totally unrealistic in the world of project financing. Perhaps this is one of the

5 Concluding Remarks

This paper has considered the phenomenon of blackmail in a simple dynamic framework. In particular, we have considered the question of how the potential surplus due to the victim's privacy is going to be shared between the victim and the blackmailer. We have shown that there is a blackmail equilibrium which gives a precise prediction how much money the blackmailer will get by extorting the victim. Interestingly, and against intuition, it depends only on the external reward, and not the victim's valuation of his privacy. Furthermore, we have shown that the blackmailer gets more money by extorting the victim than by selling his piece of information directly to the interested third party. This follows from the fact that in the former case the victim pays a blackmail premium to the blackmailer for not taking the victim's money and revealing the information anyway. It was also shown that even if the blackmailer is able to move first by announcing a blackmail demand, the blackmailer is unable to get more money from the victim than he does in the case when the victim moves first. This counter-intuitive result follows from the fact that in each period, irrespective of who moves first, the blackmailer always moves last - i.e., after the victim has handed over the blackmail payment. Despite the blackmailer's demand, the victim pays only as much as he would pay when moving first, and after this optimal payment the blackmailer prefers to suppress the information. In this sense, the model has one feature of a last mover disadvantage, and the blackmailer is unable to fully exploit the victim.

In future it might be worthwhile to examine thoroughly how an introduc-

upper limit of the payment the firm has to pay.

tion of the coexistence of a rational and an insane blackmailer would change the results of the present paper. Here we have considered only a "rational blackmailer" who does not reveal information if it is unprofitable for him. In practice, of course, there may also exist "insane blackmailers" who will reveal information even if it is costly. However, it is a well-known fact in game theory literature that the introduction of an insane player may change the results greatly. Most obviously, here the victim is worse off, and both blackmailer types are better off. However, this potential extension does not add much to the analysis of reputation and imperfect information by Kreps and Wilson (1982). The main difference with Krep-Wilson is that here the blackmailer who reveals information in the very first period will end the game immediately, which is not the case in Kreps and Wilson (1982).

A more interesting case for further study would be to develop a model that fully integrates collusion and blackmail. Also, it would be interesting to look at whether the results presented here will carry on into the bargaining literature. For example, it would be interesting to try to incorporate the idea of blackmail with the present assumptions into the alternative offers' bargaining model. How much would the blackmailer get from the victim there? What would determine the shares the bargainers get? Is there a blackmail premium? These are among the open questions left for future studies.

References

[1] Acemoglu, D. 1995: Career Concerns and Cover-ups:A Dynamic Model of Collusion, mimeo.MIT, Depatment of Economics.

[2] Basu, K., Bhattacharya, S. and Mishra, A. 1992: Notes on Bribery and the control of corruption, *Journal of Public Economics 48.* 349 - 359.

[3] Carrillo, J. 1995: Corruption in Hierarchies, mimeo. IDEI, Toulouse.

[4] Gromb, D. (1994): Renegotiation of Debt contracts, mimeo. MIT, Sloan School of Management.

[5] Hepworth, 1980: Modern Blackmail, *Scottish Journal of Sociology 4.* 291 - 299.

[6] Kofman, F. and J. Lawarre 1993: Collusion in Hierarchial Agency, *Econometrica 61.* 629 - 656.

[7] Konrad, K. and S. Skaperdas 1997: Extortion, *Economica 65,* 461 - 477.

[8] Konrad, K. and S. Skaperdas 1998: Credible Threats in Extortion, *Journal of Economic Behavior & Organization 33*, 23 - 39.

[9] Kreps, D and R. Wilson 1982: Reputation and Imperfect Information, *Journal of Economic Theory 27.* 253 - 279.

[10] Laffont, J.-J. and J. Tirole, J. 1991: The Politics of Government Decision-Making: A Theory of Regulatory Capture, *Quarterly Journal of Economics 106.* 1089 - 1127.

[11] Laffont, J.-J. 1997: Collusion in Organisations, *Scandinavian Journal of Economics 99. 485 - 495.*

[12] Lambert - Mogiljansky, A. 1994: Corruption in Procurement- The Economics of Regulatory Blackmail. Research Papers In Economics 1994:4. University of Stockholm.

[13] Madson 1996: TV Drama serie presented at the BBC1.

[14] Martimort, D. 1996a: A Theory of Self-Enforceable Collusion in Organizations, mimeo. IDEI, Toulouse.

[15] Maskin, E. And J.Tirole 2001: Markov Perfect Equilibrium, I:Observable Actions, *Journal of Economic Theory ###*

[16] Mogiljanski, A. 1994: Corruption in Procurement- The Economics of Regulatory Blackmail. Research Papers In Economics 1994:4. University of Stockholm.

[17] The Oxford Thesarus. An A-Z Dictionary of Synonyms.1994.The Oxford University Press. Oxford.

[18] Tirole, J. 1992: Collusion and The Theory of Organisations. In Advance in Economic Theory. eds. J.-J. Laffont, Proceedings of the 6th World Congress of the Econometric Society. Cambridge University Press.