Reputation When Threats and Transfers Are Available*

Ernesto Dal Bó — *UC Berkeley*

Pedro Dal Bó — *Brown University*

Rafael Di Tella — *Harvard Business School*

December 30, 2005

Abstract

We present a model where a long run player is allowed to use both money transfers and threats in her attempts to influence the decisions of a sequence of short run players. We show that threats might be used credibly (even in arbitrarily short repeated games) by a long-lived player who gains by developing a reputation of carrying out punishments. This model comprises as particular cases that of a long-lived pressure group offering rewards and punishments to a series of targets (public or corporate officials) in exchange for policy favors, or that of a long-lived extorter who demands money in order not to punish. We use the model to analyze the "convicted non-payor" debate. We show that a corrupt judge has incentives to develop a reputation of passing worse than deserved sentences on defendants that do not pay bribes to her. The model brings out formal similarities between phenomena like lobbying and extortion.

JEL Classification: D72, D78, K42, L14.

Keywords: Lobbying, Extortion, Threats, Credibility, Judicial Corruption, Activism.

*Corresponding author: Ernesto Dal Bó, UC Berkeley, Haas School of Business, 545 Student Services Building #1900, Berkeley, 94720 CA. E-mail: dalbo@haas.berkeley.edu.

1 Introduction

In this paper we study a long run player seeking to affect the decisions of a finite series of short run players by using both transfers and threats. Of course, a problem with threats is that they may not be credible. We construct an equilibrium in which the long run player develops a reputation for carrying out her threats as in Kreps and Wilson (1982). But contrary to previous studies of reputation, we find that the existence of transfers allows the long run player to profit from her reputation more easily than without transfers. Even for an arbitrarily small amount of asymmetric information and a small number of short run players (as low as two), the long run player obtains positive profits from reputation. In addition the model shows how transfers and threats can be used simultaneously and highlights the formal connections between seemingly unrelated phenomena, such as lobbying and extortion. This suggests that–at least in a formal dimension–the difference between interest groups and mafias may be one of degree rather than nature.

A particular case of this model is that of a long-lived lobby or pressure group facing several public or corporate officials from whom a particular policy position is demanded. The ability of the pressure group to credibly use threats saves on bribe payments or any other kind of rewards. Pressure groups may threaten targets (a government, a firm, or individual officials) with various forms of damage. Relevant examples are physical violence, legal harassment, smear campaigns in the media, or boycotts and other forms of direct action campaigns. To the extent that carrying out such retaliations is costly, delivering on a threat is a dominated strategy in the stage game. In our repeated game, however, reputational concerns will push pressure groups to keep their word. The model helps generate a taxonomy of pressure groups and activist members depending on the strength of the punishments utilized. This could help organize discussion around the issue of what we understand by radical groups along the dimension of tactics rather than agendas. Lastly, the model predicts that reputation-enhancing elements will be valuable to the group. Examples of such elements are members

with known contacts with media and judicial players, which may aid the organization of smear and legal harassment strategies. Also, activist groups may want to enroll members with known track records in direct action campaigns. This may explain why certain groups feature the militant trajectory of their leaders prominently, and why we should expect to see a labor market for activists and lobbyists where track records matter.

Another particular case of the model is that of a long-lived judge extorting money from defendants in exchange for not passing a worse than deserved conviction. Under this interpretation our model settles a vital legal debate regarding the "convicted non-payor problem." This is the problem of defendants that did not pay bribes to a corrupt judge and got heavier-than-deserved penalties. Ayres (1997) holds that convicted non-payors should be granted new trials when the convicting judge is found to have been on the take. Other important voices in the legal world, such as Judge Posner's, disagree on the grounds of doubting that a judge taking bribes from payors should have imposed worse-than-deserved sentences on non payors. One reason for this happening indeed is informally put forward by Ayres (1997): a judge seeking to extract higher bribes from future defendants has an incentive to develop a reputation for "toughness" on those defendants that do not pay bribes. Our model formally demonstrates exactly this proposition even for cases in which the judge may have a short time horizon or a low initial reputation for toughness. The fact that the perversion of justice involved might imply death sentences is good enough a reason to want to have this debate settled.[1]

The structure of the paper is as follows. Section 2 reviews some related literature. Section 3 presents our model of influence through both rewards and punishments. Section 4 constructs a Sequential Equilibrium in which threats become endogenously credible. Section

[1]More on this in Section 6 where we deal with the application of our model to this legal debate. As for now let us borrow from Ayres (1997) who speaks about a judge found to have been extorting defendants: "*At this point it might be useful to point out that Judge Maloney sentenced more people to death than any other judge in Cook County.*"

5 presents the problem of the a "nasty" pressure group and explains the relationship between the use of threats and cheap bribes. Section 6 presents the problem of judicial extortion and describes the legal debate around it. It is shown how the model can be applied to generate predictions that settle the debate. Section 7 concludes.

2 Related literature

The traditional literature on political influence focuses on interest groups that, in their attempts to influence officials, use bribes exclusively (see for instance Bernheim and Whinston, 1986, and Grossman and Helpman, 1994). Dal Bó and Di Tella (2003) go to the opposite extreme in that they study the consequences of a pressure group not using bribes, but only threats, in its attempts to affect policymaking. The involved punishments might go from the relatively 'mild' (such as damaging the official's reputation and career prospects through, say, the creation of unfavorable rumors[2]) to the very serious, such as legal harassment which might end in jail, or to direct physical violence. Dal Bó, Dal Bó and Di Tella (2002) bridge the gap left open by the opposition between models where groups use either bribes or threats, as it integrates the use of both instruments by a pressure group. The two aforementioned papers displaying threats, however, assume that threats are credible. Threats are also important in the literature on private politics (see Baron 2001, and Baron and Diermeier 2005), concerning activist groups that seek to alter corporate policy. The models in this literature also make progress by abstracting away from credibility considerations. In this paper, the credibility of threats is no longer assumed, but is endogenously generated by administering punishments. One implication of our analysis is that previous work involving threats is

[2]Leaver (2002) shows how regulated firms can affect the decisions of regulators when able to disclose information that might damage the regulator's future employment opportunities. She finds that the shorter the regulators' office terms (the higher the career concerns) the fewer the rate of return reviews and the higher the residential electric prices across US states.

robust to a dynamic extension where threats are not assumed credible.

The model of repeated interaction we use is an extension of that in Kreps and Wilson (1982). We allow the (finitely lived) long run player an extra (continuous) instrument: money transfers (which could be either bribe offers or requests). This is appealing since the assumption that only punishments are available is restrictive; in most instances players will have the ability to make transfers, so we would like to know how this affects reputation building. The model thus extended allows us to study the combined use of rewards and threats. It also changes the problem of the long run player in a way that allows her to benefit from the possibility of building up a reputation even in short games. In Kreps and Wilson's paper, for any given value of the initial prior, the game has to be long enough for the possibility of reputation building to have a positive impact on the long run player's payoff.

Other authors have noted that extortionary pressures run into a credibility problem. Konrad and Skaperdas (1997) study a model where the extorter commits to carry out punishments by sinking costs in advance. We take the comparatively harder road of assuming that no such up front commitment moves are possible. In our model reputation arises exclusively from honoring one's promises. Smith and Varese (1994) explore a related model with only two periods, analyzing the possibility that fake mafia members might mimic real mafiosi when attempting to charge for protection. Our model provides a general multiperiod framework and has a more generic extensive form highlighting the formal connections between lobbying and the functioning of mafias.[3] Shavell and Spier (2002) study the credibility of threats for cases in which the threatening party faces either a short run incentive or a short run cost to carry out the punishment. For the first case and in an infinite horizon game in which the threat can only be carried out once, they show that the threatener can credibly commit to not carry out the punishment if she receives the payment.

[3] On the mafia, see Gambetta (1993) and Varese (1994).

We show that both rewards and punishments are used by interest groups in equilibrium and that the use of punishments saves on rewards. This can help explain a phenomenon that has puzzled political economists who have observed unequal exchange of favors to often take place in the political arena (see, for example, Tullock 1997).[4] Finally, our model is related to work in conflict theory (see for instance Garfinkel 1990, Skaperdas 1992, Powell 1993, and Hirshleifer 2001), where the possibility of coercion has traditionally been included.

3 The model

There are $N+1$ players in this game: one long run player (LR) and N short run ones, with the typical short run player denoted SR_n, $n = 1, 2, ..., N$. The long run player sequentially plays the game in Figure 1 against one short run player at a time, starting with SR_N and ending with SR_1, following the reputation literature convention.

Nature determines the type of both LR and the short run players. LR knows that every short run player she meets is "sane" with probability α and "crazy" with probability $1-\alpha$. Moreover, LR is "tough" with probability p_N and "weak" with probability $1-p_N$. The short run players observe all previous play but do not know the type of LR. They may have beliefs regarding her type and we denote the belief held by SR_n as p_n.

At stage n, both types of LR can offer a transfer (which can be positive or negative) denoted by t_n in exchange for a favor from SR_n. A sane SR_n may accept (A_n) or reject (R_n) the offer. A crazy SR_n can only reject the offer. If the sane SR_n accepts, LR earns a profit π from the deal and pays the transfer t_n (actually earns the transfer if t_n is negative). If SR_n rejects the offer, LR has access to a punishment technology (Pu_n), which, at a fixed cost c, inflicts a utility loss of size r on the short run player.

[4] For other models where little or no money changes hands in the political influence process, see Helpman and Persson (2001) and Leaver and Makris (2001).

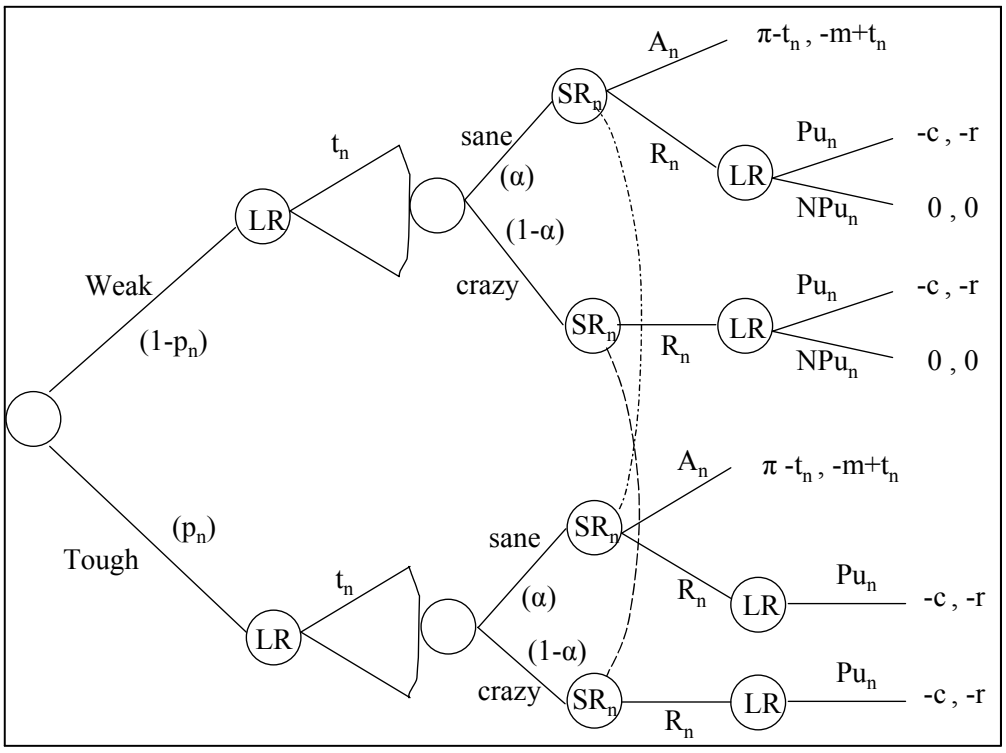

Figure 1: Stage Game

Sane short run players receive a utility of $t_n - m$ for accepting the offer of LR (where m is the moral cost of being bought or the expected cost of being caught) and a disutility of r if they are punished by LR after rejecting her offer. We assume that c and r are strictly positive. For analytical simplicity we assume that $\pi \geq m$ and $\alpha r > c$.[5]

As an example of this game, think of an industry lobby asking officials from different regions to issue a permit allowing the realization of a socially undesirable project in each region. The project yields the industry a return π for each territory where it is allowed. The industry lobby can use both bribes and threats to convince each official to issue the permit. In Section 6 we show how the model can also be used to explain the credibility of threats in blackmail and extortion. For expositional simplicity, while describing and discussing the

[5]The solution of the model without this last assumption is similar to the one we will show and is available from the authors upon request.

equilibrium we will refer to transfers as money paid by the long run player to the short run players. As we will show, however, whether the long run player offers or requests payments is endogenously determined.

4 Reputational Equilibrium

In this section we present a sequential equilibrium of the game in which a reputation effect may arise. While there may be other equilibria in which a reputation effect arises, the equilibrium presented here is of great simplicity since p_n is a sufficient statistic of the history of the game. Therefore, strategies can be written solely as a function of beliefs in each stage (and the stage number) instead of as a function of past history. In addition only LR eventually randomizes.

Consider the following profile of actions and beliefs:

Strategies:

(i) For the short run player in period n (SR_n):

When Sane:

If n=1:

A_1 if $t_1 \geq m - p_1 r$, R_1 *otherwise*.

If n>1:

If $p_n < \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$: A_n if $t_n \geq m - p_n \alpha^{n-1} \left(\frac{r}{c}+1\right)^{n-2} \frac{r^2}{c}$, R_n *otherwise*

If $p_n \geq \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$: A_n if $t_n \geq m - r$, R_n *otherwise*

When Crazy: *always* R_n.

(ii) For the long run player (LR):

When Weak:

If n=1:

$t_1 = m - p_1 r$, NPu_1.

If n>1:

If $p_n < \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$: $t_n = m - p_n \alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2}\frac{r^2}{c}$, Pu_n with probability $\frac{p_n}{(1-p_n)c}[\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r - c]$, and NPu_n with probability $1 - \frac{p_n}{(1-p_n)c}[\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r - c]$.

If $p_n \geq \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$: $t_n = m - r$, Pu_n

When Tough: Same transfers as Weak but always Pu_n.

Beliefs:

$p_{n-1} = \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$ if $0 < p_n < \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$and Pu_n

$p_{n-1} = p_n$ if A_n or $[p_n \geq \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$ and $Pu_n]$

$p_{n-1} = 0$ if NPu_n or $p_n = 0$

Proposition 1 *The former strategies and beliefs define a sequential equilibrium of the game.*

Proof. We have to check two things: first, that the beliefs are consistent with the strategies of the players and, second, that each of these strategies is a best response to other players' strategies in any stage of the game given the beliefs.

Beliefs are consistent. In this game we need only to study whether beliefs follow Bayes' rule whenever possible to check that they are consistent with the prescribed strategies.[6] In addition, beliefs will not be modified by equilibrium bribes, since in equilibrium both types of LR offer the same amount of bribe. Besides, we assume that deviations in the amount of transfers offered do not change beliefs.

We assume that once the short run players believe with certainty that the pressure group is weak they do not modify their beliefs: $p_{n-1} = 0$ if $p_n = 0$. If SR_n accepts the bribe no new information about LR is revealed and then there is no update in beliefs: $p_{n-1} = p_n$ if A_n. If SR_n rejects and is not punished, this shows that LR is weak: $p_{n-1} = 0$ if NPu_n.

[6] From Fudenberg and Tirole (1991) we know that the set of perfect Bayesian equilibria of this game coincides with the set of sequential equilibria. This implies that we only need to check that the beliefs conform to the application of Bayes' rule wherever possible.

When $p_n > 0$, punishments may modify beliefs. In any stage in which the weak LR punishes with probability $\beta_n = \frac{p_n}{(1-p_n)c}[\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r - c]$, the posterior beliefs can be calculated using Bayes' rule: $p_{n-1} = P(tough \ / \ Pu_n) = \frac{p_n}{p_n+(1-p_n)\beta_n} = \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$.

Strategies are best responses. We now check that the prescribed strategies for the short run players are best responses. Any sane short run player will accept the bribe only if its amount is enough to pay the moral cost minus the expected cost of a punishment: A_n if $t_n \geq m - P(Pu_n)r$, where $P(Pu_n)$ is the probability of punishment if the bribe is not accepted. This probability is equal to p_n if $n = 1$, while it is equal to $p_n\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2}\frac{r}{c}$ if $n > 1$, leading to the sane short run players' strategies given above. Crazy short run players can only reject offers. Hence, doing so is a best response.

We can also verify that LR is never going to offer more than the prescribed bribe since that would not affect the behavior of the short run players and would only result in a reduction of profits. Note that $\pi \geq m$, so, even at the cost of paying a full bribe m, LR would be (weakly) better off than with a rejection. Thus, LR will always prefer to pay transfers that guarantee acceptance by the sane officials (according to their equilibrium strategy) rather than offering lower amounts that trigger rejections.

Now we check that the prescribed strategy for LR is a best response given every history. To do this we proceed through a number of steps. First, note that a weak LR does not punish rejections in the last period: punishing costs c and yields no return.

Second, we study the case of $n > 1$. We describe a feature of the equilibrium in order to simplify the analysis. Regardless of the plays before n, if the equilibrium strategy prescribes that a weak LR should randomize at n, the strategy will also prescribe randomizations for the following stages until $n = 2$ or until the LR's type is revealed. The reason for this is that if at n a randomization was expected ($p_n < \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$) and the outcome of this randomization was Pu_n, then the posterior $p_{n-1} = \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r} < \frac{c}{\alpha^{n-2}\left(\frac{r}{c}+1\right)^{n-3} r}$ if $\alpha\left(\frac{r}{c}+1\right) > 1$, which is implied by the restriction initially imposed on α. Therefore, if $p_n < \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$, a

weak LR should randomize after a rejection at any stage between n and 2 (or until its type is revealed). We can now calculate inductively a weak LR's payoff in equilibrium for any such period: $U_n^{PG}(p_n) = n\alpha(\pi - m) + p_n\alpha^n\left(\frac{r}{c}+1\right)^{n-1} r$. If SR_n rejects the bribe, LR may punish him, triggering the updating of beliefs, or not punish him, revealing its type. As we calculate in the following equations, both actions yield the same payoff for a weak LR, so that randomization is a best response:

$U_n^{LR}(Pu_n \,/\, R_n,\ p_n) = -c + U_{n-1}^{LR}(p_{n-1}) = -c + (n-1)\alpha(\pi - m) + p_{n-1}\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r =$

$= -c + (n-1)\alpha(\pi - m) + \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r = (n-1)\alpha(\pi - m).$

$U_n^{LR}(NPu_n \,/\, R_n,\ p_n) = 0 + U_{n-1}^{LR}(0) = (n-1)\alpha(\pi - m).$

Now we check that a weak LR is willing to punish rejections with probability one when $p_n \geq \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$. In order to do this, it is enough to analyze the case in which $\frac{c}{\alpha^{n-2}\left(\frac{r}{c}+1\right)^{n-3} r} \geq p_n \geq \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$, i.e. the last stage in which a weak LR punishes a rejection with probability 1. This is so because in the previous periods a weak LR would have an even higher incentive not to reveal its type. If SR_n rejects the bribe a weak LR's payoffs are:

$U_n^{LR}(Pu_n \,/\, R_n,\ p_n) = -c + U_{n-1}^{LR}(p_n) = -c + (n-1)\alpha(\pi - m) + p_n\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r.$

$U_n^{LR}(NPu_n \,/\, R_n,\ p_n) = 0 + U_{n-1}^{LR}(0) = (n-1)\alpha(\pi - m)$

Therefore, Pu_n is the best response given that $p_n \geq \frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$. ■

In equilibrium, a weak LR always pays the short run players the minimum transfers they will accept, and the sane short run players always accept. In the last period a weak LR does not punish a rejection and then must pay a transfer equal to $m - p_1 r$. If $n > 1$, the optimal action of LR depends on the beliefs of the short run player. When the prior of the short run player is greater than $\frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2} r}$, LR punishes for sure a rejection. If the prior probability is lower than that level, a weak LR randomizes. When the outcome of the randomization is not punishing, the long rund player's type is revealed, leading to a posterior $p_n = 0$. When the randomization results in punishment, the short run players' belief that LR is tough is

updated upwards and LR pays lower transfers in the future[7].

Just as it would happen in the model by Kreps and Wilson (1982), if the prior is above certain level, LR wants to punish (fight, in their setting) for sure. If it is below, in equilibrium LR punishes with some probability that allows it to build a reputation for being tough. The farther a short run player is (in the order of play) from the one who will play last, the more likely it becomes that a given initial prior will be high enough for LR to be willing to punish with probability one. Moreover, for any positive initial prior p_N, there is a game which is long enough –i.e. where N is large enough– for a weak LR to be willing to punish for sure.

Figure 2 shows a typical path of play (for $\alpha = 1$, $m = 3$, $c = 1$, $r = 1.3$ and $N = 10$). It shows the dynamics of equilibrium transfers, beliefs and the critical value $\frac{c}{\alpha^{n-1}\left(\frac{r}{c}+1\right)^{n-2}r}$ from Proposition 1. In periods 10 to 5, the prior is high enough that the long run player finds it advantageous to punish with probability one, in order to maintain a reputation for toughness. In this case, the long run player would punish with a probability less than one in case the short run player rejects the offer after period 5. This decrease in the probability of punishment results in an increase in the equilibrium transfer. As the game nears its end, the probability of punishment decreases and the transfer increases.

Figure 3 shows the off-equilibrim path of play that arises if the official in period 3 rejects the transfer (in cases when $\alpha < 1$ this could be part of equilibrium play when the short run player in period 3 is crazy). Panel (a) shows the off-equilibrium path when randomization by the pressure group results in punishment, while panel (b) shows the case in which randomization results in no punishment. Punishment causes in an increase in the prior of the pressure group being tough and results in lower transfers compared with the case in which the pressure group does not punish and its type is revealed.

One interesting difference with the Kreps-Wilson paper is that in our setting the prior

[7] As demonstrated in the proof above, after a randomization has taken place, the upward update is not large enough to make LR willing to punish the next SR for sure.

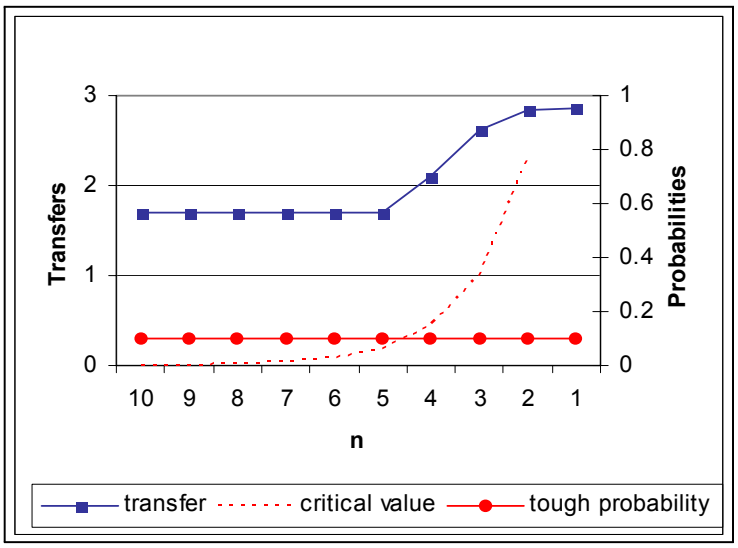

Figure 2: Equilibrium play

can –while positive– be arbitrarily low and the long run player will still receive some benefits from the fact that it is punishing with positive probability. The reason is that punishing with positive probability lowers the bribe that buys a favorable decision from sane short run players. From our Proposition, the equilibrium transfer by a weak long run player who enjoys a very low prior on toughness in period 2 $t_2 = m - p_2 \alpha \frac{r^2}{c}$. It is clear in this expression that no matter how low p_2 is, the long run player still benefits from the perception that he might be tough. Graphically, this effect can be seen in the equilibrium level of transfers represented in Figure 2. Note that the transfer in period 2 is lower than that in period 1, precisely because at that point the interest group is punishing with some probability and hence obtaining a "discount" on the price of favors. This does not happen in the basic chain store model without side payments. In such model, if the prior is lower than a threshold level entry occurs and the incumbent does not benefit at all from the fact that it might punish entrants. The bottom line is that in the basic chain store game a low enough initial prior means that the expected payoff of the long run player is zero. When transfers are available, the expected payoff of the long run player is positive, no matter how low the initial prior is

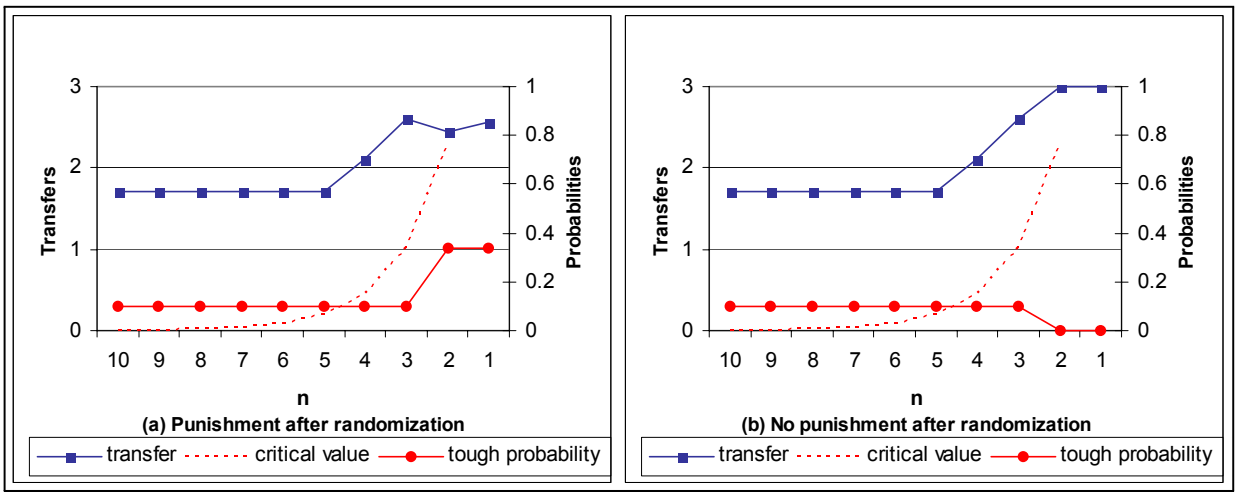

Figure 3: Play in the event that the short run player rejects offer in period 3

-or, equivalently, no matter how short the game is. So while in Kreps and Wilson's model, given a prior, the game has to be long enough for some reputational benefit to accrue for sure to the long run player, in our set up the prior can be arbitrarily low and the game arbitrarily short (say with only two periods), and there still exists a reputational benefit: the long run player pays lower transfers. The implication is that it is not necessary for a long run player to be likely to stay around for a long time in order to reap some benefits in terms of low transfers.

It is important to note that the long run player uses transfers and threats simultaneously to provide the desired incentives to the short run players. In fact, if the power of threats is high enough the transfers are negative. The long run player may not only obtain a favor with value π from the short run player but also a monetary transfer when its retaliation power r is large relative to the cost m of yielding to the group's demand. Recall that in a typical period n, and depending on the level of the prior, the transfer is $t_n = m - p_n \alpha^{n-1} \left(\frac{r}{c} + 1\right)^{n-2} \frac{r^2}{c}$ or $t_n = m - r$. That is, whether the long run player pays bribes or extracts extortionary payments is a result of the reputational equilibrium and not a given. In this way, the equilibrium studied here highlights the formal connections between seemingly

unrelated phenomena, such as lobbying and extortion. Mafias differ from lobbies in that they have a great capacity to inflict harm (r is high and c is low) and in many occasions offer something in exchange (for instance, enforcement), making m small. It can easily be seen that large r and small c and m tend to make payments negative.

5 Political influence

Direct relabeling of the model allows to think about interest groups (long run players) influencing sequences of government officials (short run players). Buying influence by transferring money to the officials is possible if the interest group will fully compensate officials for the disutility of doing it a favor. Such quid pro quo is (comparatively) simple to enforce, because the trade of money for a policy can be arranged to take place simultaneously. Using threats may save on transfer payments, but delivering a future punishment on an official that has refused to trade may be costly, and, hence, not credible. However, if the pressure group deals with more than one official, the results from the previous section show that there is an equilibrium in which the group develops a reputation for retaliating on those officials who reject bribes, and pays low bribes. Similarly, activist groups targeting firms may obtain large changes in corporate policy in exchange for very scant public congratulation from the group (see also Baron and Diermeier 2005 on why campaigns are typically negative). The real payoff to the corporation is the non-realized punishment.

It has been argued that the bribes paid to officials in the public sector are usually of negligible value in comparison to the benefits they secure. Tullock (1980) expressed a puzzle regarding rent seeking efforts that appear too small relative to the favors they secure. This came to be known as the "Tullock paradox". The model developed in this paper allows us to put forward an explanation for why politicians would require small compensation in exchange for their favors. In our view, the process of political influence is not limited to the

offer of rewards in the form of gifts, courting, or bribes. Pressure groups may also want to give incentives through threats. When, for example, a group is influential enough so as to damage the future career of a politician, the latter may be inclined to concede favors to such group just to avoid its enmity. The existence of threats–even when in some cases they could be tacit–can explain why politicians are willing to sell their favors cheap.

While credibility of threats is no longer assumed, in our model a pressure group may develop a reputation for honoring its threats allowing the group to pay cheap bribes. Note that the conditions for a pressure group to be able to profit from the use of threats are extremely weak. The only condition is that exists a small amount of asymmetric information. In our model, contrary to the case of Kreps and Wilson (1982), a pressure group can profit from developing a reputation even in a short horizon game –as short as two periods– for any arbitrarily small amount of imperfect information.

This is an interesting feature of our model since it implies that it is not necessary for a pressure group to be likely to stay around for a long time in order to reap some benefits in terms of cheap bribes. If the environment creates a positive prior in the officials' minds that a group is "nasty", then this group has incentives to punish honest officials with positive probability even when interacting with a small number of them. Of course, when a group is known to be influential and can be expected to face many officials, the potential for cheap bribes is larger.

6 Judicial Extortion

Our model of transfers and threats can be applied to the general problem of extortion, and to the particular issue of judicial blackmail. We understand by this a situation in which a judge faces a sequence of defendants and can ask for a bribe in order not to impose a worse-than-deserved penalty. However, imposing excessive penalties on those defendants

who refuse to pay bribes can be expected to be costly for the judge. For example, departing from strict criteria of justice may expose the judge to possible investigations of judicial abuse from the rest of the judicial structure. Thus, a credibility problem arises. After a defendant refuses to pay the requested bribe, the judge does not have a short run incentive to impose a larger-than-fair sentence. However, if there is some amount of imperfect information on the judge's type, our model shows that the judge has an incentive to build a reputation for toughness, no matter how low the prior on the judge's toughness or how short her time horizon.

This form of judicial corruption has recently been described by Ayres (1997), who has first-hand experience on the topic. Ayres (1997) begins with *"On January 25, 1990, I stood in a Cook County Circuit Court and accused the presiding judge, the Honorable Thomas J. Maloney, of extortion".*

Perhaps the most interesting part of Ayres (1997) is where he goes on to analyze the "convicted non-payor" problem.[8] That is, the problem of convicted defendants that did not pay bribes and got heavy sentences. He views this as a serious problem. In his words, *"The United States is now grappling with the convicted non-payor problem created by Judge Maloney's pattern of corruption. William Bracy and Roger Collins were convicted of murder and sentenced to death in a bifurcated jury trial over which Judge Maloney presided. [T]he defendants are seeking a new trial arguing that Judge Maloney's corruption in other cases gave him an incentive to be biased against defendants who did not pay him–in part to avoid suspicion that he was on the take.*"

Ayres mentions two reasons why Judge Maloney had an incentive to convict defendants who did not have a deal (nor they discussed a deal) with Maloney: *"1) to divert the attention of the prosecutor and of the electorate and 2) to create a reputation as a tough judge so as to more easily extract money from defendants who did pay."* The second argument is closely

[8] The paper also discusses the acquitted payor problem.

related to our model. As our bribes and threats model shows, the judge does have an incentive to be tough on those who do not pay in order to extract higher payments from future defendants.

Judge Richard Posner, who intervened in the case, noted that a rule for automatic reversal in cases such as this may have large consequences because it "*would thus require the invalidating of tens of thousands of civil and criminal judgements, since Judge Maloney alone presided over some 6,000 cases during the course of his judicial career...*".[9] More importantly, he does not see the merit in the argument that the judge's acceptance of bribes could have affected other cases where bribes were not paid. Judge Posner states *"The fact that Maloney had an incentive to favor prosecution in cases in which he was not bribed does not mean that he did favor the prosecution in such cases more than he would have done anyway."*[10] Ayres (1997) is critical of this reasoning (*"It is striking to hear one of the parents of law and economics argue that incentives don't on the margin affect behavior"*). Our model formally demonstrates Ayres' second argument above to be true.

In order to see this, we can consider the model in section 3 with the long run player being the judge and the short run players being the defendants. If we assume that the judge does not ask for non pecuniary contributions from the defendants and that the defendants do not face costs besides the amount of the transfer from the extortion payments, we can simplify the model by making $\pi = m = 0$. Now r denotes the disutility imposed on the defendant by a worse-than-deserved sentence, c is the cost the judge faces when giving a worse-than-deserved sentence, and $-t_n$ is the extortionary payment made by the defendant n. The rest of the parameters have the same interpretation as in Section 3. All mixing probabilities and equilibrium strategies in this game are as in Section 4. Then it is easy to see that "sane" defendants are willing to pay a positive amount of money to the corrupt judge to avoid a worse-than-deserved sentence. More importantly, the amount that "sane" defendants are

[9] See Bracy v. Gramley, 81 F.3d 689; 1996 U.S. App.

[10] See Bracy v. Gramley, 81 F.3d 689; 1996 U.S. App.

willing to pay is increasing on the defendants belief that the judge is "tough," which gives the judge an incentive to punish those defendants that do not pay. The judge will develop a reputation for punishing those who do not pay by *actually* passing worse-than-deserved sentences. Note that with probability $1-\alpha$ each defendant will be "crazy" and thus punished in equilibrium. Hence, in equilibrium non payors will be convicted with worse than deserved sentences. Clearly the discovery that a judge has been on the take should trigger the revision of previous sentences.[11]

7 Conclusion

We study a set up where private coercion against individuals and organizations is feasible in the attempt to affect their behavior, but where the use of coercive methods runs into a credibility problem. We provide a model where a long-lived player faces a finite sequence of short run players. We characterize a sequential equilibrium where the long run player develops a reputation for honouring his threats of punishing the short run players who do not accept his payment offers or requests. We show that the inclusion of a continuous choice variable like the money transfer allows the long run player to benefit from reputation even in arbitrarily short games. We apply the model to analyze both lobbying and extortion, highlighting the formal similarities between the two phenomena. Whether an organization with some coercive capability makes or extracts payments (or other kinds of favors) depends on the amount of damage it can inflict at a given cost. In other words, whether the long run player behaves as a lobby or a mafia is an endogenous result in our model, not a given.

The first application of the model explains how lobbies may develop a reputation for retaliating against officials that do not do favors to them. This reputation saves on bribes

[11] The decision of the The United States Court of Appeals for the Seventh Circuit that expressed the view that the accused had not shown good cause for discovery was reversed by the Supreme Court. However, the Supreme Court did not consider the reputation effect. See 520 U.S. 899; 117 S. Ct. 1793; 138 L. Ed. 2d 97.

or other rewards that lobbies may direct to officials. The setup we have presented offers some tools to think about pressure groups trying to affect both public and corporate policy, and it can suggest some areas for future enquiry. Our model emphasizes the importance of the beliefs that targets may hold regarding the commitment of the group to carrying out its threats. Therefore, the model suggests focusing attention on the mechanisms by which pressure groups construct such reputation. The most obvious way is by exercising punishments. But there might be complementary strategies that will help raise the prior p. In the case of activist groups, one possibility is to incorporate activists with known records of direct action–this can help groups start the game with a prior that is high enough so that reputation-preserving punishments can be utilized in equilibrium with probability one. This will produce the maximum possible savings on rewards. In fact, the public relations strategy of certain activist groups engaging in direct action (such as, for example, Shepherds of the Sea) gives prominence to the trajectory of its leaders. This may be a way to shape the prior targets hold on the probability that the group is willing to punish. Also, this suggests that the trajectories of some activists and lobbyists should resemble those we observe in more standard labor markets: we should expect individual track records (experience) to matter and command a price that reflects some of the credibility value that these track records lend groups.

The second application of the model can be used to settle the vital legal debate on the "convicted non-payor" problem that arises from judicial extortion. The implication of our model is that past trials conducted by judges that are found to be on the take should be revised.

References

Ayres, I. (1997). "The Twin Faces of Judicial Corruption: Extortion and Bribery," *Denver University Law Review*, 74.

Baron, D. (2001). "Private Politics, Corporate Social Responsibility and Integrated Strategy," *Journal of Economics & Management Strategy*, 10, 7-45.

Baron, D. and D. Diermeier (2005). "Strategic Activism and Non-Market Strategy," mimeo Stanford GSB.

Bernheim, D. and M. Whinston (1986). "Menu Auctions, Resource Allocation and Economic Influence," *Quarterly Journal of Economics* 101(1), p. 1-31.

Dal Bó, E., Dal Bó, P. and R. Di Tella (2002). *"'Plata o Plomo?':* Bribe and Punishment in a Theory of Political Influence," Working paper No. 28, Brown University.

Dal Bó, E. and R. Di Tella (2003). "Capture by Threat," *Journal of Political Economy* 111(5), 1123-54..

Fudenberg, D. and J. Tirole (1991). Game Theory, MIT Press.

Gambetta, D. (1993). *The Sicilian Mafia: The Business of Private Protection*, Mass.: Harvard University Press.

Garfinkel, M. (1990), "Arming as a Strategic Investment in a Cooperative Equilibrium", *American Economic Review* 80, 50-68.

Grossman, G. and E. Helpman (1994). "Protection for Sale", *American Economic Review* 84, 833-50.

Helpman, E. and T. Persson (2001). "Lobbying and Legislative Bargaining," *Advances in Economic Analysis & Policy*, 1(1).

Hirshleifer, J. (2001). The Dark Side of the Force: Economic Foundations of Conflict Theory. Cambridge University Press.

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

Kreps, D. and R.Wilson (1982). "Reputation and Imperfect Information," *Journal of Economic Theory* 27, 253-79.

Leaver, C. (2002). "Bureaucratic Minimal Squawk: Theory and Evidence," mimeo, UCL.

Leaver, C. and M. Makris (2001). "Passive Industry Interests in a Large Polity," mimeo, University of Bristol.

Powell, R. (1993). "Guns, Butter, and Anarchy," *American Political Science Review* 87(1), 115-32.

Shavell, S. and Spier, K.E. (2002) "Threats Without Binding Commitment," *Topics in Economic Analysis & Policy*, 2(1), Article 2.

Skaperdas, S. (1992), "Cooperation, Conflict, and Power in the Absence of Property Rights", *American Economic Review* 82, 720-39.

Smith, A. and F.Varese (2000). "Payment, Protection and Punishment: The Role of Information and Reputation in the Mafia," Oxford Department of Sociology working paper 2000-03.

Tullock, G. 1997. "Where Is the Rectangle?" *Public Choice* 91 (April): 149-59.

Varese, F. (1994). "Is Sicily the Future of Russia?: Private Protection and the Rise of the Russian Mafia, *Archives Europeenes de Sociologie*, 35, 2, 224-59.