Discussion Paper #1139
September 28, 1995

"Threats without Binding Commitment"

Steven Shavell

Kathryn Spier

Northwestern University

Discussion Paper No. 1139

THREATS WITHOUT BINDING COMMITMENT

by

Steven Shavell and Kathryn Spier

September 28, 1995

THREATS WITHOUT BINDING COMMITMENT

Steven Shavell and Kathryn Spier*

September 28, 1995

This paper explores the power of threats in the absence of binding commitment. The threatener cannot commit to carry out the threat if the victim refuses payment, and cannot commit *not* to carry out the threat if payment is made. If exercising the threat is costly to the threatener, then the threat cannot succeed in extracting money from the victim. If exercising the threat would benefit the threatener, however, then the threat's success depends upon whether the threat may be repeated. In the equilibrium of a finite-period game, the threat is carried out and the victim makes no payments. In an infinite-horizon game, however, it is an equilibrium for the victim to make a stream of payments over time. The expectation of future payments keeps the threatener from exercising the threat.

1. INTRODUCTION

Threats are often used successfully to extract money or other things of value. A famous historical example concerns the Vikings, who in 991 stated that they would invade and plunder the English unless tribute was paid. The English submitted to the Viking demands and made a succession of payments of rather extraordinary magnitude.[1]

* Harvard Law School and Kellogg Graduate School of Business, Northwestern University, respectively. Shavell acknowledges the support of the John M. Olin Center for Law, Economics, and Business at Harvard Law School, and Spier acknowledges summer support from Kellogg.

[1] To raise the funds, the English King Ethelred imposed what may have been the first English tax paid in money, called the *Danegeld*; see the discussion of *Danegeld* in Hodgkin (1906), pp. 381-383. Kipling (1940, p. 658) described these events in verse. The first several lines of his "Dane-Geld" read as follows:

It is always a temptation to an armed and agile nation,
To call upon a neighbour and to say:--
"We invaded you last night--we are quite prepared to fight,
Unless you pay us cash to go away."

And that is called asking for Dane-geld workers who threaten a strike unless employment conditions are improved, and by manifold other situations. Where a threat is costly to execute, a threat without a binding commitment would not be credible in the model, so the threat would be unsuccessful.

The case in which executing a threat does not involve a cost for the threatener is approximately valid in a variety of circumstances. For example, suppose that a blackmailer threatens to reveal embarrassing information (the cost of revelation might be mere postage if the blackmailer is not worried about being caught) or that an extortionist threatens to burn down a building (the cost may be only that of a match if he is not worried about apprehension). Here, a threat without commitment may be credible because it would (just) be rational for the threatener to carry out his threat if and only if he is not paid. Therefore, the threat may be successful.

The situation in which proceeding with a threatened act would benefit the threatener is also often apposite. Suppose that a blackmailer would profit from revealing information (the victim might be a public figure and the information might be sold to the media) or suppose that a country (like Denmark) threatening to invade another would gain by appropriating the victim's land and wealth. In situations like these, the problem of the threatener is not that his threat is empty (the problem that the threatener faces when executing the threat is costly). Rather, the threatener's problem is that he will have an incentive to carry out his threat even if he is paid (the blackmailer's problem is that he will have an incentive to sell his information even if he is paid by the victim). Because this means that the victim will not prevent the threatened act by paying, he will not pay.[3] The threatener cannot overcome this problem in a single (or finite) period setting, and his threat will therefore fail in this version of the model.[4]

---

[3] Accordingly, the threatener would like to commit *not* to act if he is paid, rather than, in the case where the threat is costly to execute, to commit *to* act if he is not paid.

[4] This point, and the possible solution to the threatener's problem in the repeated threat context to be described, is discussed informally in Shavell (1993).

to make to the threatener, after which the threatener must decide whether to carry out the threat. If the threat is carried out, the game ends; the victim suffers a loss, v, and the threatener gains b (in addition to the stream of past payments, p_s, where $s \leq t$). If the threatener does not exercise the threat, the game proceeds to period t+1, and the stage game is repeated.

The analysis is broken into two parts. First, we consider a game with a known finite horizon. Next, we consider the infinitely repeated game. This latter case is formally equivalent to the situation where the game will end with a known probability in each period; the discount factor may be interpreted as the probability that the game will continue into the following period.

2.1 Finite Horizon

Suppose that the game has a known finite horizon, T. To start, if the threatener's gain from carrying out the threat is strictly positive, $b > 0$, then there is a unique subgame-perfect equilibrium where no payments are made and the threat is carried out immediately. This outcome can be derived by backwards induction. In the last period, the threatener would strictly prefer to carry out the threat regardless of the previous payments made by the victim. Knowing that his payments cannot influence the threatener's actions, the victim would pay $p_T = 0$ in period T. Working backwards, in period $T-1$ the threatener anticipates receiving $p_T = 0$ in the following period and therefore carries out the threat immediately. (He prefers to receive a payoff of b in period $T-1$ than wait a period to receive the same payoff.) Anticipating this, the victim pays $p_{T-1} = 0$. Repeating this argument verifies that the victim never pays a cent and the threatener carries out the threat at his first opportunity.

When $b > 0$, this unique subgame-perfect equilibrium is clearly inefficient. Since the victim's loss exceeds the threatener's gain, it would be better for the players to negotiate around the inefficiency and avoid the dead-weight loss, $v-b$, associated with carrying out the threat. In the

Proposition 1: Suppose that the horizon, T, is finite.

i. *If $b > 0$, there is a unique subgame-perfect equilibrium. The victim pays nothing to the threatener, and the threatener carries out the threat in period 1. The payoffs to the victim and the threatener, respectively, are $(-v, b)$.*

ii. *If $b = 0$, then for each $\theta \in [0, 1]$ there is a subgame-perfect equilibrium (which satisfies elimination of weakly dominated strategies) where the victim pays $p^*(\theta) = [(\delta^{T-1}-\delta^{T})/(1-\delta^{T})]\theta v$ to the threatener in each period, and the threatener never carries out the threat. The payoffs to the victim and the threatener, respectively, are $(-\delta^{T-1}\theta v, \delta^{T-1}\theta v)$.*

iii. *If $b < 0$, there is a unique subgame-perfect equilibrium. The threatener never carries out the threat and the victim makes no payments. The payoffs to the victim and the threatener, respectively, are $(0, 0)$.*

Proof:

The cases of $b < 0$ and $b > 0$ are easily proven by backwards induction, as described in the text.

For the case of $b = 0$, the following strategies support the subgame-perfect equilibrium described in the proposition for a given θ: If $p_t < p^*(\theta)$ in some period t, then the threatener waits and exercises the threat in period T with probability $\theta \in [0, 1]$. If $p_t \geq p^*(\theta)$ for all t, then the threatener does not exercise the threat. The victim pays $p_t = p^*(\theta)$ if $p_s \geq p^*(\theta)$ for all $s < t$, and pays $p_t = 0$ if $p_s < p^*(\theta)$ for some $s < t$.

First, we check that the victim's strategy is a best response to the threatener's strategy. Assuming that the victim has not paid less than $p^*(\theta)$ in the past, the victim can do no better than pay $p^*(\theta)$ in each subsequent period. Paying $p_t > p^*(\theta)$ does not reduce the chance that the threat is carried out (if the victim keeps on schedule by paying $p^*(\theta)$, the probability is zero). By paying the victim's loss exceeds the threatener's gain. We begin by observing that this inefficient outcome is also an equilibrium of the infinite horizon game when $b > 0$. Since the victim believes that the threatener will carry out the threat at his first opportunity, he has no incentive to pay the threatener. Since the threatener does not expect to receive any money from the victim, it is rational for him to carry out the threat.[9] We will refer to this equilibrium as the *stationary equilibrium*.[10]

However, when $b > 0$ there are also many *efficient* subgame-perfect equilibria with successful threats; payments are made by the victim, and threats are never carried out in equilibrium. In particular, we can use the inefficient stationary equilibrium to construct a set of efficient equilibria with a constant payment stream over time, $p_t = p^*$, where $p^* \in [\ [(1-\delta)/\delta]\ b,\ (1-\delta)v\ ]$. The motivation is as follows: If the victim deviates by paying less than p^*, the threatener believes that the victim will pay nothing in the future, and the victim believes that the threatener will carry out the threat at his first opportunity. In other words, after a deviation, the continuation equilibrium automatically "switches" to the inefficient stationary equilibrium. In this way, the victim is punished for paying less than p^*.

With these strategies, we see that any division of the surplus, $v - b$, is possible in equilibrium. When the per-period payment is at its lower bound, $p^* = [(1-\delta)/\delta]\ b$, the threatener does no better than in the stationary equilibrium. His present discounted payoff is b, and the victim captures the surplus, $v - b$. When payment is at the upper bound, $p^* = (1-\delta)v$, the threatener is extracting all of the surplus (his payoff is v); the present discounted value of the payment stream equals what the victim would lose if the threat were actually exercised.

---

9 The following strategies support this equilibrium: The victim pays $p_t = 0$ regardless of the history of the game, and the threatener exercises his threat with probability one at each decision node.

10 An equilibrium is stationary when the players' strategies are the same at each node of the game, both on and off of the equilibrium path. The stationary equilibrium described in the text is the only stationary equilibrium of the game when $b > 0$.

First we will construct upper and lower bounds on the per-period payment, p^*. Clearly, the threatener would strictly prefer to carry out the threat if $b > \delta p^* + \delta^2 p^* + \ldots$. The victim would not be willing to pay p^* in each period if $v < p^* + \delta p^* + \delta^2 p^* + \ldots$, or if the loss from having the threat carried out is smaller than he present discounted value of the payments.[11] Rearranging these expressions tell us that any per-period payment must fall within a certain range: $[(1-\delta)/\delta]\, b \leq p^* \leq (1-\delta)\, v$.[12]

The following strategies support the subgame-perfect equilibrium for each p^* in this range. The threatener does not carry out the threat in period t if $p_s \geq p^*$ for all $s \leq t$. If $p_s < p^*$ for some $s \leq t$, then the threatener carries out the threat immediately. The victim pays $p_t = p^*$ if $p_s \geq p^*$ for all $s < t$, and pays $p_t = 0$ if $p_s < p^*$ for some $s < t$.

The victim cannot do better than pay p^*: By paying $p > p^*$ he does not increase the chance of acceptance (it is already unity), and by paying $p < p^*$ he is (weakly) worse off since the threat will be carried out, and $p + v \geq p^*/(1-\delta)$.[13] (The left hand side denotes the payoff from deviating, and the right hand side is the present discounted value of the equilibrium payments.) So long as $p_s \geq p^*$ for all $s \leq t$ the threatener will not carry out the threat because he expects payments to continue in the future and $b \leq p^*[\delta/(1-\delta)]$. If $p_s < p^*$ for some $s \leq t$, then the threatener expects no

[11] The timing of the game specifies that in each period the victim decides upon the payment, after which the threatener chooses whether or not to carry out the threat. If the threatener does not carry out the threat, then the next payment he will receive is in the *following* period. This is why the first term in the threatener's inequality is discounted by δ. If the victim refuses to pay, the threatener will carry out the threat in the same period; that is why, in the victim's inequality, neither v nor the first payment are discounted.

[12] Note that a range exists only when $\delta > b/v$. If the discount factor were very small, then the payers would be myopic. The threatener would prefer to exercise the threat and gain b than to wait for future payments from the victim.

[13] This is true because $p^* \leq (1-\delta)\, v$. Note that if $p^* = (1-\delta)\, v$ then the victim is indifferent between continuing to make payments and paying $p = 0$. For all other cases, however, the victim is strictly worse off when he deviates and pays $p < p^*$.

demand exceeded the true harm the victim would suffer from the threat or else because the victim would incorrectly surmise that the threatener would bear a cost from carrying out his threat.

(b) *Emotional motivation.* A plausible emotional motivation provides a reason to believe that a threat might be credible without a binding commitment. Suppose that a threatener will become angry with the victim if his demand is rejected, and thus enjoy carrying out his threat, whereas the threatener will not be angry (and possibly will be favorably disposed toward the victim) if the victim pays him. If this is so and the victim understands that, the threatener's psychological makeup will itself lead him to carry out his threat if and only if he is not paid, so that his threat will be credible.[14]

(c) *Social undesirability and possible illegality of threats.* We considered the question how threats may succeed but did not emphasize the issue of their influence on social welfare. The influence of threats on social welfare may come about in part through ex ante effects: activity on the part of potential victims to avoid becoming prey to threats, and activity on the part of potential threateners to position themselves so as to be able to carry out threats. Such activities may represent a social waste (the activities of extortionists who would threaten to burn down a store if not paid off) or may constitute a social benefit (suppose victims curtail their socially undesirable, possibly criminal acts so that they cannot be blackmailed). Other avenues through which threats affect social welfare are the making of payments to threateners and the actual execution of threats (which may come about through asymmetric information, as just discussed, and also because a threatener would gain from his threat, as in Proposition 1i). These effects too may be socially desirable (for instance, the blackmail payment made by a criminal might be desirable as it would tend to deter crime) or undesirable. In consequence, it can be argued that some types of threat are fairly clearly socially

[14] The main point of this paragraph is stressed by Hirshleifer (1987).

REFERENCES

Ginsburg, Douglas H. and Paul Shechtman, "Blackmail: An Economic Analysis of the Law," 1993, *University of Pennsylvania Law Review*, 141: 1849-1876.

Klein, Daniel B. and Brendan O'Flaherty, 1993, "A Game-Theoretic Rendering of Promises and Threats" *Journal of Economic Behavior and Organization*, 21: 295-314.

Hepworth, Mike, 1975, *Blackmail: Publicity and Secrecy in Everyday Life*, Routledge, London.

Hirshleifer, Jack, 1987, "On the Emotions as Guarantors of Threats and Promises," 307-326, in John Dupre, editor. *The Latest on the Best: Essays on Evolution and Optimality*. MIT Press, Cambridge, MA.

Hodgkin, Thomas, 1906, *The History of England From the Earliest Times to the Norman Conquest*, Longmans, Green, and Co., London.

Jacoby, Neil H., Peter Nehemkis and Richard Eels, 1977, *Bribery and Extortion in World Business: A Study of Corporate Political Payments Abroad*, Macmillan, New York.

Kipling, Rudyard, 1940, *Rudyard Kipling's Verse, 1885-1926*, Doubleday, New York.

Milburn, Thomas W. and Kenneth H. Watman, 1981, *On the Nature of Threat*, Praeger, New York.

Posner, Richard A., 1993, "Blackmail, Privacy, and Freedom of Contract," *University of Pennsylvania Law Review*, 141: 1817-1843.

Schelling, Thomas C. 1960, *The Strategy of Conflict*, Harvard University Press, Cambridge.

Shavell, Steven, 1993, "An Economic Analysis of Threats and Their Illegality: Blackmail, Extortion, and Robbery," *University of Pennsylvania Law Review*, 141: 1877-1903.