

# Prohibitions on Punishments in Private Contracts\*

Philip Bond

Andrew F. Newman

Northwestern University

University College London and CEPR

Incomplete and Very Preliminary

March 23, 2003

## Abstract

We study legal restrictions on private contracting in the form of limitations on the severity of non-monetary punishments. We locate the rationale for such restrictions in externalities that parties impose on future relationships: punishments that lower an agent's future productivity may lower social welfare, and the agent may not take this into account. These externalities assume two forms: (1) future principals' interests are not taken into account by private parties; and (2) consonant with much of the legal literature, agents who are boundedly rational take insufficient account of their "future selves" and may need protection.

In the first instance, we derive results on the dependence of socially inefficient contracting on the relative bargaining powers of principals and agents, on growth rates and uncertainty about productivity, and on the number of trading partners. For the second case, we focus on over-confidence and note that

---

\*Both authors thank the Institute for Advanced Study for hospitality and financial support over the academic year 2002-03.

its effects are ambiguous: although over-confidence may lead agents to accept contracts they ought to reject, it actually reduces the private need to engage in severe punishments.

## 1 Introduction

A great deal of recent research in contract theory has focused on what has become known as “contractual incompleteness.” By this, economists mean that certain contingencies are impossible to include in private contracts.<sup>1</sup> However, a second — and perhaps equally important — limitation on private contracting has received much less attention: severe restrictions are placed by the state on the punishments that can be imposed on a party that breaches a contract. That is, a contract cannot stipulate corporal punishment or imprisonment. Instead, parties can make use of only monetary incentives, which by their nature are limited. And certain physical assets, such as houses, are typically also protected. Strikingly, this restriction is a historically recent innovation. In this paper we consider what may motivate this restriction on private contracts. Our main argument is that non-monetary punishments impose externalities on future trading parties, and (in the case of boundedly rational individuals) on one’s future self.

The restriction on the punishments that private parties can threaten one and another with is an important one. Without such a restriction it is possible to achieve the optimal frictionless outcome in many contracting models. Under many circumstances, welfare losses due to agency problems and credit constraints would disappear (or nearly so) if arbitrary punishments were possible.<sup>2</sup>

---

<sup>1</sup>See, e.g., Hart (1995) for a survey.

<sup>2</sup>See, e.g., Mirrlees (1999) and Mookherjee and Png (1989) for cases in which the possibility of large punishments would effectively eliminate the agency problem.

At least informally, it is widely argued that restrictions on punishments can be justified on ethical grounds. This argument seems at best incomplete, because it overlooks the fact that the state regularly uses imprisonment as a punishment for other offenses, even ones of a non-violent nature. Thus a debtor who consumes his loan instead of investing it and repaying his creditor cannot be imprisoned; while in a directly analogous setting a taxpayer who fails to pay his “debt” to the government may quite well suffer just such a punishment.

At the same that the state persists in its use of non-monetary punishments, avenues by which contracting parties can at least partially replicate their use have been closed. Debtors’ prisons no longer exist. Indentured servitude is banned. Related, personal bankruptcy prevents a creditor from exercising a claim on future earnings. Even specific performance is rarely imposed on a breaching party, and the use of non-compete clauses in labor contracts is subject to severe restrictions (and in California, is banned altogether).

As noted, constraints on the use of non-monetary punishments are a historically fairly recent phenomenon. In describing the enforcement of indentured servitude contracts in early colonial America, Galenson (1984) notes that “[i]n 1612, the colony’s governor dealt firmly with some recaptured laborers: ‘Some he appointed to be hanged Some burned Some to be broken upon wheles, others to be staked and some to be shott to death.’ ” Morgan (1975) considers physical punishments the main incentive device used to motivate indentured laborers. And as Chwe (1990) notes, “whipping Asian laborers was standard practice on Hawaiian sugar plantations well into this century.”

What then can account for the modern restriction that private contracts stipulate only monetary punishments? In this paper, we view the prime distinguishing property of monetary punishments as being the fact that they do not disrupt an individual’s future productivity. In contrast, non-monetary incentives can have severe

effects. Imprisonment and capital punishment are the clearest examples, since they absolutely eliminate an individual's future output. Corporal punishment, by causing bodily harm, will likewise tend to reduce an individual's future productivity. For want of a better word, we describe such punishments as *exclusionary*: in one way or another, they act to exclude an individual from future productive activity.

Given this distinction between monetary and exclusionary punishments, in this paper we explore the following two — related — explanations for the simultaneous restriction on the use of exclusionary punishments in private contracts with their ongoing use in the criminal system.

First, when a contract stipulates an exclusionary punishment it imposes an externality on the future trading partners of the punished individual. To take an extreme example, ancient Athens is said at one point to have imprisoned so many of its farmers for non-repayment of debt that there was insufficient food to feed its population. Of course, an individual who agrees to be bound by the prospect of an exclusionary punishment partially takes these lost opportunities into account. We show that the externality is the greatest, and thus regulation against exclusionary punishments most likely to be justified, when the party to be incentivized has a strong bargaining position today, will have a weak bargaining position tomorrow, when growth rates are high and/or variable, and when the number of future trading partners is high.

Second, it is widely held that individuals are too prone to agree to exclusionary punishments because they overestimate the probability that they successfully perform the required task, and so underestimate the probability that the punishment will actually be imposed. That is, boundedly rational individuals impose an externality on their future selves. We demonstrate that this argument, though intuitive, is subject to important caveats. First, although over-confidence leads to greater use of exclusionary punishments, it also diminishes the size of the punishment necessary: overconfident agents need lower powered incentives. So it is by no means clear that

over-confidence increases the social cost of exclusionary punishments. Second, if an agent is overconfident today he is likely to be overconfident tomorrow. As such, although he underestimates the probability of the exclusionary punishment being imposed, he also overestimates the cost of the punishment: he thinks his future surplus from working is higher than it really is. Again, the net effect of these two offsetting biases is ambiguous.

Our paper is closest to the small literature that has examined a much milder restriction on private contracts, namely the non-enforceability of penalty clauses for breach. Aghion and Bolton (1990), Chung (1992), and Spier and Whinston (1995) are leading examples, all of which seek to account for this restriction as stemming from the fact that penalty clauses, while privately optimal, may be socially undesirable. In each case the social undesirability stems from the fact that penalty clauses can be used to deter entry into an industry. Also related is Diamond and Maskin (1979) who study the effects of penalty clauses on search and breach intensity in the labor market. Conceptually, our approach shares some elements of commonality with Camerer *et al* (2002), who seek to account for when apparently paternalistic policies (of which restrictions on non-monetary punishments is one example) can be justified on the grounds that they help boundedly rational agents more than they hurt rational agents. Finally, Chwe (1990) characterizes conditions under which the use of physical violence to provide incentives is privately optimal; in contrast, our focus is on when such punishments are privately optimal but socially sub-optimal.

## 2 Model

As outlined in the introduction, we are interested in whether or not contracting parties will privately agree to a greater use of “exclusionary” punishments than is socially optimal. On the one hand, exclusionary punishments incentivize an agent to work.

On the other hand, they impose a cost on the agent and his future trading partners. To make the analysis as transparent as possible, we use the simplest model capable of capturing this trade-off.

There are two periods and three individuals — one agent,  $A$ , and two principals,  $P_1$  and  $P_2$ . In each period  $t \in \{1, 2\}$  the agent  $A$  can either work for principal  $P_t$ , or can operate a backyard technology that produces  $w$ . If he works for principal  $P_t$ , he can either work or shirk. If he shirks, the principal receives an output of  $H > 0$  with probability  $p$ , and an output of 0 with probability  $1 - p$ . Working raises the probability of output  $H$  by  $\Delta p$ , but requires effort  $B_t$  from the agent. Throughout, we assume that working is socially efficient, i.e.

$$\Delta p B_t \geq H \tag{1}$$

The agent and both principals are risk neutral. The agent is restricted to have non-negative consumption. Principal  $t$  can induce the agent to work by offering a “bonus contract”  $x_t$  payable when output  $H$  is realized. (Without loss, we can assume the agent is paid nothing when the output 0 is realized.) Moreover, principal  $P_1$  has the additional option of taking some action that serves to deny the agent access to the period 2 labor market. As discussed in the introduction, corporal punishment, imprisonment and non-compete clauses all fall within this class. Accordingly, at date 1 the contract is a pair  $(x_1, \pi)$ , where  $\pi$  is the probability that the agent is excluded in period 2 if the low output is realized. (Again without loss, we can assume the agent is never excluded if the high output is realized.)

For simplicity, assume that there is no discounting, and that the agent cannot store any payments received in period 1 in order to ease the second period incentive problem. Let  $S$  be the agent’s expected utility in period 2, provided he is not excluded. The agent’s period 1 incentive constraint is thus

$$(x_1 + \pi S) \Delta p \geq B_1 \tag{IC_1}$$

We assume that in both periods the project is only possible if the principal supplies a unit of capital. Thus the principals' individual rationality constraints are

$$(p + \Delta p)(H - x_t) \geq 1 \quad (\text{P}_t\text{-IR-W})$$

if the contract induces the agent to work, and

$$p(H - x_t) \geq 1 \quad (\text{P}_t\text{-IR-S})$$

otherwise. In order to focus on the interesting case in which the contracting parties use exclusionary contracts, we assume throughout that there is no way to supply the agent with purely monetary incentives in period 1 that both induce him to work, and satisfy principal 1's individual rationality constraint (P<sub>1</sub>-IR-W):

$$(p + \Delta p) \left( H - \frac{B_1}{\Delta p} \right) - 1 < 0 \quad (2)$$

Finally, since some surplus is available at both dates, we need to specify how this surplus will be split. We adopt the standard randomized "take-it-or-leave-it" offers framework. That is, in period  $t$  with probability  $\theta_t$  the agent proposes a contract to principal  $P_t$ , who either accepts or rejects. Similarly, with probability  $1 - \theta_t$  the principal  $P_t$  proposes a contract to the agent, who either accepts or rejects.

Before proceeding to the analysis, we should note that although it is natural to think of the principals as being employers and the agent as being an employee, other interpretations are possible. In particular, it is straightforward to interpret the principals as lenders, and the agent as a borrower, with the face value of the loan set to  $H - x_t$  in period  $t$ .

### 3 Contracts and tomorrow's trading partners

We start by illustrating the basic externality at work: when the agent and principal  $P_1$  agree to use an exclusionary contract ( $\pi > 0$ ) to incentivize the agent, they are

imposing a cost on principal  $P_2$  who no longer gets a share of the surplus in period 2. Because of this, if left unregulated the agent and principal  $P_1$  will use exclusionary contracts more than is socially optimal.

Throughout this section, we maintain the same assumptions about technology in period 2 as we do for period 1: in particular, since exclusionary penalties are unavailable in the last period, it is impossible to give the agent incentives to work in period 2 and simultaneously satisfy principal 2's individual rationality constraint (P<sub>t</sub>-IR-W). That is, we assume that

$$(p + \Delta p) \left( H - \frac{B_2}{\Delta p} \right) - 1 < 0$$

Moreover, we assume that

$$pH - 1 - w \geq 0$$

so that the backyard technology is even less productive than shirking for the principal. Given this, the agent's period 2 surplus is

$$S = \theta_2 (pH - 1) + (1 - \theta_2) w = w + \theta_2 (pH - 1 - w)$$

Observe that  $pH - 1 = w + \frac{S-w}{\theta_2}$ . The overall social surplus when the agent shirks in period 1 is

$$SW_{no-inc} = pH - 1 + \left( w + \frac{S - w}{\theta_2} \right)$$

while the social surplus when he works, with a contract specifying a probability of exclusion  $\pi$ , is

$$SW_{inc}(\pi) = (p + \Delta p) H - 1 - B_1 + (1 - (1 - p - \Delta p) \pi) \left( w + \frac{S - w}{\theta_2} \right)$$

Thus

$$SW_{inc}(\pi) - SW_{no-inc} = \Delta p H - B_1 - (1 - p - \Delta p) \pi \left( w + \frac{S - w}{\theta_2} \right)$$



### 3.1 Agent makes offer in period 1

First, consider the case in which the agent makes the contract offer in period 1. If he offers a contract with enough incentives to work, he sets  $x_1$  as high as possible while satisfying principal  $P_1$ 's individual rationality constraint, (P<sub>t</sub>-IR-W), i.e.,

$$x_1 = H - \frac{1}{p + \Delta p} \quad (3)$$

and the probability of exclusion as low as possible while satisfying his own incentive constraint (IC<sub>1</sub>), i.e.,

$$\pi S + x_1 = \frac{B_1}{\Delta p} \quad (4)$$

So

$$\pi = \frac{1}{S(p + \Delta p)} \left( (p + \Delta p) \left( \frac{B_1}{\Delta p} - H \right) + 1 \right) \quad (5)$$

which is strictly positive by assumption (2). The agent's utility from such a contract is

$$\begin{aligned} U_{inc} &= (p + \Delta p) x_1 - B_1 + (1 - (1 - p - \Delta p) \pi) S \\ &= (p + \Delta p) H - 1 - B_1 + S - (1 - p - \Delta p) \pi S \end{aligned}$$

On the other hand, the agent can simply propose  $x_1 = H - \frac{1}{p}$  and  $\pi = 0$ , which by assumption (2) fails to supply sufficient incentives to induce him to work. The agent's utility is then

$$U_{no-inc} = px_1 - 1 + S = pH - 1 + S$$

Thus the agent's gain from offering a contract with a enough incentives to get him to work is the increase in expected output (since he makes the offer, he is the residual claimant), net of the cost of working  $B_1$ , and the expected cost of exclusion:

$$U_{inc} - U_{no-inc} = \Delta p H - B_1 - (1 - p - \Delta p) \pi S$$

So

$$\begin{aligned}
(U_{inc} - U_{no-inc}) - (SW_{inc}(\pi) - SW_{no-inc}) &= (1 - p - \Delta p) \left( \pi \left( w + \frac{S - w}{\theta_2} \right) - \pi S \right) \\
&= (1 - p - \Delta p) \left( \frac{1}{\theta_2} - 1 \right) \pi (S - w) > 0
\end{aligned}$$

That is, the agent's private gains from using an exclusionary contract always exceed the social gains. In slightly more detail,

**Lemma 1 (*Agent's bias*)**

*The agent views the gain from supplying incentives using an exclusionary contract as higher than the social gains by an amount*

$$\left( \frac{1}{p + \Delta p} - 1 \right) \left( (p + \Delta p) \left( \frac{B_1}{\Delta p} - H \right) + 1 \right) \left( \frac{1}{\theta_2} - 1 \right) \left( 1 - \frac{w}{S} \right) \quad (6)$$

*The bias is more pronounced when the agent's period 2 bargaining power is lower and when his period 2 "outside option"  $w$  is lower. The bias tends to 0 as the agent's period 2 bargaining power approaches 1.*

**Proof of Lemma 1:** The expression for the agent's bias in favor of using an exclusionary contract (6) is obtained simply by substituting (3) into the above expression for  $(U_{inc} - U_{no-inc}) - (SW_{inc}(\pi) - SW_{no-inc})$ . For the effect of changing the  $\theta_2$  and  $w$ , note that

$$\left( \frac{1}{\theta_2} - 1 \right) \left( 1 - \frac{w}{S} \right) = \frac{(1 - \theta_2)(pH - 1 - w)}{w + \theta_2(pH - 1 - w)}$$

which is decreasing in  $\theta_2$  and  $w$ . **QED**

Will the agent's bias actually lead him to propose an exclusionary contract when it is socially inefficient to do so?<sup>3</sup> The answer is yes. From above, the exclusionary

---

<sup>3</sup>In principle, it is possible that the bias tends to zero as the social inefficiency from the exclusionary contract tends to zero.

contract is socially efficient if and only of

$$\Delta p H - B_1 \geq (1 - p - \Delta p) \pi \left( w + \frac{S - w}{\theta_2} \right)$$

where we know that  $\pi$  is a linear and decreasing function in  $B_1$ . Define  $B_0$  as the value of  $B$  at which pure wage incentives can be used to incentivize the agent, i.e. (2) holds at equality. So there exists  $B^* > B_0$  such that an exclusionary contract is socially efficient if and only if  $B \in [B_0, B^*]$ . See Figure 1.

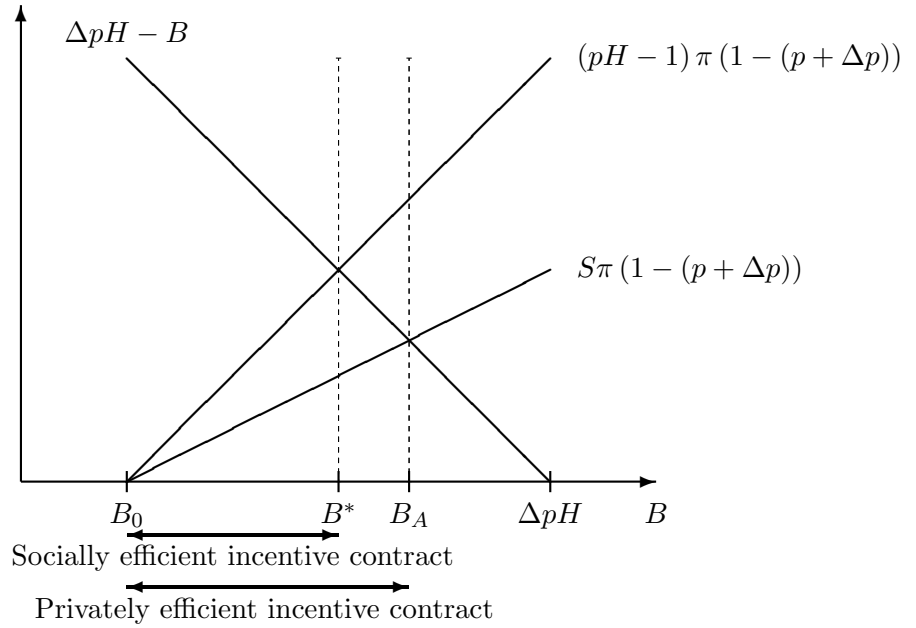


Figure 1: Use of incentive contract when agent makes offer in period 1

Similarly, an exclusionary contract is attractive to the agent if and only if

$$\Delta p H - B_1 \geq (1 - p - \Delta p) \pi S$$

So again, there exists a  $B_A > B_0$  such that an exclusionary contract is used by private contracting parties if and only if  $B \in [B_0, B_A]$ . And from Lemma 1 we know that  $B_A > B^*$ . So we have established:

**Lemma 2 (*Misuse of exclusionary contracts*)**

If  $B \in (B^*, B_A]$  and the agent makes the contract proposal, then an exclusionary contract is used even though it is socially inefficient.

### 3.2 Principal makes offer in period 1

Thus far, we have focused on the case in which the agent makes the contract offer in period 1. We now turn to an analysis of the other case, in which the principal  $P_1$  makes the offer.

When the principal  $P_1$  offers a contract  $(x_1, \pi)$  that supplies sufficient incentives to induce the agent to work, his utility is  $V_{inc} = (p + \Delta p)(H - x_1) - 1$ , while if the contract does not offer sufficient incentives to work,  $V_{no-inc} = pH - 1 - w$ . Thus

$$V_{inc} - V_{no-inc} = \Delta p H - (p + \Delta p)x_1 + w$$

The agent's utility from a contract  $(x_1, \pi)$  that induces him to work is

$$\begin{aligned} & (p + \Delta p)x_1 - B_1 + (1 - (1 - p - \Delta p)\pi)S \\ = & (p + \Delta p)(x_1 + \pi S) - B_1 + (1 - \pi)S \end{aligned}$$

Out of the various possible exclusionary contracts, certainly the principal  $P_1$  will offer the one that sets the agent's incentive constraint (IC<sub>1</sub>) to equality. The principal  $P_1$  sets  $x$  as low as possible subject to the agents individual rationality constraint being satisfied,

$$(p + \Delta p)(x_1 + \pi S) - B_1 + (1 - \pi)S \geq w + S$$

Substituting in for  $x_1 + \pi S = B_1/\Delta p$  (i.e. (IC<sub>1</sub>) at equality) gives

$$x_1 \geq w + \frac{(1 - p)B_1}{\Delta p} \tag{7}$$

Moreover, clearly the probability of exclusion must be chosen to lie below one, i.e.

$$x_1 \geq \frac{B_1}{\Delta p} - S \tag{8}$$

Our main result is that regardless of whether the individual rationality constraint (7) or the incentive constraint (8) binds, the principal only proposes using an exclusionary contract in circumstances in which the agent proposes such a contract when it is his “turn” to make the offer.

At first sight, this result might appear surprising: after all, the principal  $P_1$  does not care directly about whether or not the agent is excluded from working in period 2. Somewhat informally, the reason is as follows. Suppose we want to increase principal  $P_1$ 's expected payoff by \$1, while continuing to provide the agent with an incentive to work. To do this, we need first to decrease the payment  $x_1$  by  $\$ \frac{1}{p+\Delta p}$ ; while to preserve the agent's incentive to work, we need to increase  $\pi S$  by the same amount. The net effect on the agent's welfare is  $-\$1 - \$ \frac{1-p-\Delta p}{p+\Delta p} = -\$ \frac{1}{p+\Delta p}$ . So giving principal  $P_1$  an extra dollar of consumption while incentivizing the agent costs more than a dollar; while when the agent shirks, consumption can be transferred one-for-one. Consequently it is quite possible for the agent to prefer to use an incentive contract when he gets all the surplus, but for the principal to prefer to use a no-incentive contract when he gets the surplus; while the reverse is not true. See Figure 2.

**Lemma 3 (*Principal 1's bias is smaller*)**

*The principal  $P_1$  proposes an exclusionary contract only if the agent does.*

**Proof of Lemma 3:** There are two cases to consider, corresponding to whether or not the individual rationality constraint (7) or the incentive constraint (8) binds.

First, consider the (easier) case in which the incentive constraint (8) binds. Here, the principal  $P_1$  offers  $x_1 = \frac{B_1}{\Delta p} - S$ . Since it is the incentive constraint that binds we have  $\frac{pB_1}{\Delta p} \geq w + S$ . So

$$V_{inc} - V_{no-inc} = \Delta p H - B_1 + (p + \Delta p) S - \frac{pB_1}{\Delta p} + w$$

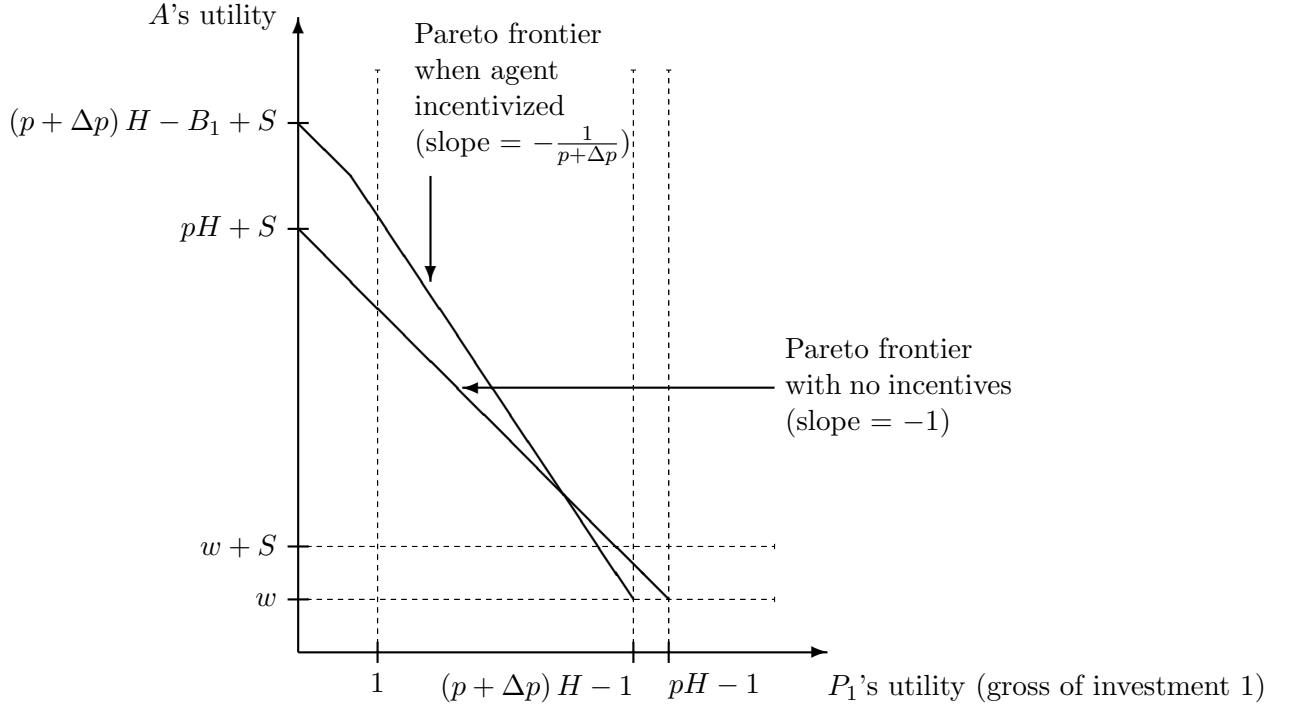


Figure 2: Pareto frontiers for incentive and no-incentive contracts

$$\begin{aligned}
 &\leq \Delta p H - B_1 - (1 - p - \Delta p) S \\
 &\leq U_{inc} - U_{no-inc}
 \end{aligned}$$

Second, consider the case in which the individual rationality constraint (7) binds. Here,  $x_1 = w + \frac{(1-p)B_1}{\Delta p}$  and  $\pi S = \frac{B_1}{\Delta p} - x_1$ . In this case,

$$\begin{aligned}
 V_{no-inc} &= U_{no-inc} - (w + S) \\
 V_{inc} &= (U_{inc} - (w + S))(p + \Delta p)
 \end{aligned}$$

These equalities can basically be read directly off Figure 2; alternatively, they can easily be algebraically verified. Clearly if  $U_{inc} < w + S < pH - 1 + S = U_{no-inc}$  then the agent will never propose an exclusionary contract, and nor will principal  $P_1$  since

$V_{inc} < 0$ . On the other hand, if  $U_{inc} \geq w + S$  then we have

$$V_{inc} - V_{no-inc} = U_{inc} - U_{no-inc} - (1 - p - \Delta p)(U_{inc} - (w + S)) \leq U_{inc} - U_{no-inc}$$

**QED**

Lemmas 1 and 3 together imply that constraints on contractual punishments are more likely to be appropriate when the agent is in a relatively powerful bargaining position today, but will not be tomorrow.

## 4 Uncertainty, growth and mobility

Thus far, we have assumed that in period 2 it is always efficient for the agent to work (but “shirk”!) for principal  $P_2$ . What happens if we relax this assumption.

Specifically, suppose now that the value of principal  $P_2$ 's project is stochastic, with a positive probability that  $pH - 1 < w$ . This uncertainty is publicly resolved after period 1, but before period 2. So the expected value of the agent's surplus in period 2, when not excluded, is now

$$S = w + \theta E[\max\{pH - 1 - w, 0\}]$$

What effect does this have on the inefficiency, or otherwise, or private contracting decisions? We focus for now on the case where the agent has all the bargaining power in period 1. On the one hand, note that

$$U_{inc} - U_{no-inc} = \Delta p H - B_1 - (1 - p - \Delta p) \left( \frac{B_1}{\Delta p} - H + \frac{1}{p + \Delta p} \right)$$

so that the agent's decision as to whether or not to use an exclusionary contract is *independent* of the surplus available in period 2. On the other hand, it is immediate from Lemma 1 that the magnitude of the agent's bias in favor of exclusionary contracts is increasing in  $S$ . Thus we have:

**Corollary 1 (*Growth and uncertainty*)**

*Both (1) an increase in the growth rate (higher expected  $pH$  in period 2), and (2) an increase in uncertainty in the value of second period  $pH$ , in the sense of second-order stochastic dominance, increases the agent's bias (relative to the social optimum) in favor of using an exclusionary contract.*

Historically, this is consistent with the fact that moves to circumvent the private use of exclusionary contracts have generally coincided with periods of increasing growth and change.

Moreover, note from the proof of Lemma 3 that when principal  $P_1$  makes an offer, then increasing surplus  $S$  makes him more likely to choose an exclusionary contract — while at the same time making it less likely that such a contract is socially efficient.

In particular, one economic and social change that may lead both to higher growth and greater uncertainty is an increase in mobility. Specifically, suppose that instead of having the opportunity just to deal with principal  $P_2$  at date 2, the agent can choose instead to deal with a third principal  $P'_2$  who has a project with expected output  $p'H'$ . In this case the agent's expected surplus is

$$S = w + \theta E [\max \{pH - 1 - w, p'H' - 1 - w, 0\}]$$

(We assume for now that the agent first decides which of principals  $P_2$  and  $P'_2$  to deal with, and then bargains over the surplus.) This gives:

**Corollary 2 (*Increased mobility*)**

*An increase in the agent's employment options at date 2 increases the agent's bias (relative to the social optimum) in favor of using an exclusionary contract.*



## 5 Over-confidence

We now turn to an alternative explanation that is often proposed for courts' refusal to enforce exclusionary contracts, namely that this is to protect agents from their own over-confidence. For example, McCormick (1935, page 601) writes:

It is a characteristic of men, however, that they are likely to be beguiled by the “illusions of hope,” and so feel so certain of their ability to carry out their engagements in future, that their confidence leads them to be willing to make extravagant promises and commitments as to what they are willing to suffer if they fail.

Many recent studies support this claim, by showing that individuals systematically overestimate their skill.<sup>4</sup>

Although the idea that contracting constraints are needed to protect agents from themselves is intuitive, in this section we will exhibit what we view as three important caveats.

First, even if the agent overestimates the benefits of the incentive contract relative to the no-incentive contract, the social costs related to him mistakenly choosing the incentive contract are reduced by over-confidence. The reason is that overconfident agents need less high-powered incentives, and so the exclusion probability  $\pi$  can be set lower.

Second, and related, if the degree of over-confidence is large enough then there may be no need for an exclusionary contract at all. Of course, overconfident agents may still sign up for incentive contracts when they should not do so; but now this represents an aggregate social benefit, since working is socially efficient.

---

<sup>4</sup>To give just one (well-known) example, individuals systematically overestimate their driving ability — see Svenson (1981).

Third, if an individual is overconfident of his ability this will lead him not just to underestimate the risk of failure today, but tomorrow also. The consequence of this is that he overestimates the value of not being excluded. This leads him to act more cautiously when considering the merits of an exclusionary contract.

We should note that the underlying logic of the idea that contract constraints protect an agent from himself is very close to the argument laid out above, that contracting constraints stop an agent imposing externalities on others. Instead of the externality being imposed on future trading partners, it is instead imposed on the agent himself.

## 5.1 Basics

Formally, we assume that an agent (incorrectly) believes that by exerting effort he can raise the probability of success by  $\Delta q > \Delta p$ . Moreover, we assume that the agent is aware that others do not share this opinion, i.e., he knows that principals believe that effort only raises the success probability by  $\Delta p$ . Finally, we assume that principals know that they are dealing with an overconfident agent, and that this fact is common knowledge.

In this section, we restrict attention (at least for now) to the case in which the agent has all the bargaining power in both periods. Thus there is no longer any need for constraints on contracting to protect future trading partners. Instead, any constraints will stem from the need to protect the agent from himself.

Let  $S$  denote the period 2 surplus (which goes entirely to the agent), and  $\hat{S}$  the agent's belief about the level of this surplus. If

$$(p + \Delta p) \left( H - \frac{B_2}{\Delta q} \right) - 1 < 0 \tag{9}$$

then it is impossible to supply the agent with incentives in period 2, and so  $S = \hat{S} = pH - 1$ . On the other hand, if (9) does not hold then incentives can be provided.

The agent offers the contract  $x_2 = H - \frac{1}{p+\Delta p}$  which is the highest wage principal 2 is prepared to pay. The agent is motivated to work, since doing so raises his expected income by  $\Delta q x_2 = \Delta q \left( H - \frac{1}{p+\Delta p} \right)$  which by the assumption that (9) does not hold is greater than the cost of working,  $B_2$ . So in this case,

$$\begin{aligned} S &= (p + \Delta p) H - 1 - B_2 \\ \hat{S} &= (p + \Delta q) \left( H - \frac{1}{p + \Delta p} \right) - B_2 = S + (\Delta p - \Delta q) \left( H - \frac{1}{p + \Delta p} \right) \end{aligned}$$

In period 1, the agent works only under a contract  $(x_1, \pi)$  that satisfies

$$(x_1 + \pi \hat{S}) \Delta q \geq B_1$$

Principal 1 accepts such a contract if and only if

$$(p + \Delta p) (H - x_1) - 1 \geq 0$$

Since the agent proposes the incentive contract, he will always offer

$$\hat{x}_1 = H - \frac{1}{p + \Delta p}$$

Moreover, whenever exclusion incentives are used (i.e.  $\pi > 0$ ) the agent's preferred contract  $(\hat{x}_1, \hat{\pi})$  satisfies

$$\hat{x}_1 + \hat{\pi} \hat{S} = \frac{B_1}{\Delta q}$$

Observe that because the agent is overconfident of success, he must be offered fewer incentives than would otherwise be the case.

## 5.2 The “classical” case

As far as we can tell, the case which proponents of over-confidence as a basis for contracting constraints have in mind is that where

$$(p + \Delta p) \left( H - \frac{B_1}{\Delta q} \right) - 1 < 0 \quad (10)$$

and  $S = \hat{S}$ . That is, even under over-confidence monetary incentives are insufficient to persuade the agent to work in period 1, and the agent correctly perceives the period 2 surplus. Under this case, the agent's perceived welfare from offering his preferred incentive contract is

$$\begin{aligned} \hat{U}_{inc} &= (p + \Delta q) \hat{x}_1 - B_1 - (1 - p - \Delta q) \hat{\pi} \hat{S} + \hat{S} \\ &= pH - 1 - B_1 - (1 - p - \Delta p) \hat{\pi} \hat{S} + \hat{S} + (\Delta q - \Delta p) \left( \hat{x}_1 + \hat{\pi} \hat{S} \right) \\ &= SW_{inc} + \left( 1 - \frac{\Delta p}{\Delta q} \right) B_1 \end{aligned}$$

while

$$\hat{U}_{no-inc} = pH - 1 + \hat{S} = SW_{no-inc}$$

Thus the agent overestimates the gains of the incentive contract (relative to any bias in his estimate of the value of the no-incentive contract) by

$$\left( \hat{U}_{inc} - SW_{inc} \right) + \left( \hat{U}_{no-inc} - SW_{no-inc} \right) = \left( 1 - \frac{\Delta p}{\Delta q} \right) B_1 > 0$$

So in this case holds, the “classical” intuition is correct: the agent underestimates the probability of failure, and so overestimates the value of binding himself using socially costly incentives. The bias gets worse as the the degree of over-confidence worsens (i.e.  $\Delta q$  rises). This is true even in spite of the offsetting effect that over-confidence actually allows for a reduction in the exclusion probability  $\hat{\pi}$ .

Finally, note that

$$\begin{aligned} SW_{inc} - SW_{no-inc} &= \Delta p H - B_1 - (1 - p - \Delta p) \hat{\pi} S \\ &= \Delta p H - B_1 - (1 - p - \Delta p) \left( \frac{B_1}{\Delta q} - H + \frac{1}{p + \Delta p} \right) \\ &= (1 - p) \left( H - \frac{B_1}{\Delta q} - \frac{1}{p + \Delta p} \right) - \left( 1 - \frac{\Delta p}{\Delta q} \right) B_1 + \frac{\Delta p}{p + \Delta p} \end{aligned}$$

Thus as  $\Delta q$  increases, the social cost from using an incentive contract when it is socially inefficient is smaller.

### 5.3 Do overconfident agents need to be excluded?

Our second caveat to this classical case for restrictions on contracting is that for  $\Delta q$  large enough it is quite possible for inequality (10) not to hold, even though (IC<sub>1</sub>) does. In this case, there is simply no need to use exclusion as a punishment, and  $\hat{\pi} = 0$ . Sticking for now with the case where  $S = \hat{S}$ , we have

$$\begin{aligned} \left(\hat{U}_{inc} - SW_{inc}\right) + \left(\hat{U}_{no-inc} - SW_{no-inc}\right) &= (\Delta q - \Delta p) \left(H - \frac{1}{p + \Delta p}\right) \\ &\geq \left(1 - \frac{\Delta p}{\Delta q}\right) B_1 \end{aligned}$$

from the assumption that (10) does not hold. On the one hand, the agent is now actually more biased in signing up for the incentive contract. But from the perspective of aggregate welfare, this bias is harmless — it is always efficient for the agent to work hard, and now providing the incentives for him to do so imposes no social cost.

### 5.4 Overconfident about tomorrow

Our third caveat relates to the agent's overestimate of the value of tomorrow's surplus. Consider again the case where pure monetary incentives cannot be used to induce the agent to work in period 1 (i.e. (10) holds), but suppose now that (9) does not hold so that  $\hat{S} > S$ . Now,

$$\begin{aligned} \hat{U}_{inc} &= (p + \Delta q) \hat{x}_1 - B_1 - (1 - p - \Delta q) \hat{\pi} \hat{S} + \hat{S} \\ &= pH - 1 - B_1 - (1 - p - \Delta p) \hat{\pi} \hat{S} + \hat{S} + (\Delta q - \Delta p) (\hat{x}_1 + \hat{\pi} \hat{S}) \\ &= pH - 1 - B_1 - (1 - p - \Delta p) \hat{\pi} S + S + \left(1 - \frac{\Delta p}{\Delta q}\right) B_1 + (\hat{S} - S) \end{aligned}$$

$$\begin{aligned}
& - (1 - p - \Delta p) \hat{\pi} (\hat{S} - S) \\
= & SW_{inc} + \left(1 - \frac{\Delta p}{\Delta q}\right) B_1 + (\hat{S} - S) - (1 - p - \Delta p) \hat{\pi} (\hat{S} - S)
\end{aligned}$$

while

$$\hat{U}_{no-inc} = pH - 1 + \hat{S} = SW_{no-inc} + (\hat{S} - S)$$

So

$$\begin{aligned}
& (\hat{U}_{inc} - SW_{inc}) + (\hat{U}_{no-inc} - SW_{no-inc}) \\
= & \left(1 - \frac{\Delta p}{\Delta q}\right) B_1 - (1 - p - \Delta p) \hat{\pi} (\hat{S} - S) \\
= & (\Delta q - \Delta p) \left(\frac{B_1}{\Delta q} - \hat{\pi} (1 - p - \Delta p) \left(H - \frac{1}{p + \Delta p}\right)\right)
\end{aligned}$$

The direction of the agent's bias is no longer clear. On the one hand, he underestimates the probability of exclusion. But there is an offsetting effect — he overestimates the cost that being excluded imposes on him.<sup>5</sup>

## 6 Planned extensions

- “General equilibrium” effects:  $\theta_2$  is endogenous to  $\pi$ , and  $w_1$  might be replaced by an expression involving  $\theta_1$  (since the IR constraint would be endogenous).
- More formal welfare calculations, especially in the over-confidence case.
- Renegotiation in period 2.
- Screening/signalling with overconfident and “rational” agents.
- “Coercive” bargaining.

---

<sup>5</sup>Recall that

$$\hat{\pi} = \frac{1}{\hat{S}} \left( \frac{B_1}{\Delta q} - H + \frac{1}{p + \Delta p} \right).$$

## References

- Philippe Aghion and Benjamin Hermalin. Legal restrictions on private contracts can enhance efficiency. *Journal of Law, Economics and Organization*, 6(2):381–409, Fall 1990.
- Colin Camerer, Sam Issacharoff, George Lowenstein, Ted O’Donoghue, and Matthew Rabin. Regulation for conservative: Behavioral economics and the case for “asymmetric paternalism”. *Working Paper*, 2002.
- Tai-Yeong Chung. On the social optimality of liquidated damage clauses: An economic analysis. *Journal of Law, Economics, and Organization*, 8(2):280–305, 1992.
- Michael Suk-Young Chwe. Why were workers whipped? pain in a principal-agent model. *Economic Journal*, 100(403):1109–1121, December 1990. 403.
- Peter J Coleman. The insolvent debtor in Rhode Island 1745-1828. *William and Mary Quarterly*, 22(3):413–434, July 1965.
- Peter A. Diamond and Eric Maskin. An equilibrium analysis of search and breach of contract, i: Steady states. *Bell Journal of Economics*, 10:282–316, Spring 1979. 1.
- Robert A. Feer. Imprisonment for debt in Massachusetts before 1800. *Mississippi Valley Historical Review*, 48(2):252–269, September 1961.
- David W. Galenson. The rise and fall of indentured servitude in the americas: An economic analysis. *Journal of Economic History*, 44(1):1–26, March 1984.
- Oliver Hart. *Firms, Contracts and Financial Structure*. Oxford University Press, 1995.
- Anthony T Kronman. Paternalism and the law of contracts. *Yale Law Journal*, 92: 763–798, april 1983.

- Bruce H Mann. *Republic of Debtors: Bankruptcy in the Age of American Independence*. Harvard University, 2002.
- Charles T McCormick, editor. *Handbook on the Law of Damages*. West, 1935.
- James A Mirrlees. The theory of moral hazard and unobservable behaviour: Part I. *Review of Economic Studies*, 66(1):3–21, January 1999.
- Dilip Mookherjee and Ivan Png. Optimal auditing, insurance and redistribution. *Quarterly Journal of Economics*, 104(2):399–415, May 1989.
- Edmund Sears Morgan. *American Slavery, American Freedom: : the ordeal of colonial Virginia*. Norton, 1975.
- Edwin T. Randall. Imprisonment for debt in America: Fact and fiction. *Mississippi Valley Historical Review*, 39(1):89–102, June 1952.
- Kathryn E. Spier and Michael D. Whinston. On the efficiency of privately stipulated damages for breach of contract: entry, reliance, and renegotiation. *RAND Journal of Economics*, 26(2):180–202, 1995.
- Ola Svenson. Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47:143–148, 1981.