99s-37

Optimal Justice in a General Equilibrium Model with Non Observable Individual Productivities

Pierre Lasserre, Antoine Soubeyran

Série Scientifique
Scientific Series

CIRANO
Centre interuniversitaire de recherche en analyse des organisations

Montréal
Octobre 1999

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du ministère de la Recherche, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de la Recherche, de la Science et de la Technologie, and grants and research mandates obtained by its research teams.

Les organisations-partenaires / The Partner Organizations

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

ISSN 1198-8177

Optimal Justice in a General Equilibrium Model with Non Observable Individual Properties*

Pierre Lasserre†, *Antoine Soubeyran*‡

Résumé / Abstract

Nous étudions dans un modèle d'équilibre général le rôle du système judiciaire comme instrument d'allocation des activités de production et de prédation. Les décisions individuelles se prennent compte tenu, à la fois des sentences que l'on peut escompter à la suite d'activités prohibées, et de la mesure dans laquelle, à l'équilibre, les producteurs peuvent s'approprier le fruit de leurs efforts. Comme les capacités productives diffèrent d'une personne à l'autre, il est socialement désirable de prendre en compte des considérations d'équité et d'efficacité dans la conception des institutions : la prédation est une forme de redistribution des plus productifs vers les moins productifs. Nous étudions les faits stylisés qui caractériseraient un tel système lorsque les productivités individuelles ne sont pas observables et que tant le niveau de détection que la précision des décisions de justice sont des caractéristiques coûteuses des institutions.

In this general equilibrium model, justice and police institutions are treated as a mechanism that induces individuals to extend some desirable productive effort. This determines individual encroachment activities which in turn determine the proportion of aggregate production that fails to be appropriated, and the private incentives to choose productive activities. Since individuals have different productive abilities society would ideally take both equity and efficiency into consideration in the design of its institutions: encroachment is a form of redistribution from the most talented individuals to the least talented ones. We study the stylized properties that should arise when individual productivities are not observable by the system, and when both detection levels and justice accurary are costly instruments.

Mots Clés : Justice, institutions, incitations, crime, information, observabilité, agence, équilibre général

Keywords: Justice, institutions, incentives, crime, information, observability, agency, general equilibrium

JEL: D23, D62, D63, D7, D82

---

* Corresponding Author : Pierre Lasserre, CIRANO, 2020 rue University, 25ème étage, Montréal, Qc, Canada H3A 2A5 Tél. : (514) 985-4011 Fax : (514) 985-4039 courriel : lasserrp@cirano.umontreal.ca
Financial support from theFCAR funds and the CNRS is gratefully acknowledged.

† Université du Québec à Montréal and CIRANO

‡ Université de la Méditerranée and GREQAM

1. Introduction

Consider a society made up of n individuals who may differ in productivity, and allocate their time between production, and what, for lack of a better word, we will call encroachment or cheating. By that we mean any individual activity or attitude which reduces collective output, reduces the share of total output accruing to others, or reduces others' utility. The corresponding acts may be considered licit, or illicit, depending on a society's moral standards and rules. In fact, this is one of the issues that this paper is addressing in a crude way. For example, stealing from other individuals is the typical predatory activity, but tax evasion, not working hard, and even leisure are sometimes considered as various degrees of encroachment to the extent that they directly or indirectly impinge on others' welfare. Another example is speeding, which reduces other motorists' safety and increases the drain on collective health resources. In what may constitute collective attempts to promote not only efficiency and productivity, but also equity, societies attach various levels of approval to such human actions or behaviors, and their institutions reflect these collective attitudes. Theft is punished at various degrees; petty crime may be left unpunished; leisure is often frowned upon.

In any society, encroachment occurs to some extent; it involves activities that take some production away from individual producers, directly in the case of stealing, or indirectly in the case of tax evasion or various kinds of free riding. Thus there is a problem of incomplete appropriation which reduces incentives to production. As will be made clear further below, the acuteness of this problem depends on the extent to which property rights are enforced. Furthermore, equity considerations, enforcement costs, and enforcement errors, but also information and agency issues, are as many factors that militate against or prevent complete appropriation.

These issues have received attention in the literature, as described in the excellent survey by Marceau and Mongrain (1997). In particular several scholars have tried to explain why perfect enforcement is not simply arrived at by imposing very high sanctions. Indeed, infinite penalties (death under torture followed by damnation), would reduce the incidence of crime to zero and provide perfect protection of property rights, thus maximizing production. One reason is that sanctions may be costly to impose and costs may increase with the level of sanction (Polinsky and Shavell, 1979; Shavell, 1987; Kaplow, 1990). Also, high sanctions may induce costly avoidance activities (Malik, 1990). Furthermore imposing high sanctions in an imperfect fashion may imply undesirable social costs: individuals may be imperfectly informed about the consequences of their acts (Kaplow, 1990; Bebchuck and Kaplow, 1992); innocents may be condemned (Ehrlich, 1975); juries may be reluctant to convict if the sanction is too high (Andreoni, 1991). Finally, if individuals are risk averse, Polinsky and Shavell (1979) have argued that it might be desirable to leave some crimes unpunished.

Moreover, it may be socially desirable to discriminate between individuals in crime tolerance: it is all right for a lost hiker to break into a log cabin, make fire, and save his life; Bill Gates in the same situation might be encouraged to call a helicopter. Indeed, from an equity point of view, encroachment may be viewed as the correction of an inequity in the distribution of talents or other forms of wealth across individuals. It may be desirable for institutions to allow or encourage a distribution of roles in society where more productive individuals appropriate their output only partially, while the least endowed ones partially offset their disadvantage by encroaching on others' production. This arbitrage between equity and production is a feature of the model presented below.

Whether avoiding the use of maximum sanctions is justified by costs, redistributive reasons, or otherwise, this leads to the question whether sanctions should be proportional to the gravity of crime. In particular Shavell (1991) and Andreoni (1991) describe conditions calling for monotonic sentences. In the paper by Shavell, this happens if the detection effort is identical for all crimes while their gravities differ: using uniform sentences would then lead to excessive marginal dissuasion of small crimes. Andreoni argues that various characteristics of the judiciary system make it less likely that a given crime be punished if the sentence is high; again, this justifies monotonic sentences. In our paper, a similar result arises because of self selection constraints, in a model where both detection efforts and dissuasion are endogenous.

Indeed the choice of judiciary institutions is complicated by agency issues: institutions exist in order to affect behaviors that cannot be dictated directly because of information asymmetries; ideally they should be designed in such a way as to achieve their objective subject to these informational constraints. This is the approach taken in this paper following the early lead of Mirrlees (1971). The selection problem is complicated in our model by the presence of endogenous costs of running the institutions, and by feedback effects: sentences discourage encroachment, which in turn encourage production. The optimum sentence schedule takes this equilibrium effect into account, and this explains why we obtain a result which is unusual in the agency literature: under the optimum incentive penalty schedule, the effort levels not only of less talented individuals, but also of more talented ones, are distorted relative to the full information optimum.

Property rights are costly to protect; often, it may be preferable not to establish them or, equivalently, not to enforce them. In general, even if enforcement is not subject to decreasing returns, it is likely that some incompleteness in property rights enforcement is socially desirable. Although enforcement is more often considered to be a public responsibility, as in this paper, several authors have looked at the private allocation of resources to enforcement (De Meza and Gould, 1992; Lasserre, 1994). Typically private involvement does not lead to a socially optimal allocation of enforcement resources. In this paper, we consider only public enforcement.

Institutions are also constrained by the instruments available to the justice system. Some instruments, such as fines, fullfil both a role of compensating victims or society, and a role of discouraging misbehavior; other instruments, such as imprisonment have only a disincentive effect. In order to be effective, they must be wasteful. Since compensation requires the culprit to be endowed with some wealth or productivity, we argue that the former type of instruments is more readily available to rich societies than to poor ones, and, among rich societies, that it is more accessible to homogenous or egalitarian societies than to heterogenous or unequal societies. To the extent that our aim is not to investigate how such societal characteristics are arrived at, we take the instruments available to the system as exogenously given.

The issues addressed in this paper are very similar to those considered in the literature on conflicts. Differences lay at methodological and, perhaps, philosophical levels. The literature on conflicts is resolutely game theoretic. Recent papers such as Neary (1997) analyze subgame perfect equilibria of the allocation of wealth between players without introducing any institutions. Property rights are defined by whatever power (arms) each player possesses or privately acquires.

In contrast, our paper uses a normative approach and introduces institutions ensuring that the resources devoted to property rights enforcement are public resources. We do not mean to present any convincing model of the constitutional stage nor do we claim that societies endowed with optimal institutions are the only ones that survive. However our model leads to stylized societal characteristics that may be confronted with casual observation in order to start investigating that question.

The paper is organized as follows. In the next section we present the model and its equilibrium, first in a tragedy of the commons context, then in a situation of costlessly enforced property rights which we call *Utopia*. In Section 3, we study the design of justice institutions in a world of costly enforcement and information asymmetry. Various cases arise according to the severity of the appropriation problem faced by the economy, the relative sizes of the groups that compose society, and the relative productivities of the individuals in each group. The optimal institutional arrangement is defined in each case in terms of sentences, police detection, and justice accuracy. Finally, in the conclusion, we summarize our main results and discuss the stylized characteristics which they may imply for a society.

2. The model and the challenge

2.1. Production and the use of time

Society is made up of $\underline{n}$ individual whose productivity $\underline{\theta}$ is relatively low, and $\bar{n}$ individuals whose productivity $\bar{\theta}$ is relatively high. Productivity levels determine production in the following sense. Individuals have a maximum of one unit of effort (time) to allocate between productive and unproductive activities. When an individual of productivity θ ($\theta \in \{\underline{\theta}, \bar{\theta}\}$) extends a productive effort e ($e \in [0, 1]$), his production is θe. Consequently, aggregate output is $Q = \underline{n}\,\underline{\theta}\,\underline{e} + \bar{n}\bar{\theta}\bar{e}$. On the other hand, when an individual, whatever his productivity, devotes $1-e$ units of his time to encroachment, he appropriates himself a booty of $[1-e]\,b$ which is taken from society's total output. Subject to the constraint that aggregate booties must not exceed aggregate output, the reward from encroachment is higher, the more time is devoted to that activity by an individual; however, the less time is devoted to production, the more likely the aggregate output constraint is to be binding. If it is binding, all production is taken away from producers; as a result no-one produces and the economy does not exist. This is the extreme variant of the tragedy of the commons analyzed further below.

Let the aggregate time spent encroaching on others' production be $Z = \underline{n}\,[1-\underline{e}] + \bar{n}\,[1-\bar{e}]$. We note that, since individuals each have one unit of time to allocate between production and encroachment, the total time available is $n = \underline{n} + \bar{n}$, and $Z \leq n$. A proportion α of aggregate output is taken away from private production to become the reward of encroachment

$$\alpha(\underline{e},\bar{e}) = b\frac{Z}{Q} = b\frac{n-\underline{n}\,\underline{e}-\bar{n}\bar{e}}{\underline{n}\,\underline{\theta}\,\underline{e}+\bar{n}\bar{\theta}\bar{e}} \tag{1}$$
$$0 \leq \alpha \leq 1$$

In the rest of the paper we assume without loss of generality that $n = 1$ and that $\underline{n}$ and $\bar{n}$ are population shares. At the individual level, the proportion of production which is appropriated is stochastic, but we assume that all individuals face the same odds in that respect, so that an individual who spends $e(\theta)$ units of time in production activities will expect to get $[1-\alpha]\,\theta e(\theta)$ from his own production. The combined expected income from production and encroachment is thus, for an individual of productivity θ

$$Y(e,\theta,\alpha) = [1-\alpha]\,\theta e + [1-e]\,b \tag{2}$$

Given the long-run, institutional, focus of this paper, it is reasonable to assume (and not explicitly model) that individuals are able to ensure themselves against the vagaries of encroachment, so that their utility is based on expected income. What is not insurable, however, is the fundamental risk of being born of one type rather than another. This risk will be given consideration in due course.

2.2. No institutions: the tragedy of the commons

In the absence of any institutional constraints, each individual, taking α as given, will choose e so as to maximize the utility of expected income

$$V(e,\theta,\alpha) \equiv u(Y(e,\theta,\alpha)) \tag{3}$$

Since $\frac{\partial V}{\partial e} = u'\left[[1-\alpha]\,\theta - b\right]$, one has $e = 0$ or $e = 1$ according to the sign of $[1-\alpha]\,\theta - b$. Since the pair $(\underline{e},\bar{e}) = (1,0)$ violates at least one first-order condition, this gives three candidate equilibria: $(0,0)$, or $(0,1)$, or $(1,1)$. Whether these candidates are actual equilibria depends on parameters and on the value of α that they induce. Except at threshold values of b, any equilibrium is locally stable. Proposition 1 describes the configurations that may arise according to the value of the booty.

Proposition 1 *(tragedy of the commons) For any b, effort allocation $(\underline{e},\bar{e}) = (0,0)$ is a locally stable equilibrium. Furthermore*

1. If $b < \frac{\bar{n}\bar{\theta}}{\underline{n}\,\underline{\theta}+\bar{n}\bar{\theta}}\underline{\theta}$ then $(1,1)$ is an additional, locally stable, equilibrium, with $\alpha = 0$.
2. If $\frac{\bar{n}\bar{\theta}}{\underline{n}\,\underline{\theta}+\bar{n}\bar{\theta}}\underline{\theta} < b < \min\left\{\bar{n}\bar{\theta},\underline{\theta}\right\}$ then $(1,1)$ and $(0,1)$ are two additional, locally stable, equilibria, with $\alpha = 0$ and $\alpha = b\frac{\underline{n}}{\bar{n}\bar{\theta}}$ respectively;
3. If $\min\left\{\bar{n}\bar{\theta},\underline{\theta}\right\} < b < \max\left\{\bar{n}\bar{\theta},\underline{\theta}\right\}$ then $(0,1)$ is an additional, locally stable, equilibrium, with $\alpha = b\frac{\underline{n}}{\bar{n}\bar{\theta}}$;

Proposition 1 makes clear that the nature of the appropriation problem faced by the economy changes as b increases. At low values of b society faces a coordination problem in order to avoid the *institutional trap* equilibrium $(0,0)$; however the alternative equilibrium $(1,1)$ is Pareto optimal. At intermediate values of b, the Pareto optimal allocation is also an equilibrium, but there are two alternative *institutional trap* equilibria: $(0,0)$, and a milder version of the *tragedy of the commons* $(0,1)$. At still higher values of b, the situation becomes worth as the only possible equilibria are the mild form of the *tragedy of the commons* $(0,1)$, or its extreme form $(0,0)$. The latter form is the only equilibrium when b exceeds $\max\left\{\bar{n}\bar{\theta},\underline{\theta}\right\}$: here the economy is kept out of existence by appropriation failure.

One notes that these outcomes arise not only according to the value of b relative to $\underline{\theta}$ and $\bar{\theta}$, but also according to the value of $\bar{n}$. While $\underline{\theta}$ and $\bar{\theta}$ describe the production technology, b may be viewed as a parameter describing the technology of encroachment activities and, more fundamentally, the severity of the appropriation problem faced by an economy; it determines how easy and rewarding encroachment is relative to production from an individual point of view. $\bar{n}$ is a sociological data: if the more productive group is sizeable, as, perhaps, in a developed society, then the damage from encroachment will be spread over a higher number of individuals so that appropriation failure is less likely to threaten the very existence of society.

In the rest of the paper, in order to focus on serious cases that go beyond coordination failure, we maintain the following assumption

Assumption 1 *:* $b > \min\left\{\bar{n}\bar{\theta},\underline{\theta}\right\}$

2.3. Utopia or the illusion of costless internalization

In any society, however primitive, there is a distribution of power which defines, perhaps only imprecisely or implicitly, some notion of property rights. A tragedy of the commons occurs when property rights are insufficient for a society to realize its economic potential. In the model just introduced, all individuals have the same access to encroachment, but production is penalized relative to encroachment as an activity. In that primitive social state, when deciding how much to produce, an individual does not consider the contribution of his production to others' well-being; similarly, in the choice of his encroachment activities, he does not consider the fact that his actions reduce the pool from which others predate, nor does he take into account the fact that, in the aggregate, encroachment discourage production. A society with more extensive property rights and enforcement institutions may constitute an improvement over that situation. If predators are punished for stealing, they will devote less time to that activity and turn to production; simultaneously, if producers keep a higher proportion of their output for private consumption or exchange, they will be encouraged to produce more. This may yield a Pareto improvement over the initial situation.

As Coase has emphasized, the traditional analysis of that issue is presented in a framework where the internalization of externalities is costless. As such it is as utopic as the analysis of an economy without transaction costs. Nevertheless, the 'social optimum' defined in our textbooks provides an interesting benchmark as it defines a desirable state that can perhaps be aimed at and certainly referred to. We call *Utopia* this state of reference whose analysis follows.

Assuming that any feasible time (or effort) allocation could be imposed upon any individual at no cost, what allocation would be most desirable? One way to approach the question is to ask it from behind the Rawlsian veil of ignorance, that is to say from a point of view where individuals are not yet aware of the productivity with which they will be endowed. In this mental exercise, an individual will picture himself endowed with some hypothetical productivity in society; precisely, his probability to be of type $\bar{\theta}$ is $\bar{n}$, and his probability to be of type $\underline{\theta}$ is $\underline{n}$. This is true of all individuals. Since all individuals are identical *ex ante* in this mental experiment, the social optimum is the Harsanyi solution; it may be found by maximizing the expected utility of any single individual

$$W(\underline{e}, \bar{e}) = \underline{n} V(\underline{e}, \underline{\theta}, \alpha) + \bar{n} V\left(\bar{e}, \bar{\theta}, \alpha\right) \tag{4}$$

subject to

$$0 \leq e \leq 1, \; e = \bar{e} \text{ or } \underline{e} \tag{5}$$

where α is given by (1). We will maintain the following assumption:

Assumption 2 $W_{\underline{e}\,\underline{e}}(\underline{e}, 1) \leq 0$.

Let u'_m be a weighted average across individuals of marginal utilities of income, the weights being individual shares in aggregate output

$$u'_m(\underline{e}, \bar{e}) = \frac{\underline{n}\,\underline{\theta}\,\underline{e}}{Q}\underline{u}' + \frac{\bar{n}\bar{\theta}\bar{e}}{Q}\bar{u}' \tag{6}$$

where $\underline{u}' = u'(Y(\underline{e}, \underline{\theta}, \alpha))$ and $\bar{u}' = u'\left(Y\left(\bar{e}, \bar{\theta}, \alpha\right)\right)$ with α is defined by (1). Substituting (1) into (4), one obtains, after some algebra

$$W_{\underline{e}} = \underline{n}\left[\underline{\theta}\,\underline{u}' + [u'_m - \underline{u}']\,[\alpha\underline{\theta} + b]\right] \tag{7}$$

$$W_{\bar{e}} = \bar{n}\left[\bar{\theta}\bar{u}' + [u'_m - \bar{u}']\left[\alpha\bar{\theta} + b\right]\right] \tag{8}$$

These first-order conditions illustrate the arbitrage faced by the Rawlsian planner in the Harsanyi solution. Consider type $\bar{\theta}$. For each of the $\bar{n}$ individuals in that group, increasing their effort yields an increase $\bar{\theta}$ in output which is valued at their marginal utility (the first term in the expression between brackets); of this, a proportion α is taken away by encroachment, while the increased productive effort reduces the return from encroachment by b (a loss to each type $\bar{\theta}$ individual of $-\bar{u}'\left[\alpha\bar{\theta} + b\right]$); finally, the increased productive effort has the additional effect of reducing α by an amount proportional to $\left[\alpha\bar{\theta} + b\right]$ which benefits both types and is thus valued at the population average marginal utility u'_m (thus the term $u'_m\left[\alpha\bar{\theta} + b\right]$). As written the formula emphasizes that the planner's problem boils down to an arbitrage between production and income distribution: if, as intuition suggests should be the case and as we formally prove in the Appendix (Lemma 1), the more productive types have a higher than average income, then $u'_m - \bar{u}'$ is positive so that $W_{\bar{e}}$ is unambiguously positive. The planner should set $\bar{e} = 1$. By a similar argument, $u'_m - \underline{u}'$ is then negative, so that $W_{\underline{e}}$ may be negative at high levels of $\underline{e}$: the planner may wish to set $\underline{e}$ at less than 1. Thus society would accept a level of aggregate production falling short of the theoretical maximum in order to achieve a more equal distribution of income by letting less productive individuals use encroachment to improve their lot.

Sofar, we have established that the *Utopian* solution $\left(\underline{e}^U, \bar{e}^U\right)$ is such that $\underline{e}^U \leq \bar{e}^U = 1$. In general, parameter configurations such that $\underline{e}^U = 0$ is optimal cannot be ruled out. However, it is interesting to anticipate on the next section and note at this stage that, in all circumstances where the introduction of justice institutions would possibly raise an incentive (selection) issue, the *Utopian* solution is $\underline{e}^U > 0$. This can be shown by noting that, in a system where penalties are used to induce the proper effort on the part of individuals, no penalty threat need be imposed on the most productive individuals if they choose $\bar{e} = 1$ anyway. They will do so if $V_e\left(e, \bar{\theta}, \alpha\right) > 0$ or $[1-\alpha]\,\bar{\theta} - b > 0$. Since α is decreasing in both $\bar{e}$ and $\underline{e}$, this condition is satisfied for all $\underline{e}$ if it holds for $\underline{e} = 0$ in which case $\alpha\left(0,1\right) = \frac{\underline{n}b}{\bar{n}\bar{\theta}}$. Substituting into the inequality, a selection problem ($\bar{e} < 1$ for some $\underline{e}$) may possibly arise only if $b \geq \bar{\theta}\bar{n}$. However, when[1] $b \geq \bar{\theta}\bar{n}$, setting $\underline{e} = 0$ and $\bar{e} = 1$ would cause $\underline{y}$ to exceed $\bar{y}$, which is ruled out by Lemma 1 (for details, see the proofs of Lemma 1 and Proposition 1). Consequently, in any parameter configuration involving a selection issue, the *Utopian* solution is necessarily such that $0 < \underline{e}^U$.

In the rest of this section, we will investigate under which conditions $\underline{e}^U < 1$, meaning that the *Utopian* society tolerates or encourages some encroachment activities by its less talented individuals. Given that $\bar{e}^U = 1$ and that Assumption 2 holds, a necessary and sufficient condition for $\underline{e}^U = 1$ is $W_{\underline{e}}\left(1,1\right) \geq 0$. In that case, $\alpha = 0$, $\underline{u}' = u'\left(\underline{\theta}\right)$, $\bar{u}' = u'\left(\bar{\theta}\right)$ and $u'_m = \frac{\underline{n}\,\underline{\theta}u'\left(\underline{\theta}\right) + \bar{n}\bar{\theta}u'\left(\bar{\theta}\right)}{\underline{n}\,\underline{\theta} + \bar{n}\bar{\theta}}$. Substituting these values, together with $\underline{e} = \bar{e} = 1$, into (7), a necessary and sufficient condition for $\underline{e}^U = 1$ is

$$\underline{\theta}u'\left(\underline{\theta}\right) + \left[\frac{\underline{n}\,\underline{\theta}u'\left(\underline{\theta}\right) + \bar{n}\bar{\theta}u'\left(\bar{\theta}\right)}{\underline{n}\,\underline{\theta} + \bar{n}\bar{\theta}} - u'\left(\underline{\theta}\right)\right] b \geq 0$$

or

$$b \leq \underline{\theta}\frac{u'\left(\underline{\theta}\right)}{u'\left(\underline{\theta}\right) - u'\left(\bar{\theta}\right)}\frac{\underline{n}\,\underline{\theta} + \bar{n}\bar{\theta}}{\bar{n}\bar{\theta}} \equiv b^U \tag{9}$$

If $b \leq b^U$, W has its maximum at $\underline{e}^U = \bar{e}^U = 1$. If $b > b^U$, $W_{\underline{e}}\left(1,1\right) < 0$ so that $\underline{e}^U < 1$.

The results of the *Utopia* model are gathered in the following proposition (proof components not already in the main text are in the Appendix).

Proposition 2 *(the* Utopian *optimum)*

Let b satisfy Assumption 1 *and W satisfy Assumption 2. At the* Utopian *optimum:*

1. *More productive individuals devote their whole time to production, while less productive individuals may devote some time to encroachment:* $0 < \underline{e}^U \leq \bar{e}^U = 1$*;*
2. *If b is relatively low, then it is socially optimal for less productive individuals to devote their whole time to production:* $b \leq b^U \Rightarrow \underline{e}^U = 1$*; If b is relatively high, then it is socially optimal for less productive types to devote some time to encroachment:* $b > b^U \Rightarrow \underline{e}^U < 1$*;*

[1]This provides also a complementary interpretation of Assumption 1.

The second element in the proposition implies that strict property-rights enforcement may not be desirable: when the reward from encroachment is sufficiently high, it is optimal for society that the production of the most productive individuals be incompletely appropriated, thus allowing a transfer to the least endowed through encroachment. The value of b above which this is true, as defined by (9), reflects a comparison between the reward from productive activities and the reward from encroachment for the least productive individuals. Since encroachment imposes a cost on the most productive types, it is not desirable unless its reward to the less endowed ones exceeds the reward from production by a sufficient margin. It is legitimate to ask what technological and sociological characteristics underlie b^U.[2] The answer is provided in the next proposition.

Proposition 3 *(encroachment in* Utopia*)*
The critical booty level b^U above which encroachment by the least productive individuals is desirable is defined by (9)*; b^U is higher (lower):*

1. *the more numerous the least (most) productive group* $(\frac{\partial b^U}{\partial \underline{n}} > 0;\ \frac{\partial b^U}{\partial \bar{n}} < 0)$*;*
2. *the more productive the least (most) talented individuals* $(\frac{\partial b^U}{\partial \theta} < 0;\ \frac{\partial b^U}{\partial \underline{\theta}} > 0)$*;*
3. *the lower (higher) risk aversion, or the narrower (wider) the productivity difference between the two groups; furthermore, $b^U \to +\infty$ when $u'(\underline{\theta}) - u'(\bar{\theta}) \to 0$.*

These results are fairly intuitive: when least productive individuals are numerous, the collective negative impact of their encroachment is high; only relatively high private benefits to deprived individuals will make such social cost acceptable. Similarly, when $\bar{\theta}$ is high, the most productive group is better able to cope collectively with encroachment so that it may be tolerated at relatively lower levels of b. Finally, the third item in the proposition reflects the fact that encroachment is used by society as a redistributive device, so that it lacks any interest when individuals are not risk averse or when productivity differences are small.

3. The institution design problem

3.1. The general framework

The social optimum analyzed above is *Utopian* in that its implementation raises two major issues. First the social optimum is described by the individual efforts associated with individual productivities; there are reasons to believe, as assumed further below, that the θ's are private information so that no implementation procedure should rely on their direct observation. Second, any implementation procedure requires influencing individual behaviors. In general this is likely to be costly.

It is plausible and widely believed that institutions such as police and the justice system, by defining and enforcing property rights, constitute attempts by societies to

[2] Since both fractions multiplying $\underline{\theta}$ in (9) are bigger than unity, b^U strictly exceeds $\underline{\theta}$, so that the possibility that $\underline{e} = 1$ is not ruled out by assumption A1.

improve on the *tragedy of the commons*. But, if so, how should these institutions be designed? How should they differ according to the characteristics of the societies which they regulate? What stylized facts should they incorporate and what role assignments should they induce? These are some of the questions that will be addressed now.

We will not question why most societies seem to adopt some property right assignment and some enforcement instruments as their first institutions, while such instruments as redistributive taxes and subsidies appear only in the modern area. Instead, we will take this selection of instruments as given and focus on the enforcement issue rather than on the establishment of property rights or the selection of other instruments.

Thus suppose that well defined property rights exist; to the extent that they are enforced, these rights entitle producers to keep their production. Society is endowed with enforcement institutions which we will take to consist of a police organization, and a justice system, whose combined activity is to catch and punish presumed offenders. These institutions operate under various technological and information constraints. We assume that information on individual activities, but not on individual productivities, may be acquired at a cost; however perfect information may be very costly, so that the institutions typically make errors. Thus, with a degree of accuracy related to the resources devoted to the police and justice system, institutions come up with suspects and with evaluations of their crimes or booties.

A condemnation carries a sentence which is normally, although not necessarily, related to the offense attributed to the suspect. As long as the system does not operate completely randomly, it is possible to devise the expected sentence in such a way that it reflects the expected value of the offense attributed to an individual, which is itself correlated with his time allocation decision. An individual who devoted $1-e$ units of his time to encroachment faces an expected sentence of $\phi(e)$. This expected sentence function is chosen by society together with the resources devoted to the system. It should be noted that our assumption that individual characteristics cannot be learned by the justice system is consistent with the observation that existing systems usually do not base sentences on individual characteristics, and that, even when individual characteristics are not entirely unobservable, societies are usually reluctant to allow their institutions to take them into account.

Suppose that $F(\tilde{e})$ is the sentence that applies to an individual whose effort has been determined (possibly erroneously) to be $\tilde{e}$ by the system. Suppose also, as will turn out to be endogenously the case, that only two effort levels may be observed in equilibrium: $\bar{e}$ and $\underline{e}$. What is the expected sentence faced by an individual who extends an effort of e, $e \in \{\bar{e}, \underline{e}\}$? If a proportion λ of the total population is subjected to investigation, so that a proportion $1-\lambda$ does not receive any sentence, and if the system works completely randomly, the expected sentence is independent of the individual's effort and simply reflects the distribution of types in society: $\phi(\underline{e}) = \phi(\bar{e}) = \lambda\{\bar{n}F(\bar{e}) + \underline{n}F(\underline{e})\}$. If the system is more focused someone who chose $\bar{e}$ will be more likely to receive a sentence of $F(\bar{e})$ and vice-versa. Thus if we measure the accuracy of the system by ρ, $0 \leq \rho \leq 1$, expected sentences are respectively

$$\phi(\underline{e}) \;=\; \lambda\{\rho\bar{n}F(\bar{e}) + [1-\rho\bar{n}]\,F(\underline{e})\} = \lambda\underline{E} \qquad (10)$$

$$\phi(\bar{e}) \quad = \quad \lambda\{[1-\rho\underline{n}]F(\bar{e})+\rho\underline{n}F(\underline{e})\} = \lambda\bar{E} \tag{11}$$

where $\rho = 0$ corresponds to a perfectly discriminating system, $\rho = 1$ corresponds to a completely random system, while $\underline{E}$ and $\bar{E}$ are the expected sentences faced by each type conditional on being under investigation. For any positive λ, any expected sentence may be arrived at by choosing $\underline{E}$ and $\bar{E}$ appropriately. However very low levels of λ would require very high conditional sentences to be applied with a very low probability; for any preferences involving risk aversion, expected utility is thus an increasing function of λ whenever ϕ is positive.

In general, to the extent that sentences are non negative, (10) and (11) induce a relationship between $\phi(\underline{e})$ and $\phi(\bar{e})$:

$$\phi(\bar{e}) = \phi(\underline{e}) + \lambda[1-\rho][F(\bar{e}) - F(\underline{e})] \tag{12}$$

Two special cases will be of interest later on: when $F(\bar{e}) = F(\underline{e}) = 0$, in which case $\phi(\underline{e}) = \phi(\bar{e}) = 0$; and when $F(\bar{e}) = 0 < F(\underline{e})$ in which case (12) reduces to

$$\phi(\bar{e}) = g\phi(\underline{e}) \tag{13}$$

where

$$0 \leq g = \frac{\rho - \rho\bar{n}}{1-\rho\bar{n}} \leq 1 \tag{14}$$

and $g = 1$ when $\rho = 1$ (random justice) while $g = 0$ when $\rho = 0$ (perfect system).

A sentence may amount to a pure deadweight loss, as in the case of a prison sentence, or it may involve a certain amount of redistribution or compensation to society as when fines or hard labor are involved. We assume that a proportion r of any sentence is returned to society rather than being wasted; r is assumed to be the same for all sentences. However direct compensation of victims is ruled out: fines for speeding may be used to help finance police, but not to compensate individual motorists for their increased exposure. We also assume that sentences are non negative, not only because this is a prevalent feature of all existing systems, but also because the notion of returning part of a sentence would become absurd otherwise.

The cost of maintaining and operating the institutions is $C(\lambda, \rho)$, a decreasing, convex function of ρ and an increasing, concave, function of λ such that $C(\lambda, \rho) \geq 0$. In order to finance the system, aggregate taxes T must be set at such a level that costs are covered by restitutions and taxes

$$C(\lambda, \rho) = T + r[\bar{n}\phi(\bar{e}) + \underline{n}\phi(\underline{e})] \tag{15}$$

We assume that the tax burden is divided equally across individuals so that, given the normalization $n = 1$, each individual pays $t = T$.

3.2. Sentences as a mechanism

Let us assume that *ex post* individual expected utility may be written as

$$\omega(e,\theta,\alpha) - t \equiv V(e,\theta,\alpha) - \frac{1}{\lambda}\phi(e) - t \tag{16}$$

where the coefficient $\frac{1}{\lambda}$ is introduced as a convenient way to account for the monotonic relationship between expected utility and the probability of investigation: because of risk aversion, the reduction in expected utility associated with a given level of expected sentence is more pronounced when individual sentences are high and spread over few individuals (λ close to zero) than when individual sentences are lower and the burden is spread more evenly (λ close to unity). Since individual expected utilities of at least zero may be reached even in a society plagued by the *tragedy of the commons*, ω is bounded away from zero. That constraint effectively rules out $\lambda = 0$ unless $\phi = 0$: while it is feasible for society to have positive expected sentences while letting $\lambda \to 0$, this would be worse than having no institutions as in the *tragedy of the commons*.

By choosing sentences $F(e)$, society may choose the expected sentence function ϕ. This function governs individual decisions concerning the allocation of time between production and encroachment, which in turn determines aggregate production and predation. The cost of institutions is thus determined and taxes must be levied accordingly. We take the institutional structure just described as given and seek to optimize its components. At this stage, we assume that the detection level $\lambda \in (0,1]$, and the accuracy of the justice system $\rho \in [0,1]$, are given. Later on we will investigate the choice of λ and ρ. All other societal and technological characteristic are treated as parametric.

Let us study the optimal choice of ϕ. Choosing ϕ has the same effect as specifying a menu of effort expected-sentence pairs $(e(\theta), \phi(\theta))$, from which each individual would pick his preferred pair according to his type θ. As a result, the institutions may be viewed as an incentive mechanism, and the revelation principle applies: for any possible institutional mechanism, there exists an equivalent mechanism which induces individuals to reveal their actual productivity by their choice of a pair in the menu.

As with the *Utopia* problem, the social objective is to achieve the Harsanyi solution. This requires choosing $\underline{e}$, $\bar{e}$, $\underline{\phi} \equiv \phi(\underline{e})$, and $\bar{\phi} \equiv \phi(\bar{e})$ in such a way as to maximize *ex ante* individual expected utility $\bar{n}\bar{\omega} + \underline{n\omega} - t$ where

$$\underline{\omega} \equiv \underline{V}(\alpha) - \frac{1}{\lambda}\underline{\phi}, \text{ with } \underline{V}(\alpha) \equiv V(\underline{e}, \underline{\theta}, \alpha) \tag{17}$$

and

$$\bar{\omega} \equiv \bar{V}(\alpha) - \frac{1}{\lambda}\bar{\phi}, \text{ with } \bar{V}(\alpha) \equiv V\left(\bar{e}, \bar{\theta}, \alpha\right) \tag{18}$$

There is no participation constraint: any individual will be part of society once his type is drawn. However, since individual types are not observed by the system, if sentences are used, they must be chosen in such a way that each individual willingly chooses the level of effort that the sentence is meant to induce. In particular, this requires $\underline{V}(\alpha) - \frac{1}{\lambda}\underline{\phi} \geq V(\bar{e}, \underline{\theta}, \alpha) - \frac{1}{\lambda}\bar{\phi}$ and $\bar{V}(\alpha) - \frac{1}{\lambda}\bar{\phi} \geq V\left(\underline{e}, \bar{\theta}, \alpha\right) - \frac{1}{\lambda}\underline{\phi}$. Expected sentences must also satisfy (10) and (11) which requires (12) to hold. Finally, to the extent that sentences are non negative, both $\underline{\phi}$ and $\bar{\phi}$ are non negative so that

$$\underline{\omega} \leq \underline{V}(\alpha) \text{ and } \bar{\omega} \leq \bar{V}(\alpha) \tag{19}$$

Individuals may choose effort levels other than the levels $\underline{e}$ and $\bar{e}$ that the system wants to induce. However, all levels other than $\underline{e}$ and $\bar{e}$ are easy to discourage, by letting them imply a prohibitively high expected sentence. A stylized property of justice systems, one might argue, is that the expected sentence is a non decreasing function of the gravity of the offense, i.e. $\phi(e)$ is non increasing. Imposing such a restriction might make it impossible to choose high enough sentences to discourage all effort levels other than $\underline{e}$ and $\bar{e}$. In fact this turns out not to be an issue: if $\underline{e} < \bar{e}$, it is easy to show that an expected sentence $\underline{\phi}$ that convinces type $\bar{\theta}$ (given $\bar{\phi}$) not to adopt an effort level of $\underline{e}$ also eliminates any effort in the interval $[\underline{e}, \bar{e})$ if applied over that interval; similarly if an expected sentence of $\bar{\phi}$ is imposed on the interval $[\bar{e}, 1]$ it can be shown that it will cause $\bar{e}$ to be preferred to any other effort in that interval. Thus if we anticipate on the results $\underline{e} \leq \bar{e}$ and $\bar{\phi} \leq \underline{\phi}$, we may think of the expected sentence function as a decreasing step function

$$\phi(e) \begin{cases} = \phi^1 \ , \ e < \underline{e} \\ = \underline{\phi} \ , \ \underline{e} \leq e < \bar{e} \\ = \bar{\phi} \ , \ e \geq \bar{e} \end{cases}$$

where ϕ^1 is sufficiently higher than $\underline{\phi}$ to discourage efforts below $\underline{e}$. If $\underline{e}$ and $\bar{e}$ are set equal to each other the choice of $\phi(e)$ is trivial: choose $\underline{\phi} = \bar{\phi} = 0$. If $\underline{e} < \bar{e}$, the real difficulty for society, besides the choice of $\underline{e}$ and $\bar{e}$, is to avoid adopting too high expected sentences at $\underline{e}$ and $\bar{e}$ because these sentences will actually be imposed and should not be too wasteful. We now turn to that problem.

3.3. Sentence and effort selection: setup

Given the set of constraints just described some simple substitutions and manipulations outlined in the Appendix lead to the following formulation of the institutional choice problem faced by society:

$$\max_{\underline{e}, \bar{e}, \bar{\omega}, \underline{\omega}} \left\{ [1 - \lambda r] \left[\underline{n}\, \underline{\omega} + \bar{n} \bar{\omega} \right] + \lambda r \left[\underline{n}\, \underline{V}(\alpha) + \bar{n} \bar{V}(\alpha) \right] - C(\lambda, \rho) \right\} \tag{20}$$

subject to (5), (1), (12), (19), and

$$\bar{\omega} \leq \underline{\omega} + \left[\bar{V}(\alpha) - V(\bar{e}, \underline{\theta}, \alpha) \right] \tag{21}$$

$$\bar{\omega} \geq \underline{\omega} + \left[V\left(\underline{e}, \bar{\theta}, \alpha\right) - \underline{V}(\alpha) \right] \tag{22}$$

(21) and (22) are modified forms of the self selection constraints: in order type $\underline{\theta}$ not to choose the effort meant for type $\bar{\theta}$, it must be true that $\underline{V} - \frac{1}{\lambda}\underline{\phi} \geq V(\bar{e}, \underline{\theta}, \alpha) - \frac{1}{\lambda}\bar{\phi}$, or $\underline{V} - \frac{1}{\lambda}\underline{\phi} - V(\bar{e}, \underline{\theta}, \alpha) + \bar{V} \geq \bar{V} - \frac{1}{\lambda}\bar{\phi}$; this gives (21); (22) is obtained similarly.

The boundaries of the half-spaces defined by (21) and (22) define linear relationships between $\underline{\omega}$ and $\bar{\omega}$. In order the set of feasible pairs $(\underline{\omega}, \bar{\omega})$ not to be empty it is necessary that the intercept of the first line be above the intercept of the second line: $\bar{V}(\alpha) -$

$V(\bar{e}, \underline{\theta}, \alpha) \geq V\left(\underline{e}, \bar{\theta}, \alpha\right) - \underline{V}(\alpha)$, from which it follows that

$$\bar{e} \geq \underline{e} \tag{23}$$

The choice set is compact: $0 \leq e \leq 1$, for $e = \underline{e}, \ \bar{e}$; $0 \leq \omega \leq \omega^{\max}$, for $\omega = \underline{\omega}, \ \bar{\omega}$, where the limited amount of resources in the economy guarantees the existence of the upper bound $\omega^{\max}$. The objective function is continuous. Consequently a solution exists.

Finding the solution is facilitated by examining situations where (21) and (22) are not binding or not imposed. Then it is optimal to choose $\bar{\omega}$ and $\underline{\omega}$ as high as possible by setting[3] $\bar{\phi} = \underline{\phi} = 0$ so that the problem reduces to

$$\max_{\underline{e}, \bar{e}} W^N = W\left(\underline{e}, \bar{e}\right) - C(\lambda, \rho) \text{ subject to } (5),\ (1) \tag{24}$$

In that case the *Institution* problem differs from the *Utopia* problem only by the presence of $C(\lambda, \rho)$ in the objective function. Thus when (21) and (22) are not binding the solution $\left(\underline{e}^N, \bar{e}^N\right)$ to problem (20) is the same as the solution to the *Utopia* problem. In particular $\left(\underline{e}^N, \bar{e}^N\right)$ is independent of λ and ρ and the value of b above which $\underline{e}$ becomes smaller than unity is the same in both problems. These preliminary results are gathered in the following proposition.

Proposition 4 (*selection constraints not binding)*

Let $\left(\underline{e}^N, \bar{e}^N\right)$ *be the solution to the* Institution *problem when* W *satisfy Assumption 2 and the selection constraints are* **not** *binding or* **not** *imposed; this solution has the following properties:*

1. $\bar{e}^N = 1$; *and* $\underline{e}^N$ *is defined by the condition: either* $W_{\underline{e}}\left(\underline{e}^N, 1\right) = 0$ *and* $\underline{e}^N < 1$; *or* $\underline{e}^N = 1$ *and* $W_{\underline{e}}(1, 1) \geq 0$;
2. $\underline{\phi} = \bar{\phi} = 0$;
3. $\underline{e}^N < (=)\ 1 \ \Leftrightarrow \ b > (\leq)\ b^U$, *where* b^U *is defined by* (9)
4. $\underline{e}^N = \underline{e}^U$; $\bar{e}^N = \bar{e}^U$;
5. $\phi(e) = \begin{cases} \phi^1 \ , \ e < e^B \\ \underline{\phi} = \bar{\phi} = 0 \ , \ \underline{e}^N \leq e \end{cases}$.

When (21) or (22) are binding the solution remains to be characterized. To do so, and to determine when the self-selection constraints are binding, we will proceed in two steps: first, the choice of $\bar{\omega}$ and $\underline{\omega}$ given $\bar{e}$ and $\underline{e}$; second, the choice of $\bar{e}$ and $\underline{e}$ given the solution to the first step.

[3] The only way to do so without imposing negative sentences is to set $F(\underline{e}) = F(\bar{e}) = 0$, which satisfies (12).

3.4. When and why are self-selection constraints binding?

When $\bar{e}$ and $\underline{e}$ are given, the last two terms in (20) may be treated as constants so that the problem becomes equivalent to the linear programming problem $\max_{\bar{\omega},\underline{\omega}} \underline{n}\,\underline{\omega}+\bar{n}\bar{\omega}$ subject to (21), (22), and (12). This is illustrated in Figure 1 where line $\bar{D}$, corresponding to (21), lies above line $\underline{D}$, corresponding to (22) (otherwise both constraints could not be satisfied simultaneously), and both lines have non negative intercepts (because V is increasing in θ).

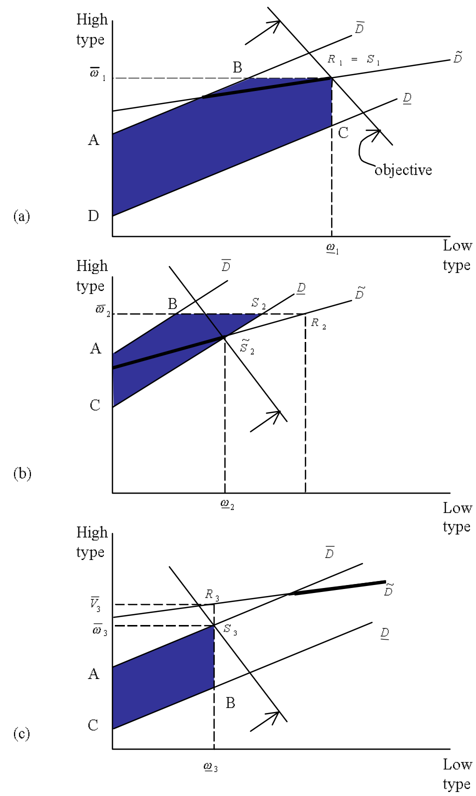

FIGURE 1: Expected post-sentence utilities for three alternative pre-sentence utility level. (a) No binding selection constraint; (b) Binding constraint on high type; (c) Binding constraint on low type.

Let R_1, R_2, and R_3 represent three possible locations for the point $\left(\underline{V}(\alpha) \geq 0, \bar{V}(\alpha) \geq 0\right)$. The corresponding solutions $(\underline{\omega}, \bar{\omega})$ to problem (20) must lie to the south-west of these points because $\phi \geq 0$, and between lines $\underline{D}$ and $\bar{D}$. This delineates the trapezoidal feasible sets in Figure 1: ABR_1CD in Figure 1a; ABS_2C in Figure 1b; AS_3BC in Figure 1c.

Furthermore, constraint (12) must also be satisfied. We show in the Appendix (Lemma 2) that, for any given level of $\underline{\phi}$, it is optimal to choose $F(\bar{e}) = 0$, which implies that (13) may be substituted for (12). Substituting $\frac{1}{\lambda}\underline{\phi} = \underline{V} - \underline{\omega}$ and $\frac{1}{\lambda}\bar{\phi} = \bar{V} - \bar{\omega}$ into (13) gives

$$\bar{\omega} = \left[\bar{V} - g\underline{V}\right] + g\underline{\omega} \tag{25}$$

This is represented by $\tilde{D}$ in each variant of Figure 1: a line of slope g through points R_1, R_2, or R_3.

Let points S_1, S_2, and S_3 represent the solution $(\underline{\omega}, \bar{\omega})$ to problem (20) in each variant of Figure 1. Each of these points must lie within the grey feasible set and on the $\tilde{D}$ line. Since the objective is increasing in $\underline{\omega}$ and $\bar{\omega}$, as indicated by the arrow, each solution is the most northeasterly point that meets these requirements.

Thus in Figure 1a, when no self-selection constraint is binding, S_1 is the same point as R_1 which means that $\underline{V}(\alpha) = \underline{\omega}$ and $\bar{V}(\alpha) = \bar{\omega}$, no sentences being paid. In Figure 1b (22) is binding so that the solution represented by $\tilde{S}_2$ is such that $\underline{V}(\alpha) > \underline{\omega}$ and $\bar{V}(\alpha) \geq \bar{\omega}$, i.e. $\underline{\phi} > 0$ and $\bar{\phi} \geq 0$. The inequalities defining the treatment of the more productive type are strict if the slope of $\tilde{D}$ is strictly positive ($g > 0$ means inaccurate justice); if $g = \rho = 0$ (perfectly accurate justice), $\bar{V}(\alpha) = \bar{\omega}$, i.e. $\bar{\phi} = 0$. Finally, Figure 1c, with its pre-sentence utility point R_3 above the $\bar{D}$ line, may be seen to be impossible since there is no point on $\tilde{D}$ which is also in the AS$_3$BC set.

Thus if the pair $(\underline{e}, \bar{e})$ is in the admissible set, only points such as R_1 or R_2 may arise. At a point such as R_1, above $\underline{D}$, it must be the case that $\underline{V}(\alpha) = \underline{\omega}$ and $\bar{V}(\alpha) = \bar{\omega} \geq V\left(\underline{e}, \bar{\theta}, \alpha\right) - \underline{V}(\alpha) + \underline{\omega}$. Substituting explicit forms for V and $\underline{V}(\alpha)$, this implies, after some algebra, $[1-\alpha]\,\bar{\theta} \geq b$ or $\frac{\partial V\left(\bar{e}, \bar{\theta}, \alpha\right)}{\partial \bar{e}} \geq 0$.As noted earlier the self-selection constraint is not binding here because the private objective of types $\bar{\theta}$ is to produce as much as possible, which does not conflict with society's objective. Crimes of such low yield may be called *petty crimes*. They are tolerated by society as a mean to achieve transfers without incurring selection costs. They have this property because they do not appeal to more productive types and, as such, do not threaten the existence of society.

The converse corresponds to points such as R_2 with $\underline{\phi} > 0$:

$$[1-\alpha]\,\bar{\theta} < b \text{ or } \frac{\partial V\left(\bar{e}, \bar{\theta}, \alpha\right)}{\partial \bar{e}} < 0 \tag{26}$$

In the Appendix, we show that such constrained cases arises only if $b \geq b^N$ where

$$b^N \equiv \max\left\{b_1, b^U\right\} \tag{27}$$

and $b_1 \leq \bar{\theta}$ is defined in the Appendix. Indeed when b is high enough, even the more productive individuals find encroachment activities privately attractive; in that case the justice system must take account of the self-selection constraint if it wants to induce different efforts on the part of each type. However, as we underline next, pooling is a feasible alternative.

Whether $\left(\underline{V}(\alpha), \bar{V}(\alpha)\right)$ is a point such as R_1 or R_2 is affected by the initial choice of $(\underline{e}, \bar{e})$. To make this dependency explicit, consider again the property of R_1 to be above line $\underline{D}$: $\bar{V}(\alpha) = \bar{\omega} \geq V\left(\underline{e}, \bar{\theta}, \alpha\right) - \underline{V}(\alpha) + \underline{\omega}$ with $\underline{\omega} = \underline{V}(\alpha)$. This is equivalent to $[1-\alpha]\,\bar{\theta}\bar{e} + [1-\bar{e}]\,b \geq [1-\alpha]\,\bar{\theta}\underline{e} + [1-\underline{e}]\,b$ or $[1-\alpha]\,\bar{\theta}\,[\bar{e}-\underline{e}] \geq b\,[\bar{e}-\underline{e}]$, which implies either, $\bar{e} = \underline{e}$, or, (if $\bar{e} \neq \underline{e}$), $[1-\alpha]\,\bar{\theta} \geq b$. Substituting (1), this implies that R_1 (neither (21) nor (22) is binding) arises if and only if either

$$\underline{e} = \bar{e} \tag{28}$$

or

$$\underline{s}\,\underline{e} + \bar{s}\bar{e} \geq 1 \tag{29}$$

where $\bar{s} = \bar{n}\frac{\bar{\theta}}{b}$ and $\underline{s} = \underline{n}\left[\frac{\bar{\theta}-\underline{\theta}}{\bar{\theta}} + \frac{\underline{\theta}}{b}\right]$. Conversely, R_2 ((22) is binding) arises if and only if

$$\underline{s}\,\underline{e} + \bar{s}\bar{e} < 1 \text{ and } \bar{e} > \underline{e} \tag{30}$$

Thus the self-selection issue may be avoided by choosing to induce the same effort on the part of both types. We call this *blind justice* and note that, unlike the other alternative, it does not require any sentence to be born by either type. In contrast a *discriminating justice* system will induce distinct behaviors ($\bar{e} > \underline{e}$) on the part of each type and require positive expected sentences. These results are gathered in the following proposition.

Proposition 5 *(constrained justice)*
Let $\rho \in [0,1]$, $\lambda \in (0,1]$, and W satisfy Assumption 2.

1. *When $b \leq b^N$, the solution to the* Institution *problem is not affected by the self-selection constraint and is described by Proposition 4.*

2. *When $b > b^N$, the solution to the* Institution *problem differs from the unconstrained optimum in such a way that either:*

   - (blind justice) $\underline{e} = \bar{e}$ *and* $(\underline{\phi}, \bar{\phi}) = (0,0)$*; or*
   - (discriminating justice) $\underline{e} < \bar{e}$, $\underline{\phi} > 0$ *and* $\bar{\phi} \geq 0$*, with* $\bar{\phi} = 0$ *if* $\rho = 0$ *and* $\bar{\phi} > 0$ *otherwise.*

The choice between a *blind justice* and a *discriminating* one will result from a comparison between the optimized social welfare functions achieved in each case. In order to do so it is necessary to characterize the solution in each case. Let us start with the simpler option, *blind justice.*

3.5. Blind justice: $b > b^N$ and $\underline{e} = \bar{e}$

When $b > b^N$, society cannot choose $\left(\underline{e}^N, \bar{e}^N\right)$ because that choice would not meet the self-selection constraint. Rather than choosing an alternative pair $\underline{e} < \bar{e}$ that would meet the constraint, society may simply choose the pooling alternative, thus adopting a *blind justice* system involving $\underline{e} = \bar{e}$.

When $\bar{e} = \underline{e}$, we may substitute $\bar{\omega} = \bar{V}$ and $\underline{\omega} = \underline{V}$ into problem (20) to get the following problem

$$\max_{e} \underline{n}\,\underline{V}\left(\alpha\left(\underline{e}, \bar{e}\right)\right) + \bar{n}\bar{V}\left(\alpha\left(\underline{e}, \bar{e}\right)\right) - C\left(\lambda, \rho\right) \tag{31}$$
$$\text{subject to (1) and } \underline{e} = \bar{e} = e$$

Since a corner solution at $e = 0$ is easily dismissed, the first-order condition is

$$-\left[\underline{n}\,\underline{V}'\underline{\theta} + \bar{n}\bar{V}'\bar{\theta}\right]\left(\frac{\partial\alpha(e,e)}{\partial\underline{e}} + \frac{\partial\alpha(e,e)}{\partial\bar{e}}\right)e + \left\{\underline{n}\,\underline{V}'\left[[1-\alpha]\,\underline{\theta} - b\right] + \bar{n}\bar{V}'\left[[1-\alpha]\,\bar{\theta} - b\right]\right\} \geq 0 \tag{32}$$

where $\underline{V}' = V'(\underline{e}, \underline{\theta}, \alpha)$and $\bar{V}' = V'\left(\bar{e}, \bar{\theta}, \alpha\right)$. Since both $\frac{\partial\alpha(e,e)}{\partial\underline{e}}$ and $\frac{\partial\alpha(e,e)}{\partial\bar{e}}$ are negative and $\left[\underline{n}\,\underline{V}' + \bar{n}\bar{V}'\right]$ is positive, the first term in (32), representing the collective gain from reduced encroachment, is positive; since $[1-\alpha]\,\theta - b < 0$ for $\theta = \bar{\theta}, \underline{\theta}$ (see the discussion of (26)), the second term, representing the sum of private individual utility losses from increased effort, is negative. If the first term dominates at $e = 1$, then $e = 1$ is the solution; otherwise the solution is $e < 1$.

Let b^B be defined by

$$\left[\underline{n}\,\underline{V}'\underline{\theta} + \bar{n}\bar{V}'\bar{\theta}\right]\frac{1}{\left[\underline{n}\,\underline{\theta} + \bar{n}\bar{\theta}\right]} + \underline{n}\,\underline{V}'\frac{\left[\underline{\theta} - b^B\right]}{b^B} + \bar{n}\bar{V}'\frac{\left[\bar{\theta} - b^B\right]}{b^B} = 0 \text{ for } e = 1 \tag{33}$$

It is shown in the Appendix (Proposition 6) that an interior solution arises for $b > b^B$ while a corner solution $e = 1$ is optimal for lower values of b.

Thus *blind justice* requests all types to devote their whole time to production when the private reward from encroachment is relatively low, but allows all types to indulge in some predation when b is high. This behavior may be induced by imposing no sentences when effort is at, or above, the prescribed level, while imposing a prohibitive sentence when effort is below the prescribed level. Since no sentences would be imposed in equilibrium, the solution is not affected by detection levels or justice accuracy.

Encroachment tolerance in a *blind justice* system is surprising since such a system is the alternative available to society when it is too costly to discriminate between types as a way to promote equity. In fact one is tempted to identify a *discriminating justice* system with a system that arbitrates between redistributive and productive considerations, while *blind justice* would be adopted as a system that focuses on production maximization. Such an interpretation would be misleading. Indeed forcing both types to produce a maximum effort level exacerbates income differences as low productivity types must rely on their own, low, production without using encroachment to improve their income. Our analysis shows that, when encroachment is attractive enough, considering productivity endowments and population distribution, it is socially desirable to reduce the output of both types in order to allow some redistribution through encroachment, thus reducing income discrepancies. Our findings are summarized in the following proposition.

Proposition 6 *(blind justice)*

Let W satisfy Assumption 2 and let e^B solve problem (20) *when $b > b^N$ and $\underline{e} = \bar{e}$; the solution has the following properties:*

1. *for $b^N < b \leq b^B$, the solution is $e^B = 1$;*

2. *for $b > b^B$, the solution is such that $0 < e^B < 1$;*

3. *b^B exists and is defined by condition* (33).

4. $\phi(e) = \begin{cases} \phi^1, & e < e^B \\ 0, & e^B \leq e \end{cases}$.

3.6. Discriminating justice: the case of perfect accuracy

Let us now study the properties of a discriminating justice system, i.e. when $b \geq b^N$ and $\underline{e} < \bar{e}$. It will be established that the effort requested from the less productive types has to be distorted, which is a familiar outcome of agency models; however we will also show that the effort of the more productive types may have to be distorted as well, contrary to what happens in standard agency situations. We start with the case of a perfectly accurate justice system: $\rho = 0$. By (14) it follows that $g = 0$ so that (13) implies that $\bar{\phi} = 0$. This corresponds to Figure 1b, when $\tilde{D}$ is a horizontal line so that points $\tilde{S}_2$ and S_2 are identical.

We may substitute $\bar{\omega} = \bar{V}$ and $\underline{\omega} = \bar{V}(\alpha) + \underline{V}(\alpha) - V\left(\underline{e}, \bar{\theta}, \alpha\right)$ into problem (20) to get

$$\max_{\underline{e},\bar{e}} \left\{ \underline{n}\,\underline{V}\left(\alpha\left(\underline{e},\bar{e}\right)\right) + \bar{n}\bar{V}\left(\alpha\left(\underline{e},\bar{e}\right)\right) - C(\lambda, 0) - [1-\lambda r]\,\underline{n}\left[V\left(\underline{e},\bar{\theta},\alpha\left(\underline{e},\bar{e}\right)\right) - \bar{V}\left(\alpha\left(\underline{e},\bar{e}\right)\right)\right] \right. \\ \left. \text{subject to } (5),\ (1),\ \text{and } (19) \right\} \tag{34}$$

Thus the self selection constraint changes the objective from that of maximizing $W^N\left(\underline{e},\bar{e}\right) = \underline{n}\,\underline{V}\left(\alpha\left(\underline{e},\bar{e}\right)\right) + \bar{n}\bar{V}\left(\alpha\left(\underline{e},\bar{e}\right)\right) - C(\lambda, 0)$, the objective under Utopia, net of the per capita cost of institutions, to that of maximizing $W^N\left(\underline{e},\bar{e}\right)$ minus the cost in terms of expected utility foregone of the sentences necessary to induce self-selection. When $\lambda r = 1$, the latter vanishes because any cost incurred by an individual when he receives a sentence is offset at the aggregate level by sentence restitution.

More precisely, the expected individual sentence differential that must be imposed to achieve self-selection is

$$\Phi\left(\underline{e},\bar{e}\right) \equiv \lambda\left[V\left(\underline{e},\bar{\theta},\alpha\left(\underline{e},\bar{e}\right)\right) - \bar{V}\left(\alpha\left(\underline{e},\bar{e}\right)\right)\right] \tag{35}$$

In the present instance, only the $\underline{n}$ low productivity individuals face sentences and the social burden of these sentences is reduced by restitutions to the tone of $\underline{n}r\Phi = \lambda r\underline{n}\left[V\left(\underline{e},\bar{\theta},\alpha\left(\underline{e},\bar{e}\right)\right) - \bar{V}\left(\alpha\left(\underline{e},\bar{e}\right)\right)\right]$.[4]

The extent to which the solution differs from $\left(\underline{e}^N, \bar{e}^N\right)$ depends on the magnitude of $[1-\lambda r]\,\underline{n}\frac{\Phi}{\lambda}$ and on the sensitivity of Φ to changes in $\underline{e}$ and $\bar{e}$. In fact, using the definition of V

$$\begin{aligned} \frac{\partial\Phi\left(\underline{e},\bar{e}\right)}{\partial\underline{e}} &= \lambda u'\left(\left[\left[1-\alpha\left(\underline{e},\bar{e}\right)\right]\bar{\theta} - b\right]\underline{e} + b\right)\left[\left[1-\alpha\left(\underline{e},\bar{e}\right)\right]\bar{\theta} - b\right] \\ &\quad -\frac{\partial\alpha}{\partial\underline{e}}\left[\lambda u'\left(\left[\left[1-\alpha\left(\underline{e},\bar{e}\right)\right]\bar{\theta} - b\right]\underline{e} + b\right)\bar{\theta}\underline{e} - \lambda u'\left(\left[\left[1-\alpha\left(\underline{e},\bar{e}\right)\right]\bar{\theta} - b\right]\bar{e} + b\right)\bar{\theta}\bar{e}\right] \end{aligned} \tag{36}$$

Discriminating justice requires that $\underline{e} < \bar{e}$ and (see (26)) $[1-\alpha]\,\bar{\theta} < b$. Thus the first

[4] In the current special case, $\bar{\phi} = 0$ so that the expected sentence differential Φ is equal to the expected sentence on type $\underline{\theta}$. The optimal expected sentence $\underline{\phi}$ would be obtained by substituting the optimal values of $\underline{e}$ and $\bar{e}$ into Φ.

term on the right-hand side of (36) is negative. As to the term on the second line, since $\frac{\partial \alpha}{\partial \underline{e}} < 0$, its sign depends on the sign of the term between square brackets: since $\underline{e} < \bar{e}$ and $\left[[1-\alpha]\bar{\theta} - b\right] < 0$, $\left[[1-\alpha]\bar{\theta} - b\right]\underline{e} + b > \left[[1-\alpha]\bar{\theta} - b\right]\bar{e} + b$, so that the bracketed term is non positive. It follows that $\frac{\partial \Phi(\underline{e},\bar{e})}{\partial \underline{e}}$ is negative so that the presence of the self selection constraint in the social problem increases the first-order condition with respect to $\underline{e}$ by substracting a negative term from it. Consequently, other things equal, an interior solution $\underline{e}^D(\lambda, 0)$ for $\underline{e}$ would be such that $\underline{e}^D > \underline{e}^N$. This corresponds to the familiar result in agency theory whereby a principal imposes a distortion to the low type in order to save on the rent that has to be given up (here the sentence that has to be imposed) to induce truthful behavior. Inducing a higher effort on the part of less productive individuals reduces the temptation for more productive types to imitate them, thus allowing society to impose lower expected sentences on least productive individuals for their encroachment activities. This effect is reinforced by the aggregate feedback effect which operates through α: inducing a higher effort on the part of less productive individuals also reduces the share of encroachment in the economy, which increases the incentive to work.

However, the analysis must go beyond this simple remark, which applies if the value selected for $\bar{e}$ is the same in the constrained case as in the unconstrained one, namely $\bar{e}^D = \bar{e}^N = 1$. We are now going to show that the optimal value of $\bar{e}$ may be interior $(\bar{e}^D(\lambda, 0) < 1)$ when the selection constraint is binding, a result not usually found in the agency literature. In fact, one can verify that $\frac{\partial \Phi}{\partial \bar{e}}$ may be positive or negative

$$\begin{aligned}\frac{\partial \Phi(\underline{e}, \bar{e})}{\partial \bar{e}} &= -\lambda u'(\bar{y})\left[[1-\alpha(\underline{e},\bar{e})]\bar{\theta} - b\right] \\ &\quad -\lambda \frac{\partial \alpha}{\partial \bar{e}}\left[u'\left(\left[[1-\alpha(\underline{e},\bar{e})]\bar{\theta} - b\right]\underline{e} + b\right)\bar{\theta}\underline{e} - u'(\bar{y})\bar{\theta}\bar{e}\right]\end{aligned} \tag{37}$$

The first term on the right-hand side is positive: increasing the effort requested from productive individuals reduces their utility, thus making it more tempting for them to imitate less productive types, which in turn will be prevented by imposing a higher expected sentence on the latters. This effect is well known in the agency literature.

However the second term, accounting for the general equilibrium effect of increasing $\bar{e}$, is negative: the reduction in α makes it less attractive for more productive types to imitate less productive ones.

Thus $\frac{\partial \Phi(\underline{e},\bar{e})}{\partial \bar{e}}$ may be positive or negative; if $\frac{\partial \Phi(\underline{e},\bar{e})}{\partial \bar{e}}$ is positive, and of sufficient magnitude, it may offset the other terms of the first-order condition for $\bar{e}$ in problem (34); an interior solution may arise, unlike the corner solution $\bar{e}^N = 1$ of the unconstrained case.

To be more precise about the circumstances under which this happens, let us rewrite problem (34) as

$$\max_{\underline{e},\bar{e}} W^N(\underline{e}, \bar{e}) - \left[\frac{1}{\lambda} - r\right]\underline{n}\Phi(\underline{e}, \bar{e}) \tag{38}$$

Since $\underline{e} < \bar{e}$, the sole possible corner solution for $\bar{e}$ is at $\bar{e} = 1$ and occurs if and only if the first-order condition for $\bar{e}$ is non negative when $\bar{e} = 1$ while the first-order condition for $\underline{e}$ is also satisfied:

$$\frac{\partial W^N(\underline{e},1)}{\partial \bar{e}} - \left[\frac{1}{\lambda} - r\right] \underline{n} \frac{\partial \Phi(\underline{e},1)}{\partial \bar{e}} \geq 0 \tag{39}$$

and

$$\frac{\partial W^N(\underline{e},1)}{\partial \underline{e}} - \left[\frac{1}{\lambda} - r\right] \underline{n} \frac{\partial \Phi(\underline{e},1)}{\partial \underline{e}} = 0 \tag{40}$$

Imagine that b is allowed to vary over $[b^N, \infty)$, the interval over which discriminating justice is a candidate solution. By Propositions 4 and 5, when $b = b^N$, $\bar{e}^N = 1$ is a corner solution with $\frac{\partial W^N(\underline{e},1)}{\partial \bar{e}} > 0$. Let $b^D(\lambda, 0)$ be some value of b in $[b^N, \infty)$ at which the solution for $\bar{e}$ becomes interior; if it exists b^D is defined by the condition that (40) holds and (39) is satisfied with equality. We show in the Appendix (Proposition 7) that, for any combination of all parameters other than b such that $\lambda r \neq 1$, there exist some value of b above which (39) is violated while (40) holds. This implies that the value of $\bar{e}$ that solves problem (38) is then lower than unity i.e. the solution is interior. Thus $b^D(\lambda, 0)$ exists and $\bar{e}^D$ is interior when $b > b^D(\lambda, 0)$; it is also easily shown that $b^D(\lambda, 0) > [1-\alpha]\bar{\theta}$.

This result is new to the agency literature where the more productive type is never asked to produce at a level that would be inefficient in a full information setup. As revealed by an examination of (37), this happens because our model allows for feedback effects via the encroachment rate α; unlike standard agency models, the social planner can affect the cost of achieving self selection by manipulating the effort levels of both the more productive and the less productive types, rather than manipulating only the effort of the less productive type. Our results are gathered in the following proposition.

Proposition 7 *(perfectly accurate, discriminating, justice)*

Let $\rho = 0$, $\lambda \in (0,1]$ and W satisfy Assumption 2. Let $\left(\underline{e}^D(\lambda,0), \bar{e}^D(\lambda,0)\right)$ be the solution when $b > b^N$ and $\underline{e} \neq \bar{e}$; this solution has the following properties:

1. *for $b^N < b \leq b^D(\lambda,0)$, the solution is such that $\bar{e}^D(\lambda,0) = 1$ and $\underline{e}^N \leq \underline{e}^D(\lambda,0) < 1$, where the inequality is strict if $\lambda r < 1$;*

2. *for $b > b^D(\lambda,0)$, the solution is such that $0 < \underline{e}^N \leq \underline{e}^D(\lambda,0) < \bar{e}^D(\lambda,0) \leq \bar{e}^N = 1$, where the inequalities are strict if $\lambda r < 1$;*

3. $\phi(e) = \begin{cases} \phi^1 \ , \ e < \underline{e}^D(\lambda,0) \\ \underline{\phi} > 0 \ , \ \underline{e}^D(\lambda,0) \leq e < \bar{e}^D(\lambda,0) \\ \bar{\phi} = 0 \ , \ \bar{e}^D(\lambda,0) \leq e \end{cases}$.

4. *$b^D(\lambda,0)$ is defined by the condition that (40) holds and (39) is satisfied with equality; it is such that $b^D(\lambda,0) \geq \max\left\{b^N, \left[1-\alpha\left(\underline{e}^N, \bar{e}^N\right)\right]\bar{\theta}\right\}$;*

5. *A sufficient condition for the existence of $b^D(\lambda,0)$ is $\lambda r < 1$.*

3.7. Inaccurate discriminating justice: $0 < \rho \leq 1$

The *discriminating justice* described sofar corresponds to a system which is perfect in the sense that an individual is sure, if investigated, to get a sentence which reflects his own effort and is unrelated to others' efforts. When $\rho > 0$, this is no longer true; the expected sentence faced by any individual is affected by the level of the sentence meant for the other type (see (13)). If low productivity types are given positive expected sentences, then the expected sentence devised for high productivity types must also be strictly positive. As when $\rho = 0$, this corresponds to Figure 1b; however line $\tilde{D}$ is now strictly upward sloping rather than horizontal. Consequently, the solution is point $\tilde{S}_2$ rather than point S_2 as when $\rho = 0$. This describes a situation where both types pay strictly positive sentences on average, with $\bar{\phi} < \underline{\phi}$ by (13) since $g < 1$. In fact the solution does not exist: when the justice system is random ($\rho = g = 1$). In that case line $\tilde{D}$ has a slope of unity and is parallel with both $\underline{D}$ and $\bar{D}$ so that it never intersects with the set ABS_2C; by Proposition 5, the sole alternative is *blind justice*.

Let us now characterize the general *discriminating justice* solution when it exists. The reader will remember that such locations for $(\underline{V}, \bar{V})$ as point R_2, below line $\underline{D}$, correspond to relatively high levels of b, when both types would prefer reducing their effort, i.e. when $[1-\alpha]\,\theta - b < 0$ for both types (see (26)).

Solving (25) and (22) for $\bar{\omega}$ and $\underline{\omega}$ gives

$$\begin{aligned}
\underline{\omega} &= \underline{V} - \frac{1}{1-g}\left[V\left(\underline{e},\bar{\theta},\alpha\right) - \bar{V}\right] \\
\bar{\omega} &= \frac{\bar{V}}{1-g} - \frac{g}{1-g}V\left(\underline{e},\bar{\theta},\alpha\right)
\end{aligned}$$

These constraints are easy to interpret. As when the justice system does not make any mistake, the difference between $\bar{\omega}$ and $\underline{\omega}$ is equal to $V\left(\underline{e},\bar{\theta},\alpha\right) - \bar{V}$, which ensures that, considering the expected sentences, types $\bar{\theta}$ are not tempted to adopt the effort meant for types $\underline{\theta}$. However, since $V\left(\underline{e},\bar{\theta},\alpha\right) - \bar{V} > 0$, the second constraint implies that $\bar{\omega}$ must now be set lower than[5], rather than equal to, $\bar{V}$. This reflects the fact that it is now impossible to impose a positive expected sentence on types $\underline{\theta}$ without sentencing types $\bar{\theta}$ by mistake occasionally; furthermore, $\bar{\omega}$ diminishes when $V\left(\underline{e},\bar{\theta},\alpha\right)$ rises: the higher the temptation for types $\bar{\theta}$ to emulate types $\underline{\theta}$, the higher the expected sentence imposed on low type to achieve self-selection, hence the more severely high types are sentenced by mistake.

Substituting these expressions into (20) and rearranging, the objective function may be written as

$$\begin{aligned}
J\left(\underline{e},\bar{e};\lambda,\rho\right) &= \underline{n}\,\underline{V} + \bar{n}\bar{V} - [1-\lambda r]\left[\underline{n} + \frac{g}{1-g}\right]\left[V\left(\underline{e},\bar{\theta},\alpha\right) - \bar{V}\right] - C\left(\lambda,\rho\right) \qquad (41) \\
&= J\left(\underline{e},\bar{e};\lambda,0\right) + \left[C\left(\lambda,0\right) - C\left(\lambda,\rho\right)\right] - [1-\lambda r]\frac{g}{1-g}\left[V\left(\underline{e},\bar{\theta},\alpha\right) - \bar{V}\right]
\end{aligned}$$

[5] To see this, note that $\bar{\omega}$ is linearly decreasing in $V\left(\underline{e},\bar{\theta},\alpha\right)$ and equal to $\bar{V}$ when $V\left(\underline{e},\bar{\theta},\alpha\right) = \bar{V}$.

where $J(\underline{e}, \bar{e}; \lambda, 0)$ is the objective function in (34), the *Institution* problem when $\rho = 0$ and the selection constraint is binding.[6] This objective is to be maximized subject to $0 \leq \underline{e} < \bar{e} \leq 1$.

Thus when ρ is strictly positive rather than null, the *discriminating justice* problem (34) is modified in two respects. First the objective function is increased by the cost saving $C(\lambda, 0) - C(\lambda, \rho)$ associated with a lower effectiveness of the system. Second the objective function is reduced by $[1 - \lambda r] \frac{g}{1-g} \left[V\left(\underline{e}, \bar{\theta}, \alpha\right) - \bar{V}\right]$. This is the increase in the cost of achieving self selection (by way of a sentence differential) due to the drop in justice accuracy. While the cost of achieving self-selection bears only on the less productive group when $\rho = 0$ (see (34)), this additional cost depends on type distribution in a more complex fashion, via g, because imperfect accuracy implies that more productive types now also experience positive expected sentences because of justice errors.

The additional cost of achieving self-selection may be reduced by further distorting $\underline{e}$ and $\bar{e}$ in such a way as to reduce $\Phi(\underline{e}, \bar{e})$; thus the optimal pair $\left(\underline{e}^D(\lambda, \rho), \bar{e}^D(\lambda, \rho)\right)$ differs from $\left(\underline{e}^D(\lambda, 0), \bar{e}^D(\lambda, 0)\right)$. However, it is straightforward to adapt the proof of Proposition 7 to show that the same patterns arise as when $\rho = 0$: for high values of b, both $\underline{e}^D(\lambda, \rho)$ and $\bar{e}^D(\lambda, \rho)$ are interior; for lower values of b, $\underline{e}^D(\lambda, \rho)$ is interior but $\bar{e}^D(\lambda, \rho) = 1$. The critical value of b, $b^D(\lambda, \rho)$, that separates the regime where $\bar{e}^D(\lambda, \rho)$ is interior from the regime where $\bar{e}^D(\lambda, \rho) = 1$, is defined by the following system in $\underline{e}$ and b^D which is a straightforward adaptation of (40) and (39) to situations where $\rho > 0$:

$$\frac{\partial W^N(\underline{e}, 1)}{\partial \bar{e}} - \left[\frac{1}{\lambda} - r\right]\left[\underline{n} + \frac{g}{1-g}\right]\frac{\partial \Phi(\underline{e}, 1)}{\partial \bar{e}} = 0 \tag{42}$$

and

$$\frac{\partial W^N(\underline{e}, 1)}{\partial \underline{e}} - \left[\frac{1}{\lambda} - r\right]\left[\underline{n} + \frac{g}{1-g}\right]\frac{\partial \Phi(\underline{e}, 1)}{\partial \underline{e}} = 0 \tag{43}$$

When $\bar{e}^D(\lambda, \rho) = 1$, one can study the effect of the modification of the objective function on the optimal level of $\underline{e}$ by examining the first-order condition for an interior solution of problem (41) after substituting (35) for $\left[V\left(\underline{e}, \bar{\theta}, \alpha\right) - \bar{V}\right]$

$$0 = \frac{\partial J(\underline{e}, 1; \lambda, \rho)}{\partial \underline{e}} = \frac{\partial J(\underline{e}, 1; \lambda, 0)}{\partial \underline{e}} - \left[\frac{1}{\lambda} - r\right]\frac{g}{1-g}\frac{\partial \Phi(\underline{e}, 1)}{\partial \underline{e}} \tag{44}$$

At the optimum, the sign of $\frac{\partial J(\underline{e}, 1; \lambda, 0)}{\partial \underline{e}}$ must be opposite to the sign of the last term on the right-hand side of (44), which is positive (see the discussion below (36)): $\frac{\partial J\left(\underline{e}^D(\lambda,\rho),1;\lambda,0\right)}{\partial \underline{e}} < 0$; since $\frac{\partial J\left(\underline{e}^D(\lambda,0),1;\lambda,0\right)}{\partial \underline{e}} = 0$ by the first-order condition for problem (34), it follows by the second-order condition that $\underline{e}^D(\lambda, \rho) > \underline{e}^D(\lambda, 0)$.

When $\bar{e}^D(\lambda, \rho)$ is interior it is harder to determine exactly what happens when ρ increases because both $\bar{e}$ and $\underline{e}$ are affected. It is straightforward to show, as we just did for $\underline{e}$, that both $\underline{e}$ and $\bar{e}$, taken separately, would be set further away from the *Utopia* optimum when $\rho > 0$ than when $\rho = 0$ if the effort of the other type was kept unchanged.

[6]When $\rho = 0$, $g = 0$ so that problems (34) and (41) are identical.

However, since both effort levels are affected, not much more may be said unless J is supermodular and has the increasing difference property (Milgrom and Shannon, 1994).

The following proposition describe the properties of a *discriminating justice* system of imperfect accuracy and compare such a system with a perfectly accurate one.

Proposition 8 *(inaccurate, discriminating, justice)*

Let $\rho \in (0,1]$, $\lambda \in (0,1]$, *and* W *satisfy Assumption 2; let* $\left(\underline{e}^D(\lambda,\rho), \bar{e}^D(\lambda,\rho)\right)$ *be the solution to the* Institution *problem when* $b > b^N$ *and* $\underline{e} \neq \bar{e}$. *This solution has the following properties:*

1. *for* $b^N < b \leq b^D(\lambda,\rho)$, *the solution is such that* $\underline{e}^N \leq \underline{e}^D(\lambda,0) \leq \underline{e}^D(\lambda,\rho) < 1$ *(where the inequalities are strict if* $\lambda r < 1$*) and* $\bar{e}^D(\lambda,\rho) = \bar{e}^N = 1$;
2. *for* $b > b^D(\lambda,\rho)$, *the solution is such that* $\underline{e}^N \leq \underline{e}^D(\lambda,\rho) < \bar{e}^D(\lambda,\rho) \leq \bar{e}^N = 1$, *where the inequalities are strict if* $\lambda r < 1$;
3. $\phi(e) = \begin{cases} \phi^1, & e < \underline{e}^D(\lambda,\rho) \\ \underline{\phi} > 0, & \underline{e}^D(\lambda,\rho) \leq e < \bar{e}^D(\lambda,\rho) \\ \bar{\phi}, \text{ with } 0 < \bar{\phi} < \underline{\phi}, & \bar{e}^D(\lambda,\rho) \leq e \end{cases}$
4. $b^D(\lambda,\rho)$ *is defined by the condition that* (42) *and* (43) *be satisfied; it is such that* $b^D(\lambda,\rho) > \max\left\{b^N, \left[1-\alpha\left(\underline{e}^N,\bar{e}^N\right)\right]\bar{\theta}\right\}$;
5. *A sufficient condition for the existence of* $b^D(\lambda,\rho)$ *is* $\lambda r < 1$.

3.8. Choosing accuracy and detection levels

The accuracy of the justice system reflects the probability for a suspect to be attributed a bounty, and to be sentenced, according to his actual encroachment activities $1-e$. It diminishes as ρ increases from zero (perfect accuracy) to one (random system). It is often argued that the incentive objectives of a justice system may be achieved by imposing high sentences with a low probability. In this model, no incentive effect is achieved unless $\rho \neq 1$, but there is also an arbitrage between the cost of the system and its accuracy. In the absence of any direct cost considerations, society *ex ante* prefers perfect accuracy because this is how self selection is achieved at least cost.

The detection level λ is the probability for any individual, whatever his activities, to be investigated. As explained earlier, in the absence of any other considerations, society *ex ante* prefers λ to be set at the highest possible level ($\lambda = 1$) because this implies the least volatility in *ex post* sentences.

How should λ and ρ be chosen, given that they jointly determine the performance of the police-justice institution and the overall cost of running the system? The optimal level of ρ and λ may be obtained by maximizing $I(\lambda,\rho) \equiv$ $\max_{e,e}\left\{\underline{n}\,\underline{V} + \bar{n}\bar{V} - [1-\lambda r]\left[\underline{n} + \frac{g}{1-g}\right]\left[V\left(\underline{e},\bar{\theta},\alpha\right) - \bar{V}\right] - C(\lambda,\rho) \mid 0 \leq \underline{e} < \bar{e} \leq 1\right\}$. Using the envelope theorem and (35), this requires the following first-order conditions to hold

$$-\frac{\partial C(\lambda,\rho)}{\partial\rho}-[1-\lambda r]\left[V\left(\underline{e},\bar{\theta},\alpha\left(\underline{e},\bar{e}\right)\right)-\bar{V}\left(\alpha\left(\underline{e},\bar{e}\right)\right)\right]\frac{\partial\frac{g}{1-g}}{\partial\rho} = 0 \quad (45)$$

$$-\frac{\partial C(\lambda,\rho)}{\partial\lambda}+r\left[\underline{n}+\frac{g}{1-g}\right]\left[V\left(\underline{e},\bar{\theta},\alpha\left(\underline{e},\bar{e}\right)\right)-\bar{V}\left(\alpha\left(\underline{e},\bar{e}\right)\right)\right] = 0 \quad (46)$$

Since $\frac{\partial g}{\partial \rho} > 0$ (see (14)), condition (45) indicates that ρ must be set at a level such that the cost saving associated with a marginal reduction in justice accuracy is offset by the increase in expected sentences that the rise in $\frac{g}{1-g}$ implies.

While justice accuracy makes it cheaper to achieve self-selection, police presence affects overall welfare only through restitutions. In fact, the welfare cost of an expected sentence of ϕ at the individual level, $\frac{\phi}{\lambda}$, is decreasing in λ because of the preference of individuals for low variability in actual sentences (risk aversion). However, the effectiveness of sentences in achieving self-selection relies on the same cost: lower expected sentences are necessary to achieve self-selection when sentences are more painful (low λ). The net effect on individual welfare of a change in λ is nul. However, the portion of a sentence returned to society is averaged over all individuals so that its welfare value is not affected by *ex post* sentence variability at the individual level, while its welfare cost at the individual level is higher, the lower λ. Consequently, to the extent that some portion of a sentence is returned to society, there is a benefit from increased detection.

4. Discussion and conclusion

Police and justice institutions designed to remedy the *tragedy of the commons* may differ substantially from the textbook *utopian* optimal system. We have modelled such institutions as a combination of endogenous detection and accuracy levels that must be set, together with sentence schedules, according to various exogenous characteristics of the society they aim to regulate while taking cost considerations and information constraints into account.

Unless there is a perfect redistribution of sentences, the latters constitute a social cost that comes in addition to the cost of operating the system. That cost can be mitigated or even eliminated by distorting the effort requested from least productive individuals away from the *Utopian* optimum. *Blind justice* (pooling) is one of the available alternatives; by requesting the same productive effort from individuals of different productivities, this formula eliminates the need for sentences to be applied in equilibrium. The other alternative, *discriminating justice* (self-selection), induces different effort levels on the part of each type of producer. This involves self-selection costs that take the form of sentences.

Self-selection costs are mitigated by distorting effort levels away from the *utopian* optimum. Thus, in a *discriminating justice* system, asking the least productive individuals to work more reduces the attractiveness for more productive types to emulate their behavior thus allowing a lower sentence to be imposed upon them; this is the usual result found in the agency literature.

However another way to reduce the temptation for most productive types to imitate least productive ones consists in requesting from the formers a lower effort than would be desirable at the *Utopian* optimum. This second form of distortion, affecting the most productive types rather than the least productive ones, would be used together with the first form at high levels of b. It comes in contrast with the standard result in agency theory according to which agents at the 'good' end of the type spectrum are typically induced to adopt the same behavior under information asymmetry as under full information.

To the agency theorist, this surprising result finds its explanation in the fact that, in this model, sentences born by one type affect other types as well, not only because of justice errors, but also because of the externality involved in the determination of the aggregate encroachment rate.

From a social welfare point of view, truncating the production of more productive individuals in a *discriminating justice* system may be rationalized by noting that self-selection is least costly when effort differences between types are minimized; in fact the cost of self selection vanishes when $\underline{e} = \bar{e}$. However, as the *blind justice* system illustrates, the optimum effort may be strictly lower than unity when both types are constrained to produce the same effort: allowing some encroachment improves the lot of least efficient types substantially, which may warrant a loss in aggregate production. Similar considerations apply under *discriminating justice* when the need to reduce the cost of self-selection requires $\underline{e}$ and $\bar{e}$ to be set close to each other. This explains why a reduced effort level may be requested from more productive types even in a *discriminating justice* system.

Whatever the detection level, the design of optimal *discriminating justice* institutions is qualitatively similar when the system is perfectly efficient as when it makes errors. At low values of b, the non observability of productivity types does not raise any selection problem so that appropriate effort levels are induced by a mere threat of sentence, with no sentence being expected to be paid in equilibrium by any type. At higher levels of b, self-selection requires that less productive types be given sentences in equilibrium so that more productive types are discouraged from adopting the effort behavior meant for less productive ones.

Furthermore, when justice is not perfectly accurate, more productive types are sometimes sentenced by mistake so that they face positive expected sentences as less productive types do. Consequently, when the self-selection constraint is binding, a perfectly accurate system ($\rho = 0$) has an advantage over an imperfect one ($\rho > 0$): expected sentences are lower and effort distortions used to reduce expected sentences also tend to be lower in such a system. However, this advantage has to be weighted against the direct cost of running the institutions which diminishes as accuracy diminishes. In fact, when the system works randomly ($\rho = 1$) direct costs are low but the justice system is unable to screen individuals so that *blind justice* is the only available alternative.

Stealing a loaf of bread is an example of what may be considered petty crime in a modern society. However the same crime was considered serious at earlier development stages of the same societies. Our model explains such changes: as implied by Proposition 3, b^N is higher when the less productive group is more numerous or when its productivity is low. Thus stealing a loaf of bread would be considered a serious crime in a less developed society and would be punished. In a society where $b \leq b^N$, the *Utopia* optimum is achieved up to the cost of institutions. Sentences are trivially increasing in crime gravity: at the crime level tolerated by society, sentences are zero; at higher levels sentences are high enough to be perfectly dissuasive but they are not applied in equilibrium.

When b is higher than b^N, the opportunity cost of honesty is high for all individuals; society must arbitrage between the *Utopian* objective of discriminating between individuals in order to improve equity, and the costs of its institutions: direct cost and cost of self-selection.

The cost of self-selection vanishes when $\lambda r = 1$, when detection is certain and sentences are not of a wasteful type, so that they may be redistributed to society. In standard agency models (Laffont-Tirole,), this would correspond formally to situations where the private cost of funds equals the social cost. It may be argued long after Montesquieu (1747|1970) that r is constrained by individual wealth, especially by the least endowed's wealth, relative to b. In a homogeneous society, r can be high because the least endowed individuals are not dirt poor (Singapore? Western Europe?). A high level of r would be manifest in in a system that relied on fines and community work, rather that emprisonment and banishment, for sentencing. Furthermore, our model shows that a high level of r makes detection more valuable and accuracy less important. Consequently systems characterized by a high r would also exhibit much policing (high λ) and less developed a law profession (low ρ).

Conversely, a heterogeneous society and its low r would be manifest in its heavy reliance on wasteful sentences such as emprisonment or the death penalty. According to our model, these caracteristics would be associated with a low level of λ, that could be manifest in an undevelopped police system, or in a police system whose intrusiveness was checked by rules that effectively protected individuals against police actions. In such a society, however, justice accuracy would play an important role in reducing self-selection distorsions, and one would expect the law profession to be highly developped.

Our variable b is a crude measurement of the reward from illicit activities. The very nature of such activities is multidimensional and changes over time. The magnitude of b reflects the gravity of the appropriation problem faced by society. The appropriation of a bag of corn or the protection of a forest tract may be easier than the appropriation of an idea, so that b evolves over time and evolves in a way that is not necessarily correlated with with the evolution of individual productivities. In societies where, and at times when, knowledge production is an important, valuable, activity, b may be considered high. In such societies, our model predicts that it might be desirable to tolerate some encroachment activities not only on the part of least productive types but also on the part of more productive ones.

BIBLIOGRAPHY

Andreoni, J. (1991) "Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?," *Rand Journal of Economics*, **22(3)**, 385-95.

Bebchuk, L.A. and L. Kaplow (1993) "Optimal Sanctions when Individuals ar Imperfectly Informed about the Probability of Apprehension," *Journal of Legal Studies*, **21**, 365-70.

Ehrlich, I. (1975) "The Deterrent Effect of Capital Punishment: A Question of Life and Death," *American Economic Review*, **65(3)**, 397-417.

Hildenbrand, W.. (1974) *Core and Equilibria of a Large Economy.* USA: Princeton University Press.

Kaplow, L.l (1990) "Optimal Deterrence, Uninformed Individuals, and Acquiring Information about Whether Acts are Subject to Sanctions," *Journal of Law, Economics, and Organization*, **6(1)**, 93-128.

Laffont, J.-J. and J. Tirole (1993) *A Theory of Incentives in Procurement and Regulation.* Cambridge, Mass.: MIT Press.

Lasserre, P. (1994) "Protection décentralisée des droits de propriété, police et justice," *Actualit conomique*, **70(2)**, 177-90.

Malik, A.S. (1990) "Avoidance, Screening, and Optimum Enforcement," *Rand Journal of Economics*, **21**, 341-53.

Marceau, N. et S. Mongrain (1997) "Dissuader le crime: un survol," miméographe, département des sciences économiques, UQAM.

Meza, D. de and J.R. Gould (1992) "The Social Efficiency of Private Decisions to Enforce Property Rights," *Journal of Political Economy*, **100(3)**, 561–80.

Milgrom, P. and C. Shannon (1994) "Monotone Comparative Statics," *Econometrica*, **62(1)**, 157-80.

Mirrlees, J.A. (1971) "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, **38**, 175-208.

Montesquieu, C.-L. (1970) *De l'esprit des lois.* Paris: Gallimard.

Neary, H. (1997) " Equilibrium Structure in an Economic Model of Conflict," *Economic Enquiry*, **35(3)**, 480-94.

Polinsky, A.M. and S. Shavell (1979) "The Optimal Tradeoff Between the Probability and Magnitude of Fines," *American Economic Review*, **69(5)**, 880-91.

Shavell, S. (1991) "Specific versus General Enforcement of Law," *Journal of Political Economy*, **99**, 1088-1108.

Proof appendix

Proof of Proposition 1: $(0,0)$, $(0,1)$ and $(1,1)$ respectively induce $\alpha = 1$, $\alpha = b\frac{\underline{n}}{\bar{n}\bar{\theta}}$, and $\alpha = 0$. $(0,0)$ is a locally stable equilibrium if and only if, when $\alpha = 1$, $V_e\left(\underline{e}, \underline{\theta}, \alpha\right) = u'\left(.\right)\left[\left[1-\alpha\right]\underline{\theta} - b\right] < 0$ and $V_e\left(\bar{e}, \bar{\theta}, \alpha\right) = u'\left(.\right)\left[\left[1-\alpha\right]\bar{\theta} - b\right] < 0$. Similarly, $(0,1)$ (respectively $(1,1)$) is a locally stable equilibrium if and only if, when $\alpha = b\frac{\underline{n}}{\bar{n}\bar{\theta}}$ (respectively $\alpha = 0$), $\left[1-\alpha\right]\underline{\theta} - b < 0$ (respectively > 0) and $\left[1-\alpha\right]\bar{\theta} - b > 0$. The threshold values of b are derived from these conditions. Stability results from continuity: a small perturbation does not reverse the inequalities. ■

Lemma 1 *At the optimum of the* Utopia *model,* $\bar{u}' < u'_m < \underline{u}'$.

Proof. $\bar{u}' \neq \underline{u}'$. Otherwise, (7) and (8) would require $\underline{e} = \bar{e} = 1$ implying $\bar{u}' < \underline{u}'$, a contradiction. This leaves two possibilities:

A. $\bar{u}' < u'_m < \underline{u}'$ or

B. $\bar{u}' > u'_m > \underline{u}'$.

Suppose that B is true. Then (7) implies $\underline{e} = 1$ so that $\underline{e} \geq \bar{e}$. Now $\underline{y} = \left[1-\alpha\right]\underline{\theta}\,\underline{e} + \left[1-\underline{e}\right]b \leq \left[1-\alpha\right]\underline{\theta}\bar{e} + \left[1-\bar{e}\right]b < \left[1-\alpha\right]\bar{\theta}\bar{e} + \left[1-\bar{e}\right]b = \bar{y}$

The first inequality follows from the fact that $\frac{\partial y}{\partial e} < 0$ (since $b > \underline{\theta}$), combined with $\underline{e} \geq \bar{e}$; the second follows from $\bar{\theta} > \underline{\theta}$. But $\underline{y} < \bar{y} \;\Rightarrow \bar{u}' < \underline{u}'$ contradicting B. ■

Proof of Proposition 2: We need to back up the proof sketched in the text for item 1: $\underline{e}^U > 0$. We show that setting $\underline{e} = 0$ and $\bar{e} = 1$ would violate $\underline{y} < \bar{y}$ (Lemma 1): when $\underline{e} = 0$ and $\bar{e} = 1$, $\underline{y} = b$ and $\bar{y} = \bar{\theta} - \frac{\underline{n}}{\bar{n}}b$; then $\underline{y} \leq \bar{y}$ requires $b \leq \bar{\theta} - \frac{\underline{n}}{\bar{n}}b$ or $b \leq \bar{\theta}\bar{n}$, contradicting Assumption 1. ■

Proof of Proposition 3: From (9) :

1. $\frac{\partial b^U}{\partial \bar{n}} = \underline{\theta}\frac{u'(\underline{\theta})}{u'(\underline{\theta}) - u'\left(\bar{\theta}\right)}\frac{\underline{n\theta}\bar{\theta} - \bar{n}\left[\bar{n}\bar{\theta} + \underline{n\theta}\right]}{\left[\underline{n\theta}\right]^2} < 0$;

2. $\frac{\partial b^U}{\partial \bar{\theta}} = \underline{\theta}\frac{u'(\underline{\theta})u''\left(\bar{\theta}\right)}{\left[u'(\underline{\theta}) - u'\left(\bar{\theta}\right)\right]^2}\frac{\underline{n\theta} + \bar{n}\bar{\theta}}{\underline{n\theta}} + \underline{\theta}\frac{u'(\underline{\theta})}{u'(\underline{\theta}) - u'\left(\bar{\theta}\right)}\frac{\bar{n}\bar{n}\bar{\theta} - \bar{n}\left[\bar{n}\bar{\theta} + \underline{n\theta}\right]}{\left[\underline{n\theta}\right]^2} < 0$;

3. Other things equal, b^U increases when $u'\left(\underline{\theta}\right) - u'\left(\bar{\theta}\right)$ diminishes. ■

Derivation of (20). The following definitions and constraints apply:

$\omega\left(e, \theta, \alpha\right) - t \equiv V\left(e, \theta, \alpha\right) - \frac{1}{\lambda}\phi\left(e\right) - t$; $C\left(\lambda, \rho\right) = T + r\left[\bar{n}\phi\left(\bar{e}\right) + \underline{n}\phi\left(\underline{e}\right)\right]$; $T = t$

$$\underline{\omega} \equiv \underline{V}\left(\alpha\right) - \frac{1}{\lambda}\underline{\phi}, \text{ with } \underline{V}\left(\alpha\right) \equiv V\left(\underline{e}, \underline{\theta}, \alpha\right) \Leftrightarrow \underline{\phi} = \lambda\underline{V} - \lambda\underline{\omega}$$

$$\bar{\omega} \equiv \bar{V}\left(\alpha\right) - \frac{1}{\lambda}\bar{\phi}, \text{ with } \bar{V}\left(\alpha\right) \equiv V\left(\bar{e}, \bar{\theta}, \alpha\right) \Leftrightarrow \bar{\phi} = \lambda\bar{V} - \lambda\bar{\omega}$$

Substituting into the Harsanyi objective, $\max \underline{n}\,\underline{\omega} + \bar{n}\bar{\omega} - t$, gives $\underline{n}\,\underline{\omega} + \bar{n}\bar{\omega} - C + r\left[\bar{n}\phi\left(\bar{e}\right) + \underline{n}\phi\left(\underline{e}\right)\right]$ from which (20) follows. ■

Proof of Proposition 4: Entirely proven in the text. ■

Lemma 2 *For any given level of* $\underline{\phi}$, *it is optimal to choose* $F(\bar{e}) = 0$.

Proof. $\underline{\phi}$ and $\bar{\phi}$ are defined respectively by (10) and (11):

$$\begin{aligned} \phi(\underline{e}) &= \lambda \{\rho\bar{n}F(\bar{e}) + [1-\rho\bar{n}]F(\underline{e})\} \\ \phi(\bar{e}) &= \lambda \{[1-\rho\underline{n}]F(\bar{e}) + \rho\underline{n}F(\underline{e})\} \end{aligned}$$

Let $\underline{F} \equiv F(\underline{e})$ and $\bar{F} = F(\bar{e})$. Then, for any given $\underline{\phi}$, $\underline{F}$ and $\bar{F}$ must satisfy

$$d\underline{F} = -\frac{\rho\bar{n}}{1-\rho\bar{n}}d\bar{F} \tag{47}$$

Society's objective is the maximization of $[1-\lambda r][\underline{n}\,\underline{\omega} + \bar{n}\bar{\omega}]$. For any admissible pair $(\underline{e}, \bar{e})$ and for any admissible $\underline{\phi}$, this requires minimizing $\bar{\phi}$ by choice of $\underline{F}$ and $\bar{F}$ subject to (47), $\underline{F} \geq 0$, and $\bar{F} \geq 0$. Differentiating (11) and substituting (47) gives

$$\begin{aligned} d\bar{\phi} &= \lambda \left\{[1-\rho\underline{n}]\,d\bar{F} - \rho\underline{n}\frac{\rho\bar{n}}{1-\rho\bar{n}}d\bar{F}\right\} \\ &= \lambda \left\{\frac{1-\rho}{1-\rho\bar{n}}\right\} d\bar{F} \end{aligned}$$

Consequently $\bar{\phi}$ is increasing in $\bar{F}$ and $\bar{F}$ must be set at zero. ■

Corollary 1 *Constraint* (12) *may be replaced with* (13).

Proof. When $\bar{F} = 0$, (12) reduces to (13). ■

Proof of Proposition 5: Item #1 was not proven in the text:

1. When $b \leq b^N$, the solution to the *Institution* problem is not affected by the self-selection constraint and is described by Proposition 4.

Proof. Proposition 4 describes the solution when the self-selection constraint is not binding. We need to show that, when $b \leq b^N$, the solution to the *Institution* problem is not affected by the self-selection constraint.

Let us study the choice of $\underline{e}$ and $\bar{e}$, given the implications (outlined in the text) of that choice on $(\underline{\omega}, \bar{\omega})$ and $(\underline{\phi}, \bar{\phi})$. The constraints on $\underline{e}$ and $\bar{e}$ ($\underline{e}$ and $\bar{e} \in [0,1]$, (23) and the proper combination of (28), (29), and (30)) are represented in Figure A1 of which three variants may arise depending on whether the line $\underline{s}\,\underline{e} + \bar{s}\bar{e} = 1$ lies to the North-east or to the South-west of $(\underline{e}, \bar{e}) = (1,1)$ and, if it lies to the South-west, on whether its intercept $\frac{1}{\bar{s}}$ is higher or lower than unity.

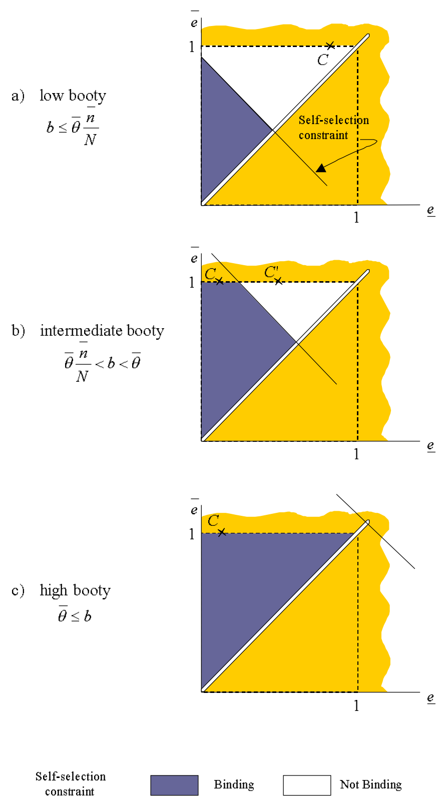

FIGURE A1: Optimal unconstrained effort pairs (C or C') and self-selection constraints.

Points $(\underline{e}, \bar{e})$ above the $\underline{s}\,\underline{e} + \bar{s}\bar{e} = 1$ line and points on the $\underline{e} = \bar{e}$ locus correspond to situations illustrated by point R_1 in Figure 1, where the selection constraints are not binding; points below the $\underline{s}\,\underline{e} + \bar{s}\bar{e} = 1$ line and away from the $\underline{e} = \bar{e}$ locus correspond to situations illustrated by point R_2 in Figure 1 where (22) is binding. Thus, when the $\underline{s}\,\underline{e} + \bar{s}\bar{e} = 1$ line lies to the South-west of $(1, 1)$, $(\underline{e}, \bar{e})$ may be located to either side of it; if $(\underline{e}, \bar{e})$ is above the line or on the diagonal, the solution is not constrained and $\left(\underline{\phi}, \bar{\phi}\right) = (0, 0)$; if $(\underline{e}, \bar{e})$ is below the line and away from the diagonal, the solution is constrained and $\underline{\phi} > 0$. Thus the properties of the solution depend on the location of the $\underline{s}\,\underline{e}+\bar{s}\bar{e} = 1$ line, itself determined by various parameters. As described in Figure A1, if $b < \bar{\theta}\bar{n}$, the intercept with the vertical axis is lower than unity; if $\bar{\theta}\bar{n} \le b < \bar{\theta}$ the intercept is higher than one but the line lies below point $(1, 1)$; if $b \ge \bar{\theta}$, the line lies above point $(1, 1)$. Let us study these alternative cases.

Proposition 4 gives the solution $\left(\underline{e}^N, \bar{e}^N\right)$ of problem (20) when no selection constraint is imposed. Since $\underline{e}^N \le 1$ and $\bar{e}^N = 1$, that pair may be represented by a point such as C in Figure A1. As implied by the discussion leading to (29), point C may or may not satisfy the constraint, depending on its position relative to the $\underline{s}\,\underline{e} + \bar{s}\bar{e} = 1$ line: if C lies above the line, or on the $\underline{e} = \bar{e}$ locus, then the constraint is met; if C is below the line and not on the diagonal, the constraint is violated. In Figure A1a, corresponding to $b \le \bar{\theta}\bar{n}$, C is necessarily above the $\underline{s}\,\underline{e}+\bar{s}\bar{e} = 1$ line so that it is admissible; in Figure A1b, corresponding to $\bar{\theta}\bar{n} < b < \bar{\theta}$, it may be below or above the line; in Figure A1c, corresponding to $\bar{\theta} < b$, it is necessarily below the line, so that the unconstrained solution is not admissible.

Let us consider the case $\bar{\theta}\bar{n} < b < \bar{\theta}$ in more detail (Figure A1b). Suppose that $b^U < \bar{\theta}$.[7] If $b \leq b^U$, by Proposition 4.4 and Proposition 2.2, $C = (1,1)$, which is necessarily above the $\underline{s}\,\underline{e} + \bar{s}\bar{e} = 1$ line, so that C is admissible. If b increases from b^U toward $\bar{\theta}$, the line moves to the right (it will contain $(1,1)$ when b reaches $\bar{\theta}$) and, by Proposition 4.3, C moves to the left of $(1,1)$. Consequently, there must exist some value of b, $b_1 \in \left(b^U, \bar{\theta}\right)$, below which the constraint is not binding and above which the constraint is binding.

On the other hand, if $b^U \geq \bar{\theta}$ (Figure A1c), the unconstrained solution is $(1,1)$, hence it is admissible for any $b \leq b^U$; if $b > b^U$, the unconstrained solution is $(\underline{e} < 1, 1)$, hence it is not admissible.

Let[8]

$$b^N \equiv \max\left\{b_1, b^U\right\}$$

We have shown that constraint (22) is not binding if $b \leq b^N$. ■

Proof of Proposition 6: Let

$$D(b) \equiv \left[\underline{n}\,\underline{V}'\underline{\theta} + \bar{n}\bar{V}'\bar{\theta}\right]\frac{1}{\left[\underline{n}\,\underline{\theta} + \bar{n}\bar{\theta}\right]} + \underline{n}\,\underline{V}'\frac{[\underline{\theta} - b]}{b} + \bar{n}\bar{V}'\frac{\left[\bar{\theta} - b\right]}{b} \text{ with } e = 1 \tag{48}$$

We will first show that $D(b)$ has the same sign as the left-hand side of the first-order condition (32) when $e = 1$. Then we will show that b^B exists and that $D(b)$ is positive at $b \leq b^B$ and strictly negative at $b > b^B$, thus completing the proofs of items #1, #2, and #3. Item #4 is proven in the text.

Using (1) one can show that

$$\left(\frac{\partial\alpha(e,e)}{\partial\underline{e}} + \frac{\partial\alpha(e,e)}{\partial\bar{e}}\right)e = -b\left[\frac{1}{\left[\underline{n}\,\underline{\theta} + \bar{n}\bar{\theta}\right]} + \frac{\left[\bar{n}\bar{\theta} + \underline{n}\,\underline{\theta}\right][1-e]}{\left[\underline{n}\,\underline{\theta} + \bar{n}\bar{\theta}\right]^2 e}\right]$$

Substituting into the first-order condition (32) yields

$$b\left[\underline{n}\,\underline{V}'\underline{\theta} + \bar{n}\bar{V}'\bar{\theta}\right]\left[\frac{1}{\left[\underline{n}\,\underline{\theta} + \bar{n}\bar{\theta}\right]} + \frac{\left[\bar{n}\bar{\theta} + \underline{n}\,\underline{\theta}\right][1-e]}{\left[\underline{n}\,\underline{\theta} + \bar{n}\bar{\theta}\right]^2 e}\right] + \underline{n}\,\underline{V}'\left[[1-\alpha]\,\underline{\theta} - b\right] + \bar{n}\bar{V}'\left[[1-\alpha]\,\bar{\theta} - b\right] \geq 0$$

When $e \to 0$ the left-hand side tends toward $+\infty$ which confirms that the solution must be strictly positive. Setting $e = 1$ and dividing by b, the left-hand side reduces to $D(b)$. $D(b)$ is decreasing; thus if b^B exists, (32) is negative for $b > b^B$, which implies that the solution to problem (31) is interior.

We still need to show that b^B exists. Since $e = 1$, $\underline{V}'$ and $\bar{V}'$ are independent of b (see definitions (2), (3), (17), (18)). Thus $D(b)$ is decreasing and continuous; since it is clearly positive at low values of b, a sufficient condition for the existence of b^B is:

[7] An examination of (9) indicates that $b^U < \bar{\theta}$ is possible for some parameter configurations.

[8] As just shown, b_1 does not exist if $b^U \geq \bar{\theta}$; we take the convention that the max is then b^U.

$\lim_{b\to\infty} D(b) < 0$. That limit may be written as

$$\lim_{b\to\infty} D(b) = \left[\frac{\underline{n}\underline{\theta}}{\underline{n}\,\underline{\theta}+\bar{n}\bar{\theta}}\underline{V}' + \frac{\bar{n}\bar{\theta}}{\underline{n}\,\underline{\theta}+\bar{n}\bar{\theta}}\bar{V}'\right] - \left[\underline{n}\,\underline{V}' + \bar{n}\bar{V}'\right]$$

This can be seen to be the difference between two concepts of average marginal utilities: production weighted average marginal utility and population weighted average marginal utility; since low productivity types weight less in production than in population, the first average gives less weight to $\underline{V}'$ relative to $\bar{V}'$ than the second one. Since $\bar{V}' < \underline{V}'$ the expression is negative. ■

Proof of Proposition 7: The claims that $\underline{e}^D(\lambda, 0) = \underline{e}^N$ and $\bar{e}^D(\lambda, 0) = \bar{e}^N$ when $\lambda r = 1$ (in items #1 and #2) follow from the fact that the objective functions for problems (20) and (34) are identical when $\lambda r = 1$. All other items were proven in the text, except for item #4, the existence of $b^D(\lambda, 0)$, and the statement $b^D(\lambda, 0) > [1-\alpha]\bar{\theta}$.

Proof. Let us start with $b^D(\lambda, 0) > [1-\alpha]\bar{\theta}$. If $b \leq [1-\alpha]\bar{\theta}$, (37) implies that $\frac{\partial\Phi(\underline{e},\bar{e})}{\partial\bar{e}} \leq 0$ which in turn implies that the inequality in (39) is strict so that $\bar{e}^D = 1$. By continuity, this remains true in a neighborhood of $[1-\alpha]\bar{\theta}$ so that, if it exists, $b^D(\lambda, 0)$ must strictly exceed $[1-\alpha]\bar{\theta}$.

Turning to the existence issue, we need to prove that, for any combination of all parameters other than b such that $\lambda r \neq 1$, there exist some value of b above which the value of $\bar{e}$ which solves problem (38) is interior. Since the constraint $\underline{e} < \bar{e}$ is not binding in that problem, we may remove it so that the conditions of application of Berge's maximum theorem are satisfied (see, e.g. Hildenbrand (1974), Corollary of Theorem 3, p. 30). Thus the solution $\left(\underline{e}^D, \bar{e}^D\right)$ is upper-semi-continuous on $[b^C, +\infty)$. Our strategy will be to show that, when $b \to \infty$, the optimal value of $\bar{e}$ is strictly interior: by upper-semi-continuity, this implies that there exists a finite value of b above which the optimal value of $\bar{e}$ is strictly interior.

First, let us show that $\lim_{b\to\infty} \underline{e}^D = \bar{e}^D$. Suppose, contrary to the claim, that there exists $\varepsilon > 0$ such that $\lim_{b\to\infty}\left[\bar{e}^D - \underline{e}^D\right] > \varepsilon$. Then $\lim_{b\to\infty}\alpha = +\infty$, which contradicts $\alpha \leq 1$.

Second, note that $\lim_{\underline{e}\to\bar{e}} \Phi(\underline{e}, \bar{e}) = 0$. Consequently, when $b \to \infty$, problem (38) approaches the following problem: $\max_e W^N(e, e)$. This problem has been shown to have an interior solution (Proposition 6). ■

Proof of Proposition 8: A simple adaptation of the proof of Proposition 7. ■

Liste des publications au CIRANO *

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

99c-1 Les Expos, l'OSM, les universités, les hôpitaux : Le coût d'un déficit de 400 000 emplois au Québec — Expos, Montréal Symphony Orchestra, Universities, Hospitals: The Cost of a 400,000-Job Shortfall in Québec / Marcel Boyer

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

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

95c-1 La réglementation incitative / Marcel Boyer

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

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

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

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

99s-36 Déterminants du recours au travail atypique : Une étude des travailleurs à statut précaire dans les organisations québécoises / Michel Tremblay, Patricia Bielman, Gilles Simard et Denis Chênevert

99s-35 Testing the Option Value Theory of Irreversible Investment / Tarek M. Harchaoui et Pierre Lasserre

99s-34 A Resource Based View of the Information Systems Sourcing Mode / Vital Roy et Benoit Aubert

99s-33 Budget Processes: Theory and Experimental Evidence / Karl-Martin Ehrhart, Roy Gardner, Jürgen von Hagen et Claudia Keser

99s-32 Tax Incentives: Issue and Evidence / Pierre Mohnen

99s-31 Decentralized or Collective Bargaining in a Strategy Experiment / Siegfried Berninghaus, Werner Güth et Claudia Keser

99s-30 Qui veut réduire ses heures de travail? Le profil des travailleurs adhérant à un programme de partage de l'emploi / Paul Lanoie, Ali Béjaoui et François Raymond

99s-29 Dealing with Major Technological Risks / Bernard Sinclair-Desgagné et Carel Vachon

99s-28 Analyse de l'impact productif des pratiques de rémunération incitative pour une entreprise de services : Application à une coopérative financière québécoise / Simon Drolet, Paul Lanoie et Bruce Shearer

99s-27 Why Firms Outsource Their Human Resources Activities: An Empirical Analysis / Michel Patry, Michel Tremblay, Paul Lanoie et Michelle Lacombe

---

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