Symmetric Third-Party Governance of Trades[*]

Alia Gizatulina[†]

December 30, 2007

Abstract

This paper studies individual incentives to pay to the fixed cost of an impartial contract enforcement institution. If agents do not collect enough resources, they could still enforce their trades through an asymmetric enforcement system, where only resourceful agents are able to punish their encounters for cheating. The main results are: (i) When the impartial system excludes non-contributors, even resourceful agents, who prefer asymmetric governance to be the only available mode for everyone in the population, would pay to the fixed cost; however there are also equilibria where two modes co-exist in equilibrium, i.e., some resourceful agents remain under asymmetric governance. (ii) The asymmetric governance mode improves upon a regime with no punishment at all, but there is always a positive rate of cheating within contracts; consequently, on aggregate it is a sub-optimal regime as compared to the impartial system under which the rate of cheating is zero.

1 Introduction

Exchanges constitute one of the main parts of economic activity. In general, most of them, while being mutually beneficial, are vulnerable to opportunistic behaviour of the trading agents themselves. When damaging behaviour from an encounter is to be expected, agents may decide not to start any trade at all at the first place. Hence, unless some punishment for misbehaviour is available to agents, the level of economic activity is likely to be inefficiently low.

When agents are able to have repeated economic interactions, in most cases reputational concerns help to enforce mutually beneficial trading strategies. However when the size of a market and distance among agents grow, the reputation-based governance usually does not work well any longer. Instead, a rule-based third party governance is to be introduced (Dixit (2003b)).

It was argued in the literature that third party enforcement institutions are likely to require an investment into up-front capacity (Li (2000)). The resources may be needed to cover "efficiency wage" of enforcers, cost of a data-base on previous cheating, investment into design of the efficient and up-to-date codes, etc. These are likely to be complementary and hence failure to cover a part of the cost is likely to result in dysfunctional enforcement institutions.

[*] This paper was named before "Endogenous Symmetric Punishment". I am very grateful to Martin Hellwig for helpful and generous in ideas discussions. I would like to thank as well Avinash Dixit, Bruno Jullien, Elena Panova and Eugenia Winschel for valuable comments and suggestions. All errors are mine.

[†] Email: gizatulina@uni-mannheim.de. Department of Economics (CDSE - GESS), University of Mannheim. D-68131, Mannheim, GERMANY. Financial support from Deutsche Forschungsgemeinschaft (DFG) and Heinrich Vetter Stiftung is gratefully acknowledged.

This paper looks at two alternative governance systems. The first one is an idealized enforcement system, i.e., it is assumed to be able to enforce trade agreements[1] perfectly and there are no incentive, cognitive or informational constraints to discover malfeasances of any trading agent, but it requires agents to cover its cost up-front[2], before any first trade takes place. This ideal system is called "the symmetric punishment system" (SPS, for short)

The second system is, in a sense, dysfunctional. It enforces agents' contracts not only depending on their true behaviour within a contract, but also as a function of both agents' "strength types". Namely, some agents are more able to obtain justice than others. In a match of two agents such that an agent i is "stronger" than agent j, agent i is able to punish dishonest behaviour of j, being "the strong" while if i cheats on j, j is not able to get compensation from i as he is "the weak" one. This system, compared to the previous one, is assumed to be costless to agents[3]. It is called "the asymmetric punishment system" (APS).

In the paper I study the following questions:

- Which incentives to behave honestly do different systems provide to agents? How incentive to behave honestly depend on agents' strength type? What is the individual valuation for either system?
- Which incentives to contribute resources to the fixed cost of the efficient system do agents have, given they are excluded and have to trade under the asymmetric system if they do not contribute?

The asymmetric system may be illustrated by the imperfect legal systems of many developing countries. For example, in India only relatively resourceful agents are able to obtain justice (an article in the *The Economist*, December 26, 2006 provides examples of this). The importance of resources for performance within the American legal system is discussed, for example, in Galanter (1978)[4]. He was the first in the legal literature to theorize different dimensions of strength bringing advantages to "the haves" over "the have-nots" within the legal process. The main idea of his paper was to argue that legal experience, financial resources and political power affect positively and non-negligibly the probability of winning the case[5].

[1] In this paper I am concerned only with contract rights, i.e. I do not consider issues of thieft or extortion (i.e. violation of property rights).

[2] In case there are still such constraints, one can argue they could be overcome by, for example, paying higher "efficiency wage" to enforcers in order to make them immune to any bribes or by hiring more investigators to search for the true state of things. But in the essence, it is only an increase of the fixed cost. That is why I directly assume that the fixed cost is already such that everything is working perfectly once it is covered.

[3] The zero marginal cost of the APS can easily be justified through the assumption that the sector contains at least two service providers and they compete in prices. Introduction of any non-zero fixed cost does not provide qualitative changes, as what matters is relative fixed costs of two punishment systems.

[4] For another example of the importance of resources in the Italian courts see Enriques (2002).

[5] His analysis has provoked a number of subsequent studies aiming to prove or disprove, in a quantified way, the role of the resource factor in the probability of winning in courts. Lempert (1999) reviews many such studies and makes a sharper statement that it is political power and financial means that are more relevant for litigants' performance in courts. As "experience", "repeatedness" or "technicity of issues" are matters that could be overcome via the employment of top lawyers. Sheehan and Songer (1992) quantitatively prove the presence of advantages of "the haves" in the United States Courts of Appeal. They show that the Federal Government is likely to win against local governments, big firms perform better than small firms, any government performs better than a firm, any individual is less likely to win than any firm, and any minority group individual or a person from the bottom of the income distribution is less likely to win than the government, a firm and any other non-minority or a relatively well-to-do individual.

This system may also be illustrated by the work of some private organizations. For example, the theoretical study by Dixit (2003a) makes a case for costly incentives of Mafia enforcers, who need not provide impartial enforcement services for free. A study by Hill (2003) and Hill (2006) gives examples of a non-negligible demand for Japanese Yakuza services (e.g., to enforce bankruptcy law, to recollect debts in general or to enforce anti-competitive agreements among colluding companies). Here as well, often the "winning" party of a trade dispute is the one who is able to hire the strongest gang. Moreover the demand for enforcement remains to be positive despite of the fact that in general Yakuza take advantage of both trading agents.

Similarly to Dixit (2003a), to model a transaction where each agent has an opportunity to profit at the expense of his trading partner, I employ the two-sided prisoner's dilemma game. That is, agents could exchange goods or services, or produce something together, but receiving a positive benefit from a transaction by both agents is conditional on both agents behaving cooperatively[6]. If an agent deviates, he perceives the highest benefit when the second agent continues to give his best to the common business. If both agents misbehave at the same time, the common value of the exchange is the lowest[7]. I assume each agent is free not to trade at all, and the outside opportunity brings him a utility not worse than a payoff in case where both misbehave to each other.

The main answers to the first question are as follows. When applied asymmetrically punishment is not too high or too low, the APS induces a strictly positive rate of honest behaviour. It is then beneficial *to every agent,* even the weakest one, as compared to the world with no punishment at all. However, participation on the market is fragile, and the market can breakdown completely if the applied punishment is too low or too high. The intuition is, if punishment is too high, weak agents get a lower utility than staying outside of the market, so they do not enter. In this case strong agents anticipate that they do not have a chance to match the weaks and obtain a rent of the strong, so they do not enter either. If punishment for cheating is too low, the weak agents start to cheat too much and the strong agents prefer to stay at home in the first place. Hence, the level of asymmetrically imposed punishment cannot be increased arbitrarily in order to provide both agents with incentives to behave honestly and so there is always non-zero rate of cheating under the APS. By contrast, in the system with symmetrically imposed punishment, the rate of cheating under a wide range of the parameters is zero. In this case everyone trades in the market.

The aggregate welfare from contracting under the system imposing punishment symmetrically is always non-inferior to the aggregate welfare where only the asymmetric punishment system is in place. However, at the intermediate values of punishment, when trade on the market under the APS does not break down, two systems cannot be Pareto ranked among themselves. Namely, if punishment is moderately high, the strong agents obtain rents from trade under asymmetric governance and prefer this to be the only mode for everyone. If punishment is moderately low, these are the weak who obtain rents under the APS as compared to the SPS.

Before proceeding to the answers to the second line of questions, there are a couple of remarks on modelling. The first remark is, the SPS actually possesses properties of an excludable public good: it is non-rival and it is technically possible to exclude agents from enforcement of their contracts. It

[6] There are many possible examples of opportunism within bilateral contracting: a buyer cannot observe immediately the quality of the good proposed by the seller, and if the latter provides him with a lemon he cannot reject it at the moment of the exchange (and by doing this provide correct incentives to the seller ex ante); the buyer himself could renege on the payment if, for example, it is to be done in a while. Essentially, for any transaction nowadays if there is a contract, it means there is a scope for one-sided or mutual misbehaviour that this contract aims to prevent.

[7] One can observe that I am only changing the names of the strategies common to the prisoner's dilemma game, in particular instead of playing "cooperative" and "defecting" strategies, here agents either play "honestly" or "cheat".

is non-rival because in equilibrium it is only a threat as no one cheats, hence no one uses its services (but this threat, however, should be still credible). So a marginal agent does not exhaust, as such, its availability to others. The second remark is, agents valuation for the SPS is likely to be private information at the ex ante stage, before matching with a trading partner.

Consequently, a problem of aggregating preferences for the SPS and collecting resources to cover its up-front, fixed cost is similar to the problem of provision of a public good under payoffs uncertainty. This has been studied already in the mechanism design literature. The main result is there is no way to elicit detailed private information about agents valuations for a public good through a Pareto-improving multilateral contract, signed voluntarily after some negotiation process (Mailath and Postlewaite (1990)). Which means voluntary contributions to the cost of a public good are zero in a large economy. That is why I study directly the admission fee game (the subscription game), which is based on exclusion from enforcement by the symmetric punishment system of those agents who do not contribute to its cost. This explains the formulation of the second question above.

The last remark is, in case of the punishment system, "exclusion" is inherently two-dimensional. Each agent could be excluded from *ability to punish* and from *ability to be punished*. Hence exclusion could be "partial" – when agents are excluded only from a possibility to punish their encounters; "full" – when agents are excluded from both, a possibility to punish their encounters and possibility of being punished by them; and "quasi-none" – when both agents are allowed to punish each other, provided that at least one of them has contributed to the cost of the symmetric system. I study results of the admission fee game under these three exclusion rules separately.

The main result for the subscription game is that when the applied exclusion rule is sufficiently severe (i.e. it is either "partial" or "full", but not "quasi-none"), there exists an equilibrium in which *all agents*, even the strongs, who prefer only the asymmetric governance mode to be available to everyone, do subscribe (i.e. contribute) for the services of the symmetric punishment system. Fearing of not being able to get any rent of the strong any longer, strong agents have incentives to pay in order to be protected and rather profit from trading opportunities with those who are under the SPS[8].

The equilibrium above is an equilibrium with extreme subscription behaviour. But there are also equilibria where not all strong agents subscribe for the SPS, if they believe that the number of those who remain under the APS is relatively large. In a hope to match these agents and obtain a rent of the strong, they remain under the APS themselves. If enough of agents share the belief that there is a non-zero fraction of agents who do not subscribe to the SPS, they do not to subscribe themselves either. This coordination issue brings about the co-existence of two governance systems, asymmetric and symmetric one (e.g. Mafia and the official, say, rather efficient legal system).

Finally, I evaluate boundaries of the aggregate welfare from the play of the admission fee game in a decentralized manner. Because the APS alone brings a positive level of payoffs to all agents, there is a benchmark level of the cost above which it is inefficient, from the social point of view, to install the SPS as it is too costly as compared to the APS. The main result is, depending on the exclusion rule, there may be both over- and under-provision of the SPS. This means, the SPS may have a very high cost a but everyone subscribes for it even if the aggregate payoffs would be higher, had everyone stayed under the APS. Similarly, the SPS may cost very few, but the subscription game fails

[8] Thus, in a sense, the paper could quantify the value of "the social contract" (Hobbes (1651), Rousseau (1762)). Here as well, by subscribing for the services of the SPS, agents conclude with a third-party a contract that it should protects them against others' cheating and at same time it should take them liable for their own misbehaviour with respect to the others. (So the difference between the aggregate gain from "the order" on the market and the cost of maintainance of the third party is equal to the value of "the social contract").

to collect even this amount. In addition, if there is co-existence of two modes in equilibrium, there is a welfare loss because once the symmetric system is in place, whichever its cost, the aggregate welfare is superior when everyone trades under the SPS.

The different exclusion rules influence to different degrees the boundaries of inefficiency. The partial exclusion rule bears the risk to over-provide the SPS only moderately (i.e. it installs the SPS only at a "moderately inefficiently high" cost) whereas the full exclusion rule may install the SPS even when its cost is very high. By contrast, the quasi-none exclusion rule, in general, under-provides the SPS.

The related literature belongs to two different streams. First, it is research on the ways economic transactions could be governed. My analysis is very close to the book by Dixit (2004). However the book does not cover explicitly the issue of collecting resources for an up-front fixed cost of the symmetric punishment. The second stream is the mechanism design literature on provision of public goods in a large population of agents. The relevant papers are Hellwig (2003), Hellwig (2007), Mailath and Postlewaite (1990) and Norman (2004). Given that the type of the problem studied here is a particular one, namely exclusion from punishment is inherently two-dimensional, it makes the results available in the literature for public goods provision not to be directly applicable and justifies the explicitness of the current analysis.

The paper is structured as follows. The next section studies individual behaviour and resulting payoffs under asymmetric and symmetric punishment systems. Section 3 gives the description of the mechanism design tools and exposes the results for the admission fee game. Section 4 contains some discussions, and, in particular, it discusses results for the monopolistic asymmetric punishment system. Section 5 briefly presents conclusions.

2 Transactions under Different Governance Modes

2.1 Bilateral transactions without any governance

An economy consists of a continuum of agents of mass 1. Agents match in pairs one-shot to trade with each other on some markets, and as it was justified already in the introduction, each game, played bilaterally, has the strategic properties of the prisoner's dilemma game.

The next matrix presents in a simplified way this strategic situation. H stands for playing "honest" and C for "cheating" behaviour:

$$\begin{array}{|c|c|c|} \hline & H & C \\ \hline H & 1,1 & 1-\nu, 1+\nu \\ \hline C & 1+\nu, 1-\nu & 0,0 \\ \hline \end{array} \tag{1}$$

Here ν is the stake of cheating and it is such that $\nu > 1$. The only equilibrium strategy here is to misbehave for both agents.

2.2 Punishment

Under *asymmetric governance* agents are heterogenous in ability to punish dishonest behaviour of their encounters. This is captured by the parameter θ which is distributed in the population according to a continuous *cdf* $F(\theta)$ with the support $\left[\underline{\underline{\theta}};\overline{\overline{\theta}}\right]$, $\overline{\overline{\theta}} - \underline{\underline{\theta}} > 0$. If an agent i matches with an agent j such that the realization of types is $\theta_i > \theta_j$, an agent i is called the *strong (*he has $\theta = \overline{\theta}$

within the match*)*, he is able to punish for sure the deviation of j, if it occurs. Agent j is then the *weak* (he has $\theta = \underline{\theta}$ within the match) and if the strong agent cheats him, by contrast, he is not able to obtain a compensation.

The payoff matrix of the transaction is as follows, the row agent is the strong and the column agent is the weak:

$strong\backslash weak$	H	C
H	$1, 1$	$1 - \nu + \rho, 1 + \nu - \rho$
C	$1 + \nu, 1 - \nu$	$0, 0$

. (2)

The exogenously set parameter ρ reflects the level of punishment[9] imposed on a cheating player[10].

Note that neither player can be punished if both agents have deviated, i.e. if the state $\{C, C\}$ has occurred. This assumption can be justified by resource constraint of agents (e.g. when nothing is produced together, nothing can be taken from each other).

Both agents observe each other's type at the moment of match and prior to the choice of their optimal strategies within a transaction.

Hence, potentially, every agent can be in one of two possible games: where he is strong and where he is weak. The ex ante probability (prior to the realization of matching uncertainty) to be in either game is coming from agents' type θ. The higher is θ_i, the more agent i is likely to be in the game where he is strong as the probability to be of a higher type than agent j is given by $F(\theta) = F_j(\theta_i) = \Pr(\theta_j < \theta_i) = \int_{\underline{\theta}}^{\theta_i} dF_j(\theta)$.

I assume there the fee agents pay for the service of asymmetrically imposed punishment is equal to zero[11]. I discuss in the section 4 the results when there is a monopolistic provider of the asymmetric punishment services who makes take-it-or-leave-it offers.

As for the *symmetric governance system*, by definition, it is able to punish cheating of any agent, independently of his whatsoever identity. The payoffs matrix for a match is then:

$strong\backslash weak$	H	C
H	$1, 1$	$1 - \nu + \rho, 1 + \nu - \rho$
C	$1 + \nu - \rho, 1 - \nu + \rho$	$0, 0$

(3)

Again, as in the previous case, I assume that if both agents cheat on each other there is no way to punish both of them simultaneously. The fixed cost of this system is equal to K.

In the following section I provide with equilibrium strategies chosen by agents under each governance mode.

[9] I incorporate directly the punishment into the payoff matrix, although there is certainly some multistage negotiation/litigation process behind. Once there is a complete information about the individual behaviour within a transaction, taking the reduced form is w.l.o.g.

[10] This parameter may consist of several parts in reality, for example it may be a probabilistically imposed punishment, where only with some probability π ($0 \leq \pi \leq 1$) a fine R is imposed. Probability π may reflect the general quality of asymmetric punishment system whereas R could reflect, for example, discretion at which the punishment is imposed (as R can be higher or lower than the cheating stake ν). Then ρ can be thought about as $\rho = \pi R$. However, under assumption of complete information among all agents about types and given that π and R are exogenous parameters, it is also without any loss that one can analyze directly for ρ.

[11] The reason of why I assume away all cost considerations is to distill out the net impact of *asymmetrically* imposed punishment on agents utility.

2.3 Individual Behaviour under Two Punishment Modes

2.3.1 Equilibrium Payoffs from Transactions under Asymmetric Punishment

Suppose only the system with asymmetrically imposed punishment is available to everyone. Agent i plays H with probability $\zeta_i(H)$ and C with the complementary probability $\zeta_i(C) = 1 - \zeta_i(H)$. Exploring the matrix (2) one can see there are three possible types of equilibria, depending on the values of the parameters:

Case 1 $\rho > \nu$

Under such constellation, there is no pure strategy equilibrium in the game specified in (2). The equilibrium mixed strategies are:

$$\zeta_i(\{H\}, \underline{\theta}) = \frac{1-\nu+\rho}{1+\rho}; \zeta_i(\{C\}, \underline{\theta}) = \frac{\nu}{1+\rho} \tag{4}$$

for the behaviour of an agent i when he is weak.

And when agent i is the strong type:

$$\zeta_i(\{H\}, \overline{\theta}) = \frac{\nu-1}{\rho-1}; \zeta_i(\{C\}, \overline{\theta}) = \frac{\rho-\nu}{\rho-1}. \tag{5}$$

The punishment and the deviation stake have opposite effects on probability of honest behaviour from the strong and from the weak. The higher is ρ, the more honestly behaves the weak and the less honestly behaves the strong. The higher is ν the more honest behaviour comes from the strong and the less comes from the weak. As $\rho \to \nu$ an agent who is the strong behaves honestly with probability going to 1. But such ρ deteriorates the incentives of the weak agent to behave honestly.

Corresponding utilities to be the weak and to be the strong are:

$$EU^A(\underline{\theta}) = (1+\nu-\rho)\,\frac{\nu-1}{\rho-1} \tag{6}$$

$$EU^A(\overline{\theta}) = (1+\nu)\,\frac{1-\nu+\rho}{1+\rho} \tag{7}$$

Thus, any agent of type θ has an expected utility from entry in to the market, before learning the type of his match:

$$EU^A(\theta) = F(\theta) \cdot EU^A(\overline{\theta}) + (1-F(\theta)) \cdot EU^A(\underline{\theta}) \tag{8}$$

$$= F(\theta) \cdot (EU^A(\overline{\theta}) - EU^A(\underline{\theta})) + EU^A(\underline{\theta})$$
$$= F(\theta) \cdot \tfrac{2\rho\nu(\rho-\nu)}{(\rho^2-1)} + (1+\nu-\rho)\,\tfrac{\nu-1}{\rho-1}$$

The expected utility is strictly increasing in θ as $\rho > \nu > 1$.

Case 2 $\nu - 1 < \rho < \nu$

In this case the unique, pure strategies equilibrium is $(\zeta_i(\{H\}, \overline{\theta}), \zeta_j(\{H\}, \underline{\theta})) = (1, 0)$. That is an agent who is the strong behaves honestly but the weak agent always cheats. The expected utility from entering into the market is

$$EU^A(\theta) = 1 + (1 - 2F(\theta))(\nu - \rho) \tag{9}$$

It is decreasing with θ.

Case 3 $\rho < \nu - 1$

Here, the punishment is so low that it does not preclude any misbehaviour from either agent and the unique, pure strategies equilibrium is $(\zeta_i(\{H\}, \overline{\theta}),\ \zeta_j(\{H\}, \underline{\theta})) = (0,0)$. The expected utility from participating in the market is $EU^A(\theta) = 0$.

2.3.2 Entry into contracting when only asymmetric punishment is available

I am interested in trading situations where the initial decision, to participate in the market is voluntary. For example, relatively weak agents are free not to enter into the market at all and hence not to submit themselves under unpunished acts of the strong. Or inversely, this may be the strong who would prefer not to deal with cheating weak agents and so not to enter at all. Consequently one should study the question how large is the demand for the APS services given the parameters of the model. This is defined by the gains relative to staying outside of the market.

I assume that not participating at all in the market brings the utility equal to zero to any agent[12], independently of his θ. Because the expected utility from participating in the market is increasing with type θ when $\rho > \nu$, presumably the agents with low θ would be first who would prefer to take rather outside opportunity. Similarly when $\rho < \nu$, strong agents are the first who would stay outside of contracting.

It happens that non-participation by a subgroup of agents has population-wide implications and actually *no agent at the end enters the market if for a subgroup of agents it is not beneficial.* The following proposition provides with the details of this result.

Proposition 4 *There exist two levels of punishment* $\rho^{\min} = \nu - 1 > 0$ *and* $\rho^{\max} = \nu + 1 < \infty$ *such that:*

- if $\rho \in \left[\rho^{\min}, \rho^{\max}\right]$ *the equilibrium payoffs for all types under asymmetric punishment are strictly greater than individual payoffs for all types under no punishment;*

- if $\rho \notin \left[\rho^{\min}, \rho^{\max}\right]$ *equilibrium payoffs are zero for all types, as they are in the absence of punishment.*

Proof.

1. Case $\rho > \nu$: The expected utility from entry into the market is increasing monotonically in type. Hence I will check incentives of the lowest entering type (it is unique). For any (ρ, ν) define this type θ^e, and he is found from the identity $EU^A(\theta^e) \equiv 0$. Namely

   $$F(\theta^e) = \frac{(\rho+1)(\nu-1)(\rho-\nu-1)}{2\nu\rho(\rho-\nu)}.$$

   One can see that the participation is full, as r.h.s is below zero, if $\rho < \nu + 1 \equiv \rho^{\max}$. Now I show that for $\rho > \rho^{\max}$ there is *no* any agent who would enter.

   Suppose $\rho > \rho^{\max}$. Then there is a fraction of agents, namely $F(\theta^e)$ who take the outside option. The remaining fraction $(1 - F(\theta^e))$ reassess the utility from participating in the market given (ρ, ν). The expected utility of an agent $\theta \in \left[\theta^e, \overline{\theta}\right]$ is:

   $$EU^{A\prime}(\theta) = (F(\theta) - F(\theta^e))EU(\overline{\theta}) + (1 - F(\theta))EU^A(\underline{\theta}).$$

[12] I assume also that when agents expect an outcome $\{C, C\}$ to happen in the market (which provides with payoff zero), they prefer rather to stay at home.

I search again for an agent such that $EU^{A\prime}(\theta) \equiv 0$, call him $\theta^{e\prime}$:

$$F(\theta^{e\prime}) = F(\theta^e)\frac{(1+\nu)(1-\nu+\rho)(\rho-1)}{2\rho\nu(\rho-\nu)} - \frac{(1+\nu-\rho)(\nu-1)(\rho+1)}{2\rho\nu(\rho-\nu)}$$

Now it can be shown that $\rho > \rho^{\max}$ is sufficient condition for $F(\theta^{e\prime}) > F(\theta^e)$ to happen. I.e. if ρ is too high a new fraction of agents would prefer to take the outside option. This argument is repeated for any new θ^e, hence no agent enters.

2. Case $\rho < \nu$: From (9) it follows that entry type is defined by (again the monotonicity of *EU* implies that there is the unique cut-off):

$$F(\theta^e) = \frac{1+\nu-\rho}{2(\nu-\rho)} \tag{10}$$

It holds that $F(\theta^e) < 1$ if $\rho < \nu - 1 \equiv \rho^{\min}$ i.e participation is partial (for $F(\theta^e) \leq 1$ it is enough to find under which condition $\frac{1+\nu-\rho}{2(\nu-\rho)} \leq 1$). If ρ is too low it is not profitable to enter for a fraction of agents of *high* type ($\theta \in \left[\theta^e, \overline{\overline{\theta}}\right]$). It can be shown similarly to the case with $\rho > \nu$ that full unraveling occurs, but starting with the strongest type.

■

The intuition behind is when the lowest agent loses from participating in the market, and so he does not enter, the next to him agent does not have anymore any tiny chance to be the strong, and at most he can be only the weakest one. So, being now the lowest type, he strictly prefers not to enter either. And this process of taking outside option will continue till the moment where the very last, the strongest agent of the population remains alone, and so he does not enter either. The similar unraveling happens if the punishment is too low and the strongest agent has a negative utility from participating.

Now I shall summarize for all constellations of (ρ, ν) the results for the individual behaviour within a match and for participation decision of an agent of type θ:

Summary 5 *In the subgame of trading under the asymmetric punishment system there are following equilibria depending on (ρ, ν).*

- *if $\rho < \nu - 1$, the unique pure NE is $(\zeta_i(\{H\}, \overline{\theta}), \zeta_j(\{H\}, \underline{\theta})) = (0, 0)$, the rate of participation in the market is zero;*
- *if $\rho \in (\nu - 1, \nu)$ there is full participation with the unique equilibrium* $(\zeta_i(\{H\}, \overline{\theta}), \zeta_j(\{H\}, \underline{\theta})) = (1, 0)$;
- *if $\rho \in (\nu, \nu + 1)$ there is full participation, no pure strategy NE, the mixed strategies are defined in (4),(5);*
- *if $\rho > \nu + 1$ the rate of participation is zero, there is a NE in mixed strategies as in (4),(5)*

Corollary 6 *There is no ρ that would induce simultaneously both agents to play honestly, i.e. in this system of asymmetric application of the punishment ρ (> 0) the equilibrium $(\zeta_i(\{H\}, \overline{\theta}), \zeta_j(\{H\}, \underline{\theta})) = (1, 1)$ is unachievable.*

The optimal level of ρ, when at most it can be applied asymmetrically, should be set low to maximize the social welfare (although, counterintuitively, the cost of this would be a higher level of cheating behaviour in equilibrium):

Remark 7 *When the asymmetric punishment system is active, an increase of ρ within the range $[\nu - 1, \nu + 1]$ leads to an increase of the total share of agents who behave honestly. The aggregate (interim) social welfare however decreases.*

At $\rho = \nu + 1$ the lowest (i.e. the weakest) agent in the distribution of θ obtains the expected utility of contracting equal to zero (whereas the strongest agent is at his best possible utility over the range of $\rho \in [\nu - 1, \nu + 1]$).

The main conclusion this section comes to is that a system with asymmetric punishment, where only some agents deviations are punished, could be still beneficial individually and collectively as compared to a world when no punishment for misbehaviour is available.

2.3.3 Equilibrium Payoffs from Transactions under Symmetric Punishment

Analyzing the game in (3) one can see that the unique equilibrium when the punishment is imposed symmetrically is to play honest for both agents provided that $\rho > \nu$. The resulting expected utility from the matching under symmetric punishment is

$$EU^S(\theta) = 1 \tag{11}$$

for each θ. Consequently it is *the way* the punishment is applied that allows achieve an equilibrium in which *both* agents would behave honestly.

Note, there is a uniform gain for everyone from appearance of the symmetric punishment when the only alternative regime is to trade under no punishment at all.

2.4 Comparing the Payoffs under Different Punishment Modes

One can observe that switching to a regime where for all agents only the symmetric punishment system is available has distributional implications. When $\rho \notin [\rho^{\min}, \rho^{\max}]$ each agent, even the strongest one benefits from appearance of the symmetric punishment system. But when $\rho \in [\nu, \rho^{\max}]$ and there is a switch agents with relatively high θ incur a loss in utility whereas agents with low θ gain (this can be seen comparing (8) and (11)).

On aggregate, when $\rho \notin [\rho^{\min}, \rho^{\max}]$ and $\rho \in [\nu, \rho^{\max}]$, the SPS brings higher aggregated payoffs than the APS. In the remaining area of the parameters two systems are equivalent.

Consequently the social welfare boundary between the asymmetric and symmetric punishment systems is defined by the fixed cost K of the latter. The formal claim of this is in the proposition.

Proposition 8 *There exists a threshold level of the cost $\overline{K}$ such that if the actual cost of the SPS is above $\overline{K}$, the aggregate welfare is higher when this system is not installed.*

Proof. Consider the following constellations in (ρ, ν) plane

1. $\nu < \rho < \nu + 1$

   The aggregate interim social welfare under asymmetric punishment is (for any $F(\theta)$) $SW^A = \frac{\rho(\rho-\nu)+\nu^2-1}{(\rho^2-1)} > 0$. The aggregate social welfare under symmetric punishment is $SW^A = 1$.

When $\nu < \rho < \nu + 1$ it holds that $\frac{\rho(\rho-\nu)+\nu^2-1}{(\rho^2-1)} < 1$. Hence the highest cost $\overline{K}$ is defined from $SW^S - K \geq SW^A$. It is equal $\overline{K} = \frac{(\rho-\nu)\nu}{(\rho^2-1)}$, $0 < \frac{(\rho-\nu)\nu}{(\rho^2-1)} < 1$.

2. $\nu - 1 < \rho < \nu$

   Here, under asymmetric punishment, the interim social welfare is equal to 1 too. It is never efficient to install the symmetric punishment system from interim perspective[13]. (But ex post efficient threshold of $\overline{K}$ in this case is $\overline{K} = \frac{1}{2}$).

3. $\rho \notin [\nu - 1, \nu + 1]$

   The asymmetric punishment is not active. Hence the social welfare under absent symmetric punishment is $SW^0 = 0$. The threshold level is then $\overline{K} = SW^S - SW^0 = 1$ (both interim and ex post).

■

This proposition says that when $\rho \notin \left[\rho^{\min}, \rho^{\max}\right]$ and asymmetric governance is valueless, as long as $K \leq 1$ it is beneficial to install the symmetric punishment. When $\rho \in [\nu, \rho^{\max}]$ as long as $K \leq \frac{(\rho-\nu)\nu}{(\rho^2-1)}$ it is beneficial to install the symmetric punishment. Otherwise it is better on aggregate to remain with no or with the asymmetric punishment system only.

Proposition 9 *In any constellation of (ρ, ν), as long as $K < \overline{K}$ the social welfare is lower when two punishment modes, the asymmetric and symmetric one, coexist in equilibrium as compared to the social welfare when only the symmetric punishment regime is available.*

The proof is rather straightforward and omitted (a linear combination of payoffs from APS and SPS is inferior to the payoffs available when everyone is under SPS).

3 Contribution Games Under Incomplete Information About Types

Given the results of the above section, one could analyze which game is the best suited to aggregate private preferences for the costly symmetric punishment system, given that the type of agent θ is private information at the stage of contribution game (i.e. before the matching uncertainty is realized). I will search for a fully decentralized game, which additionally to aggregating preferences would simultaneously collect resources for the fixed cost K of the SPS[14].

First of all, recall that agents never use the SPS in equilibrium when it is available effectively. Hence it is only a threat, and is non-rival. Next, it is technically feasible to exclude individuals from any enforcement of their agreements and contracts. Given this and the fact that there is a fixed cost K to be covered, the symmetric punishment system is equivalent to an excludable public good. That is why I may borrow from the existing mechanism design literature some results on feasibility of provision of a public good in a large population of agents.

[13] But recall that the lower ρ is, the higher is the equilibirum rate of cheating behaviour.

[14] Note that the fixed cost assumption can be endogenised. In particular it can be shown that there exists a bargaining game and its equilibrium, such that each agent who is strong within a match *ex post* is able to convince the enforcing agent to rule in his favour the contract dispute at no cost to him, if there are competing enforcers.As the result, at ex ante stage, in order to counteract the aggregate advantages of all agents who are strong at the ex post stage, the population of agents have to pledge to the enforcement agents the whole surplus available to the strong agents ex post. This means that the cost of the symmetric punishment system is actually equal to $K = \frac{1}{2}(EU^A(\overline{\theta}) - 1)$. Whether $K \lesseqgtr \overline{K}$ depends on the parameters ρ and ν.

3.1 General Result for Provision of a Public Good under Incomplete Information[15]

The general result states that in quasilinear environments there is no way to construct a game within a large population of agents where in equilibrium each agent would contribute voluntary according to his valuation. That is a Pareto-improving, multilateral contract with side-transfers including compensations that would be accepted voluntary is not implementable in the second best world.

Lemma 10 *If agent's type θ is private information, in an economy with a large number of agents, there is no way to install the symmetric punishment via a contribution game with type dependent payments and achieve the first best allocation.*

Proof. See proof to the proposition 2.3 in Hellwig (2007). ■

The lemma implies that there is no way to withdraw individual information from the agents by proposing a menu of incentive compatible, individually rational and hence type-dependent contributions bound to differentiated probabilities of provision of a public good.

The intuition for this result is that in a large economy, no agent is pivotal. It is known that in a quasilinear[16] environment there is a unique mechanism that could collect type-dependent, incentive compatible contributions. This is Vickrey-Clarke-Groves mechanism (VCG)[17] which, essentially, imposes a tax on every agent for his either report about his type. The tax is constructed in a way that it makes an agent i pay to the rest of the group an amount of money equal to the sum of changes in their utilities, with changes coming from the fact that there is a change in the outcome once an agent i has reported his preferences. This makes an agent i internalize fully the impact of his report on the rest of the group and hence to be truthful.

However as the number of agents grows, the probability that any agent's report about his preferences over a public good, however big or however small it is, would have any impact on the probability of provision of a public good is next to zero. Hence, the tax cannot vary much either, nor can it, as a result, provide with the incentives to tell the truth.

For a problem with non-excludable public goods it means that the probability of their voluntary provision in a large population is zero[18]. For a problem with excludable public goods it means that the only feasible contribution is a type-independent constant payment taking the form of an admission ticket. Hence individual exclusion should be applied, even if it is inefficient given that the public good is on place[19].

[15] Because utility is transferable across agents and installation of a system imposing symmetric punishment increases the aggregate welfare when K is not too high, ideally under complete information about agents' types, there exists a vector of side-transfers among agents such that those who lose from introduction of the symmetric punishment system are compensated. A social contract having this outcome could be implemented via some multilateral bargaining procedure. But such payment scheme is obviously not viable when information about every individual's type θ is unobservable.

[16] And so it is in a linear environement, as here.

[17] The uniqueness of VCG for quasilinear environments was established by Green and Laffont (1979).

[18] More details on this result can be found in Mailath and Postlewaite (1990).

[19] There is an additional remark that should be made about feasibility of a voluntary contribution scheme. Here I assume the cost K is actually proportional to the size of population (which is an infinite number of agents of mass 1). Otherwise there would be a scope for an incentive compatible provision of a public good via a voluntary contributions scheme. As it is shown in Hellwig (2003), in a large population of agents, if the cost of a public good is negligible compared to the size of the economy, the public good can be provided with a non-zero probability. The intuition for this is that in such environment a per capita contribution rate goes to zero and so agents can contribute indifferently the requested ε.

3.2 An Admission Fee Mechanism for Provision of the Symmetric Punishment System

Given the above results, and given that I am interested in a decentralized mechanism where agents contribute voluntary, in the following I am studying the admission fee game.

The admission fee mechanism is a direct mechanism, only agents payoffs types are needed to make a (constrained efficient) decision over individual allocations. It stipulates the probability of provision of a public good, individual payments and probability of individual admission to the public good given a profile of equilibrium reports about valuation types (see for example Hellwig (2007) or Norman (2004)).

The formal definition of the admission fee mechanism for provision of the symmetric punishment system is:

Definition 11 *The admission fee mechanism to collect resources to cover K is $\Gamma = (\Theta, g(\cdot))$ where $\Theta = \prod_{i \in I} \Theta_i$ is a message space of all agents and $g(\cdot) = (y, t, \alpha)$ is the vector of outcome functions:*

$y : \Theta \to [0,1]$ *probability of emergence of the symmetric punishment system given the profile of agents messages;*

$\mathbf{t} : \Theta \to \Re_+$ *vector of individual contribution functions;*

$\mathbf{e} : \Theta \to \{n, p, f\}$ *exclusion rule (qualitative variable);*

$\boldsymbol{\alpha} : \Theta \to [0,1] \times [0,1]$ *vector of individual admission probabilities for a given exclusion rule.*

By the very nature of the punishment system agents can be excluded via two channels. An agent can be excluded from a possibility *to punish* his encounter (if the latter cheats him) and he can be excluded from a possibility *to be punished* by an encounter whom he has cheated. Consequently, instead of stipulating a single valued probability of admission to the symmetric punishment for given agent's report, the mechanism should stipulate the probability of admission to both possibilities, i.e. to punish and to be punished.

As the result, an agent i given his report about his type $\widetilde{\theta}_i$ could be excluded in three possible ways:

1. where he is admitted to a neither option (i.e. he cannot punish and cannot be punished). This is "full" exclusion;

2. where he can punish the encounter's misbehaviour but cannot be punished for his own misbehaviour towards an encounter (who is called not admitted); symmetrically where he cannot punish his encounter (who is called admitted) but can be punished by an encounter (who is then called admitted). This is "partial" exclusion.

3. where an agent who is not admitted to the symmetric punishment can still punish and be punished by his encounter if the latter is admitted. This is "quasi-no" exclusion.

In more details the rules have following clauses. The full exclusion rule:

(1) when i is admitted to the symmetric punishment and he meets j who is not admitted to the symmetric punishment, neither agent can punish misbehaviour of his encounter;

(2) when i is admitted to the symmetric punishment and he meets j who is admitted as well, either agent is punished in case he misbehaves;

(3) when i is not admitted to the symmetric punishment and he meets j who is admitted, neither agent is punished for any misbehaviour;

(4) when i is not admitted to the symmetric punishment and he meets j who is not admitted to the symmetric punishment, both are to deal within either asymmetric punishment system or none.

By contrast, the partial exclusion rule consists of:

(1) when i is admitted to the symmetric punishment and he meets j who is not admitted to the symmetric punishment, i can punish misbehaviour of j but agent j cannot punish misbehaviour of i if it occurs;

(2) when i is admitted to the symmetric punishment and he meets j who is admitted as well, both agents are punished in case when misbehaviour of either occurs;

(3) when i is not admitted to the symmetric punishment and he meets j who is admitted, i is punished for his misbehaviour, but agent j, if he misbehaves towards i is not punished;

(4) when i is not admitted to the symmetric punishment and he meets j who is not admitted to the symmetric punishment, both agents are to deal within either asymmetric punishment system or none.

The "quasi-none" exclusion rule:

(1) *when i is admitted to the symmetric punishment and he meets j who is not admitted to the symmetric punishment, i can punish misbehaviour of j and j can punish misbehaviour of i if it occurs;*

(2) when i is admitted to the symmetric punishment and he meets j who is admitted as well, both agents are punished in case when misbehaviour of either occurs;

(3) when i is not admitted to the symmetric punishment and he meets j who is admitted, either agent is punished in case he misbehaves;

(4) when i is not admitted to the symmetric punishment and he meets j who is not admitted to the symmetric punishment, both are to deal within either asymmetric punishment system or none.

Thus, for a given exclusion rule, the admission probabilities vector for each individual i contains two values $\alpha_i = (\alpha_i^p(\cdot), \alpha_i^{bp}(\cdot))$ with α_i^p being the probability that an agent i can punish an agent j and α_i^{bp} is the probability that an agent i can be punished by an agent j. For a match of two agents obviously the following equalities should hold $\alpha_i^p = \alpha_j^{bp}$ and $\alpha_i^{bp} = \alpha_j^p$.

As the individual types are unobservable, by the revelation principle, an allocation prescribed by the mechanism should be incentive compatible. An allocation $(y(\theta(\cdot)), s(\theta_i(\cdot)), t(\theta_i(\cdot)))$ achievable in the Bayesian Nash Equilibrium of the direct mechanism $\Gamma(\cdot)$ is incentive compatible if for each agent i the following holds:

$$EU_i^P(\widetilde{\theta}_i, \widetilde{\theta}^*_{-i}; \theta_i) - t^*(\widetilde{\theta}_i, \widetilde{\theta}^*_{-i}; \theta_i) \geq EU^P(\widetilde{\theta}'_i, \widetilde{\theta}^*_{-i}; \theta_i) - t^*(\widetilde{\theta}'_i, \widetilde{\theta}^*_{-i}; \theta_i) \tag{12}$$

for any profile of reports $\widetilde{\theta}^*$, types $\theta_i, \theta_i \neq \theta'_i$ and θ^*_{-i}; and under the realized punishment $P \in \{0, A, S\}$.

Given that each agent is free not to enter into contracting at all under any governance mode, an allocation induced by the provision mechanism is required to satisfy a contracting participation constraint:

$$EU_i^P(\widetilde{\theta}_i, \widetilde{\theta}^*_{-i}; \theta_i) - t(\widetilde{\theta}_i, \widetilde{\theta}^*_{-i}; \theta_i) \geq 0 \tag{13}$$

under the realized punishment $P \in \{0, A, S\}$.

Finally, the allocations should be compatible with the following budget constraint:

$$\int_{\Theta^t=\{\theta:t(\theta)>0\}} t(\theta)dF(\theta) \geq K \tag{14}$$

Timing of the model:

1. Nature distributes θ among agents, every agent observes his θ_i;
2. Agents decide on their most preferred punishment mode out of $P \in \{0, A, S\}$ given K, θ, ρ and ν and expected payoffs from entry into the market under two different punishment modes;
3. A mechanism to install the symmetric punishment stipulating a set of admissible strategies of all agents and corresponding outcomes is proposed[20];
4. An agent i accepts or not the mechanism;
5. Those who accept the mechanism, play it, available punishment systems become known, agents make their transaction under the most preferred available punishment mode and obtain the final payoffs.

3.3 Results of the Admission Fee Mechanism

In the following, first of all, I will characterize what is "the best" achievable equilibrium of the admission fee game depending on the applied exclusion rule. The "bestness" is defined by two criteria: 1. Whether the budget constraint can be met by individually rational contributions; 2. What is the highest achievable rate of subscription for the SPS, as by Proposition 9, once the SPS is on place, the higher is rate of subscription, the higher is the social welfare.

Afterwards I shall provide with the results about the entire set of possible equilibria of the admission fee game for a given exclusion rule and I will evaluate the boundaries of efficiency of different equilibria.

Additional remark is, I will provide with the results of the admission fee mechanism for two different strategic situations. First, agents decide on contribution when an outside opportunity is to trade under the APS. In the second they consider contribution strategies when the outside option is no-punishment regime. The aim is to find out how the availability of the APS to agents may hinder or may help with installation of the SPS.

3.3.1 Contribution strategies when the outside option is to trade under the APS

The valuation for the regime where everyone trades under symmetric punishment[21] as compared to the regime where only the asymmetric system is available, is monotonic in agents strength type.

[20] The mechanism designer in this case may be the agents themselves.

[21] To be precise on how is measured "valuation" for the symmetric punishment system $v(\theta, y) = EU^S(\theta) - EU^P(\theta)$ for $P \in \{0, A\}$, where $EU^P(\cdot)$ defines expected utility from the match under the alternative punishment system (asymmetric one or none). From inspection of (8) one can see that the valuation depends on whether the asymmetric punishment system is on place and how many other agents are admitted to it, as it depends on probability to be strong or weak $F(\theta)$ which in turn depends on decision of the others to be under asymmetric punishment system. For $\nu < \rho < \nu + 1$ if the asymmetric punishment system is active the higher is θ the lower would be $v(\theta)$. When there is no asymmetric punishment, individual valuation for the symmetric one is high for all agents, independently of θ and equal to 1 $v(\theta) = EU^S(\theta) - EU^P(\theta) = 1 - 0 = 1$)

In particular when $\nu < \rho < \nu + 1$ it is decreasing in θ; when $\nu - 1 < \rho < \nu$ it is increasing in θ. Consequently, one could simplify the analysis of the optimal mechanism. If $\nu < \rho < \nu + 1$, it would be natural to search for a threshold agent $\widehat{\theta}$ such that only agents below $\widehat{\theta}$ are admitted to the symmetric punishment and charged a user fee $t^* = \frac{K}{\int_{\underline{\theta}}^{\widehat{\theta}} dF(\theta)}$. If $\nu - 1 < \rho < \nu$ the analysis is actually irrelevant, as it is not efficient to try to install the SPS[22]. Hence, in the following, I will restrict analysis only to the constellation $\nu < \rho < \nu + 1$.

Before providing with the results, I introduce an additional piece of notation – $F(\widehat{\widetilde{\theta}})$. This term denotes the total number of agents who claim themselves in equilibrium to be of type $\theta < \widehat{\theta}$ for a given $\widehat{\theta}$. In a "non-truthful" equilibrium this number is different from the actual cross-sectional number of agents below[23] $\widehat{\theta}$, which is then denoted $F(\widehat{\theta})$.

The key result of the paper is presented in the following proposition:

Proposition 12 *The highest collectible amount of resources for the SPS and corresponding rate of subscription depending on the exclusion rule are as follows:*

- $t^* \int dF(\theta) = 1$ *and the rate of subsciption is 1 when the exclusion rule is full;*
- $t^* \int dF(\theta) = 1 - EU^A(\underline{\theta})$ *and the rate of subscription is 1 when the exclusion rule is partial;*
- $t^* \int dF(\theta) = \frac{1}{4}(1 - EU^A(\underline{\theta}))$ *and the rate of subscription is* $\frac{1}{2}$ *when the exclusion rule is quasi-no exclusion.*

Here $t^* \int dF(\theta)$ denotes equilibrium aggregate amount of contributions. Feasibility of the SPS is then defined whether actual cost K is below or above this amount.

Proof. The proof is simple. For each case I will consider incentives of one agents to subscribe for the SPS, given his type θ, when all the remaining agents subscribe (i.e. they claim valuations to be $\theta < \widehat{\theta}$). In particular I search what is the highest payment t that the remaining agent would agree to pay.

1. For the full exclusion rule this constraint is: $F(\widetilde{\widehat{\theta}}) \cdot 1 - t \geq 0$, hence when $F(\widetilde{\widehat{\theta}}) = 1$, $t^* \leq 1$. That is why $t^* \int dF(\theta) = 1$.

2. For the partial exclusion rule, similarly: $F(\widetilde{\widehat{\theta}}) \cdot 1 + (1 - F(\widetilde{\widehat{\theta}}))EU^A(\overline{\theta}) - t \geq EU^A(\underline{\theta})$, and so when $F(\widetilde{\widehat{\theta}}) = 1$ any $t^* \leq 1 - EU^A(\underline{\theta})$ is individually rational. The aggregate resources $t^* \int dF(\theta) = 1 - EU^A(\underline{\theta})$.

3. For the quasi-no exclusion rule the threshold maximizing collectible resources is at the intermediate values of $F(\widetilde{\widehat{\theta}})$. Namely, an agent indifferent between joining the SPS and staying outside is defined by $1 - t = F(\widetilde{\widehat{\theta}}) + (1 - F(\widetilde{\widehat{\theta}}))EU^A(\underline{\theta})$. This defines $F(\widetilde{\widehat{\theta}}) \leq \frac{1 - EU^A(\underline{\theta}) - t}{1 - EU^A(\underline{\theta})}$. The total amount $t \cdot \frac{1 - EU^A(\underline{\theta}) - t}{1 - EU^A(\underline{\theta})}$ is at maximum when $t = \frac{1 - EU^A(\underline{\theta})}{2}$. Substituting for $F(\widetilde{\widehat{\theta}})$ gives 1/2 participation rate result. ■

Hence, under two out of three exclusion rules, the admission fee game could have an equilibrium where the SPS is provided with probability one and all agents, even strong, who prefer to live in a world where only the APS is available for everyone, contribute to its cost.

[22] However when punishmnet is too low - the rate of cheating increases. If one modelled disutilities because of ethical sides of cheating, the efficiency analysis would have changed here.

[23] This implies implicitly that when agents mispresent their types they do so "continuously", i.e. there is no "holes" in an interval of types who mispresent their preferences.

The reason for this is that the SPS treats those who has not contributed to its fixed costs either as "weaks", i.e. it does not enforce agent demand for punishment of those who are under the SPS (under the partial exclusion rule) or as "nothing", i.e. it does not enforce any type of a claim against these agents, whether to punish or to be punished (under the full exclusion rule). Those who sign up for protection by the SPS are stronger than agents with a very high θ who remain under the APS. So the previously strong agents have now reduced chances to obtain associated rents and this induces them to subscribe for the SPS as well. Obviously feasibility of exclusion of those who do not subscribe is crucial for the result.

When the rule is quasi-no exclusion, an equilibrium with the full rate of subscription is impossible. This is because if too many agents are under the SPS, the remaining agents match them with a high probability and obtain the full protection payoffs without contributing anything for the SPS. This dilutes their own incentives to contribute, and hence they free-ride.

3.3.2 Contribution strategies when the outside option is to trade in a regime with no punishment

When $\rho \notin [\nu - 1, \nu + 1]$ no one wishes to trade under the APS. Hence the valuation for the SPS is uniformly equal to 1, even for the strong agents as they cannot benefit from their strength anyhow.

Consequently, within the admission fee game agents could be asked a simple message, say from $\{0, 1\}$. If agent says 0 he is excluded. If he says 1 he is admitted and when matched to those who are excluded, their trades are governed according to the exclusion rule in effect.

The main results, depending on the exclusion rule, are as follows:

Proposition 13 *The highest collectible amount of resources for the SPS and corresponding rate of subscription depending on the exclusion rule are as follows:*

- $t^* \int dF(\theta) = 1$ *and the subscription rate is full when the exclusion rule is full;*
- $t^* \int dF(\theta) = \frac{EU^A(\overline{\theta})^2}{4(EU^A(\overline{\theta}) - EU^A(\underline{\theta}) - 1)}$ *and the subsciption rate is* $\min\left\{\frac{EU^A(\overline{\theta})}{2(EU^A(\overline{\theta}) - EU^A(\underline{\theta}) - 1)}, 1\right\}$ *if the exclusion rule is partial;*
- $t^* \int dF(\theta) = \frac{1}{4}$ *and the subscription rate is* $\frac{1}{2}$ *when the exclusion rule is quasi-no exclusion.*

Proof. Suppose there is a subset of agents $(-i)$, call it S^*_{-i} of a mass $F(S^*_{-i}) \in [0, 1]$, who subscribe for the services of the symmetric punishment system.

Under the full exclusion rule, agent $i's$ best reply would be to subscribe as well if $1 - t^* > 0$. Moreover as long as $t^* < 1$ subscription is a (weak) dominant strategy. Hence for any $K \leq 1$ the symmetric punishment is provided with probability one.

Under the partial exclusion rule, for a given t agent i is indifferent between subscribing and not when the fraction $F(S^*_{-i})$,defined below, subscribes[24]:

$$F(S^*_{-i}) \cdot 1 + (1 - F(S^*_{-i})) \cdot EU^A(\overline{\theta}) - t \geq F(S^*_{-i}) \cdot EU^A(\underline{\theta}) \qquad (15)$$

This gives $F(S^*_{-i}) \leq \frac{EU^A(\overline{\theta}) - t}{EU^A(\underline{\theta}) + EU^A(\overline{\theta}) - 1}$ and the aggregate "incentive compatible" resources $t \cdot (\frac{EU^A(\overline{\theta}) - t}{EU^A(\underline{\theta}) + EU^A(\overline{\theta}) - 1})$. Maximizing this with respect to t and substituting back brings the equilibrium $F(S^*_{-i})$ at which is the collected resources are at maximum.

[24] One should not be surprised by finding utilites of the strong and the weak. The partial exclusion by SPS has exactly the same strategic impact on agent's behaviour in mixed matches (i.e. of those who are admitted and those who are not) as APS has for strongs and weaks.

Under the quasi-no exclusion rule, an agent indifferent between two regimes is $1 - t \geq F(S^{*}_{-i}) \cdot 1$. Proceeding as in the case with the partial exclusion one obtains the results. ■

Hence, when no one wishes to enter into the market and trade under the APS, the admission fee game bring different results. In particular, under the partial exclusion rule they may be no longer the full rate of participation. Under the partial exclusion rule agents have incentives to free-ride on those who have subscribed for the symmetric punishment. As under the partial exclusion rule they are not excluded from being punished and hence they obtain a positive utility for free. Under the quasi-no exclusion rule agents have incentives to free-ride, and even to a larger extent. By contrast, the full exclusion guarantees the first best efficiency result because joining the SPS is a dominant strategy for any agent.

Note when the APS is too inefficient and the market breaks down, agents willingness to pay for the SPS increases (this one can see when comparing the highest collectible amount of contributions). But the most important conclusion of this subsection is that agents could pay "against their will" for the symmetric punishment if those with whom they trade do pay and protected by the latter.

The following section demonstrates in details necessary element for the above result, namely, agents should share right beliefs about others' behaviour.

3.4 Efficiency of the Admission Fee Game

In this subsection I study whether there are other equilibria, in addition to the full subscription equilibrium of the previous section. For these equilibria I evaluate the corresponding aggregate welfare.

The first departure from efficiency happens when there are equilibria where two punishment modes co-exist. As according to the result of the proposition 9, the social welfare is not at the optimum. The second departure from efficiency appears, according to the proposition 8, if the SPS is on place while its cost is above the benchmark $\overline{K}$. This subsection studies whether either type of inefficiency could arise in equilibria of the admission fee game.

Results for the partial exclusion rule:

It happens that under the partial exclusion rule, for each stipulated admission threshold $\widehat{\theta}^{*}$ there could arise a continuum of subscription equilibria. The result depends on beliefs agents share about the subscription behaviour in the population.

Proposition 14 *In the admission fee game based on the partial exclusion rule, for a given* $K \in (0, 1 - EU(\underline{\theta}))$ *each threshold* $\widehat{\theta}$ *induces three possible types of equilibria:*

(1) a truthful equilibrium where for a given threshold $\widehat{\theta}$ *agents report truthfully their types* θ*, all agents* $\theta < \widehat{\theta}$ *are admitted to the symmetric punishment and all agents* $\theta > \widehat{\theta}$ *remain under asymmetric;*

(2) a continuum of non-truthful equilibria where for a given $\widehat{\theta}$*, the actual rate of subscription is defined by some* $\widetilde{\widehat{\theta}} \in \left[\widehat{\theta}, \overline{\theta}\right]$ *.*

(3) a fully pooling equilibrium where all agents report their types to be $\theta > \widehat{\theta}$*, i.e. the subscription rate is zero.*

Proof. The multiplicity of equilibria is due to the fact that agents beliefs about the actual rate of participation determine their own subscription strategies. Depending on what is the common belief about the subscription rate, for a given $\widehat{\theta}$, there may be different actual subscription rates, including the rate of subscription meant by the mechanism, when $\widehat{\theta}$ was chosen.

Firstly I shall prove the possibility of the truthful equilibrium. Suppose all agents report their types truthfully. Then the optimal $\widehat{\theta}$ for a given level of the cost K is defined by IC condition

$$1 \cdot F(\widehat{\theta}) + (1 - F(\widehat{\theta})) \cdot EU(\overline{\theta}) - t \geq EU(\underline{\theta}). \tag{16}$$

merged to the budget balance requirement $t \cdot F(\widehat{\theta}) = K$:

$$F(\widehat{\theta})^2(1 - EU(\overline{\theta})) + F(\widehat{\theta}))(EU(\overline{\theta}) - EU(\underline{\theta})) = K. \tag{17}$$

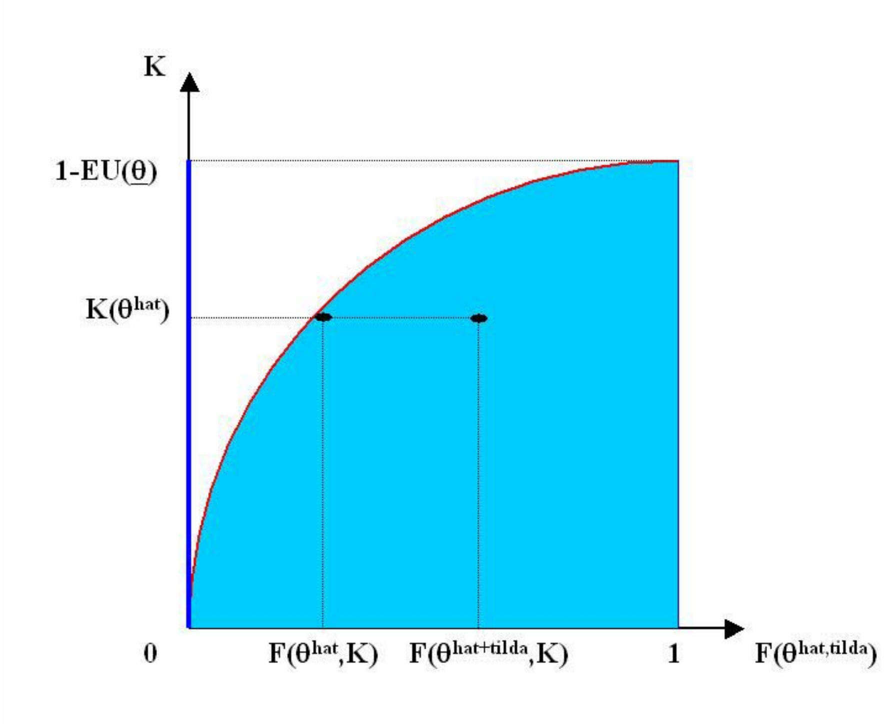
It is easy to check that indeed agents with $\theta > \widehat{\theta}$ state truthfully their types and remain under the APS.

The second equilibrium comes from the observation that agents beliefs are free and so if every agent expects (by whichever reason) that there is a subgroup of agents, say an interval $\left[\widehat{\theta}, \widetilde{\widehat{\theta}}\right]$ who report their types to be $\theta < \widehat{\theta}$, any isolated agent with $\theta \in \left[\widehat{\theta}, \widetilde{\widehat{\theta}}\right]$ is better off from reporting his type to be $\theta < \widehat{\theta}$ rather than saying the truth, as:

$1 \cdot F(\widetilde{\widehat{\theta}})^2 + F(\widetilde{\widehat{\theta}}) \cdot (1 - F(\widetilde{\widehat{\theta}})) \cdot EU(\overline{\theta}) - K > F(\widetilde{\widehat{\theta}})EU(\underline{\theta})$ when $F(\widetilde{\widehat{\theta}}) > F(\widehat{\theta})$ and keeping the cost K constant[25]. Consequently even totally uninformative equilibrium with $F(\widetilde{\theta}) = 1$ can occur for any given $\widehat{\theta}$.

The third equilibrium arises when i believes all other agents report a low valuation for the symmetric punishment, it is not installed and hence i does not strictly gain from reporting any positive valuation (= self-fulfilling). ∎

The results of the proposition are illustrated in the following Figure:



[25] This is because the LHS $F(\widehat{\theta})^2(1 - EU(\overline{\theta})) + F(\widehat{\theta})(EU(\overline{\theta}) - EU(\underline{\theta})) \equiv K$ is increasing in $F(\widehat{\theta})$. Hence when K remains fixed but $F(\widehat{\theta})$ increases, the LHC of the participation constraint remains valid for a given $F(\widehat{\theta})$.

The blue area on the picture is for the set of non-truthful subscription equilibria for a given level of the cost K, $y^* = 1$; the red line is the truthful equilibrium for a given cost K, $y^* = 1$, in the white area $y^* = 0$.

This picture essentially means the following. Take any cost $K < 1 - EU^A(\underline{\theta})$ and stipulate an incentive compatible admission threshold $\widehat{\theta}^*$, then the number of agents who in equilibrium could claim high valuation ($\theta < \widehat{\theta}^*$) ranges from $F(\widetilde{\widehat{\theta}}) = F(\widehat{\theta}^*)$ till $F(\widetilde{\widehat{\theta}}) = 1$.

Subscription behaviour that would arise as an equilibrium for a given threshold $\widehat{\theta}$ depends on the (common) beliefs agents share. If all agents believe that all believe that a fraction $F(\widehat{\widetilde{\widehat{\theta}}}) > F(\widehat{\theta}^*)$ would claim high valuation and subscribe for symmetric governance, for each single agent with $\widehat{\widetilde{\widehat{\theta}}} > \theta > \widehat{\theta}^*$ it is profitable to subscribe for the symmetric punishment (provided $F(\widehat{\widetilde{\widehat{\theta}}}) \cdot t(\widehat{\widetilde{\widehat{\theta}}}) \leq K$, i.e. the symmetric punishment system is to be in place). The opposite holds as well, if an agent from $\left[\widehat{\theta}^*, \widehat{\widetilde{\widehat{\theta}}}\right]$ believes that everyone else in $\left[\widehat{\theta}^*, \widehat{\widetilde{\widehat{\theta}}}\right]$ remains to trade under the APS, for him it is optimal not to join the SPS as well.

This strategic pattern leads to overprovision inefficiency:

Proposition 15 *There exists a level of cost $\widehat{K}^{PE}$ with $\widehat{K}^{PE} > \overline{K}$, such that under the partial exclusion rule, for $K \in \left[\overline{K}, \widehat{K}^{PE}\right]$ the symmetric punishment is installed, even though it is welfare decreasing.*

The proof is rather straightforward. Consider an equilibrium where the entire population subscribes (for a given cost K inefficiency even higher when a fraction of agents remain under the APS while the costly SPS is on place). In equilibrium where everyone subscribes agents are insensitive to the actual level of contribution t^* provided $t^* \leq 1 - EU^A(\underline{\theta})$ as everyone's behaviour depends only on beliefs. Then, even if the actual cost is $K = 1 - EU^A(\underline{\theta}) \equiv \widehat{K}^{PE}$ agent i subscribes and pays $t^* = 1 - EU^A(\underline{\theta})$ if he believes all others are under protection of the SPS. The highest potential efficiency loss due to over-provision is then equal to $\widehat{K}^{PE} - \overline{K} = \frac{\nu\rho(\rho-\nu)}{\rho^2-1} > 0$ as $\rho > \nu$.

The over-provision result may be interpreted also as a feasibility of extraction of agents' surplus from trade by a punishment system, provided it could threat to exclude those who do not wish to give up the surplus.

To conclude, the decentralized play of the admission fee game could bring about both types of inefficiencies: two modes may be in demand simultaneously in equilibrium by different groups of agents and the SPS may be installed when it is too costly or not installed when it is efficient to have it.

Results for full exclusion rule:

The full exclusion rule as well has a continuum of subscription equilibria.

Proposition 16 *In the admission fee game based on the full exclusion rule, for a given $K \in (0, 1)$, each threshold $\widehat{\theta}$ induces three possible types of equilibria:*

(1) a truthful equilibrium where for a given threshold $\widehat{\theta}$ agents report truthfully their type θ, all agents $\theta < \widehat{\theta}$ are admitted to the symmetric punishment and all agents $\theta > \widehat{\theta}$ remain under asymmetric;

(2) a continuum of non-truthful equilibria where for a given $\widehat{\theta}$, the actual rate of subscription is defined by some $\widetilde{\widehat{\theta}} \in \left[\widehat{\theta}, \overline{\theta}\right]$.

(3) a fully pooling equilibrium where all agents report their types to be $\theta > \widehat{\theta}$, i.e. the subscription rate is zero.

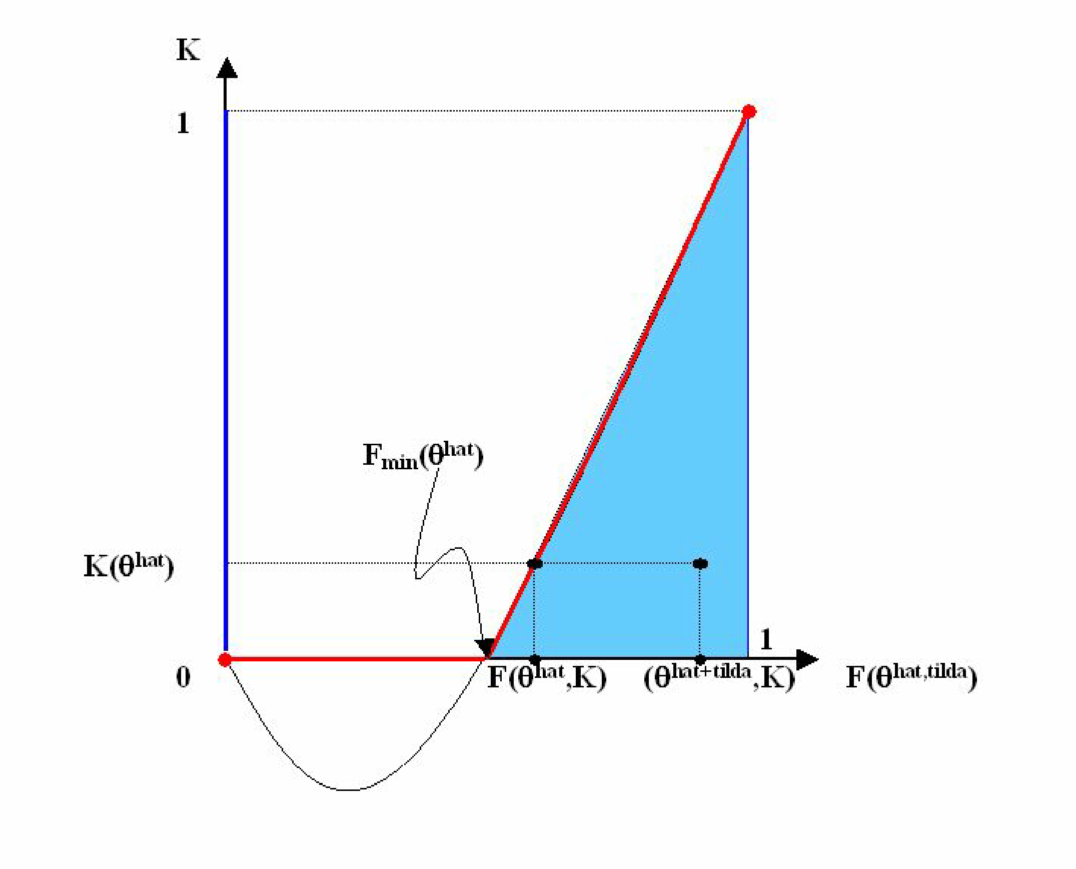
Proof. The proof is very similar to the case with partial exclusion rule. The truthful equilibrium is defined by equation (merging IR and BB requirement)

$$(1 - EU(\overline{\theta})) \cdot F(\widehat{\theta})^2 + EU(\underline{\theta}) \cdot F(\widehat{\theta}) - K = 0. \tag{18}$$

The only difference is that for y to be positive, for a relatively high K, there should exist a common belief $F(\widehat{\theta})$ that satisfies this equation.

The proof of untruthful equilibria, i.e. that it may happen $F(\widehat{\widetilde{\theta}}) \neq F(\widehat{\theta})$ quite the same in the case with partial exclusion. If all agent follow some untruthful equilibrium strategy, every remaining agent has incentives to play it as well, due to complementarity of participation decision. ■

The sets of equilibria under full exclusion rule for each K are depicted on the following figure:



The red line defines the truthful equilibrium for a given K. The blue area corresponds to the set of possible equilibria (defined by $F(\widehat{\widetilde{\theta}})$) for each level of the cost K, $F_{\min}(\widehat{\widetilde{\theta}})$ is the minimal critical mass of agents to whom it is profitable to subscribe (given agents believe that at least fraction $F_{\min}(\widehat{\widetilde{\theta}})$ subscribe).

Contrary to the previous case, with the partial exclusion rule, agents beliefs about the subsciption behaviour should be sufficiently optimistic. I.e. they should believe that there is a minimal number (”a critical mass ”) of agents who join the SPS, otherwise installation of the SPS may fail completely. The reason for this threshold is as follows. Under the full exclusion rule, if there are too few people who are under the SPS, it is with a high probability that an agent who joins the SPS meets the one who is under the APS. In this case, according to the full exclusion rule neither agent is imposed any punishment. Consequently the only payoffs that both agents can expect are zero. Anticipating this, if an agent believes too few of others subscribe, he prefers to remain under the APS which gives him a higher expected utility. The opposite holds as well. If many agents are to join the SPS, remaining agents would join it and would have even an increased willingness to pay (as compared to the case with the partial exclusion). That is why the highest collectible amount of resources is the higher under the full exclusion rule ($= 1$) as compared to the partial exclusion rule ($= 1 - EU^A(\underline{\theta})$).

One can observe as well, the co-existence of two modes in equilibrium is possible under the full exclusion rule too.

Similarly to the case with the partial exclusion rule the admission fee mechanism with the full exclusion rule can over-provide the public good, i.e. it may install it even when it is welfare decreasing.

Proposition 17 *There exists a level of cost $\widehat{K}^{FE}$ with $\widehat{K}^{FE} > \overline{K}$, such that under the partial exclusion rule, for $K \in \left[\overline{K}, \widehat{K}^{FE}\right]$ the symmetric punishment is installed, even though it is welfare decreasing.*

The logic of the proof is similar to the one in the proposition 15. The highest individually rational payment in the equilibrium where all agents subscribe for the symmetric punishment system services is $t^* = 1$. Hence if the actual cost is $K = 1 \equiv \widehat{K}^{FE}$ the symmetric punishment is provided, even if it is inefficient (from the proposition 8 it follows that $\widehat{K}^{FE} > \overline{K}$). The departure from efficiency is up to $\widehat{K}^{FE} - \overline{K} = 1 - \frac{(\rho-\nu)\nu}{(\rho^2-1)} > 0$.

Results for the quasi-no exclusion rule:

Proposition 18 *In the admission fee game based on the quasi-no exclusion rule, for each K such that $K \leq \frac{1}{4}(1 - EU^A(\underline{\theta}))$ there exist two associated thresholds $\widehat{\theta}$ resulting in truthful revelation. However both threshold imply co-existence of the SPS with the APS in equilibrium. When $K > \frac{1}{4}(1 - EU^A(\underline{\theta}))$, the only equilibrium is all agents report $\theta > \widehat{\theta}$ for any $\widehat{\theta}$ and so $y^* = 0$.*

Proof. An agent indifferent between joining the SPS and remaining under the APS is defined by the IC condition $1 - t = F(\widehat{\theta}) + (1 - F(\widehat{\theta}))EU^A(\underline{\theta})$.

Merging this with the budget balance requirement defines

$$F(\widehat{\theta})(1 - EU^A(\underline{\theta})) - F(\widehat{\theta})^2(1 - EU^A(\underline{\theta})) = K \tag{19}$$

The LHS is a concave function with a maximum at $F(\widehat{\theta}) = 1/2$. From here follows that the maximum collectible amount is $\frac{1}{4}(1-EU^A(\underline{\theta}))$ and that there are two possible participation equilibria for $K \leq \frac{1}{4}(1-EU^A(\underline{\theta}))$. Namely, (19) has two solutions for $F(\widehat{\theta})$; both belong to $(0, 1)$ and so define two incentive compatible participation thresholds. By proposition 9 I chose the highest (SW is increasing in the number of agents who are under the SPS).

When $K > \frac{1}{4}(1 - EU^A(\underline{\theta}))$ there is no such IC payment that once aggregate would match this level of the cost. ■

The intuition to a relative sparseness of the equilibrium set is that under quasi-no exclusion rule each agent's willingness to join the SPS is decreasing in the number of agents whom he believe to be under the SPS. This is because the higher is share of agents who are under the SPS the higher is the probability for i to match with them and obtain for free the utility equal to 1. As encounter's subscription to protection extends the full protection to agent i, agents are willing to free-ride on others subscriptions. At the same time, when others do not subscribe each agent prefer to join the SPS and obtain rather a sure utility of 1. These gives rise to only two possible subscription equilibria: with low rate of participation (and high individually rational t) and with high rate of the participation (and reduced t), and there are no equilibria with extreme participation rates as it was with other exclusion rules.

As the result, the quasi-no exclusion rule has an opposite welfare implications. Namely there is, in general, underprovision of the SPS.

Proposition 19 *Under the quasi-no-exclusion rule there exist a level of the cost* $K \in \left[\frac{1}{4}(1 - EU^A(\underline{\theta})), \overline{K}\right]$, *such that it is valuable to introduce the SPS but the admission fee game fails to do it.*

The proof is omitted, it is straightforward from the proof of the previous proposition.

Conclusions to the admission fee game with the quasi-no exclusion rule are 1. There is likely to be under-provision the SPS, i.e. agents fail to collect resources for the SPS even for intermediate levels of the cost K; 2. There is always a part of agents who remain under the APS even if the SPS is in place.

Hence, the decentralized play of the admission fee game could bring quite inefficient results and reduce the aggregate welfare as compared to the world where only the APS is available for everyone.

4 Some Extensions and Discussions

4.1 Results when there is a monopolistic provider of asymmetric punishment

Everywhere I assumed that the marginal costs of using the ASP were zero due to competition among providers. If there is a monopolistic provider who can make take-it-or-leave-it offers for the price of his services to resolve a dispute the results change a bit.

In particular, assume that the price that a monopolistic provider charges is ϕ. Then the payoff matrix from trading can be modified as follows:

$strong \backslash weak$	H	C
H	$1, 1$	$1 - \nu + \rho - \phi, 1 + \nu - \rho$
C	$1 + \nu, 1 - \nu$	$0, 0$

(20)

i.e. agents pay a monetary fee ("the individual marginal cost of asymmetric punishment") only when the punishment ρ is imposed after equilibrium plays and it is only when a weak agent cheats on the strong one. This would affect the equilibrium strategies in the following way (under assumption that $\rho > \nu > 1$):

$$\zeta_i(\{H\}, \underline{\theta}) = \frac{1 - \nu + \rho - \phi}{1 + \rho - \phi}; \zeta_i(\{C\}, \underline{\theta}) = \frac{\nu}{1 + \rho - \phi} \tag{21}$$

$$\zeta_i(\{H\}, \overline{\theta}) = \frac{\nu - 1}{\rho - 1}; \zeta_i(\{C\}, \overline{\theta}) = \frac{\rho - \nu}{\rho - 1}. \tag{22}$$

As the result the expected payoff from contracting under asymmetric punishment would be

$$EU^A(\theta) = F(\theta)\frac{2\rho\nu(\rho-\nu)+\nu\phi(\nu+1-2\rho)}{(\rho-1)(1+\rho-\phi)} + \frac{(1+\nu-\rho)(\nu-1)}{\rho-1} \tag{23}$$

It can be shown that the result of full or zero participation rate in trade under asymmetrically imposed punishment carries over into the case with non-zero marginal costs. However there appears a condition on the maximal price that asymmetric enforcer can collect from agents (given that their outside option is not to trade at all). In particular, it should hold simultaneously that $\rho < \nu + 1$ and $\phi < \rho + 1$ and $\phi < \frac{2\rho(\rho-\nu)}{2\rho-\nu-1}$.

When $\rho \in [\nu, \nu + 1]$ the third condition implies the second. Hence there is an upper boundary on the price that a monopolistic provider can charge $\overline{\phi} \equiv \frac{2\rho(\rho-\nu)}{2\rho-\nu-1}$.

In this case he extract the entire surplus of the strong agent and the utilities of any strong and any weak agent would be equal, i.e.

$$EU^A(\overline{\theta}) = EU^A(\underline{\theta}) = \frac{(1+\nu-\rho)(\nu-1)}{\rho-1}. \tag{24}$$

Hence at the ex ante stage all agents expect to have the same level of utility, independently of the type θ, $EU^A(\theta) = \frac{(1+\nu-\rho)(\nu-1)}{\rho-1}$. As the result all agents would value equally the symmetric punishment system.

The highest efficient level of the cost K of the symmetric punishment system would be $\overline{K} = 1 - EU^A(\underline{\theta}) = \nu\frac{\rho-\nu}{\rho-1}$. It can be seen immediately that the partial exclusion rule is the most (second-best) optimal exclusion rule. As the admission fee mechanism would always result in the full subscription if the admission fee is $t^* \leq 1 - EU^A(\underline{\theta})$. Moreover to subscribe is a dominant strategy for any agent, i.e. he would subscribe for the symmetric punishment independently of his type θ and of strategies of the others.

4.2 Discussion of implications for the legal system

One could derive some implications on the optimal dynamics of the exclusion rules applied by an official legal system in order to win over the shadow sector of an economy. Of course the discussion here takes a very simplistic view of how the legal systems and the black markets function in reality. But at the same time, simplicity allows to distill out the general idea.

First of all, the admission fee could be thought about as a tax that an entrepreneur has to pay for being a legal entity. By becoming legal, he would be entitled to enforce contracts via the official court system. An alternative for an entrepreneur is to produce his goods/services in a black market and to enforce agreements with a help of, say, not impartial private agents.

His decision to become legal is likely to depend on: the sizes of the legalized production sector and of the black market; what is his strength type for enforcement via the private enforcement system and how many other, weaker, agents are in the black market still; what are the details of enforcement of contracts between those who are in the legal sector and the ones who are in the black market (as he might match and trade before observing the fact that his partner is under protection by the official system).

The different exclusion rules could be thought about as follows.

The full exclusion rule would be representative for a legal system that considers a contract made with non-legalized entity to be void and not subject to any consideration. Given the beliefs pattern described in the previous section, from the one side the coordination failure may be very severe, i.e. the whole economy may stay within the black market. From the other side, if the legal sector is already large, applying the full exclusion rule may force the remaining agents to join and be willing to pay to become legal.

The partial exclusion rule would be representative for a legal system which imposes a very high fine on those who are on the black market and are not paying the tax, once they are discovered. Hence in case where it is the legal entrepreneur who cheats the one who is on the black market, the latter simply does not apply for any punishment to the official system in fear of fines (and so he is excluded); at the same time, when it is the black market entrepreneur who misbehaves towards the legalized one, the legal system's codes may be sufficiently rich in clauses allowing the legal entrepreneur to sue the black market entrepreneur (e.g. under some general civil rights).

The quasi-no exclusion rule would represent a legal system in which an agent who is a legal entity is both under full protection and under full liability for his behaviour within a contract, even with respect to those who are on the black market. Of course for the full liability of agents from the legal sector to be effective one need additionally that the agents who are on the black markets do not fear to apply for justice. This could be obtained by some leniency programs for those who were on the black market.

Then, one can see that in a dynamic perspective, for an official legal system in order to attract the whole population under its protection, it is better to start with the quasi-no exclusion rule. In this case the "weakest" agents are likely to be the first to join. Afterwards, once the weakest become legalized, it is efficient to start to apply partial exclusion rule, in order to attract the agents of intermediate types. As they are now "the weakest" ones on the black market, their valuation of the protection by the official system is likely to increase. Finally, when the legal sector has attracted those types and has become of sufficiently large size, it would be beneficial to use the full exclusion, as even the "strongest" types would find it finally profitable to pay taxes and trade under the efficient official system.

5 Conclusions

In the paper I have studied how a population of agents could install the common good – an efficient third party enforcement intermediary. It becomes available only once agents collect a necessary up-front amount of resources for its fixed cost. Otherwise agents could trade under an alternative enforcement regime where they are heterogenous in punishing ability. Anticipating the payoffs from trade under the asymmetric punishment systems agents decide on their optimal strategy within the contribution game.

The main results of the paper are:

- The asymmetric punishment system alone could provide to every agent, even the weakest one a positive level of utility. Moreover the rate of honest behaviour is higher as compared to a world with no punishment at all. The higher is the punishment the higher is the rate of honest behaviour in the population but payoffs of the weakest agents worsen proportionally. And because of agents' freedom not to trade at all, the punishment applied asymmetrically cannot be increased indefinitely. As the result there is always a non-zero level of cheating in the equilibrium under the APS.

  If the punishment is larger than the cheating stake, relatively strong agents prefer the world where only asymmetric punishment is available to everyone. But at the same time, under such parameters constellation the symmetric punishment system brings higher aggregated payoffs than the asymmetric one, provided the cost of the SPS is not too high.

- In a world where individual strength types are unobservable at the stage when the contribution game is played, the only way to collect resources for the fixed cost of the SPS is to apply exclusion to those who do not contribute.
- The admission fee game is, in general, successful in collecting resources for the SPS. Even the strong agents may contribute to the cost of the SPS, provided that the exclusion rules are sufficiently strong. Moreover the punishment system is able to extract agents entire surplus from trade by threatening with exclusion.

  If the exclusion rule is by contrast too mild, i.e. it is quasi-no exclusion rule, there is likely to be under-provision of the SPS. Agents may fail to collect resources to cover the fixed cost even in states when it is only moderately high and it is efficient to install the SPS.
- However there are as well equilibria where relatively strong agents remain under the APS and relatively weak join the SPS. That is two modes could co-exist. For this to happen it is necessary that agents hold a commonly known belief that a fraction of others remain under the APS. Beliefs about subscription behaviour within the population are crucial to determine the equilibrium subscription rate.
- When there is a monopolistic provider of the asymmetric punishment who charges a fee and extracts the entire ex post surplus of the strong agents, the admission fee game is more likely to have an equilibrium where all agents subscribe for the SPS and the demand for services of the APS is equal to zero. This is because the strong agents do not have rents under the monopolistic APS. But, again the exclusion rule should be either partial or full, but not the mild rule of quasi-no exclusion.

The more general result this paper comes to is the admission fee game as a decentralized mechanism need not perfectly aggregates individual preferences for a punishment mode. Together with a very efficient subscription equilibrium there are other inefficient ones that may arise.

Further research should be devoted to endogenize the fixed cost K. One possible way would be explicit modelling of the agents who provide the punishment services and their incentives (see footnote 15). There are of course other ingredients of the cost K. If the SPS is the official legal system, the design of up-to-date codes is likely to be costly. The paper by Gennaioli and Shleifer (2007) given an example of an argument about costly codes. The common law system produces for free, by precedent, efficient rules and codes, (even when the judges are biased, the inefficiencies are aggregated away). By contrast, the centralized civil law system, at least in theory, has to put efforts to design the codes explicitly in advance. This part, distinguishing the origins of the rules and who contribute to their efficiency, could be modeled as well.

Of course studying explicitly the composition of the cost K may allow to consider other, more optimal mechanisms for emergence of the SPS. Nevertheless this paper hopefully still clarifies some matters underlying the emergence of an ideal governance system and potential ways of modelling this problem.

References

Dixit, A. (2003a): “On Modes of Economic Governance,” *Econometrica*, 71, 449–481.

——— (2003b): “Trade expansion and contract enforcement,” *Journal of Political Economy*, 111, 1293–1317.

——— (2004): *Lawlessness and Economics: Alternatives Modes of Governance.* Princeton University Press, Princeton.

Enriques, L. (2002): “Do Corporate Law Judges Matter? Some Evidence from Milan,” *European Business Organization Law Review*, 3, 765–821.

Galanter, M. (1978): “Why the ’Haves’ Come Out Ahead: Speculations on the Limits of Legal Change,” *Law and Society Review*, 9, 95–160.

Hellwig, M. (2003): “Public-Good Provision with Many Participants,” *Review of Economic Studies*, 70, 589–614.

——— (2007): “The Provision and Pricing of Excludable Public Goods: Ramsey-Boiteux Pricing versus Bundling,” *Journal of Public Economics*, 91, 511–540.

Hill, P. (2003): “Heisei Yakuza: Burst Bubble and Botaiho,” *Social Science Japan Jounral*, 6, 1–18.

——— (2006): “The Japanese Mafia, Take Two: Postscript to the Paperback Edition,” *Working Paper, Department of Sociology, University of Oxford.*

Hobbes, T. (1651): *Leviathan.* available at http://oregonstate.edu/instruct/phl302/texts/hobbes/leviathan-contents.html.

Lempert, R. (1999): “A classic at 25: Reflections on Galanter’s ”Haves” article and works it has inspired,” *Law and Society Review*, 38.

Li, J. (2000): “The benefits and costs of relation-based governance: and explanation of the East-Asian miracle and crisis,” *Working Paper City University of Hong Kong.*

Mailath, G., and A. Postlewaite (1990): “Asymmetric information bargaining problems with many agents,” *Review of Economic Studies*, 57, 351–367.

Norman, P. (2004): “Efficient Mechanisms for Public Goods with Use Exclusions,” *Review of Economic Studies*, 71, 1163–88.

Rousseau, J.-J. (1762): *Du Contrat Social ou Principes du Droit Politique.* available at http://fr.wikisource.org/wiki/Du$_{c}ontrat_{s}ocial$.

Sheehan, R., and D. Songer (1992): “Who Wins on Appeal? Upperdogs and Underdogs in the United States Courts of Appeals,” *American Journal of Political Science*, 36, 235–258.