Edinburgh School of Economics

Discussion Paper Series

Number 148

The Emergence of Institutions

Santiago Sanchez-Pages (University of Edinburgh)
Stephane Straub (University of Edinburgh)

Date
December 2006

Published by
School of Economics
University of Edinburgh
30 -31 Buccleuch Place
Edinburgh EH8 9JT
+44 (0)131 650 8361
http://www.ed.ac.uk/schools-departments/economics

THE UNIVERSITY *of* EDINBURGH

The Emergence of Institutions[1]

Santiago Sánchez-Pagés and Stéphane Straub[2]

December 2006

[1]We thank Andrzej Baniak, Eric Brousseau, Paco Candel-Sánchez, Richard Langlois, John Moore, Luis Garicano, Peter Grajzl, József Sákovics, Jonathan Thomas, Charles Vellutini and seminar audiences at Edinburgh, Birmingham, CEU Budapest, Esnie 2006, and the 11th Coalition Theory Workshop in Warwick, for their useful comments.

[2]Both authors: Economics, University of Edinburgh, William Robertson Building, 50 George Square, EH8 9JY, Edinburgh, United Kingdom. E-mail: ssanchez@staffmail.ed.ac.uk, and stephane.straub@ed.ac.uk.

Abstract

This paper analyzes how institutions aimed at coordinating economic interactions may appear. We build a model in which agents play a prisoners' dilemma game in a hypothetical state of nature. Agents can delegate the task of enforcing cooperation in interactions to one of them in exchange for a proper compensation. Two basic commitment problems stand in the way of institution formation. The first one is the individual commitment problem that arises because an agent chosen to run the institution may prefer to renege ex post. The second one is a "collective commitment" problem linked to the lack of binding agreements on the fee that will be charged by the centre once it is designated. This implies first that a potentially socially efficient institution may fail to arise because of the lack of individual incentives, and second that even if it arises, excessive rent extraction by the institution may imply a sub-optimal efficiency level, explaining the heterogeneity of observed institutional arrangements. An institution is less likely to arise in small groups with limited endowments, but also when the underlying commitment problem is not too severe. Finally, we show that the threat of secession by a subset of agents may endogenously solve part of the second commitment problem.

Keywords: Institution, Coordination, State of nature, Secession.
JEL classification codes: C72, D02, O17, Z13.

"As for 'philosophical history', it involved accounting for the development of beliefs, practices, theories, and institutions on the basis of natural causes or principles, when actual records and reports of witnesses were lacking."

Ian Simpson Ross, *The Life of Adam Smith* (1995).

1 Introduction

Since Adam Smith, the question of how different types of institutions support the efficiency of economic exchanges in varying social and historical environments has been a central one in social sciences. More recently, it has also become a central theme in Economics, with the recognition that market and non-market institutions are key in supporting and enhancing the growth potential of economic interactions (Olson, 1965; North, 1990; Bardhan, 2005).

As Bardhan (2005) points out, to date the literature on the economic analysis of social and political institutions have focused mainly on their role as protectors of property rights. A more neglected role of institutions is to correct the coordination failures or commitment problems that sometimes plague the most basic type of economic interactions. These problems, that can remain even if property rights are secure, are likely to be critical for an economy at the initial stages of its development. As showed by Aoki et al. (1997), the prominent role of the state in the East Asian development process (through the intervention in the capital markets, the establishment of technological standards or the use of contingent transfers), including the economic transition of Japan after WWII (see Okazaki, 1997), demonstrates that formal institutions can be crucial in economic development, not only by protecting individual property rights, but also by inducing and enforcing coordination when private mechanisms to do so are absent or underdeveloped.

Previous works on the role of institutions as coordination devices have mainly explored two related lines of enquiry. First, they have analyzed in great detail the internal functioning of specific institutional arrangements, sometimes relating them to relevant game-theoretical mechanisms. Examples are found in the economic history literature[1] with Greif's (1993) study of the coalition supporting the interactions of Maghribi traders with their distant agents in the 11th century, Milgrom, North and Weingast's (1990) analysis

[1]See Greif (1997) for a survey of the economic history literature that relies on microeconomic theory to study institutions.

of merchant courts at the Champagne fairs of the 12th and 13th centuries; in the development literature with for example the analysis of market institutions in Africa in Fafchamps (2004) or in Asia by MacMillan and Woodruff (1999, 2000); and in the law literature with Lisa Bernstein's (2001) account of the private legal framework that rules the US cotton industry or Bernstein (1992) and Richman (2006) accounts of private arrangements in the diamond industry.

Second, a few economic contributions analyze how informal or personalized relationship-based institutions may coexist with more formal, anonymous mechanisms, and how the transition from one to the other may occur (e.g. Kranton, 1996, and Dixit, 2004). This has also been an important topic in social anthropology. For example, Ensminger (1992) describes the century-long process through which changes in the environment finally triggered the Orma tribe in Kenya to abandon their constitutional authority from rule by a collective council of elders and to recognize the authority of the modern Kenyan nation-state.

As most of these contributions describe institutions already in place, or the transition between existing systems, very little has been said on the environmental and individual factors that lead to the emergence, possibly at varying levels of efficiency, of these institutional frameworks.

The aim of this paper is to model the process through which such institutions may or may not arise. Following our previous discussion, we see an institution as a coordination device, a body in charge of enforcing agreements or conventions that ultimately increase the efficiency of economic interactions between agents. In the words of Greif (1997), it is a non-technological set of constraints on behavior, which are self-enforcing. This self-enforcing nature of institutions is modelled through a game agents play in an hypothetical state of nature. Depending on the existing incentives, players' actions will eventually lead to the establishment of this coordinating mechanism as an equilibrium of that game. Therefore, our analysis of the process of institution creation can be relevant to different economic contexts, such as a tribe developing formal trade exchanges[2], a group of firms and clients in a given industry, a country in transition to an industrialized economy, or even a set countries faced with some international coordination issue.

[2]See Attali (2003) for examples of the introduction of witnesses or legimitator certifying the validity of exchanges in early societies of Africa, aboriginal Australia or precolombian Nicaragua among others.

Our point of departure is an economy in which the value of each individual's endowment is enhanced by interacting with others. Such payoff-enhancing relationships can arise in the context of a commercial exchange in which specific endowments are transferred according to each agent's needs, a productive venture in which complementary skills are put together to create additional value, or even intellectual or artistic exchanges.

In its simplest form, this is an informal economy: It lacks any institution in charge of ensuring the efficiency of bilateral relationships. In this state of nature, interactions take the form of a simple prisoners' dilemma game. Mutual cooperation would make it possible to attain objectives that are beyond the reach of individual agents on their own, but being opportunistic is a dominant strategy and in equilibrium very low payoffs are realized. This is a source of inefficiency. Agents would like to find a way to coordinate at the Pareto efficient outcome and ensure that it is enforced in any bilateral interaction.

In our model, agents can create an institution that will ensure mutual cooperation and thus enhance the value of the bilateral relationships that take place under its auspices. This body is akin to a judicial or political mechanism in charge of the definition and enforcement of efficient rules for social interaction. More precisely, it can be thought of as a reduced form for a set of norms, modes of behavior and beliefs of the types described in Greif (1993) and Milgrom et al. (1990) for example.

For such formal institution to arise, agents need to delegate to one of them the task of running it. But although formal interactions are more efficient, several problems stand in the way of institution formation. An institution is costly to set up since the delegate must relinquish her ability to interact with other agents, and must be properly compensated in exchange. Therefore, our model endogenizes the rise of a ruler from a population of identical individuals, in contrast with other works in the literature that exogenously impose its existence (Acemoglu, 2003; Acemoglu et al., 2004) and compare the scenarios with and without ruler (Grossman, 2002; Moselle and Polak, 2001). On the other hand, when the institution arises, agents have to decide whether to abide by its norms of interaction or not; in other words, they must decide whether to become formal or not; whenever two formal agents meet, the institution can guarantee that the efficient outcome will result. However, in order to enjoy this right, agents must pay a fee that constitutes the source of revenue for the institution.

We explore several procedures of institution formation and characterize

under which circumstances they will be successful. We make special emphasis on the implications of these different processes for efficiency and overall welfare.

In a nutshell, we find that a decentralized process of institution formation is plagued by two commitment problems. The first one is simply the individual commitment problem that arises when the revenue that can be raised by the agent chosen to act as the institutional centre is insufficient, and she prefers to renege ex post and fall back to informality, securing thereby a better payoff. Indeed, it is the potential rent associated with the eventuality of becoming the center that motivates agents to participate in the process of institution formation, so this rent has to be high enough to provide the right individual incentives.

The second one, which we label "collective commitment" problem, is linked to the fact that agents are not able to agree ex ante in an enforceable way on the fee that will be charged by the centre once it is designated.

Both limitations on commitment have implications for efficiency. The first aspect implies that an institution will not arise for some values of the parameters, despite being potentially welfare enhancing. This is in particular the case in intermediate size economies and when the extent of the coordination problem is rather limited. The intuition is that when the level of trust in the state of nature is relatively high so cooperation is only a mild issue, the outside option in which no institution emerges is more attractive, making it more difficult for the institution to arise.

The second aspect of lack of collective commitment implies that even when an institution emerges, it may do so at a sub-optimal level of efficiency, i.e. with an excessive level of the fee charged to agents in order to compensate the central agent. This happens for low levels of trust in the state of nature, because in that case a revenue-maximizing institution is able to set a high fee compared to the first-best level.

Together, these two commitment problems generate serious inefficiencies in the process of institution formation. We show that exogenously imposed commitment along each one of these two dimensions alone would reduce the scope for inefficiencies, but that the first-best institution emerges only when both problems can be solved simultaneously.

We then examine several devices that may help to solve these commitment problems endogenously. The first one is agents' use of trigger strategies to sustain cooperation in repeated interactions (e.g. Acemoglu, 2003). This potentially solves the second type of inefficiency, by forcing the implemen-

tation of a fee closer or equal to the first-best level. However, the question of how such collective punishment strategy can be implemented in a state of nature in which no coordination device exists remains open.

The second potential improvement, which again limits the ability of the centre to charge a sub-optimal level of the fee, is the threat of secession by a subset of agents. Precisely because our starting point is an institutionless society, it is plausible to assume that no group within it will be satisfied if it receives less than what it could get by forming its own mini society. To deter blocking, the institution should thus charge a fee that cannot be improved upon by any coalition. In that sense, the threat of secession may help to alleviate the inefficiency linked to a too high fee. However, this effect only operates for a limited parameter space; a big population size and high levels of trust in the state of nature make it very attractive to become a central agent and therefore create too strong incentives to secede.

The remainder of the paper is as follows. Next Section presents the model and its basic elements. In Section 3 we characterize the equilibrium level of formality, given that the institution has arisen, and the first best fee from the viewpoint of a social planner. Section 4 explores different procedures of institution formation characterized by varying degrees of commitment and of the freedom given to the participating agents. In Section 5 we analyze the stability of these rules against coalitional deviations. Section 6 offers a discussion of the results and concludes. Proofs are relegated to the Appendix.

2 The Model

Consider an economy populated by $N + 1$ agents, who have an initial endowment of value ω (representing a combination of skills, time and goods). Agents' interactions in this economy are described by the basic game G in Figure 1.

		Player j: C	Player j: NC
Player i	C	x,x	-z,z
	NC	z,-z	0,0

Figure 1: Basic game

Agents are randomly matched against one another and play G. Payoffs in the matrix represent the return per unit of endowment invested in the interaction. We assume that $z > x > 0$. The strategies C and NC denote "cooperative" and "not cooperative" respectively. C stands for a cooperative behavior that can create added value, and NC stands for a opportunistic behavior that makes the agent take advantage of a cooperative partner but yields zero returns if she is not.

The game G admits a unique Nash equilibrium, (NC, NC), that is Pareto inferior to (C, C). This game is aimed at capturing the natural gains from cooperation that exist in human interactions but also the possibility of distrust and opportunism that lead to Pareto inferior outcomes.[3]

The scenario in which individuals are matched and play G without any interference is assumed to be the status-quo of the economy. In order to solve the problems of miscoordination, agents can set up an institution. It will be in charge of enforcing the efficient outcome in any interaction that takes place under its auspices. This institution arises when agents delegate to one of them, who we will call the *center,* the task of running it. The central agent must relinquish her ability to interact with other agents but she will be compensated in exchange. At this point, we deliberatively remain vague about how this delegation process is carried through since the main body of the paper (Section 4 below) amounts to discussing several procedures of institution formation.

[3]In accordance with most of the literature on these topics, we have chosen this game as the simplest way to illustrate the type of social situations we want to analyze. Admittedly, there are many other games that may capture the trade-offs we study here.

If the institution arises, agents have to decide whether to abide to its norms of interaction, that is to become formal, or not to do so and remain informal. In the same spirit as Basu's (2000) *civic norms,* we assume that the institution is able to restrict the set of strategies agents can choose from when interacting with other formal agents, thanks to the center acting as a coordinating device.[4] In that case, they play strategy C, ensuring that the outcome of the interaction will be always (C, C). However, in order to become formal, agents have to pay a fixed fee $a \leq \omega$, that can be understood as an entry fee or a lump-sum tax that enables them to reward the center for her activity and thus to interact under the institutional umbrella. Below we will also discuss at length how the level of the fee a is fixed.

We will admit a richer description of the payoff x in G and assume that it depends on the efficiency of the institutional mechanism that in turn is a function of the level of agents' contribution a.[5] Hence, the per-person unit return from an interaction between two formal agents is

$$v_{K,a}^{F} = x\left(a\right), \tag{1}$$

where the superscript F denotes "Formal" and $x(\cdot)$ satisfies $x_a > 0$ and the standard Inada condition, $\lim_{a\to 0} x_a\left(a\right) = \infty$, holds. One reason for assuming that $x_a > 0$ may be that the institution becomes more efficient when endowed with more resources, as it is able to invest more in physical or relational supporting infrastructure, as in the case of diamond clubs described in Richman (2006).

When at least one of the two interacting agents is informal, the institution has no power to enforce the efficient outcome and the game G is played without any further restriction on the strategy space. Informal agents thus avoid paying the fee but their interactions yield lower returns. Still, in this state

[4]Milgrom et al. (1990), Kandori (1992) and Greif (1993) among others, describe different ways in which such coordination can be achieved in repeated games, even if the same players only meet occasionally. Such a mechanism is implicitly assumed here, but we refrain to model it for tractability purposes because our focus is on how it emerges in the first place.

[5]We could envision making x also a function of the proportion of agents $\frac{K}{N}$ contributing to it. However, it is not clear theoretically how this will affect x, so we abstract from this complication. Indeed, a higher proportion could have a positive effect because of network externalities for example, but congestion could also lead to a negative effect in a range of parameters (see Kranton, 1996).

of nature, agents may occasionally cooperate with each other despite the absence of material incentives to do so or of any formal institution enforcing efficient outcomes.[6] Hence, we characterize the level of trust or cooperation in the society under the state of nature by a parameter $\alpha < 1$, which is an initial condition of our economy and that may in turn depend upon culture, expectations and the specific type of interactions considered. More specifically, we assume that agents play the (C, C) outcome with probability α and (NC, NC) otherwise.[7] In that case, the per-person unit return from the interaction between a formal and an informal agent or between two informal agents is therefore

$$v_{K,a}^{I} = \alpha x\left(a\right), \tag{2}$$

where the superscript I denotes "Informal". We will assume that $x(0) > \frac{1}{\alpha}$ to ensure that participating in a completely informal economy always dominates the autarchic situation in which agents do not interact and simply consume their endowments.

Given that agents are randomly matched and that they are assumed to be risk neutral, the expected payoff of a formal agent when $K \geq 2$ agents are formal is:

$$\begin{aligned} V_{K,a}^{F} &= \frac{K-1}{N-1}\left(\omega - a\right) v_{K,a}^{F} + \frac{N-K}{N-1}\left(\omega - a\right) v_{K,a}^{I} \\ &= \frac{K-1}{N-1}\left(\omega - a\right) x\left(a\right) + \frac{N-K}{N-1}\left(\omega - a\right) \alpha x\left(a\right). \end{aligned} \tag{3}$$

We assume that an institution becomes active if at least two agents are formal, so the probability of formal exchanges is strictly positive.

Finally, the central agent, who gives up interacting with the rest of agents, receives the fees paid by all formal agents. Hence, her payoff is given by

[6] In fact, there is substantial experimental evidence showing that subjects are willing to cooperate and trust others in prisoners' dilemma-like settings much more often than what the theory predicts. See for instance Marwell and Ames (1981) or Dawes and Thaler (1988) among many others. This likelihood of cooperation is also often referred to as a measure of "social capital" in theoretical contributions based on the prisoners' dilemma (Routledge and von Amsberg, 2003; Durlauf and Fafchamps, 2004). We return to this interpretation in the final discussion.

[7] An alternative interpretation of the parameter α, in line with the literature on informality, is the level of free-riding that informal agents can make on formal institutions. See for example Loayza (1995), Marcouiller and Young (1995), Choi and Thum (2002) and Azuma and Grossman (2002).

$$V^C_{K,a} = K(a - c),$$

where c is the enforcing cost she incurs for ensuring efficiency in each transaction undertaken under her auspices.[8]

Figure 2 summarizes the timing of the game described above.

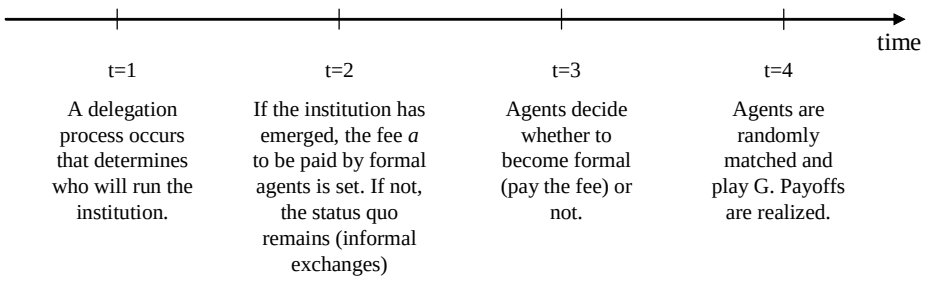

Figure 2: Timing of the game

We have thus constructed a game in four stages. In the first stage, that we will make explicit in Section 4, agents set up the institution. Then, the institutional fee a is set. In the third stage of the game, agents decide whether to become formal or not. In the last stage, they are paired with another interacting agent in society and play G, eventually resorting to the institution set up earlier.

3 The equilibrium level of formality

3.1 Existence and Stability

Given this basic framework, the first question that arises concerns the existence and stability of different configurations. Assume that K agents are formal, $N - K$ are informal, and that, without loss of generality, the $N + 1^{th}$ agent is devoted to institutional work. Given a fee a, this division of agents between formality and informality can be supported in equilibrium if and only if no agent is willing to deviate and change her status.

[8] Assuming that the cost is instead a proportion of the fee received does not change our results.

A formal agent will not be willing to deviate and become informal as long as $V^F_{K,a} \geq V^I_{K-1,a} = \omega v^I_{K-1,a}$. After some transformations, this can be written:

$$a \leq \omega \left(1 - \frac{(N-1)\,\alpha}{K-1+(N-K)\alpha}\right) \equiv a(K).$$

Similarly, an informal agents receives a payoff:

$$V^I_{K,a} = \omega v^I_{K,a},$$

and will not wish to become formal as long as $V^I_{K,a} \geq V^F_{K+1,a}$, which yields:

$$a \geq \omega \left(1 - \frac{(N-1)\,\alpha}{K+(N-K-1)\alpha}\right) \equiv a(K+1).$$

Note first that $0 < a(K) < \omega$ for all $K > 1$ and that given our assumption above stating that the institution remains inactive if $K = 1$, $a(1) = 0$. In any case, the equilibrium level of formality will clearly depend upon the properties of $a(\cdot)$.

The next Proposition characterizes the conditions under which there exists a level of the institutional fee a that can support a certain amount of formal agents as the equilibrium of the subgame in stage 3.

Proposition 1 *$a(\cdot)$ is strictly increasing in K. Therefore, either N or 0 formal agents can be supported in equilibrium.*

In this setting, only corner equilibria can arise, i.e. full formality or full informality. Note that when no more than one agent becomes formal, $v^F_{1,a} = v^I_{K-1,a} = \alpha x\,(0)$ for any a that the central agent might have set. When full formality prevails, $a(N) = \omega\,(1-\alpha)$. We will assume that $c < \omega\,(1-\alpha)$. Otherwise even the highest fee compatible with full formality cannot cover the running costs of the institution.

The following Proposition characterizes the equilibria that can arise in this subgame for each possible level of the fee a.

Proposition 2 *For a given level of the fee $a \in [c, \omega]$,*

(i) Informality can be supported in equilibrium for all $a \geq 0$.

(*ii*) *Full formality can be supported in equilibrium only if* $a \leq a(N)$.

The proof follows from the arguments above. This Proposition shows that a new coordination problem arises when the institution emerges. Paying a fee compatible with full formality may not compensate the cost of becoming formal when everybody else is informal. Hence, both full formality and informality can be sustained in equilibria for the same level of the fee. Figure 3 depicts the profile of equilibria as a function of the fee a.[9]

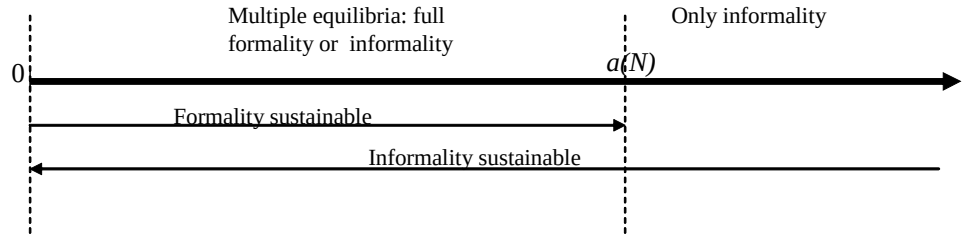

Figure 3: Profile of equilibria

3.2 The first best institutional fee

In the remainder of this Section, we characterize the optimum fee from the viewpoint of a hypothetical utilitarian central planner willing to maximize the total sum of agents' utilities. Under full formality, the planner's instrument is the fee a that agents must pay in order to enjoy the benefits of formal interactions. She will have to compare the maximum welfare attainable in this scenario with the (fixed) level of social welfare under complete informality.

In the case of full formality, the constrained maximization problem of this planner can be written as:

$$\begin{aligned} \max_{a} \; W^F &= N[(\omega - a)\, x(a) + (a - c)] + \omega \\ s.t. \qquad a &\leq a(N). \end{aligned}$$

[9]For the range of fees $[0, a(N)]$ there also exists a mixed strategy equilibrium in which agents become formal with probability $p(a) = \frac{\alpha}{1-\alpha}\frac{a}{\omega - a}$. Although this can in principle support an intermediate level of formality, the revenue raised by the institution in this equilibrium is maximized at $a = a(N)$, and $p(a(N) = 1$. Hence, in this case full formality would arise too.

The above program yields a solution a^*:

$$\frac{\omega - a^*}{a^*} = \frac{1}{\varepsilon_{a^*}} \left(\frac{x(a^*) - 1}{x(a^*)} \right), \tag{4}$$

where ε_{a^*} is the elasticity of x with respect to a.

Since the fee set must not override agents' incentives to remain formal, a^* cannot be higher than the maximum fee compatible with full formality. Hence, the planner chooses to implement full formality with a fee equal to

$$a^F = \min\{a^*, a(N)\}.$$

This implies that the solution to the planner's problem will be a corner solution, i.e. $a^* \geq a(N)$, as long as $\alpha \geq \alpha^*$ where α^* satisfies[10]

$$\alpha^* = \frac{x(\omega(1-\alpha^*)) - 1}{\omega x_a(\omega(1-\alpha^*))}. \tag{5}$$

As α increases, formal agents have stronger incentives to defect and this must be compensated with a lower fee if full formality is to be maintained. This decreases $a(N)$ and the room for an interior solution shrinks. However, the effect of an increase in the endowment ω is ambiguous: It relaxes the constraint but it also changes the objective function by making interactions more profitable.

On the other hand, the planner can leave the economy in a state of full informality. In that case, total welfare is just

$$W^I = (N+1)\omega\alpha x(0). \tag{6}$$

Under full formality, social welfare is a function of the fee actually fixed. There may exist values of the parameters for which even the maximum welfare attainable under formality is below the welfare under informality. This is characterized by the following Proposition.

Proposition 3 *Social welfare under informality is higher than under full formality for any value of the fee a if*

$$x(0) > \frac{1}{N+1}\left(\frac{1}{\alpha} + N\frac{\left(\omega - a^F\right) x(a^F) + a^F - c}{\alpha\omega}\right) \equiv \underline{x}(N, \omega, \alpha). \tag{7}$$

Moreover, the lower bound $\underline{x}(N, \omega, \alpha)$ is increasing in both the population size N and agents' initial endowment ω and decreasing in the status-quo level of trust α.

[10] It is straightforward to show that such fixed point exists.

Hence, for small and relatively poor economies (low N or ω) full formality needs not be the most desirable outcome. Similarly, if under the status-quo, the problems of miscoordination are not very severe (high α), setting an institution may be too costly relative to the gains it can bring. In that case, an utilitarian planner may prefer to implement informality.

Let us finally characterize as well the level of the fee that maximizes the welfare of the set of interacting agents alone; it will prove to be useful in Section 5. This fee solves

$$\begin{aligned} \max_{a} \quad & N\left(\omega - a\right) x(a) \\ s.t. \quad & a \leq a(N). \end{aligned}$$

The above program yields a solution a^{**} characterized by the first order condition

$$\frac{\omega - a^{**}}{a^{**}} = \frac{1}{\varepsilon_{a^{**}}}. \tag{8}$$

Therefore it is clear that $a^{**} < a^{*}$. By the same token, there exist a threshold α^{**} such that the solution to this problem is interior whenever $\alpha \geq \alpha^{**}$. It is straightforward as well to show that $\alpha^{*} < \alpha^{**}$.

4 Emergence of the institution

Since no central coordination device exists before the members of a society actually create one, any effort to set up an institution that will enforce cooperation has to proceed in a decentralized way. In this Section, we analyze this process, highlighting in particular how commitment problems affect the efficiency of the emerging institution or block its emergence despite its potentially welfare enhancing effect.

We define a *procedure of institution formation* as a fair lottery over the set of agents who freely participate in it for a given fee a. This lottery designates the agent who subsequently will be in charge of running the institution. The fee a can be set either before the lottery takes place or afterwards.

A lottery appears to be the simplest mechanism to choose the individual in charge of the institution in the absence of an explicit coordination device. A lottery is the simplest way to formalize a situation in which individuals are all alike and hence there is no reason *a priori* why they should not be

equally likely to end up in charge of the institution. A similar outcome for instance would result from a bidding procedure, in which all agents would bid for the right to become the center, and the winner would therefore be chosen randomly among them.[11] Also, considering that the basic interaction game described above is repeated many times, a possible implementation would be a scenario in which there is a rotation each period among members of society to act as the center, or equivalently a new center is randomly drawn each period.[12]

In this general framework, different procedures of institution formation are possible depending on the different degrees of commitment available both at the individual and at the collective level, the natural benchmark being a fully decentralized process with no commitment whatsoever.

At the individual level, agents must decide simultaneously whether to participate *ex-ante* or not in the lottery that will designate who will run the institution. We assume that an agent who does not participate in the lottery is subsequently excluded from the possibility of becoming formal. Hence, given a level of the fee a, the institution can arise only if

$$\frac{1}{N+1}(N(a-c)+\omega)+\frac{N}{N+1}(\omega-a)\,x(a)\geq\omega\alpha x(a), \tag{9}$$

where the left hand side shows the expected payoff from participating, as the sum of the center's and the agents' payoffs respectively weighted by their corresponding probabilities, and the right hand side is the payoff from unilateral deviation. In equilibrium, it is easy to show that either all or no agent will participate in the process.

All the different processes of institution formation that we will discuss impose this basic participation constraint. Still, agents who accept to participate in the process may change their mind *ex-post* depending on the outcome of the lottery. Therefore, when there is no commitment at the individual level, an *ex-post* participation constraint needs to be imposed as well. This requires that agents should be given guarantees that once they discover their role, they will not prefer to fall back into informality. As this *ex-post* requirement is always satisfied for an agent who does not become the

[11] We will discuss this issue more in detail in Section 6.

[12] See also Morgan (2000) for an application of lotteries to reduce free-riding in public good provision.

center[13], the relevant *ex-post* participation constraint, given a certain level of the fee a, is

$$N(a-c)+\omega \geq \omega\alpha x(0). \tag{10}$$

This defines a minimum level of the fee $\underline{a} \equiv \omega\frac{\alpha x(0)-1}{N}+c$, below which the agent chosen to be the center would prefer to give up and the whole economy would collapse to informality.

The benchmark assumption of no commitment implies that collective choices are not possible and that the central agent has total freedom to set the fee once she takes up her role. In that case, she will behave as a revenue maximizing monopolist with no constraint on the fee to be set beyond her own self-interest.

However, we will also contemplate the possibility of the fee a being chosen collectively and that this choice may be binding. In this case, agents will set a fee that maximizes total welfare behind the veil of ignorance, that is, before the outcome of the lottery is realized.[14]

Table 1 summarizes the possible combinations of assumptions:

[13]It is obvious that $(\omega-a)x(a) \geq \omega\alpha x(a)$ for any a not greater than the upper bound on a, which is $a(N)=\omega(1-\alpha)$.

[14]Admittedly there may be other processes. The ones considered here are polar cases.

	Limited commitment (ex post participation constraint)	**Strong commitment (ex ante participation constraint)**
Center maximizes revenue (sets *a*) ex post	1. Agents' only commitment is to participate in the lottery ex ante. The center may refuse to cooperate ex post and is free to set *a*.	2. Agents commit ex ante to participate in the lottery and not to renege ex post if chosen as the center.
Fee *a* set ex ante	3. Agents commit ex ante to participate in the lottery. If chosen as the center, they may renege, but have no freedom to set *a* if they accept to fulfill their role.	4. Agents commit ex ante to participate in the lottery and not to renege ex post if chosen as the center. Furthermore, the center has no freedom to set *a* ex post.

Table 1: Assumptions on degree of commitment

Next we explore these different scenarios, starting with the natural benchmark, the "No commitment" case.

4.1 No Commitment

Under the "No commitment" or fully decentralized procedure of institution formation, the fee is freely set by the central agent. Moreover, in addition to the ex-ante participation constraint, the ex-post one must be imposed. We know from Section 3 that the maximum fee that the institution can charge is $a(N)$. Therefore, agents will participate only if the two following conditions hold:

$$\frac{1}{N+1}(N(a(N)-c)+\omega)+\frac{N}{N+1}\left(\omega-a(N)\right)x(a(N))\geq\omega\alpha x(a(N)), \tag{11}$$

$$N(a(N)-c)+\omega\geq\omega\alpha x(0), \tag{12}$$

which are simply the result of rewriting the ex ante lottery participation constraint (9) and the ex post constraint of the center (10) by replacing a with $a(N)$. These two conditions are necessary for the institution to arise. Note that when $a = a(N)$, trading agents are indifferent between formality and informality. Therefore, (11) can be rewritten as:

$$N(a(N) - c) + \omega \geq \omega\alpha x(a(N)), \tag{13}$$

from which it is evident that (11) is a stronger constraint[15], so if it is not satisfied, the economy will remain in a state of informality.

Finally, we need to establish which fee will be set by the institution in equilibrium. The multiplicity of equilibria described in Proposition 2 complicates matters because it translates in a multiplicity of fees that can be supported in equilibrium. We will focus on the most natural equilibrium of this game of institution formation.

Proposition 4 *If condition (13) holds, there exists a SPE of the fully decentralized procedure of institution formation that implements full formality under the fee* $a(N)$.

There are two possible sources of inefficiency in this scenario. On the one hand, full formality is not implemented when (13) does not hold, despite the fact that it may still be efficiency enhancing. This is the case when parameters are such that the level of individual welfare obtained under formality

$$W^F_{a(N)} = \frac{1}{N+1}(N(a(N) - c) + \omega) + \frac{N}{N+1}(\omega - a(N))\, x(a(N)),$$

dominates the level of welfare under full informality but is not high enough to induce ex ante participation in the lottery.

Corollary 1 (Non-emergence of efficient institutions) *Under the fully decentralized procedure, a potentially welfare enhancing institution does not arise if and only if*

$$\omega\alpha x(0) \leq W^F_{a(N)} \leq \omega\alpha x(a(N)). \tag{14}$$

[15]Of course, this is only true for $a = a(N)$ and needs not be verified for lower values of the fee.

Such inefficiency may arise for economies of intermediate size and when the status-quo level of trust α is sufficiently high.

The lower bound in (14) determines when formality is more efficient than informality, whereas the upper bound establishes when formality is implementable. Within these bounds, the institution is welfare enhancing but it does not emerge.

Corollary 1 shows that the first type of inefficiency is more likely to occur in economies of intermediate size and with limited coordination problems (high α). In the first place, it occurs if the size of the population is not small enough for informality to be superior, but not big enough for the institution to arise. The reason why N has to be large enough for the institution to arise comes from the fact that the center's expected revenue is increasing in N, so there is a minimum population size above which the prospect of becoming the center gives agents enough incentive to participate in the lottery.

On the other hand, the range of parameters for which a welfare enhancing institution does not arise expands as α increases. At the heart of this result is the fact that high status-quo trust makes the outside option of informality more attractive and undermines the dominant position of the revenue-maximizing institution. This is an interesting result: We should observe the emergence of formal institutions in societies plagued with coordination problems and low levels of informal trust, while inefficiencies are more likely to arise in societies with relatively high level of trust. The fact that inefficiencies are less costly to agents implies that bearing the cost involved in solving them is not incentive compatible at the individual level, despite being socially efficient.

Note also that the fact that the central agent is offered the possibility of maximizing revenue when setting a does not always suffice to ensure the emergence of an efficient institution. On the contrary, even if full formality is implemented, the fee set by the central agent may be too high so the first best cannot be attained. A necessary condition for this second type of inefficiency is a low enough degree of trust in bilateral interactions, i.e. $\alpha < \alpha^*$, that implies $a^F = a^* < a(N)$ (see Section 3.2).

Corollary 2 (Implementable first best) *When condition (13) holds, the first best fee a^F can be implemented in a SPE of the fully decentralized procedure of institution formation:*

(*i*) *For high enough levels of status quo-trust, i.e. $\alpha \geq \alpha^*$.*

(*ii*) *For relatively low levels of status-quo trust (i.e.* $\alpha < \alpha^*$*) if* $a^F = a^* \geq \underline{a}$.

The intuition for this result is easy to grasp. When welfare is increasing over the range of fees compatible with formality or when the level of status-quo trust α is sufficiently high, the planner would like to set the highest fee compatible with full formality (i.e., $a^F = a(N)$). In that case, the center's incentives are aligned with social welfare and the first best can be attained by means of the decentralized procedure.

In case (*ii*), the multiplicity of equilibria described in Proposition 2 makes it possible to support the first best in equilibrium. Here, full formality can be supported in equilibrium for any fee a' in the interval $[\underline{a}, a(N)]$ by agents' use of trigger-like strategies of the following class:

$$F = \begin{cases} 1 & \text{if } a \leq a' \\ 0 & \text{otherwise} \end{cases}.$$

Then, the first best can be implemented when $a' = a^*$. However, it is not clear how in a state of nature that we define as completely noncooperative, agents can coordinate in the use of these strategies, making it unlikely that the first best be sustained when $\alpha < \alpha^*$. One context in which this could be envisioned is when the implementation of an institutional mechanism is supported by external advice, so such strategies can be exogenously suggested to players. We will come back to this point below.

4.2 Partial Commitment

While the no commitment case appears to be the natural benchmark of our economy, it is useful to consider how the outcome of the procedure of institution formation varies when some degree of commitment is introduced along each of the two dimensions considered above: Individual commitment and a binding collective choice of the fee.

Of course, this raises the question of how such a commitment is secured and enforced. We have some sort of a chicken-and-egg problem here: We started in an institutionless world, where there was a basic problem of enforcing coordination in bilateral relations. The possibility of commitment in the present case would however indicate the existence of perhaps a multilateral

mechanism capable of enforcing it.[16] After showing briefly how commitment may improve efficiency in the institution formation process under each of the possible combinations of assumptions considered above, we discuss how it may be enforced: In Section 5, we analyze in more detail a mechanism that may endogenously support some degree of individual or collective commitment despite full decentralization, namely the threat of secession by a coalition of agents.

As mentioned, introducing commitment at the individual level amounts to assume that agents do not renege ex post, whatever the outcome of the lottery. Therefore, only agents' *ex-ante* participation constraint (9) needs to be satisfied (case 2 in Table 1). On the other hand, at the collective level, commitment arises if the fee a is fixed by all participating agents before the actual running of the lottery and this choice is binding (case 3). Finally, combining the two yields the possibility of full commitment (case 4).

Case 2. First assume that agents are able to commit to set up the institution if chosen, so the *ex-post* participation constraint (12) is dropped, but that the center retains total freedom to set the fee. Therefore, only condition (11) must hold. Since we know from case 1 that condition (11) is stronger than (12), it is obvious that this does not introduce any change with respect to the benchmark no-commitment case. This case shows that a stronger individual commitment is only useful if accompanied by some degree of collective commitment on the choice of the fee (see case 4 below).

Case 3. Consider now the case in which a binding choice of the fee a is made by agents in advance to the lottery, but individual agents cannot commit *ex-ante* not to renege *ex-post* in case they are chosen to run the institution. Then, society will choose a fee that maximizes social welfare subject to the ex post participation constraint, that is, a fee high enough to compensate the central agent. This imposes that it is at least greater than $\underline{a} \equiv \omega \frac{\alpha x(0)-1}{N} + c$. Obviously, individual incentives must still be taken into account so the fee chosen has to be compatible with full formality (hence below $a(N)$). Once this holds, society will implement a fee as close as possible to the first best.

Proposition 5 *The collective choice of the fee implements full formality if*

[16] Greif (1993) shows how multilateral reputation mechanisms (where punishments are inflicted by the whole community) can be cheaper and more effective in enforcing mercantile contracts than bilateral mechanisms.

and only if $\underline{a} \leq a(N)$. *In that case, the fee set is* $a = \max\{\underline{a}, a^F\}$ *and the first best is achieved if and only if* $\underline{a} \leq a^F$.

First, it is important to note that the collective choice of the fee makes the implementation of the institution no easier than under the fully decentralized procedure, as it still requires $\underline{a} \leq a(N)$. However, this type of commitment makes the institution more efficient when implementable, because the first best is now more likely. On the other hand, even if the first best cannot be attained, i.e. $a(N) > \underline{a} \geq a^*$, there is an improvement with respect to the same case under the fully decentralized procedure, since the fee chosen is $\underline{a}$ instead of (the higher) $a(N)$.

As mentioned, one way collective commitment to a fee could be implemented at this point is through the use of trigger-like strategies. Note, however, that a repeated version of the game, with true (i.e. history-dependent) strategies in which the center is in charge for various periods, would not improve upon the one-shot version. The simple reason for this is that both deviations and punishments would arise at the beginning of each period, so the incentive constraint on the central agent in the repeated game would be exactly the same as in the one-shot version.

Case 4. Finally, consider the case where there is no *ex-post* participation constraint (strong commitment) and agents meet and agree in advance to the lottery that they should implement the utilitarian first best.[17] It is easy to see that in this case the efficient outcome is always implemented as stated in the following Proposition.

Proposition 6 *When both individual and collective commitment are possible, full formality is implemented if and only if informality does not maximize welfare, i.e.* $x(0) \leq \underline{x}(N, \omega, \alpha)$. *Moreover, the first-best is always attained.*

The intuition is straightforward: When $x(0) \leq \underline{x}(N, \omega, \alpha)$, the first best fee a^F is high enough to ensure that the *ex-ante* participation constraint (9) is satisfied. Therefore, individual incentives do not stand in the way of efficiency in this case and formality is implemented whenever it is efficient.

To summarize, when considering the decentralized institution formation process, the inability to constrain the center to chose a specific level of fee

[17] While in the no commitment case discussed in the previous section the *ex-post* participation constraint was irrelevant as it was implied by the *ex ante* one, this may of course not be the case when $a < a(N)$.

(lack of collective commitment) is a strong reason for the occurrence of inefficiencies (case 1). As this limit is relaxed, potential inefficiencies are reduced, as shown by case 3. Moreover, when the ability to set fees *ex-ante* is combined with individual commitment, the first best is always implementable.

The next Section discusses a decentralized mechanism through which collective commitment may be enforced.

5 Secession

In this Section we consider the possibility that a coalition of agents secedes from society to run their own institution. Our aim is to characterize under which conditions a central institution will be secession-proof and to analyze the impact of the threat of secession on welfare.

Since our starting point is the state of nature where no commitment is possible, the concept of secession-proofness has a clear importance. An institution can hardly be called self-enforcing if a group of agents operating under it can improve its situation by withdrawing and later applying the same procedure of institution formation used by the society as a whole.

Specifically, our analysis of secession will concentrate on the decentralized procedure of institution formation, assuming that it will be employed both by the whole population and subgroups intending to withdraw. Next we analyze when the threat of secession can prevent the emergence of a single institution and its effect on efficiency.

Let us first state our definition of blocking:

Definition 1 *Denote by a_N the fee set by the institution. A coalition formed by S interacting agents is a blocking coalition if and only if*

$$(\omega - a_N)x(a_N) < \frac{1}{S}(S(a(N) - c) + \omega) + \frac{S-1}{S}\left(\omega - a(N)\right)x(a(N)). \quad (15)$$

That is, our concept of blocking implies that no group of agents should prefer (in expectation) to withdraw from society and apply among them the fully decentralized procedure of institution formation. This is a relatively strong requirement.[18] Note that when a coalition contemplates the possibility

[18] Alternatively, we could have imposed a weaker criterion, as in Howe and Roemer (1981), in which a coalition is blocking whenever it can *guarantee* a higher payoff to its members

of secession, it recognizes that the fee that will be set in the hypothetical new institution must be itself self-enforcing. We have in this case picked $a(N)$, the (sometimes unique) equilibrium fee we have at length considered in the previous sections.[19]

Definition 2 *A fee a_N is said to be secession-proof if it does not spawn any blocking coalition.*

Secession-proof fees are natural focal points in the process of institution formation: Members of no group should receive less than what they could expect to obtain from creating a mini society under the same rules. Such fees can thus be said to be in the *core* of that particular procedure of institution formation.[20]

Given that we are analyzing the case of no commitment, we assume that the central agent will set the maximum possible secession-proof fee. Secession thus imposes new and natural constraints on the fee that the institutional agent can charge. Notice that, if full formality is not implementable when secession is not an option, this will continue to be the case when secession is possible; since the revenue of the central agent cannot increase, secession thus cannot help potentially welfare enhancing institutions to emerge.

The first question that arises is whether the set of secession-proof fees is empty or not. It is easy to check that the payoff of a coalition contemplating the possibility of withdrawing is increasing in its size S. Therefore, for a fee to be secession-proof it is enough to satisfy condition (15) for $S = N$. On the other hand, the fee that maximizes agents' welfare can be either a^{**} or $a(N)$. Let us assume, for the sake of exposition, that $\min\{a^{**}, a(N)\} > \underline{a}$. Hence the set of secession-proof fees is non-empty if and only if

$$\frac{1}{N}((N-1)(a(N)-c)+\omega)+\frac{N-1}{N}\alpha\omega x(a(N)) \leq (\omega - \min\{a^{**}, a(N)\})x(\min\{a^{**}, a(N)\})$$

If this condition is not met, we should reasonably expect the emergence of more than one institution. When $a(N) > a^{**}$ this expression implicitly

[19]Note that for all S, $a(N) = a(S) = \omega(1-\alpha)$, so we stick to the current notation for simplicity.

[20]As any core-related concept, our definition of blocking only takes into account one-step secessions. We do not consider the possibility of further blocking once a new society is formed. The set of secession-proof fees defined here is thus minimal in this sense.

defines a threshold on the population size, denoted by $N_0(\alpha, \omega)$ such that a^{**} is secession-proof whenever $N \leq N_0(\alpha, \omega)$. Similarly, when, $a(N) < a^{**}$ the threshold

$$N_1(\alpha, \omega) \equiv \frac{\alpha\omega x(a(N)) - \omega}{a(N) - c} + 1,$$

can be defined as the maximum population size that is compatible with $a(N)$ being secession-proof. Note that if $N \leq N_1(\alpha, \omega)$ the threat of secession has no bite.

The next Proposition summarizes the conditions in terms of the population size N and the level of status-quo trust α under which secession is a real possibility.

Proposition 7 *The set of secession-proof fees is non-empty if and only if*

$$N \leq \begin{cases} N_0(\alpha, \omega) & \text{if } \alpha \leq \alpha^{**} \\ N_1(\alpha, \omega) & \text{otherwise} \end{cases}.$$

Moreover, the threshold $N_0(\alpha, \omega)$ attains a minimum at $\alpha = \alpha^* (< \alpha^{**})$ whereas $N_1(\alpha, \omega)$ is increasing in α.

The main reason for blocking in this model is thus the prospect of becoming the center in the new mini society. When the size of the population is sufficiently big, the center obtains an extremely high payoff and this creates strong incentives to withdraw. As a matter of fact, notice that the condition $N > N_1(\alpha, \omega)$ can be rewritten as

$$(N-1)(a(N) - c) + \omega > \alpha\omega x(a(N)),$$

so $a(N)$ stops being secession-proof whenever the central agent of the new institution can obtain a higher payoff than the rest of the agents.

Figure 4 depicts the regions characterized by this the threshold in the parameter space.

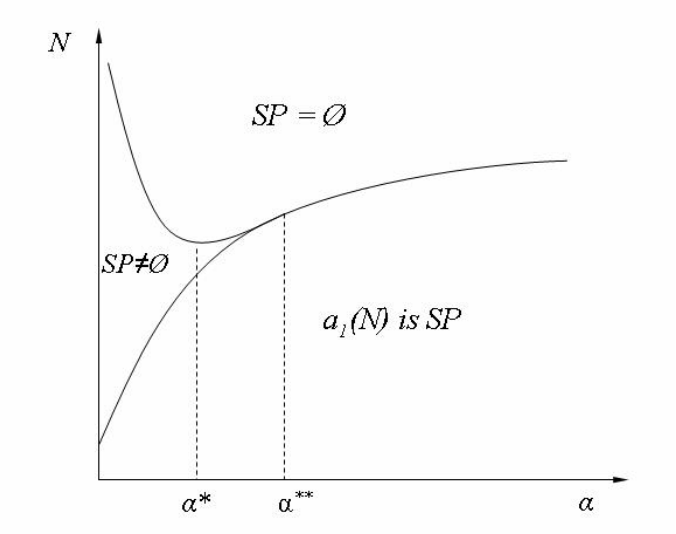

Figure 4: The set of Secession-proof fees

When the level of status-quo trust is sufficiently small (i.e. $\alpha < \alpha^{**}$) and the population size is intermediate (i.e. $N \in (N_1(\alpha, \omega), N_0(\alpha, \omega))$, it may be still possible for the institution to avoid secession by increasing agents' welfare via a reduced fee. In that case, secession can help to alleviate the inefficiency produced by a too high fee compared to the case where secession is not possible. But outside this case, secession is a real threat that can render impossible the emergence of one institution comprising all agents in society.

The natural question that now arises is whether the impossibility of a single institution matters from an efficiency perspective. The answer of course depends on the particular rules of secession and coalition formation to be considered. Here we will assume that whatever this process is, any division of the population in several smaller societies is stable only if all groups can set a secession-proof fee.

Formally, a *coalition structure* is a division of the population into a collection $C = \{C_1, ..., C_K\}$ of disjoint coalitions of generic size $S_k \geq 3$. It is straightforward to extend our previous definition of secession-proof fees to the case of subgroups: We will say that a coalition structure C is secession-proof if all coalitions in it set a (possibly different) fee that does not spawn a blocking coalition within them. Here, we will concentrate on the case of $\alpha \geq \alpha^{**}$ for simplicity, meaning that in any secession-proof coalition struc-

ture all groups must employ $a(N)$ (because it is then the only self-enforcing fee).

Next we show that if one considers secession-proof coalition structures as the natural outcome of any process of coalition formation (or secession), the impossibility of a single institution is negative from a social point of view.

Proposition 8: *When $\alpha \geq \alpha^{**}$, the total sum of payoffs under the single institution is at least as big as under any secession-proof coalition structure.*

As mentioned before, the incentives to secede come from agents' prospect of becoming the center of the new mini-society; recall that when the fee is $a(N)$ it is only the central agent who extracts positive rents. However, this is socially wasteful because it leads to an unnecessary proliferation of institutions. Obviously, this conclusion makes abstraction from the possibility that the coordination job of the center in smaller groups may entail lower transaction costs, i.e. lower c in our model. If that is the case, the above conclusion may require qualification.

6 Discussion and Conclusion

We have built a model in which agents from a population are randomly matched to play a prisoners' dilemma game. In the hypothetical state of nature, such interactions are plagued by inefficiencies, as the only Nash equilibrium is the configuration in which agents play non cooperatively, leading to a Pareto inferior outcome. Formalizing an idea implicit in some of the existing literature on institutions, we have assumed that agents can delegate the task of enforcing cooperation in interactions to one of them (the institution) in exchange for a proper compensation. Examples of multilateral mechanisms that can enforce such cooperation are found in many strands of literature, including Economics, Sociology and Law.

The contribution of this paper is to focus on the process through which such mechanism may actually emerge in a context in which no previous coordination device exists. More specifically, our aim is to determine whether this mechanism arises whenever it is potentially welfare enhancing, and when it does, whether it is as efficient as it could possibly be.

In a world in which no commitment is possible, i.e. individuals cannot commit in advance to a future behavior, be it the participation in the institution or the level of the fee they would charge if chosen to be the center, it

appears that the main motivation to participate in the process of institution formation is the potential rent associated with being a revenue maximizing center.

In this context, the model yields a clear answer to both questions above. In the first place, there exists a region in the parameter space in which a potentially welfare enhancing institution does not arise. This is because individual and social incentives are not aligned, as to some extent each individual fails to internalize the cost that he imposes on others by opting out of a potential institutional arrangement.

Such an inefficiency is more likely for societies of intermediate size. Groups that are too small are optimally left to the informal type of interaction. Although this is not explicitly in the model, an additional intuitive reason here is that within small enough groups, over time bilateral meeting between two specific individuals are more frequent and coordination is therefore more likely to rely on simple reciprocal trust; in the terms of our model, it may be that N and α are inversely related. On the other hand, as the number of individuals grows large enough, the rent associated with being in charge of running the coordinating institution becomes large enough to ensure that it will emerge.

Moreover, a welfare enhancing institution may fail to emerge if the gap between the payoff from non cooperation and cooperation is not very large, that is, if what we called trust in the state of nature is high enough. Because the outside option is not that bad, agents are more reluctant to engage in a costly institutional process. This intuitive negative correlation between institutions and the level of trust sheds light on one of the fundamental identification problem that arises in the empirical literature on social capital (see Durlauf and Fafchamps, 2006). Indeed, it seems to be the case that when formal institutions are weak, social capital (understood for example as trust in our model) substitutes for them. When formal institutions grow stronger, a process that often occurs along the path of development, some form of social capital may be destroyed or become less important (see Routledge and von Amsberg, 2003, for theoretical examples of such effects). We may therefore observe a negative correlation between measures of trust for example and social or economic outcomes, but rather than reflecting some causal link, it is the result of a fundamental endogenous link between social capital and more formal institutional forms, of the type uncovered in our model.

Second, our model makes a step towards understanding the observed heterogeneity of institutions. Indeed, even when the institution emerges, the

commitment problem mentioned above implies that it may arise at various level of efficiency, and in particular that it may be suboptimal, in the sense that it will charge a fee that is above the welfare maximizing level. Again, this is because of the absence of a collective commitment device to set the institutional fee in advance, which allows the chosen center to adopt a revenue maximizing strategy.

However, contrary to the previous one, this type of inefficiency is more likely to happen for low levels of trust, i.e. when the gap between non cooperative and cooperative payoffs is large. So different societies face different potential problems. When trust is low, a welfare enhancing institution is likely to arise but will probably be too extractive in nature. In a sense, this is the price to pay for coordination to be enforced in a context in which the loss from non cooperation is large. On the other hand, when trust is high, an institution may not arise, but if it does, it is more likely to be efficient. Indeed, because the gain from formal coordination are relatively low in that case, an institution that would be too extractive is unlikely to be individually incentive compatible in the first place.

We then show that the two types of inefficiencies stem from the lack of individual and collective commitment. However, there is a fundamental asymmetry here, in the sense that individual commitment to remain inside the institution even if not chosen to be the center would not change the results above if not accompanied by some degree of collective commitment to the fee that will be charged ex post. On the other hand, collective commitment goes some way towards solving the second type of inefficiency, the excessive level of rent extraction, and if accompanied by individual commitment to participate whatever the ex post assignment of roles, it does restore the first best.

The question of course is how commitment may arise endogenously in a world in which no coordination device or authority exist ex ante. We show that the threat of secession by subgroups of agents may generate such collective commitment, at least when the level of trust is low enough and the number of agents not too large. On the other hand, as this number becomes large enough, secession becomes unavoidable, resulting in a multi-institution world. In the basic version of our model, this always reduces welfare compared to a unique central institution. However, we indicate that transaction cost considerations may introduce a trade-off here, if for example coordination in smaller groups is characterized by lower such costs. Endogenizing these transaction costs is an interesting area for future research and would

make it possible to better understand situations characterized by multiple institutional layers.

Finally, the reader may argue that we assuming identical agents is at odds with reality. |First, because we were interested in other, mainly environmental, factors that may hinder or foster the emergence of institutions, abstracting from exogenously imposed individual heterogeneity allowed to better identify the effect of such factors. Moreover, this heterogeneity can be self-explanatory of the individual differences that we observe once institutions have arisen. Instead of assuming them for a start, we have analyzed here how the process of institution formation actually creates these differences. It is clear, however, that individual heterogeneity represents an interesting avenue for further research and in the future we intend to explore the impact of endowment inequality on the results of the present paper. This may have interesting implications, in particular in the field of development economics.

References

[1] Acemoglu, D. 2003. Why Not a Political Coase Theorem? Social Conflict, Commitment and Politics. *Journal of Comparative Economics*, **31**, 620-652.

[2] Acemoglu, D., J. Robinson and T. Verdier. 2004. Kleptocracy and Divide-and-Rule: A Model of Personal Rule. *Journal of the European Economic Association Papers and Proceedings*, **2**, 162-192.

[3] Aoki M., K. Murdock, and M. Okuno-Fujiwara. 1997. Beyond the East Asian Miracle: Introducing the Market Enhancing View. In M. Aoiki, H. Kim and M. Okuno-Fujiwara, Eds., *The Role of Government in East Asian Economic Development: Comparative Institutional Analysis*. Oxford University Press: Oxford.

[4] Attali, J., 2003. *L'Homme nomade*. Fayard.

[5] Azuma, Y. and H. Grossman. 2002. A Theory of the Informal Sector. NBER working paper 8823.

[6] Bardhan, P.K. 2005. *Scarcity, conflicts and cooperation: essays in the political and institutional economics of development*. MIT Press: Massachusetts.

[7] Basu, K. 2000. *Prelude to Political Economy*. Oxford University Press: Oxford.

[8] Bernstein, L. 2001. Private Commercial Law in the Cotton Industry: Creating Cooperation through Rules, Norms, and Institutions. *Michigan Law Review*, **99,** 1724-1788.

[9] Bernstein, L. 1992. Opting out of the Legal System: Extralegal Contractual Relations in the Diamond Industry. *Journal of Legal Studies*, **21**, 115-157.

[10] Choi, J.P. and M. Thum. 2002. Corruption and the Shadow Economy. CESifo Working Paper 633.

[11] Dawes, R.M. and R.H. Thaler. 1988. Cooperation. *Journal of Economic Perspectives*, **2**, 187-197.

[12] Dixit, A. 2004. *Lawlessness and Economics: Alternative Modes of Governance*. Princeton University Press.

[13] Durlauf, S. and M. Fafchamps. 2006. Social Capital. In *Handbook of Economic Growth*, P. Aghion and S. Durlauf, Eds., North Holland: Amsterdam.

[14] Ensminger, J. 1992. *Making a Market. The Institutional Transformation of an African Society*. Cambridge University Press: New York.

[15] Fafchamps, M. 2004. *Market Institutions in Sub-Saharan Africa*. MIT Press: Massachusetts.

[16] Greif, A. 1993. Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition. *American Economic Review*, **83**, 525-48.

[17] Greif, A. 1997. Microtheory and Recent Developments in the Study of Economic Institutions through Economic History. In D.M. Kreps and K. F. Wallis, Eds., *Advances in Economic Theory (vol. 2)*. Cambridge University Press: New York.

[18] Grossman, H. 2002. "Make us a king": Anarchy, Predation, and the State. *European Journal of Political Economy*, **18**, 31-46.

[19] Howe, R.E. and J.E. Roemer. 1981. Rawlsian Justice as the Core of A Game. *American Economic Review*, **71,** 880-895.

[20] Kandori, M. 1992. Social Norms and Community Enforcement. *Review of Economic Studies*, **59**, 63-80.

[21] Kranton, R. 1996. Reciprocal Exchange: A Self-sustaining System. *American Economic Review*, **86**, 830-851.

[22] Loayza, N., 1996. The Economics of the Informal Sector: A Simple Model and some Empirical Evidence from Latin America. *Carnegie-Rochester Conf. Series Public Policy*, **45**, 129-162.

[23] Marcouiller, D. and L. Young. 1995. The Black Hole of Graft: The Predatory State and the Informal Economy. *American Economic Review*, **85**, 630-646.

[24] Marwell, G. and R. Ames. 1981. Economists Free Ride, Does Anyone Else? *Journal of Public Economics*, **15**, 295-310.

[25] McMillan, J. and C. Woodruff. 2000. Private Ordering under Dysfunctional Public Order. *Michigan Law Review*, **98**, 2421-2458.

[26] McMillan, J. and C. Woodruff. 1999. Dispute Prevention without Courts in Vietnam. *Journal of Law, Economics & Organization*, **15**, 637-658.

[27] Milgrom, P., D. North and B. Weingast. 1990. The Role of Institutions in the Revival of Trade: The Medieval Law Merchant, Private Judges and the Champagne Fairs. *Economics and Politics*, **1**, 1-23.

[28] Morgan J. 2000. Financing Public Goods by Means of Lotteries. *Review of Economic Studies*. **67**, 761-84.

[29] Moselle B. and B. Polak. 2001. A Model of a Predatory State. *Journal of Law, Economics, and Organization*, **17,** 1-33.

[30] North, D. 1990. *Institutions, Institutional Change, and Economic Performance*. Cambridge University Press: Cambridge.

[31] Okzaki, T. 1997. The Government-Firm Relationship in Postwar Japanese Economic Recovery: Resolving the Coordination Failure by Coordination in Industrial Rationalization. In M. Aoiki, H. Kim and

M. Okuno-Fujiwara, Eds., *The Role of Government in East Asian Economic Development: Comparative Institutional Analysis*. Oxford University Press: Oxford.

[32] Olson, M. 1965. *The logic of Collective Action*. Harvard University Press: Massachusetts.

[33] Richman, B.D. 2006. How Community Institutions Create Economic Advantage: Jewish Diamond Merchants in New York. *Law and Social Inquiry*, **31**, 383-420.

[34] Ross, I.S. 1995. *The Life of Adam Smith*. Clarendon Press: Oxford.

[35] Routledge B. and J. von Amsberg. 2003. Social Capital and Growth. *Journal of Monetary Economics*, **50**, 167-193.

A Appendix

Proof of Proposition 1. Since $a(K)$ is increasing in K, only corner configurations can prevail, in the sense that no intermediate number of formal agents $0 < K < N$ can be supported as an equilibrium of this stage game. Suppose that $a \leq a(K)$ so no formal agents wants to deviate. Then, since we also have $a < a(K+1)$, informal agents would deviate and become formal, leading to full formality. Similarly, if $a \geq a(K+1)$, which is the necessary condition to sustain $N - K$ informal agents, formal agents would have an incentive to defect to informality, leading to an equilibrium with only informal agents. ■

Proof of Proposition 3. The condition (7) comes from just comparing the welfare under full formality with expression (6). On the other hand

$$\frac{\partial \underline{x}(N,\omega,\alpha)}{\partial N} = \frac{1}{(N+1)^2}\left(\frac{\left(\omega - a^F\right)x(a^F) + a^F - c}{\alpha\omega} - \frac{1}{\alpha}\right),$$

where we make use of the fact that, regardless of whether the solution is interior or not, a^F does not depend on N. It can be shown that $\frac{\partial \underline{x}(N,\omega,\alpha)}{\partial N} > 0$.

On the other hand,

$$\begin{aligned}\frac{\partial \underline{x}(N,\omega,\alpha)}{\partial \omega} &= \frac{N}{N+1}\frac{a^F(x(a^F)-1)+c}{\alpha\omega^2} \\ &\quad +\frac{N}{N+1}\frac{1}{\alpha\omega}\frac{\partial a^F}{\partial \omega}(-x(a^F)+\left(\omega-a^F\right)x_a(a^F)+1).\end{aligned}$$

Note first that the expression in brackets in the second term is the FOC of the planner's problem and hence it is nonnegative. Second, if $a^F = a(N)$, $\frac{\partial a^F}{\partial \omega} = 1-\alpha > 0$ and then is clear that the lower bound $\underline{x}(N,\omega,\alpha)$ is increasing in ω. On the other hand, when $a^F = a^*$ the bracketed term is equal to zero since the FOC of the planner's problem is binding.

Finally,

$$\begin{aligned}\frac{\partial \underline{x}(N,\omega,\alpha)}{\partial \alpha} &= \frac{1}{N+1}(-\frac{1}{\alpha^2}-N\frac{\left(\omega-a^F\right)x(a^F)+a^F-c}{\alpha^2\omega} \\ &\quad +\frac{N}{N+1}\frac{1}{\alpha\omega}\frac{\partial a^F}{\partial \alpha}(-x(a^F)+\left(\omega-a^F\right)x_a(a^F)+1).\end{aligned}$$

Again, when $a^F = a(N)$ then $\frac{\partial a^F}{\partial \alpha} = -\omega > 0$ and $\underline{x}(N,\omega,\alpha)$ is decreasing in α and; when $a^F = a^*$ the second term is equal to zero. ■

Proof of Proposition 4. If the center sets any fee not greater than $a(N)$, all agents will be formal. Hence, a revenue maximizer center will choose $a(N)$. When (11) holds, this ensures the emergence of the institution.

To address the potential multiplicity of equilibria, consider the following strategy on the side of agents:

$$F = \begin{cases} 1 & \text{if } a = a' \\ 0 & \text{otherwise} \end{cases}$$

It is clear that this strategy can be supported in equilibrium whenever $a' \leq a(N)$. If condition (11) holds, full formality can be supported if agents use this strategy for $a' = a(N)$. ■

Proof of Corollary 1. The comparative statics on N can be derived by noting that $W^F_{a(N)}$ is increasing in N, while the upper and lower limits do not depend on N (since $a(N) = \omega(1-\alpha)$). Rewriting $W^F_{a(N)} = \frac{N}{N+1}\left[a(N)-c)+(\omega-a(N))\,x(a(N))\right]+\frac{\omega}{N+1}$, the derivative with respect to N is given by

$$\frac{\partial W^F_{a(N)}}{\partial N} = \frac{\omega\left(\alpha x\left(a\left(N\right)\right) - \alpha\right) - c}{(N+1)^2},$$

which is positive since by assumption $\omega(1-\alpha) > c$ and $\alpha x\left(a\left(N\right)\right) > 1$.

On the other hand, the effects of the level of status-quo trust α can be estimated in the following way. Differentiating

$$W^F_{a(N)} = \frac{N}{N+1}\left[\omega - c + \alpha\omega\left[x(\omega(1-\alpha)) - 1\right]\right] + \frac{\omega}{N+1},$$

with respect to α, we get that

$$\frac{\partial W^F_{a(N)}}{\partial \alpha} = \frac{N}{N+1}\omega\left[x\left(\omega(1-\alpha)\right) - 1 - \alpha\omega x'\left(\omega(1-\alpha)\right)\right],$$

while the derivative of the upper bound is given by:

$$\frac{\partial\left[\omega\alpha x(a(N))\right]}{\partial\alpha} = \omega\left[x\left(\omega(1-\alpha)\right) - \alpha\omega x'\left(\omega(1-\alpha)\right)\right].$$

Since $\frac{\partial[\omega\alpha x(a(N))]}{\partial\alpha} > \frac{\partial W^F_{a(N)}}{\partial\alpha}$, and $W^F_{a(N)} > \omega\alpha x(a(N))$ for α close to 0 (the right hand side then tends to 0), we deduce that there is a threshold value $\underline{\alpha}$ such that formality is only implemented through the decentralized procedure if $\alpha < \underline{\alpha}$. Note that depending on the value of the parameters, it might be the case that $\underline{\alpha} > 1$, so no inefficiency arises. ■

Proof of Proposition 7. Recall from our discussion in Section 3 that there existed a value of the status quo trust denoted by α^{**} such that $a^{**} \geq a(N)$ whenever $\alpha \geq \alpha^{**}$. In that case, $\min\{a^{**}, a(N)\} = a^{**}$ and the threshold $N_0(\alpha,\omega)$ applies. By the Implicit Function Theorem,

$$\frac{\partial N_0(\alpha,\omega)}{\partial\alpha} = N(N-1)\omega\frac{1 - x(a(N)) + \alpha\omega x_a(a(N))}{\alpha(\omega x(a(N)) - \omega - c)}.$$

Note that the denominator is the FOC of the utilitarian planner problem. We know that when $\alpha < \alpha^*$ then $a(N) < a^*$, and the numerator is negative (positive otherwise).

Similarly, for $\alpha > \alpha^{**}$, $N_1(\alpha,\omega)$ becomes the relevant threshold and

$$\frac{\partial N_1(\alpha,\omega)}{\partial\alpha} = \omega\frac{x(a(N))(\omega - c) - \alpha\omega x_a(a(N))(a(N) - c) - \omega}{(a(N) - c)^2}.$$

Since in this case, $a(N) < a^{**}$, then $x(a(N)) > \alpha\omega x_a(a(N))$ so the denominator has a positive sign. Note as well, that this derivative evaluated at $\alpha = 0$ is positive, and that the denominator is decreasing in α. Hence, $N_1(\alpha, \omega)$ is everywhere increasing in α. ■

Proof of Proposition 8. When C is secession-proof the total sum of payoffs is simply

$$\begin{aligned} W_C^F &= \sum_{k=1}^{K}[(S_k - 1)(a(N) - c) + \omega + (S_k - 1)\alpha\omega x(a(N))] \\ &= (N + 1 - K)(a(N) - c + \alpha\omega x(a(N))) + K\omega. \end{aligned}$$

This expression is clearly decreasing in K, the number of coalitions in C. Therefore, the total sum of payoffs under any secession-proof coalition structure can never be greater than under the single institution (they are equal if the single institution is secession-proof itself). ■