The Emergence of Institutions*

Santiago Sánchez-Pagés and Stéphane Straub†

Abstract

This paper analyzes how institutions aimed at coordinating economic interactions may emerge. Starting from 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. Both individual and collective commitment problems stand in the way of institution formation. These problems imply first that a potentially efficient institution may fail to emerge and second that if it emerges, it may do so inefficiently. We show that big and untrustworthy societies are more likely to support institutions whereas their emergence is most difficult in small and trusting societies. However, institutions tend to be inefficient in less trusting societies. Finally, we show that the threat of secession by a subset of agents may alleviate the latter problem.

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

*We thank Andrzej Baniak, Eric Brousseau, Paco Candel-Sánchez, Avinash Dixit, Jan Fidrmuc, Francesco Giovannoni, Chris Kingston, Richard Langlois, Patrick Legros, John Moore, Boaz Moselle, Luis Garicano, Peter Grajzl, Emmanuel Raynaud, József Sákovics, Jonathan Thomas, Charles Vellutini, Irenaeus Wolf and seminar audiences at Edinburgh, Birmingham, CEU Budapest, Esnie 2006, ISNIE 2007, and the 11th Coalition Theory Workshop in Warwick, for their useful comments. Sánchez-Pagés also acknowledges financial support from the Spanish Ministry of Education and Science, grant SEJ2005-02079.

†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.

"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

1.1 Motivation and overview

To date the literature on the economic analysis of social and political institutions have focused mainly on their role as protectors of property rights.[1] 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.

The contributions along this line of enquiry have modeled institutions as a set of self-enforcing constraints on behavior, usually relating them to narrative descriptions derived from historical sources.[2] In the words of Dixit (2004), a distinction can be drawn depending on how the enforcement of coordination or cooperation is attained: On the one hand, models of self-governance, understood as equilibria sustained by some type of relation-based multilateral mechanism (punishment, communication, etc.)[3], and on the other hand models of formal rule-based institutional arrangements, whereby some qualified agent(s) takes on the role of solving coordination problems among the rest of the population.[4]

All these works describe institutional arrangements already in place, but very little has been said on the factors that lead to the emergence of these institutions in the first place. The aim of this paper is precisely to model the process through which they may or may not arise. In doing so, we focus specifically on the class of rule-based coordination mechanisms, in which a delegation process gives rise to a specialized institution responsible for enforcing some type of behavior by the agents that compose society. We

[1] See Bardhan (2005).

[2] See Greif (1997).

[3] See Greif (1993), Kandori (1992) and Ellison (1994) inter alia. A more detailed literature review is carried out below.

[4] Milgrom, North and Weingast (1990), Dixit (2003a).

focus therefore on the emergence of an institution capable of supplying this type of rule-based enforcement in the context of "individualist" societies, as Dixit (2003a) put it. We assume that, although social groups may exist, interactions mostly involve strangers and cooperation cannot rely on the type of multilateral mechanisms, norms and communication that prevail in the context of more "collectivist" societies such as families, or small and homogenous ethnic or religious groups.

Our point of departure is an economy in which the value of each individual's endowment is enhanced by interacting with others. In its simplest form, this is an informal economy that lacks an 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 benefit both parties but being opportunistic is a dominant strategy and in equilibrium very low payoffs are realized. Agents would like to remedy this inefficiency by finding a way to coordinate at the Pareto efficient outcome and ensure that it is enforced in any bilateral interaction.

The self-enforcing nature of institutions is modelled through a game agents play in this hypothetical state of nature. Depending on the existing incentives, players' actions will eventually lead to the establishment of a formal coordinating mechanism as an equilibrium of the game. 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 of the institutional intermediaries described in Milgrom et al. (1990) or Dixit (2003a) for example. For the sake of tractability, and because we want to focus on the potential emergence of this body, we choose not to model explicitly its internal functioning.

For such formal institution to arise, agents need to delegate to one of them the task of running it. The 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. Then, if the institution arises, agents have to decide whether to become formal and abide by its norms of interaction or not. Whenever two formal agents meet, the institution can guarantee that the efficient outcome will result. However, in order to enjoy this benefit, 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 impact of these different processes on efficiency and welfare.

In a nutshell, we find that individuals' motivation to participate in the process of institution formation is the possibility to be benefited with the rent falling to the center's share. Because of this and the fact that they do not fully internalize the social benefit linked to their participation in the process, 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 center is insufficient, and she prefers to renege ex post, leading society to fall back into informality. The second one, which we label "collective commitment" problem, is linked to the fact that agents may not be able to write binding agreements on the fee that will be charged by the center once it is designated.

Both limitations on commitment have implications for efficiency. The first aspect implies that due to the lack of individual incentives an institution may not arise despite being potentially welfare enhancing. This is in particular the case when the extent of the coordination problem is limited. The intuition is that when the level of trust in the state of nature is relatively high, the outside option in which no institution emerges is more attractive and agents' willingness to pay in order to create the institution is lower. On the other hand, the lack of collective commitment implies that even if an institution emerges, it may do so at a sub-optimal level of efficiency because the fee finally charged may be too high from a social point of view. This happens for low levels of trust in the state of nature, because in that case the institution is able to set a high fee compared to the first-best level.

We show that small societies with high levels of trust are less likely to support the emergence of institutions than big ones with low levels of trust, but if institutions do emerge, they are more likely to be inefficient in the latter type of societies.

Exogenously imposed commitment along each one of the 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 constrain the center's ability to extract too much resources in repeated interactions.

The second potential improvement, which again limits the ability of the center to charge a sub-optimal level of the fee, is the threat of secession by a subset of agents who could be better off by forming their own mini society. To deter blocking, the institution must charge a fee that cannot

be improved upon by any coalition. 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, we review the related literature. Section 2 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. In Section 5 we endogenize commitment looking at threshold strategies and analyzing the stability of institutional rules against coalitional deviations. Section 6 offers a discussion of the results and concludes. Proofs are relegated to the Appendix.

1.2 Literature review

Previous works on the role of institutions as coordination devices have mainly explored two related lines of enquiry. First, they have analyzed the functioning of specific existing institutional arrangements, in the light of both empirical accounts and game-theoretical modelling. Examples are found in the economic history literature[5] 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 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 by 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 the private arrangements in the diamond industry analyzed by Bernstein (1992) and Richman (2006).

As discussed above, a useful distinction can be made between relation-based type of mechanisms that rely on multilateral enforcement, such as the ones described in Greif (1993) and modelled for example in Kandori (1992) and Ellison (1994), and formal rule-based type of institutions characterized by bilateral enforcement of interactions (Milgrom et al., 1990, Bernstein,

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

2001, and Richman, 2006 among others fall into that category). Relation-based multilateral enforcement generally prevails in relatively homogenous groups, while bilateral rule-based enforcement mechanisms are more likely in the context of relatively anonymous interactions in large groups.[6]

Contributions opening the black box of institutions sustaining rule-based enforcement include the analysis of judges' role in the Champaign fairs by Milgrom et al. (1990), account of Genoese traders in Greif (1997), or of a specialist in violence able to commit to the use of its military capabilities to retaliate upon deviating agents in Bates and al. (2002). This shows that such institutions are potentially very diverse and can rely not only on coercion, as in the "Hobbesian" approach of Bates et al. (2002), but also on more subtle forms of persuasion, e.g. as a "Humean" *political entrepreneur* or government, able to persuade others to take a particular action or alter their beliefs about this action's consequences, as in Taylor (1982) or Basu (2000).

Further examples of contributions describing formal institutions of the type we posit here can be found in various contexts. These include tribes developing formal trade exchanges: Attali (2003) provides examples of the introduction of witnesses or legitimators certifying the validity of exchanges in early societies of Africa, aboriginal Australia or precolombian Nicaragua among others. Mafia and organized crime examples are also relevant, as for example Gambetta's (1993) account of the role of Peppe in ensuring that cattle breeders and butchers don't cheat each other when transacting in animals. The prominent role of the state in the East Asian development process (see Aoki et al., 1997) or in 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. Accounts from the diamond and the cotton industry mentioned above show that similar issues may arise for a group of firms in a given industry faced with coordination problems.

Second, a few economic contributions have analyzed how informal or personalized relationship-based institutions may coexist with more formal, rule-based 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

[6] See Greif (1994). Dixit (2003b) and Li (2003) provide theoretical foundations.

finally triggered the Orma tribe in Kenya to move from a rule by a council of elders to the recognition of the authority of the modern Kenyan nation-state.

Finally, our research is also related to contributions concerned with the emergence of the State. The concept of the State as an entity solving state of nature coordination problems is a long-standing one, as exemplified by Taylor (1982) or Basu (2000). Bates, Greif and Singh (2002) provide numerous historical examples to support their theoretical account of why agents may seek to attribute the monopoly of violence to a delegate, a canonical State, in charge of ensuring peace and enabling higher levels of production. 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 look at how its presence shapes economic outcomes or compare the scenarios with and without ruler (Grossman, 2002; Moselle and Polak, 2001). However, rather than drawing conclusions for specific type of interactions or environments, our analysis aims at uncovering general principles that can help us understand the process of institutional creation in different economic and social contexts.

2 The Model

Consider an economy populated by $N + 1$ agents, who have an initial endowment 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	NC
Player i	C	*x,x*	*-z,z*
	NC	*z,-z*	*0,0*

Figure 1: Basic game

Agents are anonymous to each other. They are randomly and bilaterally matched 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 strategy C stands for a cooperative behavior that can create added value, and NC stands for an opportunistic behavior that allows the agent to take advantage of a cooperating partner but yields zero returns otherwise.

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. As is common in the literature, we have chosen this game as the simplest way to illustrate the type of coordination problems that occur in many social situations, but many other games could capture the trade-offs we study here, like for instance a coordination game with two equilibria, one Pareto-superior to the other.

The scenario in which individuals are randomly and bilaterally matched and play G without any interference is assumed to be the status-quo of the economy. In order to solve the problem of opportunism, agents can set up an institution with the power to enforce cooperation and ensure that the efficient outcome (C, C) results from any interaction between agents under its oversight. Given that we are mainly interested in studying when such institutions can arise, and that in doing so we want to keep the analysis tractable, in what follows we just consider a reduced form of the actual process the institution employs to enforce cooperation.

This institution arises when agents delegate to one of them, who we will call the *center,* the task of running it. The central agent relinquishes her ability to interact with other agents, so she must 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.

If the institution arises, agents have to decide whether to abide by it, that is to become formal, or not to do so and remain informal. However, in order to become formal and interact under the institutional umbrella, they have to pay a fixed fee $a \leq \omega$, that can be understood as an entry fee or a lump-sum tax that rewards the center for her activity. 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.[7] Hence, the per-person unit return from an interaction between two formal agents is

$$v_a^F = x(a), \tag{1}$$

where the superscript F denotes "Formal" and $x(\cdot)$ satisfies $x_a > 0$, $x_{aa} < 0$, and the standard Inada condition, $\lim_{a\to 0} x_a(a) = \infty$. We thus assume that the institution becomes more efficient when endowed with more resources, as it is able to monitor better its members' behavior or to invest more in physical or relational supporting infrastructure, as for example 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 interference. Informal agents thus avoid paying the fee but their interactions yield lower returns. They will sometimes result in (NC, NC) being played. Still, in this state of nature, agents may also occasionally cooperate with each other despite the absence of material incentives to do so or of any formal institution enforcing coordination.[8] Otherwise, any of the other two possible outcomes (C, NC) and (NC, C) might also be played.

As a result, the per-person *expected* unit return when G is played between a formal and an informal agent or between two informal agents is

$$v_a^I = \alpha x(a), \tag{2}$$

where the superscript I denotes "Informal".[9] We refer to the parameter $\alpha < 1$, which captures the expected return from interactions in a world

[7] We could envision making x also a function of the proportion of agents $\frac{K}{N}$ contributing to it. However, it is unclear how this would affect x. Indeed, a higher proportion could have a positive effect because of network externalities for example, but congestion could also lead to a negative effect (see Kranton, 1996).

[8] 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) and that some players are unconditional cooperators/defectors (Andreoni and Samuelson, 2006). 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.

[9] An alternative interpretation of the parameter α, in line with the literature on informality, is the level of free-riding that informal interactions can make on formal institutions.

without institution, as the level of trust or cooperation in the society under the state of nature. To keep the model tractable, this is assumed to be an initial condition of our economy that depends upon culture, expectations and the specific type of interactions considered.[10] Note that a full characterization of α could be derived as the equilibrium outcome in a framework in which agents have varying degree of trustworthiness, for example if a fraction of them are unconditional cooperators/defectors, while others are conditional cooperators (see Dixit, 2003a).

Finally, we 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.

Under risk neutrality, the expected payoff of a formal agent when $K \geq 2$ agents are formal is then:

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

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

$$V_{K,a}^{C} = K(a - c),$$

where c is the fixed enforcement cost per formal agent, linked for example to the need to record and maintain information on its behavior, maintain proper infrastructure, etc. Note that while this formulation assumes that this cost is independent of whether the formal agents ends up interacting with another formal agent or not, in our setting it will be strictly equivalent to a formulation with a cost contingent on the nature of the transaction since, as we will see below, either all or no agent become formal. The alternative assumption would not substantially change our results.

See for example Marcouiller and Young (1995), Choi and Thum (2002), Azuma and Grossman (2002) and Straub (2005). Note that the way we define α can accommodate the fact that informal agents free-ride only imperfectly on the institution even when playing (C, C), so they get $\beta x(a)$, where $\beta \in (0, 1)$.

[10] As mentioned in the final discussion, it is intuitive for example that α might be inversely related to N.

Figure 2 summarizes the timing of the game described above.

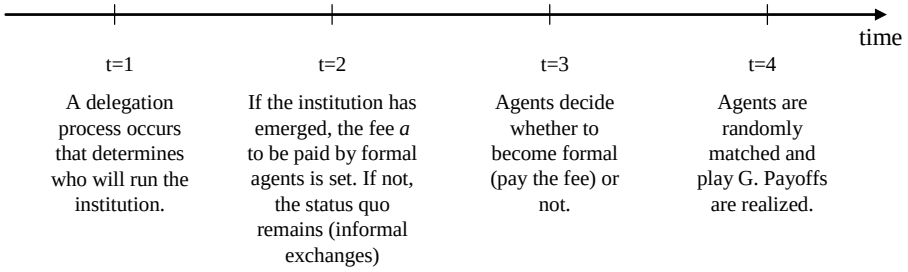

Figure 2: Timing of the game

We have thus constructed a game in four stages. In the first stage, agents set up the institution (the game ends there if they don't succeed in doing so). 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.

A formal agent will not prefer to deviate and become informal as long as $V^F_{K,a} \geq V^I_{K-1,a} = \omega v^I_a$. 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 agent receives a payoff:

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

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 $K = 1$, $a(1) = 0$.

The equilibrium level of formality will 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, $v^F_a = v^I_a = \alpha x\,(0)$ 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 could not 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,*

(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 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. For the rest of the paper, we will mainly focus on the equilibrium in which the institution forms. We see it as more natural because at that point of the game all agents have decided to participate in the process, a central agent has been chosen and the fee has already been set. Still, we will leave a further discussion on the informal equilibrium to Section 5.

Figure 3 depicts the profile of equilibria as a function of the fee a.[11]

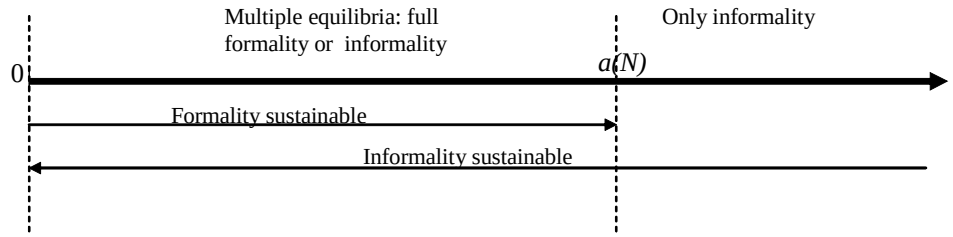

Figure 3: Profile of equilibria

3.2 The first best institutional fee

In the remainder of this Section, we characterize the optimum fee from the viewpoint of a hypothetical central planner willing to maximize the total sum of agents' utilities. The planner compares the maximum welfare attainable in the scenario in which agents pay a fee to enjoy the benefits of formal interactions 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:

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

Given our assumptions on $x(\cdot)$, this objective function above is concave and hence the first order condition of the program implicitly characterizes an interior solution a^* such that:

$$\omega - a^* = \frac{x(a^*) - 1}{x_a(a^*)}. \tag{4}$$

The fee 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

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

[11] 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 $a = a(N)$ so $p(a(N)) = 1$. Hence, in this case full formality would arise as well.

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:[12]

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

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. On the other hand, 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.

Hence, for small and relatively poor economies (low N or ω) full formality does not need to 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.

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

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, given a 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.

In the absence of an explicit coordination device, a lottery is the simplest mechanism to choose the individual to be in charge of the institution. It is so because all agents are alike and hence there is no reason why *a priori* they should not be equally likely to end up in charge of the institution. Moreover, a lottery also constitutes the reduced form of more complex mechanisms, like for instance auctions or contests: If they were used, the equilibrium would be typically symmetric and a winner should be therefore randomly chosen.

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, *ex-ante* agents must decide simultaneously whether to participate or not in the lottery that will designate who will run the institution. Hence, given a level of the fee a, the institution can arise only if

$$\frac{1}{N+1}(N(a-c)+\omega)+\frac{N}{N+1}\left(\omega-a\right)x(a)\geq\omega\alpha x(a), \tag{8}$$

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 next rely on 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, once they discover their role, agents should not prefer to fall back into informality. As this *ex-post* requirement is always satisfied for an agent who does not become the center[13], the center's *ex-post* participation constraint, given a certain level of the fee a, is the relevant one:

$$N(a-c)+\omega \geq \omega\alpha x(0). \tag{9}$$

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 into informality.[14]

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. 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.[15] Table 1 below 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] Note that because all agents are alike and either all or none participate in the lottery ex ante, we don't have to worry about the case of an agent not participating in the lottery but willing to become formal ex post.

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

	No individual commitment (ex post participation constraint)	**Strong individual commitment (ex ante participation constraint)**
No collective commitment : 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.
collective commitment : 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 the 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, the fee is freely set by the central agent. Hence, 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:

$$\frac{1}{N+1}(N(a(N)-c)+\omega)+\frac{N}{N+1}\left(\omega-a(N)\right)x(a(N))\geq\omega\alpha x(a(N)), \quad (10)$$

$$N(a(N)-c)+\omega\geq\omega\alpha x(0), \quad (11)$$

which are simply the result of rewriting the ex ante lottery participation constraint (8) and the ex post constraint of the center (9) 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, (10) can be rewritten as:

$$N(a(N)-c)+\omega\geq\omega\alpha x(a(N)), \quad (12)$$

from which it is evident that (10) is a stronger constraint[16], 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.

Proposition 4 *If condition (12) 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 (12) 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

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

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*

$$\omega\alpha x(0) \leq W^F_{a(N)} \leq \omega\alpha x(a(N)). \tag{13}$$

Such inefficiency occurs in economies of intermediate size and when the status-quo level of trust α *is sufficiently high.*

The lower bound in (13) determines when formality is more efficient than informality, whereas the upper bound establishes when formality is implementable. Between 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

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

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. As in Bates et al. (2002), the agent endowed with the institutional role reneges it if she is not able to extract enough resources from the system.

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. We should then observe the emergence of formal institutions in societies plagued with coordination problems and low levels of informal trust, while informal exchanges are more likely to prevail 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. In conclusion, even if the central agent is able to maximize revenue when setting a, the emergence of a desirable institution is not ensured.

But even if full formality is implemented, the fee set by the central agent may be too high and the first best may not be attained. A necessary condition for this second type of inefficiency to occur is a low enough degree of trust in bilateral interactions, i.e. $\alpha < \alpha^*$, that implies $a^F = a^* < a(N)$.

Corollary 2 (Implementable first best) *When condition (12) holds, the first best fee a^F can be implemented in a SPE of the fully decentralized procedure of institution formation for high enough levels of status quo-trust, i.e. $\alpha \geq \alpha^*$.*

The intuition for this result is easy to grasp. When welfare is increasing over the range of fees compatible with formality or, in other words, when the level of status-quo trust α is sufficiently high, the planner would like to set the highest fee possible (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. Otherwise, the emergent institution will tend to be inefficient.

In conjunction, these two corollaries therefore imply that different societies are characterized by different inefficiencies. Institutions are more likely to emerge in societies with low levels of trust, but if they do, they will tend

to be too extractive relative to the socially optimal outcome. On the other hand, more trustworthy societies may find it difficult to generate formal coordinating institutions, but if they succeed, these are more likely to be efficient.

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 mechanism capable of enforcing it. After showing briefly how commitment may improve efficiency in the institution formation process under each of the possible combinations of assumptions considered in Table 1 above, we discuss how it may be enforced: In Section 5, we analyze in more detail two mechanisms that may endogenously support some degree of collective commitment despite full decentralization.

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 (8) 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 to run it, so the *ex-post* participation constraint (11) is dropped, but that the center retains total freedom to set the fee. Therefore, only condition (10) must hold. Since we know from Case 1 that condition (10) is stronger than (11), it is obvious that this does not introduce any change with respect to the benchmark no-commitment case. This 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 must be at least greater than $\underline{a}$. Obviously, agents' incentives must still be taken into account so the fee chosen has to be compatible with full formality (hence not above $a(N)$). 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 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 to be attained. On the other hand, even if that cannot be the case, 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 $a(N)$, and thus closer to the social optimum.

Case 4. Finally, consider the case where there is no *ex-post* participation constraint (strong individual commitment) and agents meet and agree that they should implement the first best.[17] It is easy to see that then the efficient outcome is always implemented.

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

[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 $a < a(N)$.

(8) 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 (lack of collective commitment) is a strong reason for the occurrence of inefficiencies, and one that cannot be alleviated by introducing individual commitment (Case 2). As this limit is relaxed, potential inefficiencies are reduced, as shown by Case 3. Finally, when the ability to set fees *ex-ante* is combined with individual commitment (Case 4), the first best is always implementable.

The next Section discusses two decentralized mechanisms through which some degree of commitment may be enforced.

5 Endogenous commitment

We have just showed that ex-ante collective commitment tends to increase the likelihood of the institution arising and hence efficiency. The next step is to endogenize it. In this Section we consider two mechanisms to achieve this: We first explore the use of threshold strategies in the benchmark version of our game. Then we study the possibility of coalitional secession.

5.1 Threshold strategies

In the discussion above, we assumed away the possibility that agents can employ a reversal to informality as a threat to the central agent. Recall from Proposition 2 that full informality can also be supported in equilibrium for any fee in the interval $[0, a(N)]$. Hence, this multiplicity of equilibria can make such threat credible. More formally, consider the following threshold strategy to be used by agents:

$$\mathcal{F} = \begin{cases} 1 & \text{if } a \leq a' \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

where $a' \leq a(N)$. If agents use these strategies, the central agent's best response under no commitment is then to choose a'. This threat thus enlarges the set of possible fees that can be supported in equilibrium and this opens the door to a welfare improvement. Still, the ex-post constraint must be satisfied, so this profile can only implement fees greater than $\underline{a}$.

Proposition 7 *When condition (12) holds and $a^* \geq \underline{a}$, the first best fee can be implemented in a SPE of the fully decentralized procedure of institution formation for relatively low levels of status-quo trust, i.e. $\alpha < \alpha^*$.*

Notice that the use of this type of strategies leads to a scenario very similar to the one in Case 3, because they act as a collective commitment device to choose and enforce a given level of the fee. Again, as in that case, even if the first best cannot be attained (because $a(N) > \underline{a} \geq a^*$), threshold strategies can help to reduce the inefficiency associated with a too extractive central agent.

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, which seem to involve some degree of multilateral agreement. 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.

5.2 Secession

Let us now 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 a 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 subgroup of agents can improve its situation by withdrawing and later applying among them the same procedure of institution formation used by the society as a whole.

Specifically, our analysis of secession will concentrate on the secession-proofness of the decentralized procedure of institution formation, assuming that it will be employed both by the whole population and any subgroup intending to withdraw. Then, 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 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 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 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 solves

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

[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

[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.

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

$$\omega - a^{**} = \frac{x(a^{**})}{x_a(a^{**})}. \tag{16}$$

Therefore it is clear that $a^{**} < a^*$. Again, 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 $\alpha^* < \alpha^{**}$. Hence, the level of the fee that maximizes the welfare of the set of interacting agents is either a^{**} or $a(N)$. Let us assume, for the sake of exposition, that $\min\{a^{**}, a(N)\} > \underline{a}$.

The set of secession-proof fees is thus non-empty if and only if

$$\begin{aligned} & \frac{1}{N}((N-1)(a(N)-c)+\omega) + \frac{N-1}{N}\alpha\omega x(a(N)) \\ \leq \; & (\omega - \min\{a^{**}, a(N)\})x(\min\{a^{**}, a(N)\}). \end{aligned}$$

If this condition is not met, we should expect the emergence of multiple institutions. When $a(N) > a^{**}$ this expression implicitly 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.

The next Proposition summarizes the conditions, in terms of the population size N and the level of status-quo trust α, under which secession may occur.

Proposition 8 *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 these thresholds in the parameter space.

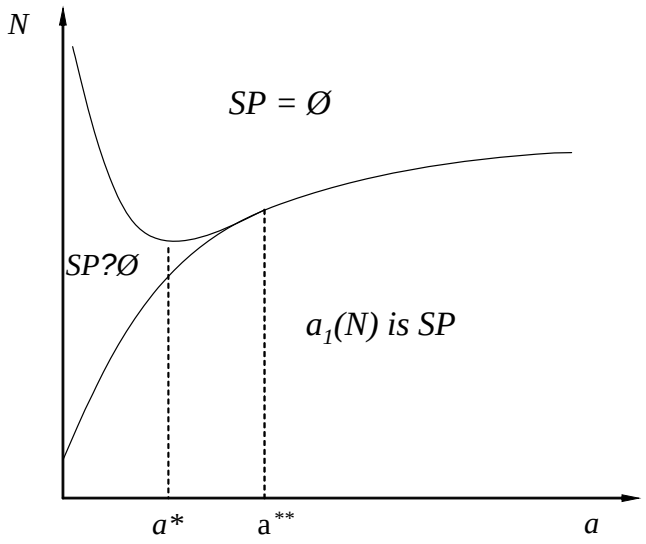

Figure 4: The set of secession-proof fees

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 still be possible for the institution to avoid secession by charging a fee below $a(N)$. 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 region of parameters, secession is a real threat that renders 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 into 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 structure all groups will set $a(N)$ since it is the unique secession-proof 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 9 *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, a trade-off arises and the above conclusion may require qualification.

6 Discussion and Conclusion

The contribution of this paper is to focus on the process through which institutions aimed at enforcing cooperation may actually emerge in a context in which no coordination device previously existed. 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.

We have built a model in which economic interactions take the form of a prisoner's dilemma game. In a hypothetical state of nature, agents from a population are randomly matched to play this game without any further interference and hence non-cooperation and inefficiency ensue. 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 this type of mechanisms can be found in the Economics, Sociology and Law literatures.

In a world of no commitment, in which 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, 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 clear answers to both questions above. First, there exists a region in the parameters 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 the 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 made explicit in our model, an additional intuitive reason for this may be that N and α are inversely related. On the other hand, as the number of individuals grows, 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 arise 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 the costly process of institution creation. This intuitive negative correlation between the likelihood of the emergence of formal institutions and the level of trust sheds light on one of the fundamental identification problems that arise 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 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, it may do so at various levels of efficiency, and in particular it may be suboptimal, in the sense that it will charge a fee that is above the welfare maximizing level. This is due to 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 gains 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 run the institution would not change the results above unless it is accompanied by collective commitment on the fee that will be charged ex post. On the other hand, collective commitment goes some way towards solving excessive rent extraction, and if accompanied by individual commitment, 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 spheres.

Another question relates to how the present conclusions are affected by the characteristics of the institution under analysis, considered here as a black box: Whether it is a purely informative instance, or one endowed with the power to reward or punish agents, whether it may incur in cheating or extortion, etc. This would allow for a better mapping between real world institutions and the theoretical mechanisms unveiled here, although it would also certainly represent a challenge in terms of analytical tractability.

Finally, we have assumed identical agents because we were interested in other, mainly environmental, factors that may hinder or foster the emergence of institutions. 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] Andreoni and L. Samuelson. 2006. Building Rational Cooperation. *Journal of Economic Theory*, **127**, 117–154.

[4] 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.

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

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

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

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

[9] Bates, R., A. Greif, and S. Singh. 2002. Organizing Violence. *Journal of Conflict Resolution,* **46**, 599-628**.**

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

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

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

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

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

[15] Dixit, A. 2003a. On Modes of Economic Governance. *Econometrica*, **71(2)**, 449-481.

[16] Dixit, A. 2003b. Trade Expansion and Contract Enforcement. *Journal of Political Economy*, **111**, 1293-1317.

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

[18] Ellison, G. 1994. Cooperation in the Prisoner's Dilemna with Anonymous Random Matching. *Review of Economic Studies*, **61**, 567-588.

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

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

[21] Gambetta, D. 1993. *The Sicilian Mafia: The Business of Protection*. Harvard University Press: Massachusetts.

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

[23] Greif, A. 1994. Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies. *Journal of Political Economy*, **102**, 912-50.

[24] 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.

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

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

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

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

[29] Li, S. 2003. The Benefits and Costs of Relation-based Governance: An Explanation of the East Asian Miracle and Crisis. *Review of International Economics*, 11, 651-667.

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

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

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

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

[34] 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.

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

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

[37] 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.

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

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

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

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

[42] Straub, S. 2005. Informal Sector: The Credit Market Channel. *Journal of Development Economics*, **78**, 299-321

[43] Taylor, M. 1982. *The Possibility of Cooperation.* Cambridge University Press: Cambridge.

A Appendix

Proof of Proposition 1. Since $a(K)$ is increasing in K, only corner configurations can prevail, i.e. 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). Taking the derivative of the right hand side with respect to N yields

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

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 $\frac{\partial \underline{x}(N,\omega,\alpha)}{\partial N} > 0$.

Similarly,

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

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 it 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 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 exists 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). ■