

Club-in-the-Club: Reform under Unanimity*

Erik Berglöf[†], Mike Burkart[‡], Guido Friebel[§], and Elena Paltseva[¶]

Abstract

In many organizations, decisions are taken by unanimity giving each member veto power. We analyze a model of an organization in which members with heterogeneous productivity privately contribute to a common good. Under unanimity, the least efficient member imposes her preferred effort choice on the entire organization. The threat of forming an “inner organization” can undermine the veto power of the less efficient members and coerce them to exert more effort. We also identify the conditions under which the threat of forming an inner organization is executed. Finally, we show that majority rules effectively prevent the emergence of inner organizations.

Key words: organizations, club good, voting rules, EU integration

JEL codes: D2, D7, P4

*We would like to thank the Editor and an anonymous referee, Tore Ellingsen, Sergei Guriev, Chloe Le Coq, and seminar participants in Berkeley, Copenhagen (KU), Moscow (CEFIR), Munich, Oxford, Stockholm (SITE) and Toulouse (IDEI). All remaining errors are our own.

[†]EBRD, ECGI and CEPR

[‡]Stockholm School of Economics, London School of Economics, CEPR, ECGI and FMG

[§]Goethe-Universität Frankfurt and CEPR

[¶]SITE, Stockholm School of Economics and New Economic School, corresponding author: elena.paltseva@hhs.se, SITE, Stockholm School of Economics, Sveavägen 65, PO Box 6501, SE-113 83 Stockholm, Sweden, +46 8 736 9688.

1 Introduction

Coordination of individual actions is the core problem that any society must solve to assure the well-being of its members. The greater part of the economics literature focuses on markets; arguably, organizations are an equally important coordination mechanism. When studying organizations, economists typically presuppose the existence of a governance system consisting of rules, penalties or transfers. Club theory, for instance, assumes that there is a system of transfers, taxes and entry fees that can be used to make members of a club behave in accordance with the common interest (e.g., Cornes and Sandler, 1996). Similarly, organization theory presupposes the existence of a principal who coordinates the members of an organization through the use of various monetary and non-monetary instruments (e.g., Milgrom and Roberts, 1992).

In many circumstances, cooperation and organizations exist even if there is no comprehensive governance system. Agents who share a common goal can form a “loosely knit” group. For example, sovereign states may come together to coordinate their actions in specific areas, such as economic policies, protection of the environment, or defence. In such situations, there is *a priori* no structure in place that determines how decisions are taken: The member states must first sort out how to decide. There is initially no other decision rule but unanimity, as pointed out by Rousseau.¹

Unanimity grants each member of an organization a veto right, thereby protecting her against coercion or what de Tocqueville (1835) called the “tyranny of the majority”. But the flipside of unanimity is slow and inflexible decision-making and underprovision of the common good. Heterogeneity is key here: Members who are less committed or less productive can veto any proposal to increase contributions (“effort”) to the common good. The problem of holding back other, more productive, members becomes particularly severe when there are complementarities between the members’ contributions. In the presence of such “weakest-link” effects, a member who invests too little, limits the amount

¹“Indeed, if there were no prior convention, where, unless the election were unanimous, would be the obligation on the minority to submit to the choice of the majority? How have a hundred men who wish for a master the right to vote on behalf of ten who do not? The law of majority voting is itself something established by convention, and presupposes unanimity, on one occasion at least” (Rousseau, 1762).

of the common good for the entire organization.

We argue that organizations operating under the unanimity rule can, nonetheless, provide more common goods than what their least committed members would prefer. The mechanism that can overcome the veto power of the least committed members is the threat of forming an inner club, or “club-in-the-club”.

To develop our argument we analyze the provision of a common good by an organization in which decisions are taken by unanimity. Members jointly produce the good, with each member’s effort as the inputs; effort should be broadly interpreted as any costly contribution to a common good. The members differ in terms of their effort cost. As each member has veto power, the common good provision is determined by the “weakest link”, that is, the member with the highest cost of effort. To capture the essence of the “weakest link”, we assume that the good is produced with a Leontief technology. This precludes more productive club members from compensating (i.e., substituting) effort underprovision by less productive members. For the same reason, we exclude transfer mechanisms that can induce club members to exert more effort than they individually prefer.

The mere possibility of forming a club-in-the-club can, however, increase the amount of the good provided by the entire organization. The inner club admits only high effort providers. When staying outside is costly or when the inner club is associated with deadweight losses, weaker members increase their effort in order to avoid that the inner club forms. The more committed members may in turn refrain from forming the inner club, once all members increase their effort. Thus, the threat of forming a club-in-the-club limits the leverage less committed members have by virtue of their veto power, and unanimity does not necessarily lead to stagnation.²

The club-in-the-club threat may also be executed. If so, an inner club provides its members an additional good, but imposes a deadweight loss on the original organization. We show that an inner club can emerge when the members are sufficiently heterogeneous

²Alternative reasons why less committed members may refrain from executing their veto rights such as reputation or log-rolling rely on repeated interaction and low time-discount rates. By contrast, our mechanism functions in a static model.

and the deadweight loss is sufficiently small. In such a case, more productive members prefer higher effort with fewer contributors to lower effort with more contributors. Putting it differently, the effort level the less committed members are willing to exert to avoid the formation of an inner club is not sufficient for compromise. The two results above imply that the club-in-the-club may lead to *more* integration - when the threat of an inner club increases the organization-wide effort, or *less* integration - when an inner club forms in equilibrium.

We finally look at the effect of a majority rather than a unanimity rule. Under a majority rule the decisive member is more productive than under unanimity, suggesting that a majority rule is a remedy against disintegration. This is indeed the case, and supermajority rules often suffice to prevent inner clubs. The required majority threshold depends on the characteristics of the organization and is lower when members are more heterogenous.

The logic of our theory applies to what we call loosely-knit organizations, that is, organizations that do not (yet) have a governance structure in place that resolves or alleviates incentive problems. The leading example is the European Union. In Section 7, we argue that important episodes of European integration are in line with the main ideas of our theory. On a more general level, our theory considers organizations with the following features: (i) members can engage in more than one purpose; (ii) inner clubs can be formed that pursue a purpose beyond the initial one, and the inner club's members can decide who is admitted; (iii) no member can be excluded from the initial club.

In principle, one could think of other loosely-knit organizations the model could apply to. For instance, home owners may form a residential community association to improve safety or leisure facilities; groups of concerned citizens may form an NGO to protect the environment or fight racism; national sport leagues may want to organize a European-wide tournament. The challenge for these loosely knit groups and organizations is similar – to find instruments to incentivize the members to contribute to the common goal.

However, the model fits these applications less well. First, the above organizations tend to lack the manifold of purposes the EU can engage in (such as trade, social issues,

research, common currency). Multi-dimensionality of purpose and the potential to keep some members out of inner clubs are, however, necessary ingredients for inner clubs to be an incentive device. Second, many other organizations that cover a broad range of purposes have governance structures in place to overcome the obstacles owing to the unanimity rule which is so important in the EU context. Thus, it seems to us that the best field of application (beyond the EU example) consists of other multi-country associations like the Southern Common Market (MERCOSUR).

Our theory is related to the literature on secession. Our mechanism has common features with the one in Buchanan and Faith (1987) on “internal exit” as an alternative to “voting with one’s feet”. In their theory, the optimal tax rate is the one that maximizes revenues of a ruling coalition under the constraint of not triggering secession of other members. In our theory, there is no ruling coalition and no taxes; everyone has a veto right about club-wide contributions, but no veto right exists concerning the possibility of forming an inner club. In Bolton and Roland (1997) secession involves lost economies of scale in public good provision, but avoids the “tyranny of the majority” by creating more homogenous political entities.³ In our theory, inner clubs make it possible to increase efficiency in the outer club and there is no tyranny of the majority. Rather, in the absence of the possibility of inner clubs, any individual is a tyrant as he can impose his preferred contribution on the entire group. In general, our focus is not so much on secessions – that is, the complete separation of federations – but rather on the creation of costly internal structures. Further, we emphasize the potential function of the inner organization threat as a mechanism to discipline less committed members. This contrasts with Gradstein (2004) who argues that secession rights, while protecting minority rights, involve inefficiencies in bargaining processes. In our model, internal threats can increase efficiency (because they can induce higher effort), or decrease it (as the formation of an inner club entails a deadweight loss). An additional distinction is that the above papers consider majority voting, while our main argument concentrates on the unanimity rule – the natural rule for organizations with highly incomplete constitutions.

³Bordignon and Brusco (2001) point out that constitutionally defined secession rights involve a trade-off: They reduce the cost of an actual break-up ex post, but they increase the likelihood of break-up.

The idea of an inner club and its impact on the members of the initial organization can also be found in the literature on cartel formation and its applications to the public goods and the formation of international environmental agreements (see e.g., d'Aspremont et al., 1983; Thoron, 1998; Diamantoudi and Sartzetakis, 2006; Weikard, 2009). This literature analyses cartel stability based on the incentives of the cartel members to leave and of the outsiders to join. Similarly, the emergence of an inner club in our setting is stable with respect to the participation decisions of the insiders and the outsiders. However, the cartel stability literature predominantly considers homogenous agents and universally assumes transferrable utility within the cartel - that is, that the cartel maximizes the sum of its members' payoffs. These two assumptions result in a unique outcome associated with a cartel of a given size. We consider a setting with heterogeneous members and do not allow for transfers. In this setting there is no justification for the utilitarian welfare perspective chosen by the cartel stability literature. Instead, we characterize the entire set of outcomes compatible with the existence of an inner club. Furthermore, the threat of an inner organization is a key point of our analysis, but is not addressed by the cartel formation literature.

Our model indicates that organizations may choose to abandon unanimity and subject their members to the will of the majority. This result is related to a growing literature analyzing how constitutions form, in particular, what determines the voting rules of a society. Aghion and Bolton (2003) identify a trade-off between minority protection and flexibility. To adapt to changes, a society must offer transfers to some individuals to prevent them from exercising their veto right. Hence, a society may under the veil of ignorance decide to replace unanimity by some type of majority voting. Messner and Polborn (2004) take a complementary view and show why societies may opt for supermajorities rather than simple majority voting. In their model, young people, who vote today over tomorrow's decision rule, anticipate that they will benefit less from reforms when they are old. Hence, they want to have more power about future reforms, which gives them an incentive to agree on a supermajority rule. Erlenmaier and Gersbach (2004) argue that first best outcomes can be achieved under unanimity, provided that it

is supplemented by a number of constitutional provisions, such as bundling of projects. Compared to all these papers, the structure of our model is more parsimonious, in particular, as we are excluding side payments. In addition, we focus on the effects of inner group formation on the efficiency of an organization in the absence of constitutional rules, i.e., under voluntary cooperation.

Our theory is related to the long-standing policy debate about the institutional structure of the EU. Possible reform models are discussed in Dewatripont et al. (1995). Our theory resembles the variable geometry and flexible integration models. In variable geometry, the EU would be separated in a core group and a periphery. In flexible integration, a country could be in the core for certain policy fields, and in the periphery for others. As we only allow for one club-in-the-club, our theory bears more resemblance with variable geometry, but flexible integration is also in the spirit of our theory, if more than one policy dimension were to be considered.

Dewatripont et al. (1995) are primarily concerned with finding the right governance structure to effectively deal with different policy dimensions. Our main interest is how to get members to contribute to reform efforts in the absence of an elaborate governance structure. Another difference to Dewatripont et al. (1995) lies in the considered reform opportunities. We examine an opportunity that provides the inner club members with a benefit, but imposes an externality on all members of the initial club. Dewatripont et al. (1995) investigate other opportunities, for instance, when members have different degrees of risk aversions, and hence some may be more willing to experiment than others. In that case, there would be positive externalities rather than negative ones, because more risk averse members can free-ride on the experimental efforts of the other members, which is quite a different (albeit interesting) story.

Harstad (2006) investigates how flexible cooperation (organization members can decide on the speed of integration) compares to rigid cooperation (all members go at the same speed). In his framework, flexibility has the advantage of protecting members with high costs from being forced to contribute to a public good. At the same time, flexibility gives rise to smaller contributions to the public good and/or free-riding, as positive ex-

ternality of the public good is enjoyed also by the non-participating members. Our paper is complementary: it considers negative externalities from an inner to an outer club and argues that the threat of forming an inner club can have a disciplining role, pushing members in the outer club to contribute more. In Dixit (2003) this role is played by network externalities. Owing to these externalities, agents may sequentially adopt an innovation (or join an organization) even though the introduction is not in their collective interest. That is, adoption is individually rational, unless agents can coordinate their actions. In our model, weaker members are in a similar situation – they would prefer the threat of forming an inner organization not to exist. In addition, stronger members can execute the threat and form an inner organization, a possibility not explored by Dixit.

The paper is organized as follows. Section 2 outlines and solves the basic model in which inner clubs are not an option. Section 3 introduces this possibility and examines the impact that a threat of an inner club has on the initial organization. Section 4 derives the conditions under which an inner club forms and characterizes the equilibrium outcomes. Section 5 discusses key assumptions. Section 6 analyses organizations operating under majority rules. Section 7 discusses European Integration as an illustration of our theory. Concluding remarks are in Section 8. Formal proofs are relegated to the Appendix.

2 The Curse of Unanimity

We consider an organization with N members, who produce a common good.⁴ The provision of the good increases in the size of the organization and in the effort e of the members. Inspired by Leontief partnership models (e.g., Vislie, 1994), we assume that the amount of the good is determined by the smallest effort in the organization, scaled by the size of the organization: $N \min[e_1, e_2, \dots, e_N]$.

The utility of each member increases in the consumption and decreases in effort. The benefit from consumption is the same for all members, whereas the effort cost differs

⁴Our interest is how an existing organization responds to new challenges for which its members have different preferences. Hence, we abstract from the question of whether any given member has an incentive to leave the organization or whether outsiders would like to join.

across members. Member $i \in N$ has effort cost $\theta_i e^2/2$, and the type parameter θ_i is distributed on the support $[\underline{\theta}, \bar{\theta}]$. Furthermore, the productivity difference between any two adjacent members is the same. We refer to $\underline{\theta}$ as the most productive or “strongest” type, and to $\bar{\theta}$ as the least productive or “weakest” type. Assigning rank 1 to the strongest type $\underline{\theta}$, the cost parameter of the member with rank i is

$$\theta_i = \underline{\theta} + \frac{i-1}{N-1} (\bar{\theta} - \underline{\theta}). \quad (1)$$

Given that the good is produced with a Leontief technology, member i 's payoff is

$$y(\theta_i, e) = N \min\{e_1, \dots, e_N\} - \theta_i e_i^2/2.$$

As the members have different costs, their preferred amount of common good differs. Hence, some members could offer side payments to others in order to influence their effort choices. However, we abstract from transfer payments in order to focus on the threat of forming a “club-in-the-club” as a mechanism to overcome the opposition of individual members against reform proposals.⁵

Production of the good is modelled as a two-stage game. In the first stage, members vote on a minimum effort level in the club and in the second stage each member simultaneously exerts an effort. Individual effort levels are verifiable and each member commits herself to exert - at least - the effort level agreed upon in the voting stage. That is, underprovision is infinitely punished, but the voting outcome is not binding from above. The asymmetry reflects our interest in the constraints that unanimity imposes on organizations. However, unilateral overprovision is never an equilibrium outcome due to the Leontief technology.⁶

In standard voting procedures agents vote over pairs of alternatives and the winner in one round is posed against another alternative in the next round. Under the unanimity rule, this procedure may easily fail to generate a unique winner. Further, the outcome of

⁵Frequently, transfers are not a feasible option because of a commitment problem, like in e.g., Acemoglu and Robinson (2000) in the context of democratization, or in Fearon (1995) in the context of war.

⁶Due to the Leontief technology, there is also no loss of generality in assuming that the organization votes on a common (minimum) effort level as opposed to a menu of type-contingent efforts.

the unanimity vote is highly sensitive to the order in which proposals are put to the vote, as well as the default option in case none of the alternatives receives unanimous support. That is, there is no robust unanimity voting procedure, and the literature has not agreed on a standard modelling approach.

Motivated by the interest in the impact that the weakest member has on the club production, we propose a procedure that parallels that of the continuous-time ascending-bid auction (Milgrom and Weber, 1982).⁷ An uninterested agent (“auctioneer”) proposes a sequence of continuously increasing effort levels $\{e\}$ starting with the initial level $e = 0$. After each proposal agents decide whether or not to vote in favour of a further increase in the common effort level. Once a member “leaves the auction” by voting against an increase, she cannot “return” by supporting any subsequent proposals. Under the unanimity rule, voting stops once a single member exits the vote. Accordingly, the option to withdraw from the voting gives veto power to each member. After the voting stage, members simultaneously choose their effort and the good is produced.

In this game, Nash equilibria are outcomes in which all members exert some common effort $e \in [0, N/\bar{\theta}]$, where $N/\bar{\theta}$ is the effort maximizing the payoff of the weakest type. More precisely, any effort $e \in [0, N/\bar{\theta}]$ can be supported in an equilibrium where at least two members withdraw from the vote at some $e^V \in [0, e]$ and where all members choose the same effort level e in production. Because of the Leontief technology, unilateral overperformance ($e_i > e$) is never profitable. In turn, no member would want to withdraw prior to e^V , as $e^V \leq N/\bar{\theta} \leq N/\theta_i$, where the latter is member i 's preferred effort choice. The formal proof is relegated to the Appendix.

It is well known that input games for a team with a Leontief technology have a continuum of Nash equilibria and that these equilibria can be Pareto-ranked. This also holds for our voting game: All members prefer the Pareto-dominant equilibrium with $e = N/\bar{\theta}$ which we use as a benchmark in the subsequent analysis.

Proposition 1 *Under unanimity, the weakest member of the organization executes her*

⁷Our procedure is not robust either. For instance, a decending order would favor the productive members.

veto power, holding back the entire organization at her privately optimal choice.

Proposition 1 captures the idea that unanimity voting may result in the weakest member blocking any attempt to increase organization-wide effort. In principle, unanimity could well favour stronger rather than weaker members of an organization. For example, more productive (and wealthier) members would exercise their veto power if the organization were to vote on redistribution and not on effort. However, we follow the wide-spread view that unanimity tends to protect weak members and to slow down reforms (e.g., Erlenmaier and Gersbach, 2004).

3 Undermining Veto Power

We now show how the veto power of weaker members can be undermined by the threat of some members to form an “inner organization”. This threat may lead to three different types of outcomes: i) “initial organization”, the equilibrium outcome with no inner organization and no additional effort; ii) “reformed” organization, the outcome with no inner organization but higher organization-wide effort; iii) “divided” organization, the outcome with a club-in-the-club. Here we analyze the first two outcome types and relegate the analysis of divided organizations to the next section.

Each member can freely decide whether she wants to join the inner club. To keep the analysis tractable, we abstract from the possibility of multiple inner organizations and allow for at most one inner organization. Furthermore, the inner organization must have at least two members ($n \geq 2$). This is a natural restriction because an inner organization provides a public - rather than private - good to its members.

Instead of adopting a multi-task framework (Holmström and Milgrom, 1991) which would view inner and outer club efforts as substitutes, we assume a negative externality in consumption.⁸ An inner organization with n members reduces the utility of consuming the outer club good for all N agents by λn with $\lambda \geq 0$. The term λn characterizes a

⁸An alternative would be to assume that there is a lump-sum benefit of being in the inner club. This provides similar results, but a less rich set of possible inner club formations.

deadweight loss (or a negative externality) associated with the existence of an inner club.⁹

For symmetry, we assume the production technology of the inner organization to be the same as the one of the outer organization. Membership in the inner organization generates additional per-capita benefits of $n(e_{In} - e_{Out})$, where e_{In} (e_{Out}) denotes the minimal effort exerted by anyone who is a member of the inner (outer) organization.

We use the term “constellation” for a partitioning of members into an inner organization with $n \leq N$ members, together with the associated effort levels in the outer and the inner organization. The payoff of type i who is a member of both the inner and the outer organization is

$$y_i = n(e_{In} - e_{Out}) + Ne_{Out} - \lambda n - \theta_i e_i^2/2. \quad (2)$$

The payoff of type j who is only a member of the outer organization is

$$y_j = Ne_{Out} - \lambda n - \theta_j e_j^2/2.$$

In general, the formation of an inner organization is sensitive to how agents coordinate, for instance, who determines the effort level e_{In} and/or the size n of the inner club. We intentionally abstract from specific coordination mechanisms and let Nature choose e_{In} .¹⁰ This allows us to identify all constellations that can be supported as Nash equilibrium outcomes. These constellations constitute the constraints of a potential decision-maker under an arbitrary agenda-setting procedure.

The production of the outer and possible inner club goods takes place in three stages. In the first stage, members vote on the minimum effort of the outer organization. As before, voting follows the ascending procedure under the unanimity rule. In the second stage, Nature draws e_{In} . Following the logic of the model, we restrict the possible draws of nature to $e_{In} > e_{Out}$. Having observed e_{In} , all members have the option to simultane-

⁹An example of such an externality is the impact that the euro has on equity trades. Coeurdacier and Martin (2009) show that the European monetary union has made the non-euro Scandinavian countries less attractive for equity holders of euro countries.

¹⁰Endogenizing the choice of e_{In} poses both technical and severe conceptual problems. If the size of the inner club were given, one can think of a procedure generating e_{In} compatible with this size. However, there is no obvious rationale for either selecting a particular inner club size, or choosing the inner club size prior to the amount of the inner club good. Last but not least, any procedure that simultaneously determines n and e_{In} would be highly arbitrary.

ously subscribe to join the inner club. By subscribing, a member commits to exert e_{In} . Otherwise, she gets infinitely punished. In the final stage, all members simultaneously choose their effort and the club goods are produced.

We make two further simplifying assumptions whose implications are discussed in Section 5. First, Nature's draw e_{In} is strictly binding in the sense that inner club members have to exert exactly e_{In} , neither less nor more. Second, each member's decision to withdraw from the voting is non-strategic in the sense of ignoring its impact on the subsequent subscription decision of other members.

The latter assumption pins down a unique voting outcome, as each member i withdraws at her preferred effort level N/θ_i . So the weakest member ends the voting by exiting at $e^V = e_{Out} = N/\bar{\theta}$.

If Nature draws a moderate level of e_{In} , there exists a Nash equilibrium where all members subscribe to join the inner organization and exert exactly e_{In} . Consider the choice of the weakest member when all other members subscribe to e_{In} . If she also subscribes, she should exactly match e_{In} . If she abstains from joining, an inner club of size $N - 1$ forms. She then sets the outer-club effort to her most preferred level $e_N^* = N/\bar{\theta}$. This option entails lower disutility of effort but also lower consumption and in addition the deadweight loss $\lambda(N - 1)$. By comparing respective payoffs

$$Ne_{In} - \bar{\theta}e_{In}^2/2 \geq N(N/\bar{\theta}) - \lambda(N - 1) - \bar{\theta} (N/\bar{\theta})^2/2 \quad (3)$$

and using the constraint $e_{In} > e_{Out}$, we get that the weakest type prefers to join the inner organization for all efforts

$$e_{In} \in \left(N/\bar{\theta}, N/\bar{\theta} + \sqrt{2\lambda(N - 1)/\bar{\theta}} \right].$$

If type N joins the inner club, so do all other types because they have lower effort cost. Consequently, there exists an equilibrium with all N members choosing the same effort level e_{In} as long as e_{In} does not exceed $e^{RO} = N/\bar{\theta} + \sqrt{2\lambda(N - 1)/\bar{\theta}}$.

Proposition 2 *Reformed organizations can emerge for any $\lambda > 0$ and $e_{In} \in (N/\bar{\theta}, e^{RO}]$. That is, the threat of forming an inner club undermines the veto power of the weakest*

member and increases organization-wide effort.

Unanimity is commonly viewed as preventing majorities from coercing minorities at the cost of organizational inertia or inability of adjusting. Proposition 2 shows that this view needs to be qualified: Unanimity need not be tantamount to complete protection of weaker members or, equivalently, to the inability to reform. The threat of forming an inner organization can undermine the veto power of each single member and may enable the organization to reform. To be an effective reform mechanism, two conditions must hold. First, the statutes of the organization must exempt the formation of an inner organization from unanimous approval. Otherwise, weaker members would have no reason to avoid the formation of an inner organization by exerting more effort. Rather, they could simply veto its formation. Second, the inner organization must impose some externalities on the outer organization. Otherwise, the weaker members have no incentives to increase their effort beyond their privately optimal level. The “reform potential” of an organization, measured by the difference $e^{RO} - N/\bar{\theta}$, increases with the deadweight loss associated with an inner organization and with the size of the initial organization.

4 Club-in-the-club

We have so far only looked at equilibrium constellations in which there are no inner organizations in equilibrium. We here explore the set of constellations with divided organizations that can be supported as Nash equilibrium outcomes for any given deadweight loss λ .

Assume an inner organization exists. Due to non-strategic voting, all agents who do not subscribe to the inner club exert $e_{Out} = e^V = N/\bar{\theta}$. As the inner club effort level e_{In} , drawn by Nature, is binding, all members who join the inner organization exert e_{In} .

For expositional simplicity, we only consider inner clubs with $n \in \{3, \dots, N - 2\}$, where the relevant participation constraints have the same functional form. While our setting allows for $n = 2$ or $n = N - 1$, the respective participation constraints differ slightly in these two cases (see Appendix for details).

We first establish which types are members of both the inner and outer organization.

Lemma 1 *Provided that an inner organization of size $n \in \{3, \dots, N-2\}$ is an equilibrium outcome, its members are the n most productive types $i \in \{1, \dots, n\}$.*

The benefits of the inner club membership, given by the first term in (2), are independent of the type i . In turn, the cost of exerting an extra effort is increasing in i . Hence, given the equilibrium inner club has n members, these must be the n most productive agents.

The above result implies that an inner organization forms in equilibrium if the following two constraints are satisfied:

$$n(e_{In} - N/\bar{\theta}) - \lambda \geq \theta_n [e_{In}^2 - (N/\bar{\theta})^2] / 2, \quad (4)$$

$$(n+1)(e_{In} - N/\bar{\theta}) - \lambda < \theta_{n+1} [e_{In}^2 - (N/\bar{\theta})^2] / 2. \quad (5)$$

The first condition ensures that the marginal, i.e. least productive, member of the inner organization prefers to be member in both the inner and the outer organization. The second condition ensures that the most productive member in the outer organization prefers to be member in the outer organization only. We assume that type $n+1$ does not join in case she is indifferent, which accounts for the strict inequality in the non-participation constraint (5).

We now establish the conditions for the existence of divided organizations and then characterize constellations supporting inner organizations of different size.

An increase in the size of the inner organization benefits all its members as it rises the provision of the inner public good. However, the less productive agents may find it too costly to exert the requested effort level e_{In} . Hence, any inner organization strikes a balance between size and productivity of its marginal member. This trade-off has no interior solution if the productivity differences among (two adjacent) members is relatively small. That is, when the N members are relatively homogeneous, an inner organization never forms in equilibrium.

Proposition 3 *An inner club can only emerge if agents are sufficiently heterogeneous, $N < \bar{\theta}/\underline{\theta}$.*

For the remainder of this section, we assume that the heterogeneity condition $N < \bar{\theta}/\underline{\theta}$ holds. This is, however, only a necessary condition for the existence of inner organizations. Rather intuitively, the size of the deadweight loss and the level of the inner club effort, drawn by Nature, also matter. Indeed, even when members are heterogeneous, divided organizations only exist for certain pairs of (e_{In}, λ) .

Denote by Ω the set of all pairs (e_{In}, λ) that satisfy the two inequalities

$$3(e_{In} - N/\bar{\theta}) - \lambda \geq \theta_3 [e_{In}^2 - (N/\bar{\theta})^2] / 2, \quad (6)$$

$$(N - 1)(e_{In} - N/\bar{\theta}) - \lambda < \theta_{N-1} [e_{In}^2 - (N/\bar{\theta})^2] / 2. \quad (7)$$

where (6) is type 3's participation constraint, (7) is type $N - 1$'s non-participation constraint, and $e_{In} > N/\bar{\theta}$. These constraints define the largest set of pairs (e_{In}, λ) that support a divided organization outcome.

Indeed, sufficient heterogeneity implies that the effort cost θ_n increases faster than n . Hence, if the non-participation constraint

$$n(e_{In} - e_{out}) - \lambda < \theta_n [e_{In}^2 - e_{out}^2] / 2.$$

holds for $n = 3$, it holds for all less productive types $k > 3$. That is, if type 3 does not want to be the marginal member of the inner club of size 3, no type $k > 3$ wants to be the marginal member of the inner club of size k . At the same time, if type n wants to be the marginal member of the inner club of size n , any more productive type $k < n$ has the same preference. Finally, member $N - 1$'s non-participation constraint has to be met, as we only consider inner clubs of size $n \in \{3, \dots, N - 2\}$.

Proposition 4 *Provided that types are sufficiently heterogenous, an inner club of size $n \in \{3, \dots, N - 2\}$ can form in equilibrium iff the pair (e_{In}, λ) belongs to the set Ω . Moreover, for each pair (e_{In}, λ) the size of the inner club is unique.*

Figure 1 depicts the set of pairs (λ, e_{In}) that can support divided organizations. The

outer border of the set is determined by the participation constraint of type 3 and the inner border by the non-participation constraint of type $N - 1$. Intuitively, type 3 refrains from subscribing to the inner club, if the deadweight loss λ is very high. Signing up for very high effort level e_{In} is too costly for type 3, so, again, she stays in the initial organization. Also, if the inner club effort level is not very different from the outer club one, there is little value for type 3 in joining the inner club and suffering the dead-weight loss λ . By the same logic, type $N - 1$ stays out unless λ is very small and/or e_{In} is close to the outer club effort level.¹¹

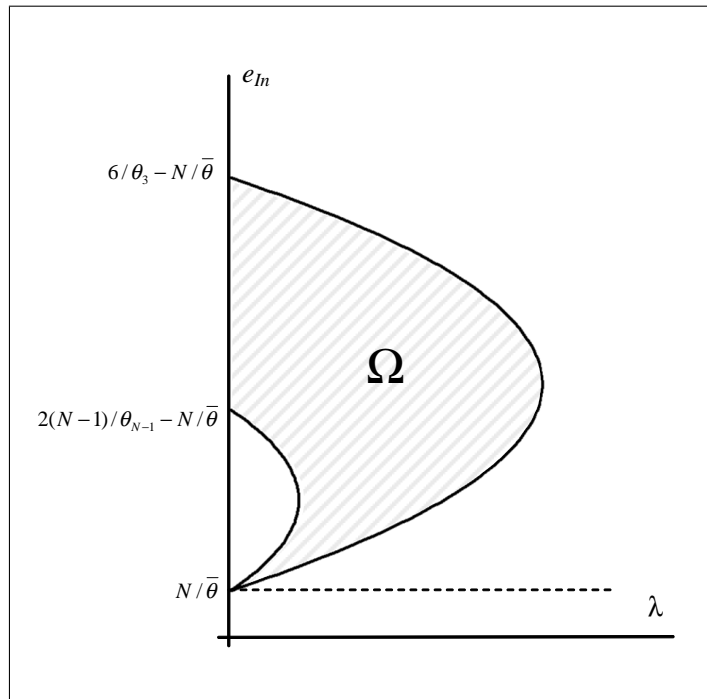


Figure 1. Set of (λ, e_{In}) compatible with divided organizations.

The uniqueness is easily understood by analyzing two inner organizations that differ in size by one member. Recall that an inner organization of size n consists of the n most productive members, as established by Lemma 1. If an effort level \tilde{e}_{In} and a deadweight loss $\tilde{\lambda}$ are compatible with an inner organization of size k , type k prefers being member of an inner club of size k to being an outsider of an inner club of size $k - 1$. But this immediately implies that an inner club of size $k - 1$ cannot be an outcome for the pair $(\tilde{\lambda}, \tilde{e}_{In})$, as type k would not stay outside. More generally, heterogeneity implies that

¹¹Setting $\lambda = 0$ in the participation constraints and rearranging yields the respective e_{In} intercepts.

if type k does not want to be the marginal member of an inner organization of size k , neither of types $k + 1, k + 2, \dots$ wants to be the marginal member of the inner club of respective size. Therefore, inner clubs with non-adjacent size cannot coexist either.

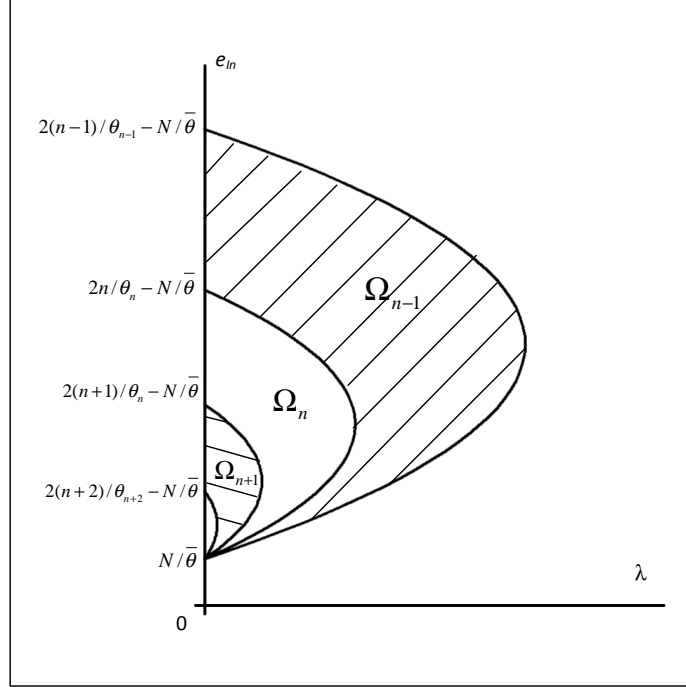


Figure 2. Inner clubs of different sizes.

The uniqueness result means that the set Ω can be partitioned into subsets Ω_n , each corresponding to all pairs (λ, e_{In}) consistent with an inner organization of size n . Each subset Ω_n is determined by type n 's participation constraint and type $n + 1$'s non-participation constraint. As can be seen at Figure 2, these subsets have an “onion-like” shape with the outer layers enclosing the pairs compatible with smaller inner organizations. Intuitively, for high λ the difference in the effort costs between the marginal inner club member n and the least productive type needs to be sufficiently high, otherwise n stays outside. For a given type distribution it means that the size of the maximum supportable inner club shrinks as λ increases. The formal characterization of Ω_n is provided in the Appendix.

The uniqueness of divided organization does not imply uniqueness of the equilibrium outcome. Indeed, the initial organization can be supported as an equilibrium for any parameter values. This follows from the assumption that an inner organization must

have at least two members. Hence, if all other members choose not to join, no single member has an incentive to deviate from this common pattern. In addition, if the effort e_{In} is not too high, the reformed organization can also exist in equilibrium. Finally, coexistence of initial, reformed and divided organizations requires moderate levels of e_{In} and deadweight loss λ , coupled with sufficient heterogeneity of members.¹²

5 Discussion

Throughout the analysis, we rely on several core assumptions to keep the model tractable. We now discuss their implications for the results. The assumption of equidistantly distributed types delivers a generic functional form for the participation constraints of the inner club members. The essential feature ensuring the formation of divided organization is the heterogeneity of types, that is, the most preferred effort of the marginal inner club member $e_n^*(n) = n/\theta_n$ is decreasing in the club size. We are confident that any distribution satisfying this property can generate divided organization equilibria. The “thickness” of the Ω_n -layers would, however, differ as compared to the equidistant distribution. The distribution has no impact on the formation of the initial and reformed organizations in equilibrium, as they are solely determined by the decisions of the least productive member.

In the model, Nature’s draw e_{In} is assumed to be binding not only from below but - unlike in the voting stage - also from above. This simplifying assumption ensures that the members of the inner club exert precisely e_{In} which for a given deadweight loss λ yields a unique divided organization. If inner club members were free to exert a higher effort than what Nature draws, inner clubs of different size could emerge in equilibrium. Suppose that a pair (e_{In}, λ) supports an inner club of size n . Then, by coordinating to work harder than e_{In} , the most productive $m < n$ members can form an inner organization in equilibrium. To see this, consider a point $(\tilde{e}_{In}, \tilde{\lambda}) \in \Omega_n$ in Figure 2. The ray along the vertical line $\lambda = \tilde{\lambda}$ starting at \tilde{e}_{In} and corresponding to an increase in e_{In} , crosses all the

¹²For details on the coexistence we refer the reader to Berglöf et al. (2009).

sets $\Omega_{n-1}, \Omega_{n-2}, \dots, \Omega_3$. Setting the inner club effort to equal exactly the Nature drawn level allows us to convey our ideas, while keeping the analysis tractable.

While we restrict our analysis to a single inner club, the logic of our model seems compatible with multiple inner organizations. For example, if Nature draws two e_{In} , a plausible equilibrium candidate is a constellation with two inner clubs, the most productive types being members of both inner clubs, intermediate types joining the “outer-inner” club and the least productive types being only in the outer organization. However, the outcomes in such an extended framework will depend on modelling details such as the assumed interaction between the deadweight loss of different inner clubs, and single vs. multiple inner club membership.

Finally, we turn to the assumption of non-strategic voting. While it is a standard assumption in many political economy models, it may be limiting in our framework. Indeed, the prospect of an inner club provides the members of the initial organization with the incentive to behave strategically in the voting stage. On the one hand, more productive types may choose to withdraw from the voting before the least productive member would pull out when voting sincerely. While this reduces the provision of the outer club good, it may induce more members to subscribe to the emerging inner club. On the other hand, less productive members may remain in the voting beyond their most preferred level. Though costly, the extra effort reduces the attractiveness of an inner club, thereby lowering the number of its potential members and the consequent deadweight loss, or even preventing its formation altogether.

Therefore, allowing for strategic voting would likely alter the effort level of the outer club and the size of the inner organization. Nonetheless, we would expect to observe the same types of organizational outcomes: divided as well as reformed and initial organizations. In addition, the game may feature an equilibrium in which all members exert an effort below the most preferred level of the least productive member.

It is worth noting that strategic voting entails certain costs but uncertain benefits. When a highly productive member withdraws early, the outer club good is provided at the lower level. At the same time, a larger inner club may or may not materialize depending

on the draw of e_{In} . More generally, the benefits of strategic voting depend on the extent to which the agent can influence or correctly anticipate the subsequent decision (i.e., the level e_{In}). Our setting abstracts from any specific agenda setting procedure and lets e_{In} be randomly chosen by Nature. In this complex environment, the benefits of strategic voting seem particularly limited, making the sincere voting assumption less restrictive than it may seem at first glance.

6 Majority Rules

In the divided organization outcome, weaker members are not forced to provide more effort than their privately optimal choice. Thus, unanimity protects weak members from the tyranny of the majority but at the price of the formation of a club-in-the-club. Many clubs may want to avoid becoming a two-class organization. One possible remedy is a majority rule since it limits the decision power of the weak members. This reduces the extent to which more productive members are held back which in turn may prevent the formation of inner clubs.

We now consider organizations operating under different majority rules $\mathcal{M}(m)$, where the majority threshold $m \in [0.5, 1)$ corresponds to the required fraction of supporting votes. As before, voting follows the ascending procedure, but under the $\mathcal{M}(m)$ majority rule it ends once a fraction $(1 - m)$ of agents has chosen to “leave the auction”.¹³ In the second stage Nature draws a (potential) inner club effort that exceeds the one voted upon in the first stage.

Under majority rule $\mathcal{M}(m)$, the organization-wide effort e_{Out}^m coincides with the best-preferred choice of its decisive member mN . For instance, the median type $0.5N$ is decisive in case of the simple majority rule $\mathcal{M}(0.5)$, and the resulting effort is $e_{Out}^{0.5} = N/\theta_{0.5N}$. As the majority threshold m increases the decisive member mN becomes less productive and the organization-wide effort e_{Out}^m declines. Hence, stronger members have more incentives to form an inner organization. If majority rules are at all effective in

¹³For notational simplicity, we abstract from the integer problem.

preventing inner clubs, they must have sufficiently low thresholds.¹⁴

Proposition 5 *Under the simple majority rule a divided organization never emerges.*

The high organization-wide effort level under the simple majority rule $\mathcal{M}(0.5)$ makes inner clubs no longer attractive even for the most productive members. That is, there are no pairs (e_{in}, λ) that support the formation of an inner organization of any size.

While the simple majority rule succeeds in preventing inner clubs, it leaves weaker members without protection against the tyranny of the majority. Clearly, a supermajority rule would coerce weaker members less. But can it still preclude the formation of an inner organization? The analysis so far has shown that divided organizations form under unanimity rule ($m = 1$), provided agents are heterogeneous. Proposition (5) establishes the inexistence of inner club equilibria under the simple majority rule ($m = 0.5$). Based on a continuity argument one may expect this result to be obtained already under a *qualified* majority rule.

Proposition 6 *For any initial organization $(\underline{\theta}, \bar{\theta}, N)$ there exists a majority threshold $\bar{m}(\underline{\theta}, \bar{\theta}, N) > 1/2$ such that no inner organization emerges under all majority rules $\mathcal{M}(m)$ with $m < \bar{m}(\underline{\theta}, \bar{\theta}, N)$. The threshold $\bar{m}(\underline{\theta}, \bar{\theta}, N)$ decreases as agents become more heterogeneous.*

The exact majority threshold depends on the characteristics of the organization. When an organization is more heterogeneous, as measured by an increase in $\bar{\theta}$ (decrease in $\underline{\theta}$), the incentives of its members are less aligned. Under a given majority rule, productive members are held back to a larger extent, which makes them more eager to form an inner organization. This tendency can be counteracted by a lower majority threshold. It increases the productivity of the decisive club member and the organization-wide effort level, thereby eliminating incentives to form an inner club.

¹⁴For the same reason as under unanimity, initial organization with all members exerting e_{Out}^m is always an equilibrium outcome under the majority rule $\mathcal{M}(m)$. Similarly, one can also support reformed organization equilibria.

7 European Integration

The evolution of the European Union (EU) provides fitting examples of our theory. To map the model into the EU experience, important concepts are: the benefit of the public good, the “effort” of the members, and the heterogeneity of the costs associated with this effort.

The benefit of the public good “European integration” has many dimensions, but the main goal has been to bring about cooperation and to assure peace. Some examples are the 1951 European Coal and Steel Community in which France, the Benelux countries, Italy and Germany coordinated their actions in these industries. On March 25, 1957, the six countries signed the Treaty of Rome creating the European Economic Community (EEC) with a view to promote trade among its member states. Further public goods were the creation of a single currency with its reduction of transaction costs in intra-European trade and the further extension of the integrated market through various enlargement waves.

Our notion of effort also has multiple interpretations in the context of the EU. For instance, to reap the benefits of European integration, countries must go through a number of adjustment processes that take the time of politicians and bureaucrats, but also impose costs on the population. Laws must be changed and harmonized; languages must be learnt; opening markets exposes firms and workers to more competition. Reaching the Maastricht criteria in particular, committed national and subnational bodies alike to budgetary austerity, often with massive consequences for the population. Probably most important is the loss of sovereignty. This is a severe concern as the referenda and discussion about the Lisbon treaty show: countries like Ireland, Poland, the Czech Republic or Germany have been substantially delaying ratification of the treaty, because public and parliament alike were concerned about sovereignty in general (the recent debate in Germany is about the role of national parliaments in EU integration decisions) or quite specific questions, such as abortion law in Ireland.

Heterogeneity between members can be treated in the model in two different ways. One could consider heterogeneity in the value associated with European integration, or

as we do in the model, one can map heterogeneity into the cost function. The modeling strategies give similar results; in reality it is not straightforward to distinguish whether one or the other would be the source of heterogeneity. An early example is the plan for integration into a European Defence Community (EDC) in 1954. The French Parliament objected ratification and thus vetoed further integration. Whether France valued common defense lower than other members or estimated the costs (the potential loss of sovereignty) higher than others, seems a question that is secondary to our model. What is important, though, is to see the heterogeneity across countries in terms of the net benefits of integration.

Beyond justifying the structure of the model, it is also important to see to what extent outcomes of our model are in line with the reality of European integration. The failure of the European Defense Community is an early example of reform efforts that got vetoed by a member. The EU then saw many blocked reforms, but during the second half of 1980s, European Commission President Jacques Delors and some of the governments of stronger member states pushed for further integration. This process resulted in the Treaty of Maastricht, which states in article 2: “This Treaty marks a new stage in the process of creating an ever closer union among the peoples of Europe.”

The core proposal to re-vitalize the EU was the creation of a common currency area with strict criteria for joining the “club-in-the-club”, the European Monetary Union. Reaching the Maastricht criteria on public debt, deficit, interest rates and inflation meant to undertake efforts for each of the aspiring membership candidates. Naturally, these efforts would be more painful for countries with larger budgetary problems, such as Belgium, Greece or Italy. However, the benefits of further integration and the creation of a joint currency would accrue to all participating members.

Arguably, the process of reaching the criteria led to a revitalization of the European integration process and a phase of growth.¹⁵ In the language of our model a group of economically stronger countries brought forward a proposal that was open to everyone. However, inclusion in the new club Euroland was only possible after exerting substantial

¹⁵The fact that some of the effort was not verifiable as the example of Greece shows, is beyond the scope of our model.

efforts. The threat of forming such an inner club that would have excluded the underperformers seems to have worked. The countries that wanted to join managed to reach the criteria.

Our model also predicts that the risk of club-in-the club formation increase when heterogeneity of members increases and that a move from unanimity to qualified majority can be a remedy. Indeed, the initial members of the European Community had quite similar aims and economic structures. Through a number of subsequent enlargement waves, the economic heterogeneity of EU members increased, thereby altering matters considerably. With the southern periphery joining, the challenge of keeping the new Union together had to be confronted as the size and use of structural funds, the state of labor markets and public administration provided ample reasons for conflict.

The Single European Act of Luxembourg (1986) can be seen as a first mild response. Here, unanimity was abandoned for many policy issues. This voting reform substantially reduced each single member's veto power. Despite such reforms, growing concerns about paralysis in the EU have spurred discussions about a two-speed Europe. Representatives of the stronger founding members, France and Germany (President Chirac and Former Foreign Minister Joschka Fischer) proposed to allow a subset of EU members to cooperate and integrate more. As in our model, larger heterogeneity increases the likelihood that inner clubs may form. As a response, the summit in Nice in 2000 explicitly set out to address the institutional problems associated with enlargement by re-weighting the allocation of votes in the Council and by extending qualified majority voting to an even larger number of areas. The 2001 intergovernmental conference in Nice was supposed to facilitate decision-making in the new larger Union and by regulating the formation of inner clubs through the instrument of "enhanced cooperation" among members. The Reform Treaty of 2007 regulates further the instrument of enhanced cooperation among sub-groups of countries, and reinforced the sole right of the Commission to formally propose such initiatives. The Reform Treaty also redefines qualified majority voting into double majority voting whereby a minimum of 55 per cent of Member States representing a minimum of 65 per cent of EU's population are required to pass legislation.

Thus, the dynamics of the European Union's voting system is well in line with the logic of our theory in which the majority thresholds decline in the heterogeneity of club members.

8 Concluding Remarks

The paper presents a theory of loosely-knit organizations. While members have a common interest, there is no governance mechanism in place that enforces contributions to the common good. Hence, organization-wide decisions must be taken unanimously, granting each member veto power. We show that there are nonetheless ways for such organizations to avoid being held back by their least committed members. The threat of forming a club-in-the-club can induce members that are less interested or less productive to contribute more to the common good than privately preferred. Key for this mechanism is that the formation of a club-in-the-club imposes a deadweight loss on all members, but benefits only those who join the inner club. Then, unanimity does not preclude reform, in the sense of all members exerting more effort than is preferred by its weakest members.

We also show that identical organizations can end up quite differently: some may stagnate at the level preferred by its weakest members, others may reform, and yet others may be divided by the formation of an inner club. Furthermore, the divided organization outcome is more likely when members are more heterogenous. To avoid this outcome an organization can adopt a majority rule. This constitutional change results in a higher organization-wide effort, and thus, often precludes the formation of an inner club. The change can be interpreted as a way of institutionalizing the reformed organization outcome, feasible under unanimity.

References

- [1] Acemoglu, D., Robinson, J. A., 2000. Why Did The West Extend The Franchise? Democracy, Inequality, And Growth In Historical Perspective, *The Quarterly Journal of Economics*, MIT Press, 115(4): 1167-1199.

- [2] Aghion, P., Bolton, P., 2003. Incomplete Social Contracts. *Journal of the European Economic Association*, 1(1): 38-67.
- [3] d'Aspremont, C., Jaquemin, A., Gabszewicz, J. J., Weymark, J. A., 1983. On the stability of collusive price leadership. *Canadian Journal of Economics* 16(1), 17-25.
- [4] Berglof, E., Burkart, M., Friebel, G., Paltseva, E., 2009. Club-in-the-Club: Reform Under Unanimity, EPRU Working Paper Series 2009-07, University of Copenhagen, Department of Economics.
- [5] Bolton, P., Roland, G., 1997. The Breakup of Nations: A Political Economy Analysis. *Quarterly Journal of Economics*, 112(4): 1057–90.
- [6] Bordignon, M., Brusco, S., 2001. Optimal Secession Rules. *European Economic Review*, 45(10): 1811-1834.
- [7] Buchanan, J., Faith, R. L., 1987. Secession and the Limits of Taxation: Toward a Theory of Internal Exit. *American Economic Review*, 77(5): 1023-1031.
- [8] Coeurdacier, N., Martin, P., 2009. The geography of asset trade and the euro: Insiders and outsiders, *Journal of the Japanese and International Economies*, Elsevier, vol. 23(2), pages 90-113.
- [9] Cornes, R., Sandler, T., 1996. *The Theory of Externalities, Public Goods, and Club Goods*. Cambridge: Cambridge University Press.
- [10] Diamantoudi, E., Sartzetakis, E. S., 2006. Stable International Environmental Agreements: An Analytical Approach. *Journal of Public Economic Theory* 8(2), 247-263.
- [11] Dixit, A., 2003. Clubs with Entrapment. *American Economic Review*, 93(5): 1824-1829.
- [12] Erlenmaier, U., Gersbach, H., 2004. Bundling and the Unanimity Rule. CEPR Discussion Paper 4808.
- [13] Fearon, J. D., 1995. Rationalist explanations for war. *International Organization*, 49(3): 379–414.
- [14] Gradstein, M., 2004. Political Bargaining in a Federation: Buchanan meets Coase. *European Economic Review*, 48(5): 983-999.
- [15] Harstad, B., 2006. Flexible Integration? Mandatory and Minimum Participation Rules. *Scandinavian Journal of Economics*, 108(4): 683-702.
- [16] Holmström, B., Paul Milgrom, P., 1991. Multi-Task Principal-Agent Problems: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics and Organization*, 7(1): 24-52

- [17] Messner, M., Polborn, M., 2004. Voting on Majority Rules. *Review of Economic Studies*, 71(1): 115-132.
- [18] Milgrom, P.R., Roberts, J., 1992. *Economics, Organization and Management*. Prentice-Hall International, Englewood Cliffs, New Jersey.
- [19] Milgrom, P. R., Weber, R. J., 1982. A Theory of Auctions and Competitive Bidding. *Econometrica*, 50(5): 1089-1122.
- [20] Rousseau, J.-J., 1762. *Contract Social*, translated by Maurice Cranston. Penguin Books, London 1986.
- [21] Thoron, S., 1998. Formation of a Coalition-Proof Stable Cartel. *Canadian Journal of Economics* 31(1), 63-76.
- [22] Vislie, J., 1994. Efficiency and Equilibria in Complementary Teams, *Journal of Economic Behavior and Organization*. 23(1): 83-91.
- [23] Weikard, H.-P., 2009. Cartel stability under an optimal sharing rule. *The Manchester School* 77(5), 599-616.

A APPENDIX

A.1 Nash equilibrium in the original game (Section 2)

Given some e^V , member i 's decision problem at the implementation stage is

$$\max_{e_i \geq e^V} (N \min\{e, e_i\} - \theta_i e_i^2/2).$$

Member i 's preferred choice $e_i^* = N/\theta_i$ exceeds e , as $e \leq N/\bar{\theta} \leq N/\theta_i$. Thus, member i always chooses $e_i = e$, since any effort $e_i - e > 0$ would be wasted. Note that the voting outcome e^V needs not to be binding as all members can choose to exert higher effort $e \geq e^V$.

At the voting stage member i 's only deviation that influences the outcome of the game is to withdraw prior to e^V . This deviation is profitable iff $e^V > N/\theta_i$. By withdrawing at $e_i^V \leq N/\theta_i$ and choosing $e_i^* = N/\theta_i$, member i attains her first best in the implementation stage. Since this applies to all members $i = 1, \dots, N$,

$$e^V \leq N/\bar{\theta} \tag{8}$$

must hold in equilibrium. Consequently, any effort $e > N/\bar{\theta}$ cannot be an equilibrium outcome. Indeed, if everyone but member N chooses e , member N 's unilateral underper-

formance ($e_N^* = N/\bar{\theta} < e$) is both profitable and compatible with the voting outcome as $N/\bar{\theta} \geq e^V$ by (8).

A.2 Proof of Lemma 1

For an inner organization of size n to exist, two types of constraints must be satisfied. First, $N - n$ members of the outer organization must prefer staying in the outer organization rather than joining the inner organization. Second, the n members must prefer to be in the inner organization.

An agent i chooses to be a member of the inner organization if the following condition holds:

$$n(e_{In} - e_{Out}) + Ne_{Out} - \lambda n - \theta_i e_{In}^2/2 \geq Ne_{Out} - \lambda(n - 1) - \theta_i e_{Out}^2/2.$$

Rearranging yields

$$n(e_{In} - e_{out}) - \lambda \geq \theta_i [e_{In}^2 - e_{Out}^2]/2. \quad (9)$$

The LHS of this constraint is independent of the type i , whereas the RHS increases in i . Thus, if condition (9) holds for type i , it must hold for all more productive types $j = 1, \dots, i - 1$.

A.3 Inner clubs of size $n=2$ and $n=N-1$

For the inner clubs of size $n \in 3, \dots, N - 2$ the participation constraints (4) and (5) only differ with respect to the size of the inner club and the marginal members' productivity. For the inner club with $n = N - 1$ members, the non-participation constraint of type $n + 1$ is different. If type N were to join the inner organization, all members would exert the same effort and an inner organization would cease to exist. Thus, the non-participation constraint of type N is

$$\bar{\theta} [e_{In}^2 - (N/\bar{\theta})^2]/2 - (N - 1)\lambda > N(e_{In} - N/\bar{\theta}).$$

A similar effect appears in the case of the inner club of size $n = 2$. As we do not allow for inner clubs consisting of one member, if type 2 does not join, the inner club fails to form. This is reflected in type 2's participation constraint

$$2(e_{In} - N/\bar{\theta}) - 2\lambda \geq \theta_2 [e_{In}^2 - (N/\bar{\theta})^2]/2.$$

These modified constraints do not substantially change the analysis, but they lead to different functional forms of the set of equilibrium effort level e_{In} .

A.4 Proof of Proposition 3

Proof by contradiction. If an equilibrium with an inner organization of size $n < N$ exists,

$$n(e_{In} - N/\bar{\theta}) - \lambda > \theta_n \left[e_{In}^2 - [N/\bar{\theta}]^2 \right] / 2$$

and $e_{In} > e_{Out}$ must hold. Setting $e_{In} = N/\bar{\theta} + \delta$ and inserting it in the first condition yields

$$\delta(n - N\frac{\theta_n}{\bar{\theta}}) > \frac{\theta_n \delta^2}{2} + \lambda.$$

This can only hold if $(n - N\theta_n/\bar{\theta}) > 0$ or, equivalently, $(n/\theta_n - N/\bar{\theta}) > 0$. Using the definition

$$\theta_n = \frac{1}{N-1} [(N-n)\underline{\theta} + (n-1)\bar{\theta}],$$

the difference $(n/\theta_n - N/\bar{\theta})$ can be written as

$$\frac{(N-1)n}{[(N-n)\underline{\theta} + (n-1)\bar{\theta}]} - \frac{N}{\bar{\theta}} > 0.$$

Rearranging yields

$$\frac{(N-n)}{\bar{\theta} [(N-n)\underline{\theta} + (n-1)\bar{\theta}]} (\bar{\theta} - N\underline{\theta}) > 0$$

which contradicts $N \geq \bar{\theta}/\underline{\theta}$.

A.5 Proof of Proposition 4

The participation constraint (4) of type θ_n , the marginal member in an inner club of size n , can be rewritten as

$$n(e_{In} - N/\bar{\theta}) - \lambda \geq \frac{\bar{\theta}(n-1) + \underline{\theta}(N-n)}{N-1} [e_{In}^2 - (N/\bar{\theta})^2] / 2$$

or, equivalently

$$n(e_{In} - N/\bar{\theta}) \left(1 - \frac{\bar{\theta} - \underline{\theta}}{N-1} \frac{(e_{In} + N/\bar{\theta})}{2} \right) \geq \frac{N\underline{\theta} - \bar{\theta}}{N-1} [e_{In}^2 - (N/\bar{\theta})^2] / 2 + \lambda. \quad (10)$$

Similarly the non-participation constraint of the type θ_{n+1} can be written as

$$[n+1](e_{In} - N/\bar{\theta}) \left(1 - \frac{\bar{\theta} - \underline{\theta}}{N-1} \frac{(e_{In} + N/\bar{\theta})}{2} \right) < \frac{N\underline{\theta} - \bar{\theta}}{N-1} [e_{In}^2 - (N/\bar{\theta})^2] / 2 + \lambda. \quad (11)$$

(As before, if type $n+1$ is indifferent, she does not join). Define a function of x and e_{In}

$$F(x, e_{In}) = x(e_{In} - N/\bar{\theta}) \left(1 - \frac{\bar{\theta} - \underline{\theta}}{N-1} \frac{(e_{In} + N/\bar{\theta})}{2} \right).$$

As the LHS of inequalities (10) and (11) coincide with $F(n, e_{In})$ and $F(n+1, e_{In})$ respectively, an inner club of size n exerting an effort e_{In} can form in equilibrium if

$$F(x, e_{In}) \geq \frac{N\underline{\theta} - \bar{\theta}}{N-1} [e_{In}^2 - (N/\bar{\theta})^2] / 2 + \lambda \quad (12)$$

holds for $x = n$, but fails for $x = n + 1$.

We begin by proving uniqueness. Consider a pair (e'_{In}, λ') . As by construction we only consider $e'_{In} > N/\bar{\theta}$,

$$1 - \frac{\bar{\theta} - \underline{\theta}}{N-1} \frac{(e'_{In} + N/\bar{\theta})}{2} < 1 - \frac{\bar{\theta} - \underline{\theta}}{N-1} \left(\frac{N/\bar{\theta} + N/\bar{\theta}}{2} \right).$$

Given that the types are heterogeneous ($N < \bar{\theta}/\underline{\theta}$),

$$1 - \frac{\bar{\theta} - \underline{\theta}}{N-1} \left(\frac{N/\bar{\theta} + N/\bar{\theta}}{2} \right) = \frac{(N-1)\bar{\theta} - N\bar{\theta} + N\underline{\theta}}{N-1} = \frac{N\underline{\theta} - \bar{\theta}}{N-1} < 0.$$

Thus, the coefficient of x in $F(x, e'_{In})$ is negative, that is, $F(x, e'_{In})$ is decreasing in x for given e'_{In} . As the RHS of (12) is a constant for given model parameters and e'_{In} , there will be at most one n such that

$$\begin{aligned} F(n, e'_{In}) &\geq \frac{N\underline{\theta} - \bar{\theta}}{N-1} [(e'_{In})^2 - (N/\bar{\theta})^2] / 2 + \lambda', \\ F(n+1, e'_{In}) &< \frac{N\underline{\theta} - \bar{\theta}}{N-1} [(e'_{In})^2 - (N/\bar{\theta})^2] / 2 + \lambda', \end{aligned} \quad (13)$$

which proves the uniqueness part.

To prove existence, it is enough to notice that conditions (6) and (7) are equivalent to:

$$F(3, e_{In}) \geq \frac{N\underline{\theta} - \bar{\theta}}{N-1} [e_{In}^2 - (N/\bar{\theta})^2] / 2 + \lambda, \quad (14)$$

$$F(N-1, e_{In}) < \frac{N\underline{\theta} - \bar{\theta}}{N-1} [e_{In}^2 - (N/\bar{\theta})^2] / 2 + \lambda, \quad (15)$$

Therefore, by the continuity of $F(\cdot)$ it follows that for any pair $(e'_{In}, \lambda') \in \Omega$, there exists a $n \in [3, N-2]$ such that the system (13) holds, which, in turn, implies that this n is the equilibrium size of the inner club. Similarly, if $(e''_{In}, \lambda'') \notin \Omega$, at least one of the conditions (14) and (15) fails. Without loss of generality assume that it is condition (14). As $F(x, e''_{In})$ is decreasing in x for given e''_{In}

$$\begin{aligned} F(3, e''_{In}) &< \frac{N\underline{\theta} - \bar{\theta}}{N-1} [(e''_{In})^2 - (N/\bar{\theta})^2] / 2 + \lambda'' \Rightarrow \\ F(n, e''_{In}) &< \frac{N\underline{\theta} - \bar{\theta}}{N-1} [(e''_{In})^2 - (N/\bar{\theta})^2] / 2 + \lambda'' \text{ for any } n > 3. \end{aligned}$$

This immediately implies that a club of size $n \geq 3$ cannot form in equilibrium as the participation constraint of n -th type, the marginal member in an inner club of size n , does not hold.

A.6 Characterization of the Ω_n Sets

We partition the set Ω into subsets Ω_n each corresponding to the pairs (e_{In}, λ) consistent with an inner organization of size n . Each of these subsets is determined by a pair of inequalities (4) and (5), or, equivalently, (10) and (11).

Proposition A1: *Provided that types are heterogenous, an inner club of size n can form in equilibrium iff the pair (e_{In}, λ) belongs to the set Ω_n , where $n \in \{3, \dots, N - 2\}$.*

Proof: The definition of Ω_n implies that all pairs $(e_{In}, \lambda) \in \Omega_n$ can be supported by an inner club of size n . We are only left to show that if a pair $(e'_{In}, \lambda') \notin \Omega_n$, then a club of size n cannot form. Indeed, in this case either type n does not want to be a member of a club of size n , or type $n + 1$ does not want to stay outside of club of size n (or both), by the definition of Ω_n .

Proposition A2: *Consider two pairs $(e_{In}, \lambda') \in \Omega$, $(e_{In}, \lambda'') \in \Omega$, $\lambda'' > \lambda'$. Then the size n' of the inner club associated with (e_{In}, λ') is at least as high as the size of the inner club n'' associated with (e_{In}, λ'') .*

Proof: If an inner club of size n' supports (e_{In}, λ') , the non-participation condition of type $n' + 1$ is given by

$$F(n' + 1, e_{In}) < \frac{N\bar{\theta} - \bar{\theta}}{N - 1} [e_{In}^2 - (N/\bar{\theta})^2] / 2 + \lambda'.$$

As $\lambda'' > \lambda'$

$$F(n' + 1, e_{In}) < \frac{N\bar{\theta} - \bar{\theta}}{N - 1} [e_{In}^2 - (N/\bar{\theta})^2] / 2 + \lambda''.$$

But this immediately implies that type $n' + 1$ wants to be an outsider of club of size n' exerting effort e_{In} also for λ'' . As $F(x, e'_{In})$ decreases in x , we conclude that $n'' \leq n'$ (with strict inequality for sufficiently large difference between λ'' and λ'). Thus, higher deadweight loss λ is compatible with clubs of smaller size.

Comments to the shape of Ω_n :

As the set Ω_n is defined by inequalities (4) and (5), the outer and the inner border of Ω_n in (e_{In}, λ) -space are given by quadratic parabolas with respect to λ -coordinate:

$$\lambda = -\theta_n e_{In}^2 + n e_{In} + \left(\theta_n \frac{(N/\bar{\theta})^2}{2} - n (N/\bar{\theta}) \right)$$

and

$$\lambda = -\theta_{n+1}e_{In}^2 + (n+1)e_{In} + \left(\theta_{n+1} \frac{(N/\bar{\theta})^2}{2} - (n+1)(N/\bar{\theta}) \right)$$

respectively. Notice that the pair $(N/\bar{\theta}, 0)$ belongs to both parabolas. Further, the symmetry axes of these parabolas are given by type n 's best preferred effort level, $e_n^*(n) = n/\theta_n$ and type $n+1$'s best preferred effort level, $e_{n+1}^*(n+1) = (n+1)/\theta_{n+1}$ respectively.

Lemma A1: *If types are heterogeneous, $N < \bar{\theta}/\underline{\theta}$, n/θ_n decreases with n .*

Proof. Subtracting $(n+1)/\theta_{n+1}$ from n/θ_n and using the definition $\theta_n = \underline{\theta} + (n-1)(\bar{\theta} - \underline{\theta})/(N-1)$ yields

$$\begin{aligned} \frac{n}{\theta_n} - \frac{n+1}{\theta_{n+1}} &= \frac{n\theta_{n+1} - (n+1)\theta_n}{\theta_n\theta_{n+1}} = \frac{[(n+1)(\theta_{n+1} - \theta_n) - \theta_{n+1}]}{\theta_n\theta_{n+1}} \\ &= \frac{1}{(N-1)\theta_n\theta_{n+1}} [(n+1)(\bar{\theta} - \underline{\theta}) - (N-1)\underline{\theta} - n(\bar{\theta} - \underline{\theta})] \\ &= \frac{1}{(N-1)\theta_n\theta_{n+1}} [\bar{\theta} - N\underline{\theta}] > 0, \text{ if and only if } N\underline{\theta} - \bar{\theta} < 0. \end{aligned}$$

■

Combined with the discussion above these results explain the shape and relative position of subsets Ω_n on Figure 2.

A.7 Proof of Proposition 5

Proof by contradiction. If an equilibrium with an inner organization of size $n < N$ exists,

$$n(e_{In} - e_{Out}) - \lambda > \theta_n [e_{In}^2 - e_{Out}^2] / 2 \quad (16)$$

and $e_{In} > e_{Out}$ must hold where

$$e_{Out} = \frac{N}{m\bar{\theta} + (1-m)\underline{\theta}}$$

is the best preferred outer-club effort of the decisive member under the majority rule $\mathcal{M}(m)$. Setting $e_{In} = N/(m\bar{\theta} + (1-m)\underline{\theta}) + \delta$ and inserting it in equation (16) yields

$$n\delta - \lambda \geq \frac{\theta_n}{2} \left(2\delta \frac{N}{m\bar{\theta} + (1-m)\underline{\theta}} + \delta^2 \right),$$

or equivalently,

$$\delta \left(\frac{n}{\theta_n} - \frac{N}{m\bar{\theta} + (1-m)\underline{\theta}} \right) \geq \frac{\delta^2}{2} + \frac{\lambda}{\theta_n} > 0.$$

This condition can only be satisfied if

$$\frac{n}{\theta_n} > \frac{N}{m\bar{\theta} + (1-m)\underline{\theta}}. \quad (17)$$

Inserting the explicit expression (1) for θ_n and rearranging yields

$$m > 1 - \frac{1}{n} \frac{(N-n)(\bar{\theta} - N\underline{\theta})}{(N-1)(\bar{\theta} - \underline{\theta})}. \quad (18)$$

As

$$\frac{(N-n)(\bar{\theta} - N\underline{\theta})}{(N-1)(\bar{\theta} - \underline{\theta})} < 1, \quad (19)$$

inequality (18) implies that the necessary condition for formation of an inner club of size n is

$$m > 1 - \frac{1}{n} \geq \frac{1}{2}, \quad (20)$$

as an inner club should have at least two members. Simple majority threshold $m = 1/2$ never satisfies condition (20). We conclude that under the majority threshold $m = 1/2$ any divided organization ceases to exist.

A.8 Proof of Proposition 6

Consider condition (18). Its RHS is increasing in n (as $N < \bar{\theta}/\underline{\theta}$). Thus, if condition (18) fails for $n = 2$, i.e., club of size 2 ceases to exist, then any larger club also ceases to exist. Therefore, no divided organization emerges as long as the majority threshold m prevents formation of divided organization of size 2. Now denote

$$\bar{m}(\underline{\theta}, \bar{\theta}, N) = 1 - \frac{(N-2)(\bar{\theta} - N\underline{\theta})}{2(N-1)(\bar{\theta} - \underline{\theta})}.$$

By condition (18) and discussion above any majority rule $\mathcal{M}(m)$ with $m < \bar{m}(\underline{\theta}, \bar{\theta}, N)$ results in no divided organization forming in equilibrium. Further, using inequality (19) one can see that

$$\bar{m}(\underline{\theta}, \bar{\theta}, N) = 1 - \frac{(N-2)(\bar{\theta} - N\underline{\theta})}{2(N-1)(\bar{\theta} - \underline{\theta})} > 1 - \frac{1}{2} = 1/2.$$

Finally, consider an increase in the agents' heterogeneity via a change in the support of the distribution of types. Then higher heterogeneity (higher $\bar{\theta}$ and lower $\underline{\theta}$) corresponds to a lower $\bar{m}(\underline{\theta}, \bar{\theta}, N)$

$$\partial \frac{\bar{m}(\underline{\theta}, \bar{\theta}, N)}{\partial \underline{\theta}} > 0, \quad \partial \frac{\bar{m}(\underline{\theta}, \bar{\theta}, N)}{\partial \bar{\theta}} < 0.$$