Optimal Voting Rules for International Organizations, with an Application to the United Nations *

Johann Caro-Burnett[†]

August 31, 2019

Abstract

I study a self-enforcing mechanism for an international organization that interacts repeatedly over time. A random shock determines which countries would be in favor of or against taking a collective action. If the organization wants to take the action, some countries may disagree with participating; therefore, incentives must be provided to such countries. I solve for the optimal stationary equilibrium, and then I show that it is equivalent to a mechanism characterized by a weighted voting rule. I study how this optimal mechanism depends on the discount factor to establish that it can rationalize random voting power. In particular, I show that within a class of parameter cases, the optimal mechanism mimics how voting power is distributed in the United Nations Security Council.

Keywords: Political Economy, Mechanism Design, Impatient Players, Security Council.

*This study is based on the first chapter of my doctoral dissertation at Yale University. I am enormously grateful to Giovanni Maggi, Eric Weese, and Mushfiq Mobarak for their guidance and valuable comments. I would also like to thank the participants in the Leitner Seminar and the Economic Development Lunch at Yale University, the participants in the Asian Meeting of the Econometric Society at the Chinese University of Hong Kong, and the participants in a number of seminars and talks at the University of Tokyo, National Graduate Institute for Policy Studies, Hitotsubashi University and Musashi University for helpful comments. All errors are mine.

†Graduate School for International Development and Cooperation (IDEC). Hiroshima University. johanncb@hiroshima-u.ac.jp.

1 Introduction

The outcomes implemented by organizations depend, roughly speaking, on the members' preferences and the decision-making process. In economic theory, it is common to assume that the preferences are exogenous. On the other hand, there is no justification to make the same assumption for the decision-making process. However, a large proportion of the literature, especially in political economy, regards the mechanism used to aggregate preferences as exogenous.[1]

In this paper, we will study the design of the decision-making process for an international organization. Indeed, the model could fit a broader class of organizations with similar characteristics, but we will focus the exposition on international organizations for two reasons: First, the structure of the model fits most international organizations, while it is more difficult to argue that the assumptions in this paper resemble other types of organizations.[2] Second, as I will show below, under certain parameter conditions, the model is very similar to how the United Nations distributes power among its members via Security Council seats, and to the best of my knowledge, there is no other theoretical model that rationalizes the United Nation's way of distributing power.

There is a wide range of voting rules in international organizations.[3] For example, the IMF, World Bank, and European Union use voting weights that are (almost) constant over time and depend on specific variables (e.g., contributions to the organization, population, GDP). On the other hand, NATO and the WTO follow a unanimity rule. Moreover, in all the above cases, each country knows the weight of its own vote, which will not change from period to period. In contrast, the United Nations follows a completely different voting rule.[4] While all the members vote in the General Assembly, the agreements

[1]See the literature review for examples of studies on endogenous voting systems.
[2]See paragraph 4 below for a description of the three main assumptions of the model.
[3]Maggi and Morelli (2006).
[4]For a more detailed explanation of the United Nations, see section 3.

reached there are not compulsory and can only serve as a recommendation to be considered by the Security Council. The Security Council votes on relevant and compulsory issues, but only a subset of the members has the right to vote.[5] Moreover, except for the five permanent members that are always part of the Security Council (China, France, Russia, the United Kingdom, and the United States), there is uncertainty over which countries will have the right to vote in the future.

In the present paper, we will study a repeated game with three elements that international organizations may often have. First, countries cannot rely on external enforcers. Thus, any set of rules they use must be self-enforcing. Second, members are heterogeneous; countries have large differences in, for example, income, military power, and natural resources. Some countries have a stronger opinion on global problems, and other countries are concerned primarily about their local issues. Third, the organization cannot use or rely on monetary transfers.[6] Note that neither the absence of transfers nor the perfectly and unrestricted use of them are realistic assumptions; in practice, organizations would be somewhere in between those two cases. However, I want to examine the provision of incentives purely by choosing the appropriate preference aggregation rule.[7]

As a preview of the results, I will solve the optimal equilibrium not only for the case of patient countries but also for the case of impatient countries. I find some general properties that hold under

[5] Members are expected to follow Security Council decisions; otherwise, they may receive sanctions from the rest of the members.

[6] This third assumption may seem the most restrictive of those in this paper. However, there are many reasons to justify the absence of transfers. First, transfers are, in general, not openly used (if used at all). For example, the United Nations Charter does not mention monetary transfers between countries as a means of compensating affected countries. There are some studies (see, for example, Kuziemko and Werker (2006)) showing that being elected as a non-permanent member of the United Nations Security Council is correlated with foreign aid. However, there is no evidence of causality. Moreover, foreign aid usually entails several restrictions. For instance, the resources may be targeted (e.g., towards health, education), or there could be implicit inefficiencies (e.g., bureaucracy, corruption). Additionally, transfers do not necessarily solve the provision of incentives in a trivial way. Any transfer has to be self-enforcing itself, so countries have to be willing to comply with any transfer prescribed by an equilibrium. This may introduce additional constraints, and as a consequence, it is beyond the scope of the present study.

[7] Some recent studies with similar environments (for instance, Azrieli and Kim (2014) and Schmitz and Tröger (2012)) also focus on the absence of monetary transfers.

the entire spectrum of discount factors. The heterogeneity of the members' preferences allows for differences in the voting weights.[8] This first result resembles some international organizations, such as the World Bank or the European Union.

Later on (as a first step to show how rotation can be optimal in the Security Council), proposition 3 shows that random voting power can be an optimal equilibrium. Moreover, to map more closely how randomness and rotation are related, we need to add more structure to the model. Therefore, in section 3, I characterize the solution under parameter conditions that mimic how the United Nations was created. Namely, I divide the set of countries into two groups. One group is composed of the creators of the United Nations, which I call the 'mechanism designers.' They have a positive Pareto weight and (loosely speaking) a relatively higher cost of complying in unfavorable states.[9] The second group has zero Pareto weight and a relatively lower cost of complying in unfavorable states. I show that for a large range of discount factors, the optimal mechanism assigns voting power only to a subset of countries. This distribution of power depends on the profile of preference shocks and resambles a council (propositions 5, 6 and 7). Moreover, as discussed in corollary 1, there is a way to attain uniformity in council size, regardless of whether the current shock is such that there is a general agreement or opinions are divided.[10] This mechanism is remarkably similar to how the United Nations rotates decision-making power among the non-permanent members of the Security Council.[11]

The content of this study can be placed in three large categories. The first two are discussed primarily in section 2: (*i*) I solve an optimal mechanism for an international organization, and (*ii*) I show that this optimal mechanism can be mapped onto voting weights. The third category is discussed

[8]As shown in Maggi and Morelli (2006), with homogeneous members, a repeated game can explain supermajority and unanimity as optimal equilibria.

[9]See lemma 5.

[10]The size of the council can actually mimic the ten rotating members of the Security Council; see remark 2.

[11]There are studies that model coalitions in two-stage voting games, which resemble a council. See Acemoglu et al. (2012) and Eguia (2011).

primarily in section 3, where (*iii*) I provide further structure (parameters) that makes the equilibrium very similar to the United Nations. Therefore, I will relate my contribution to the literature on those three categories.

Literature Review

There are several studies on endogenous decision-making rules. Some of them focus on welfare-maximizing rules, and others focus on self-selective rules. The paper most related to the present study is Maggi and Morelli (2006). They focus on efficiency and self-enforceability. One key difference from their paper is that here the members of the organization are heterogeneous. With this extension, I can provide one explanation for why some organizations use different weights for their members (such as the IMF, World Bank, and European Union) and, more important, why some other organizations have some form of randomness in their decision-making power (the United Nations Security Council). Another difference from their study is that I allow the decision variable to take values on a continuous interval. While this feature simplifies the maximization problem, it also has desirable implications. Namely, this feature can be seen as a compromise between countries. Moreover, this result resembles that of Voeten (2001), who studies the bargaining power of Security Council members as a function of outside options and identifies a compromised level of the decision variable that makes some countries indifferent between participating and not.[12]

Barbera and Jackson (2004), Koray (2000), and Lagunoff (2009) are examples of studies that focus on the stability of decision-making rules. That is, a rule is stable if it would choose itself when voted on against other decision-making rules. In Barbera and Jackson (2004), the self-stable voting rule

[12] Ticchi and Vindigni (2010) and Aghion et al. (2004) are other examples of studies on efficiency and endogenous voting.

is simple majority (or something very near simple majority). In contrast, in the present study, the voting rule is state dependent, each country has different weights, and the threshold for implementing an action is not necessarily 50% of the votes. Koray (2000) shows that a unanimous, neutral and self-selective decision-making rule is equivalent to dictatorship. Lagunoff (2009) studies a repeated game in which the choice of a decision-making rule is explicitly made. He shows that under certain conditions, the original game is equivalent to a new game with one additional artificial player that acts as a preference aggregator. They then show that social choice functions are self-selective when that artificial player is time-consistent. [13]

The second strand of related literature is on voting weights. One of the earliest studies with a structure similar to mine is Barbera and Jackson (2006). They show how the welfare-maximizing weights of representatives in a democracy depend on preferences, population distribution, and size. Ansolabehere et al. (2005) study voting weights among legislative coalitions. They find a linear relationship between parties' shares of seats and their shares of cabinet ministries. Moreover, the party that initiated the coalition gains a bonus advantage. There are two recent studies that have similar results to mine, and therefore, it is worth summarizing the key differences between those papers and mine. Azrieli and Kim (2014) and Schmitz and Tröger (2012) study a one-period game, with preferences belonging to a large set of possible values. First, they focus on incentive compatibility because the payoffs (from participating in the organization's decision) can take more than just two values. Second, they do not consider the problem of enforceability. In both studies, the results are similar to mine in the sense that the optimal decision-making rule is a weighted voting system that can be state dependent. However, in their case, this result follows because of incentive compatibility, and indeed, voting weights are always

[13] Jackson and Yariv (2015), Schmitz and Tröger (2012), and Harstad (2010) study endogenous decision-making rules and focus on large preference spaces.

state dependent. On the other hand, in my work, the analogous result follows because of enforceability. Moreover, random voting weights arise only when the discount factor has a moderate value: not high enough to sustain the Pareto-efficient allocation (which has constant voting weights) and not so low that only unanimity is self-enforceable.

Among the empirical studies, Dreher et al. (2014) and Dreher and Vreeland (2014) are the closest to my work. They analyze the determinants of elections on the Security Council, and show that GNP, population, and the number of years off the Security Council have a positive effect on the probability of being elected as a non-permanent member. Their result is very similar to the present work in two ways. First, country characteristics (such as GNP and population) should have a close connection with preferences (and, arguably, the Pareto weights), which in the present model greatly affect the voting power of a country. Second, their 'turn-taking' variable indicates that the longer a country is not elected, the more likely it is to become a member of the Security Council. This idea relates to two of the extensions of the model: imperfect monitoring and non-stationary payoffs. In those extensions, a member that currently has no voting power knows that the decisions made at the organization are poorly correlated with its own preferences; therefore, unfavorable actions are taken with a high probability. This means that with a high probability, the organization has to promise a higher voting weight in the future to such a country to secure its participation.

2 The Model

In this section, I will first describe the one-period game and then characterize the first best. Later, I will describe the repeated game[14] and find the mechanism that maximizes the weighted sum of payoffs. As it

[14]The game presented here can be seen as either a repeated game in which nature plays first or a very simple stochastic game in which the transition probabilities are independent of the current state and the actions.

is standard in the literature, I will focus on sequential equilibria in which countries condition strategies only upon publicly observable histories of actions: perfect public equilibria (PPE).[15] Moreover, I will restrict attention to stationary payoffs. This restriction is a simplifying assumption under which the (ex ante) payoffs are constant over time. Nevertheless, stationarity of payoffs does not affect the goal of this study, which is to rationalize the rotation on the Security Council. As I will show, the model still has sufficient richness to attain the remarkable results that the optimal mechanism will have (*i*) stochastic voting power and (*ii*) a council of fixed size. At the end of this section, I will discuss some extensions, including the case of non-stationary equilibria.

2.1 The Stage Game

There are N countries endowed with a binary action space; they can choose to either participate or not in a (pure) collective action. This means that, if everyone participates, the collective action is effective. Conversely, if at least one of the countries decides not to participate, the action fails, and the status quo is preserved.[16] At the beginning of the game, the state of the world realizes. This state of the world will be denoted $\mathbf{y} = (y_1, y_2, \ldots, y_N)$ and is the profile of payoffs of all members in the case in which the collective action is taken. That is, when the collective action is effective, each member receives a payoff y_i, which is *i.i.d.* across countries and periods.[17] If the action is not taken, all countries receive their status quo payoff, which is normalized to zero. Country i's payoff (y_i) is privately observed and can take one of two values. With probability p, it takes a high value $\overline{y}_i > 0$,

[15] Note that although countries' strategies will not depend on their private history, their strategies are allowed to depend on the current private information. For example, see Maggi and Morelli (2006) and the 'interim program' in Athey and Bagwell (2001).

[16] Note that this collective action can also be seen as a Leontief technology, where the organization's effective action is $x = min\{d_1, d_2, \ldots, d_N\}$, and each country can choose $d_i \in \{0, 1\}$.

[17] The independence of the shocks does not necessarily make the results of this study easier to attain. Indeed, correlated shocks can more easily rationalize a council of representatives. However, I obtain this result even with uncorrelated preferences.

and with probability $1-p$, it takes a low value $\underline{y}_i < 0$. This induces a probability distribution over preference profiles $\mathbb{P}(\mathbf{y})$ in the usual way. A country is in favor of (against) taking the action whenever its payoff is higher (lower) than the status quo payoff.

In a one-shot game, there are two equilibria. Since countries cannot be forced to take actions, the best possible equilibrium of the one-shot game is to implement the action if and only if everyone agrees. This is called unanimity. Another equilibrium is to never take any action; this is an equilibrium because of the collective action assumption. Both these equilibria satisfy the two relevant incentive conditions: Countries are willing to report their preferences truthfully and are willing to participate by taking the action whenever the organization asks them to do so. In the next subsection, we will study the Pareto-efficient allocations, which describe what outcomes can be implemented assuming that actions are enforceable.

Pareto Efficiency

As a first benchmark, let us consider all the Pareto-efficient allocations. Given a profile of Pareto weights (λ_i), we will characterize the best outcome assuming that the actions are enforceable. In any state of the world $\mathbf{y}$, the Pareto-efficient allocation is the solution to the following problem:

$$\max_{x \in \{0,1\}} x \sum_{i=1}^{N} \lambda_i y_i(\mathbf{y})$$

where $y_i(\mathbf{y})$ is the preference shock of the i^{th} country in state $\mathbf{y}$. Clearly, it is optimal to take the collective action whenever the sum in the expression above is positive and to preserve the status quo when the sum above is negative.

Lemma 1. *Let $-\sum \lambda_i \underline{y}_i$ be the (aggregated) worst possible loss from taking the collective action, and* *let $\overline{y}_i - \underline{y}_i$ be country i's gain from taking the collective action in states that favor that country. Then, the Pareto-efficient rule is to take the action if the weighted sum of the gains of all countries favored in the current state exceeds the worst possible loss, that is, whenever the following condition holds:*

$$\sum_{i|y_i(\mathbf{y})>0} \lambda_i(\overline{y}_i - \underline{y}_i) \geq -\sum_{i=1}^{N} \lambda_i \underline{y}_i$$

Proof. See Appendix A. □

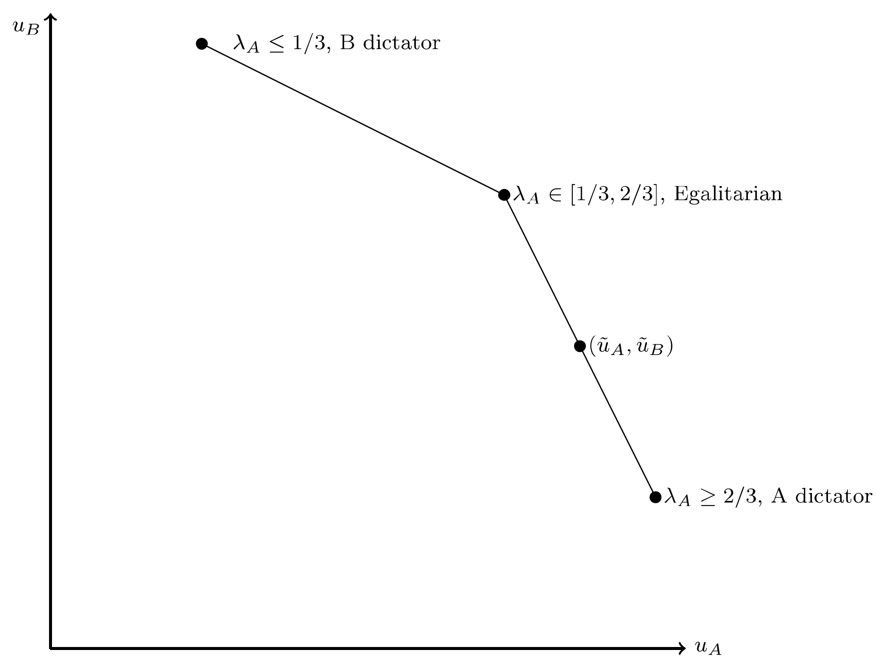
Let us make a remark on the Pareto frontier. If the decision rule is binary, the Pareto frontier consists of a finite set of points. If we allow the collective decision variable x to take values on the $[0,1]$ interval, the Pareto frontier would be convex. In either case, small perturbations in the Pareto weights do not change, in general, the Pareto-optimal decision rule. This can be seen in the following example.

Example 1. There are two countries $\{A, B\}$ with ex ante identical preferences: $\overline{y}_i = 2$ and $\underline{y}_i = -1$ and Lagrange multipliers such that $\lambda_A + \lambda_B = 1$. The worst possible loss is $-\sum \lambda_i \underline{y}_i = 1$. Let us start with one extreme case $\lambda_A = 1$. Here, it is clear that A is a dictator. In particular, the action is not taken in the state $(y_1, y_2) = (-1, 2)$. However, under an egalitarian decision rule, the action would be implemented in that state.

The next step in this example is to compute the smallest λ_A such that A is still a dictator. We solve for this by making the maximization problem indifferent between taking the action or preserving the status quo in $(y_1, y_2) = (-1, 2)$. That is the case when $\sum_{y_i(\mathbf{y})>0} \lambda_i(\overline{y}_i - \underline{y}_i) = -\sum \lambda_i \underline{y}_i$ or replacing the values $\lambda_B(2+1) = 1$, and therefore, $\lambda_B = 1/3$ or $\lambda_A = 2/3$. For any $\lambda_A = 2/3 - \epsilon$, the Pareto-efficient allocation will implement the action in state $(y_1, y_2) = (-1, 2)$ as well as states $(2, -1)$ and $(2, 2)$. This is indeed the egalitarian outcome, which is implemented not only for $\lambda_A = 1/2$ but for any λ_A in the range $[1/3, 2/3]$.

Finally, by symmetry, it is clear that $\lambda_A < 1/3$ makes B a dictator. Thus, the three relevant voting rules in the example are A-dictatorship, B-dictatorship, and the egalitarian allocation. If we restrict the decision-making rule to be discrete, there are only three efficient outcomes. However, we can obtain a convex Pareto frontier by convex combinations of A-dictatorship with egalitarian and B-dictatorship with egalitarian. This is illustrated in figure 1.

Figure 1: Expected payoffs



Let us consider the payoff profile $(\tilde{u}_A, \tilde{u}_B)$. This payoff cannot be obtained by a discrete mechanism. However, it is still efficient when $\lambda_A = 2/3$. This payoff can be obtained by tossing a coin, and then with probability 1/2 making A a dictator or with probability 1/2 implementing the egalitarian outcome. Alternatively, this outcome can also be implemented by setting $x = 1$ in states $(2, 2)$ and $(2, -1)$; $x = 1/2$ in state $(-1, 2)$; and $x = 0$ in state $(-1, -1)$. The interpretation of $x = 1/2$ is the following: In some states, the organization decides to compromise by implementing the action only partially. Henceforth, we will assume the following:

Assumption. *The choice variable x can take values on the interval* $[0,1]$.

Weighted Voting

Before we study the equilibrium of the game, let us propose an alternative and less abstract way of considering the Pareto-efficient decision rule. To do this, first let us define a weighted voting rule as a profile of weights m and a target M such that every country has a weight m_i, countries vote on whether they want to take the collective action, and the action is implemented if the sum of the weights of all members that voted in favor of taking the action exceeds a target M. Otherwise, the outcome will be the status quo.

Lemma 2. *For a given profile of Pareto weights, the efficient outcome can be implemented by a weighted voting rule.*

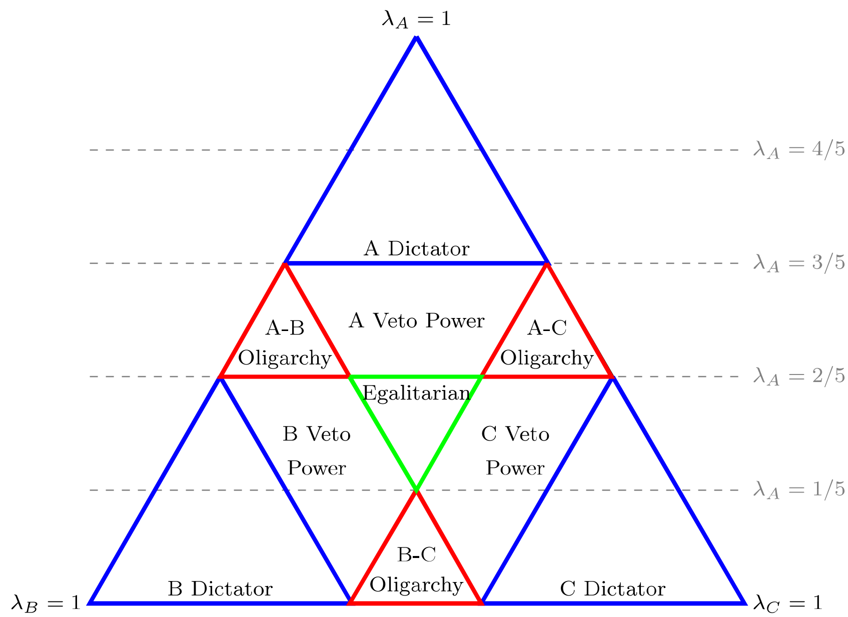
Proof. See Appendix A. □

The previous lemma can be seen as special case of Barbera and Jackson (2006). It tells us that any efficient decision rule can be achieved by weighted votes. In addition, the weights and target are not unique. Clearly, multiplying the pair (m_i, M) by a positive constant will serve the same purpose. Moreover, in general, small perturbations $(m_i+\epsilon_i, M+\epsilon_M)$ will implement the same outcome. Another property is that the class of outcomes that can be implemented by weighted votes is fairly large. Indeed, weighted votes can include some well-known examples such as the egalitarian decision rule, dictatorship, oligarchy, veto power, and one-country-one-vote.[18] The following example illustrates all

[18]The ability to implement veto power by voting weights depends on the payoff structure. Indeed, there exists veto power when the loss $\underline{y}_i$ is large relative to the gain $\overline{y}_i$. If in the next example, gains are relatively larger than losses, there would not be veto power, and instead there would be something like a 'reverse veto power.' That is, an allocation rule whereby a country has the power to unilaterally implement an action but does not have the power to veto in every state. If a country has veto power and reverse veto power, then that country is a dictator.

these points.

Figure 2: Decision rules for all combinations of Pareto weights.



Example 2. There are three countries $\{A, B, C\}$ with ex ante identical preferences: $\overline{y}_i = 2$ and $\underline{y}_i = -3$. Moreover, let us normalize the Pareto weights such that their sum equals one. To illustrate one simple case, take $\lambda_A = 1$. Then, the Pareto-efficient allocation will be to take the action if and only if country A votes *yes*. Moreover, as noted above, small perturbations in the Pareto weights do not change (in general) the decision rule. Thus, for λ_A near to one, country A will still be a dictator. Figure 2 shows all possible combinations of decision rules in this example. In the region labeled *A-B Oligarchy*, both A and B have veto power, and country C is never pivotal. In the region labeled *A Veto Power*, country A has sufficient weight to veto unfavorable decisions but not enough weight to be a dictator. Moreover, in this region, B and C can be pivotal. They change the outcome in the event that A voted *yes* and the two other countries disagree with each other.

2.2 The Repeated Game

The N members of the international organization interact repeatedly over time and discount time using a constant factor δ. From the payoff structure, there are effectively three alternatives in each period: take the collective action, preserve the status quo, or leave the organization. Therefore, we can simplify notation by keeping track of only the action implemented instead of each country's individual participation.[19] The public history in period $t > 0$ consists of the history of actions implemented and the reported preference profiles: $\mathbf{h}_t = \left\{(x_0, \mathbf{y'}_0), (x_1, \mathbf{y'}_1), \ldots, (x_{t-1}, \mathbf{y'}_{t-1})\right\}$. In the initial period, the history is the null set. For a given country i, its interim payoff in the initial period will be:

$$(1-\delta)x_0 y_{i,0} + \delta \sum_{t=1}^{\infty} \delta^{t-1}(1-\delta)E\left[x_t y_{i,t} | \mathbf{h}_1\right] \tag{1}$$

where the expectation is taken over all histories following $\mathbf{h}_1 = (x_0, \mathbf{y}_0)$. Let us recall that the status quo payoff comes from not taking the action in the current period while remaining in the organization. For simplicity, we can assume that receiving the status quo payoff forever is the same as not having the organization at all.[20] Thus, we can also set the outside option payoff equal to zero. Finally, we also assume that $p\overline{y}_i + (1-p)\underline{y_i} > 0$. This assumption ensures that, the status quo payoff is smaller than the payoff from having zero decision-making power.[21] These two assumptions avoid corner solutions but are not essential for the results of the model. In this environment, a strategy for each country is a message (declaring its preference shock) and an action that depend on the public history. Moreover, a PPE is a profile of strategies that are mutual best responses to one another.

[19] That is, instead of choosing a vector of individual decisions $(d_1, d_2, \ldots, d_N)$, which will effectively implement an action equal to $\min\{d_i\}$, we can use the final output itself. Note that this works for both discrete and continuous action spaces. Moreover, keeping track of the individual participation decisions potentially creates multiple equilibria that are payoff equivalent.

[20] However, we could make the outside option different from the status quo payoff. This is discussed in section 2.5.

[21] This would justify, for instance, all the countries that belong to the United Nations but have never been part of the Security Council.

Note that a natural candidate for the payoff-maximizing equilibrium is the Pareto-efficient allocation, together with grim trigger strategies. Moreover, grim trigger strategies are not an assumption. Indeed, the best way to provide incentives is by punishing 'off-the-equilibrium-path' behavior in the most severe yet credible way.[22] Namely, after observing a deviation from the equilibrium path, the organization is dissolved, and all members receive the status quo payoff forever. On the other hand, the Pareto-efficient allocation may not always be part of an equilibrium; thus, I am particularly interested in finding the optimal equilibrium in such situations.

2.3 A Mechanism Approach

We can regard the organization as a mechanism that collects preferences and suggests an outcome. Therefore, I will use the terms 'organization' and 'mechanism' interchangeably. To simplify notation, let us denote the recommended action by x. Therefore, $x(\cdot)$ will be a plan of actions that depends on reported preferences and the public history. Given any mechanism, a member of the organization will receive a payoff equal to the present discounted sum of the streams of all its payoffs as defined in (1). Moreover, there are three constraints to be satisfied. First, all countries must be willing to join and maintain their membership in the organization. Second, all members must be willing to participate by taking the action whenever the organization decides to do so. Third, the members of the organization should truthfully report their preferences. Since the aim of the mechanism is to provide incentives for behavior on the equilibrium path, unless explicitly mentioned, all the analysis below describes on-path histories. Thus, the organization's maximization problem is:

[22] This would not hold under imperfect monitoring.

$$\max_x (1-\delta)\sum_{t=0}^{\infty}\delta^t E\left[x_t(\mathbf{y}_t,\mathbf{h}_t)\sum_{i\in N}\lambda_i y_{i,t}\right] \tag{2}$$

subject to

$$(1-\delta)\sum_{\tau=t}^{\infty}\delta^{\tau-t}E_\tau\left[x_\tau(\mathbf{y}_\tau,\mathbf{h}_\tau|\mathbf{h}_t)y_{i,\tau}|\mathbf{h}_t\right]\geq 0, \forall i,t,\mathbf{h}_t \tag{3}$$

$$(1-\delta)x_t(\mathbf{y_t},\mathbf{h}_t)y_{i,t}+\delta\left((1-\delta)\sum_{\tau=t+1}^{\infty}\delta^{\tau-t-1}E_\tau\left[x_\tau(\mathbf{y}_\tau,\mathbf{h}_\tau|\mathbf{h}_{t+1})y_{i,\tau}|\mathbf{h}_{t+1}\right]\right)\geq 0, \tag{4}$$

$$\forall i,t,y_t,\mathbf{h}_t,\ \text{ and } \mathbf{h}_{t+1}=(\mathbf{h}_t,x_t(\mathbf{y_t},\mathbf{h}_t),\mathbf{y_t})$$

and

$$E\left[x_t(\mathbf{y}_t,\mathbf{h}_t)|\mathbf{h}_t,y_{i,t}\right]y_{i,t}\geq E\left[x_t((\mathbf{y}_{-i,t},\hat{y}_{i,t}),\mathbf{h}_t)|\mathbf{h}_t,\hat{y}_{i,t}\right]y_{i,t},\forall i,t,y_t,\mathbf{h}_t,y_{i,t}\neq\hat{y}_{i,t} \tag{5}$$

Equation (3) is the *voluntary membership* constraint. It states that after every history, the expected payoff of each member must be more desirable than leaving the organization, which I assume yields zero payoff forever. Equation (4) is the *participation* constraint. It states that after every decision made by the organization, the members must be willing to participate in the organization's decision. If the members comply, they receive an instant payoff and a continuation payoff. The sum of these two payoffs must be at least as good as the alternative, which is to not participate in the action and therefore receive the status quo payoff forever.[23] Equation (5) is the *truth-telling* condition: Members

[23]The instant payoff they receive from not taking the action is zero, and the continuation payoff is the status quo forever, as the organization is dissolved.

should truthfully report their preferences. Note that equation (5) already uses the stationary payoffs assumption; thus, the only way to provide incentives to report preferences truthfully is through the payoffs in the current period. Similarly, the equations above already use the fact that on the equilibrium path, the most stringent credible punishment is to dissolve the organization and award zero payoffs to all countries.

Lemma 3. *The following three conditions hold:*

i) The voluntary membership constraint (3) is not binding at the optimum.

ii) $x_t((y_{i,t})_{i=1}^N, \mathbf{h}_t)$ *increasing in each* $y_{i,t}$ *is a sufficient condition for (5).*

iii) The solution to (2) is an optimal equilibrium to the repeated game.

Proof. See Appendix A. □

By using parts (i) and (ii) from lemma 3, I will guess that the optimal mechanism $x_t(\mathbf{y}_t, \cdot)$ will be indeed increasing. Then, I will solve the simplified problem that maximizes (2) subject to (4). Moreover, using part (iii) in the previous lemma, we know that the solution to this simplified problem will be an optimal equilibrium of the repeated game.

2.4 Solution

To simplify the equilibrium solution, it is standard in the literature to decompose the payoffs in equation (1) into the sum of an instant payoff and a 'future' payoff of the form: $(1-\delta)x(\mathbf{y})y_i + \delta v_i(\mathbf{y})$. In addition to strategies being best responses to one another, the profile of future payoffs $(v_i(\mathbf{y}))_{i=1}^N$ must belong to the set of sequential equilibrium payoffs. Instead of using this method, I propose an 'action-based' approach, which fits more appropriately the environment considered in this study for two reasons: action is one-dimensional and we restrict attention to stationary payoff PPE. Moreover, this approach allows, in principle, to find an analytical solution. Although it may be impractical to find a closed form solution, it can easily be solved numerically.[24]

Moreover, as we are restricting attention to stationary payoff equilibria, it is possible to compute the solution to problem (2) by considering the following maximization problem:

$$\max_{\tilde{x}} \sum_{\mathbf{y}\in\mathbf{Y}} \mathbb{P}(\mathbf{y})\tilde{x}(\mathbf{y}) \sum_{i\in N} \lambda_i y_{i,t} \tag{6}$$

subject to

$$(1-\delta)\tilde{x}(\mathbf{y})y_i + \delta \sum_{\hat{\mathbf{y}}\in\mathbf{Y}} \mathbb{P}(\hat{\mathbf{y}})\tilde{x}(\hat{\mathbf{y}})\hat{y}_i \geq 0, \forall i, \mathbf{y} \tag{7}$$

$$\tilde{x}(\cdot) \text{ increasing}$$

Lemma 4. *The solutions to (6) and (2) attain the same payoff. Moreover, let* $(\tilde{x}(\mathbf{y}))_{\mathbf{y}\in\mathbf{Y}}$ *be the solution to (6). Then, there is an optimal equilibrium to the repeated game characterized by stationary actions* $x_t(\cdot, \mathbf{h_t}) = \tilde{x}(\cdot)$.

Proof. See Appendix B. □

Lemma 4 greatly simplifies the maximization problem. The result is quite obvious: since we restrict attention to stationary payoffs, even in the case of multiple equilibria, there repetition of the action schedule $\tilde{x}(\cdot)$ should be a solution. The following result is also standard in repeated games. It states that for sufficiently patient players, the efficient allocation can be implemented:

[24] Section 3 has examples of both analytical and numerical solutions.

Proposition 1. *There is a threshold δ^* such that if the discount factor exceeds that threshold, the Pareto-efficient allocation is the optimal mechanism.*

Proof. See Appendix C. □

Remark 1. *Note that if $\delta \geq \delta^*$, the solution to the problem under stationary payoffs is also the solution with unrestricted payoffs.*

The heterogeneity of the members provides a relevant extension to Maggi and Morelli (2006). In their paper, countries are homogeneous, and therefore, if the Pareto-efficient outcome cannot be implemented, the optimal mechanism is unanimity. However, when the countries are heterogeneous, it is possible to implement an outcome better than unanimity when δ is small, even arbitrarily close to 0.

Proposition 2. *For any $0 < \delta < \delta^*$, the optimal equilibrium delivers a payoff strictly better than unanimity.*

Proof. See Appendix C. □

The intuition for this is simple: Since x can take values between 0 and 1, there is always a convex combination of the Pareto-efficient allocation and unanimity that is feasible. Therefore, the optimal mechanism will deliver a payoff at least as large. Moreover, a similar result holds under some generic conditions even if x can only take discrete values.[25]

Proposition 1 characterizes the solution to the maximization problem for organizations with patient members. However, proposition 2 only states that the payoffs are somewhere in between efficiency and unanimity and that there is some degree of compromise: Some nations will forgo decision-making power to provide incentives to other nations that are tempted to quit. We can take one step further

[25] See proposition 4 in section 2.5.

and study what the equilibrium will look like on the interval $(0, \delta^*)$. To do so, let us denote $\gamma_i(\mathbf{y})$ as the Lagrange multiplier of the participation constraint in problem (6) for country i in state $\mathbf{y}$. Then, we have the following.

Proposition 3. *When* $0 < \delta < \delta^*$*:*

i) The optimal mechanism can be implemented by voting weights that are state dependent and, therefore, stochastic.

ii) The voting weights and target are given by:

$$m_i(\mathbf{y}) = (\lambda_i + \delta E[\gamma_i])(\overline{y}_i - \underline{y}_i) - (1-\delta)\gamma_i(\mathbf{y}_{-\mathbf{i}}, \underline{y}_i)\underline{y}_i \tag{8}$$

$$M(\mathbf{y}) = -\sum_i \left[(\lambda_i + \delta E[\gamma_i])\underline{y}_i + (1-\delta)\gamma_i(\mathbf{y}_{-\mathbf{i}}, \underline{y}_i)\underline{y}_i\right] \tag{9}$$

Proof. See Appendix C. □

The previous results extend Maggi and Morelli (2006) and allow for a relevant implication that was not part of their equilibrium solution: random voting power. This is relevant because it provides a first step to rationalize via an optimal mechanism the rotation of Security Council seats at the United Nations. Section 3 addresses this topic in further detail, where I restrict attention to a class of parameter conditions that resemble the United Nations. However, before we turn to such a restriction of the model, let us analyze possible extensions.

2.5 Discussion and Extensions

Outside Options Recall that we assumed that the payoffs after the organization is dissolved is the same as the status quo forever. Instead, we can assume that each member has a different outside option b_i. The introduction of an outside option does not greatly affect the equilibrium of the model, but it may add testable implications.

First, the voluntary membership restriction could be binding[26] in this case. Thus, we need to take this restriction into account:

$$\sum_{\hat{\mathbf{y}} \in \mathbf{Y}} \mathbb{P}(\hat{\mathbf{y}}) \tilde{x}(\hat{\mathbf{y}}) \hat{y}_i \geq b_i$$

In addition, the participation constraint must also be adjusted. The left-hand side represents payoffs on the equilibrium path, so it does not change; however, the right-hand side depends on two terms: first, the instant payoff of not complying, which is zero, plus the discounted payoff of not having the organization, b_i.

$$(1-\delta)\tilde{x}(\mathbf{y})y_i + \delta \sum_{\hat{\mathbf{y}} \in \mathbf{Y}} \mathbb{P}(\hat{\mathbf{y}}) \tilde{x}(\hat{\mathbf{y}}) \hat{y}_i \geq \delta b_i$$

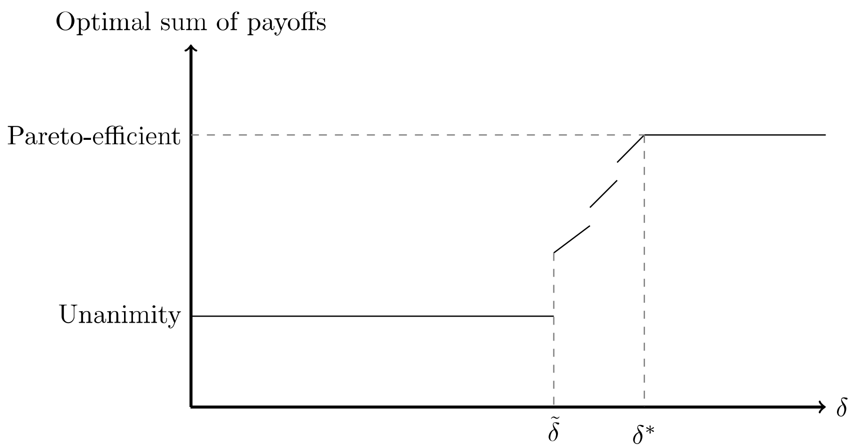
Proposition 3 will still be valid, but the voting weights and target will depend on one additional term that captures how binding the new restrictions are due to the introduction of the outside option.

Discrete Choice Mechanism Now we will study the case in which the choice variable can only be discrete: $x \in \{0, 1\}$. When this is the case, the threshold δ^* does not change, and the random voting weights property still holds for $\delta < \delta^*$. On the other hand, it should be clear that there are discontinuities in the organization's value function (the solution to (6) as a function of δ). In

[26] With the addition of an outside option, part (i) of lemma 3 is no longer true.

particular, there is a $\tilde{\delta} > 0$ that is the smallest discount factor that can sustain a payoff strictly better than unanimity. The previous points are illustrated in figure 13.

Figure 3: Optimal stationary mechanism payoff as a function of δ



Proposition 4. *The following statements hold when the choice variable is restricted to belong to* $\{0, 1\}$*:*

(i) The threshold δ^* *is the same as in the case of a continuous choice variable.*

(ii) There is another threshold $0 < \tilde{\delta} \leq \delta^*$ *such that for any* δ *below it, the optimal equilibrium is unanimity.*

(iii) Let $\underline{\delta}_i$ *be the minimum discount factor that satisfies country* i*'s participation constraint at the Pareto-efficient allocation. A necessary condition for* $\tilde{\delta} < \delta^*$ *is that there are at least two countries* i *and* j *such that* $\underline{\delta}_i < \underline{\delta}_j$*.*

Proof. See Appendix C. □

In words, part (*iii*) from the proposition above holds because if the discount factor is not large enough to implement the Pareto-efficient allocation, the organization has to provide a more favorable outcome to some countries. However, this means that some other countries will receive a payoff that is lower than the payoff they would obtain in the Pareto-efficient allocation. For those countries to still be willing to participate in the decisions of the organization, a minimum requirement is that they have some slack in their participation constraint. That is, the minimum discount factor that would make them still comply with the Pareto-efficient allocation is strictly larger than their actual discount factor $\underline{\delta}_i < \delta$. Moreover, if the threshold for all countries is the same ($\underline{\delta}_1 = \underline{\delta}_2 = \cdots = \underline{\delta}_N$), it is not possible to 'transfer' some payoff from one country to another without violating their participation constraints. In particular, this is the case when countries are homogeneous.

Imperfect Monitoring When the organization does not perfectly observe the participation of its members but instead receives imperfect signals of the compliance of each country, grim trigger strategies are no longer optimal. Let us consider two signals, one can be labeled as the *good signal* and be highly correlated with the country's participation, and the other can be labeled as the *bad signal* and be highly correlated with the country's defection. Then, in every period, the decision made by the organization x should also depend on the history of observed signals. In equilibrium, the expected discounted payoff of a member should be higher after the organization has observed a good signal, and it should be lower after the organization has observed a bad signal. This means that the voting weight of a country after a good signal should be higher than after a bad signal.

Non-Stationary Payoffs There are two major changes relative to problem (2). First, the truth-telling condition can no longer be simplified. Second, the equilibrium guess $x_t(\cdot, \mathbf{y}_t) = \tilde{x}(\cdot)$ is only true for $\delta \geq \delta^*$. This problem is much more difficult to solve and therefore is beyond the scope of this paper; however, we can say that one property of the optimal equilibrium is that the voting weights are still random for a range of values of δ, and in addition, they are history dependent. A country that has been unlucky (received negative shocks and was requested to participate by taking an action) will have accumulated decision-making power over time, so in the future this country is less likely to have negative shocks. This property is in line with empirical findings in Dreher et al. (2014), where the longer a country has not been part of the Security Council, the more likely it is to be elected in the future.

3 Applications of the Model to the United Nations

Of all the international organizations, one of the (arguably) most influential and powerful is the United Nations. It is composed of several organs (such as the General Assembly and the Security Council) and agencies (such as the IMF and the World Bank). According to its charter, the main purpose of the United Nations is to maintain international peace and security. The organ devoted to this specific task is the Security Council, which meets periodically to propose and vote on resolutions that are compulsory[27] to all members of the United Nations. However, only fifteen countries, five permanent and ten rotating (non-permanent), have the right to vote on the Security Council, from a pool of 193 members. The ten non-permanent members of the Security Council have a tenure of two years, cannot be immediately reelected, and must have the support of at least two-thirds of all the other members. In principle, this may suggest that there should be a better mechanism to choose the decisions to maintain peace, as the preferences of most of the members are being ignored (more information is usually better).

Two questions that arise from this voting setup are, first, why and under what circumstances is it optimal to ignore the opinions of the majority of the United Nations members? Second, why would members comply with resolutions on which they did not even vote? The simple answers to those

[27] Country members are expected to follow Security Council decisions; otherwise, they could receive sanctions.

questions are that (i) to secure the participation of all members, some countries have to have a high voting weight when they are tempted to defect, and (ii) members comply because the value from remaining in the organization is still high enough that a current low payoff is bearable. In this section, I will describe in greater detail the answers to these questions by mapping more closely the theoretical model to the voting system of the United Nations. First, we will discuss veto power. Then, we will find conditions that ensure a council-like voting system with rotation. Finally, we will study heterogeneity within the rotating members.

3.1 Veto Power

We saw on example 2 that a weighted voting rule can include veto power. If the Pareto weights of the five permanent members (P5) were initially very high, that could explain why they must have veto power in every period. The historical explanation for the five permanent members aligns perfectly with this assumption. The victors in WW2 decided to create an organization with the mission of preventing war while guaranteeing their own power.[28] Let us begin by studying the Pareto-efficient allocation. A member i has veto power whenever:

$$\sum_{j \neq i} \lambda_j \overline{y}_j < -\lambda_i \underline{y}_i$$

Now, to simplify the model, let us assume that there are $N_A < N$ members that have a Pareto weight of $\lambda_i = 1$, while the rest of the $N_B = N - N_A$ members have a Pareto weight of zero. This special case is intended to capture the five WW2 victors that created, organized, and later invited

[28]For example, Bourantonis (2005) stated, *"When the UN Charter was being drafted, the end of the Second World War was still in sight, with easily discernible winners and losers. The intention was for the victorious states, which were the world's great powers at the time, to exercise global leadership with a view to managing or governing the international system.... The overriding role of the Security Council reflected the strong desire of the founders of the United Nations to see it play an increasingly central role as the leading world forum for managing threats to the international order."*

other countries to join the United Nations. Therefore, it is reasonable to assume that they had all the Pareto weight. According to lemma 2, in the Pareto-efficient allocation, only the first N_A members should have a positive voting weight.

Since there are two types of countries, let us also assume that each type has a different payoff structure: the first N_A members have payoffs in $\{\underline{y}_A, \overline{y}_A\}$ and the remaining members have payoffs in $\{\underline{y}_B, \overline{y}_B\}$. Then, a member i will have veto power in the Pareto-efficient allocation if and only if $i \leq N_A$ and $(N_A - 1)\overline{y}_A < -\underline{y}_A$. Moreover, for δ smaller than but sufficiently close to δ^*, the incentive constraints will be binding only for members without veto power ($i > N_A$). In other words, having veto power ($i \leq N_A$) guarantees that a member will not be the first to have its participation constraint binding as δ decreases below δ^*. This suggests the idea that the first N_A members may still have veto power for $\delta < \delta^*$. Next, I will state conditions under which this conjecture holds under the optimal mechanism.

Lemma 5. *If the following conditions hold, veto power is guaranteed when $\delta < \delta^*$:*

i) $p^{N_B}\overline{y}_B > (1-p)^{N_B}\underline{y}_B$

ii) $N_A\overline{y}_A\overline{y}_B < \left((N_A - 1)\overline{y}_A + \underline{y}_A\right)\underline{y}_B$

Proof. See Appendix E. □

The previous lemma is indeed easier to prove as a requirement for the results in the next section. Therefore, its proof will be provided after the proofs of propositions 5, 6 and 7. Moreover, these are sufficient conditions, and not all of them are linked to each of the results. See appendix E for details.

3.2 Rotation

The next step is to study the rotation of Security Council seats when $\delta < \delta^*$, so that according to proposition 3, random voting weights are part of the optimal equilibrium. Let x^P be the Pareto-efficient allocation and x^U the unanimity allocation, and define $x^{**} \in (0,1)$ as the solution of:

$$(1-\delta)x^{**}\underline{y}_B + \delta\left(p^{N_A}x^{**}(p\overline{y}_B + (1-p)\underline{y}_B) + p^N(1-x^{**})\overline{y}_B\right) = 0 \tag{10}$$

Then:

Proposition 5. *Assume that the conditions in lemma 5 hold. Then, there is a discount factor* $0 < \delta^{**} \leq \delta^*$ *such that for any* $\delta < \delta^{**}$*, the optimal mechanism is characterized as* $x = x^{**}x^P + (1-x^{**})x^U$.

Proof. See Appendix D. □

Note that x^{**} is a scalar. In words, this result states that the decision variable in the optimal mechanism is a compromised level between the Pareto-efficient allocation and the unanimity allocation. Specifically, let us divide all states into three subsets: (i) when at least one of the first N_A country members disagrees, it can still exert its veto power; (ii) when all of the first N_A members agree and at least one of the N_B members disagrees, the action implemented is x^{**}; (iii) if everyone agrees, then $x = 1$. Moreover, it is important to note that even in the case in which all of the first N_A members agree but the remaining N_B disagree, the action is still partially implemented; this might seem counter-intuitive. However, this is indeed optimal for δ small. The reasoning is quite obvious when one considers the action-taking constraint (7). When δ approaches zero, the marginal effect that the action has on the incentive constraint via 'future payoffs' ($\tilde{x}$) is negligible. Therefore, all states are 'equally impactful' in terms of providing incentives. On the other hand, in some cases $\delta^{**} < \delta^*$, and the linear combination $x^{**}x^P + (1 - x^{**})x^U$ is no longer optimal for $\delta \in (\delta^{**}, \delta^*)$. Furthermore, note that the first N_A members still hold their veto power when $\delta < \delta^{**}$. This characterizes the optimal equilibrium for δ small.

Although the previous result is robust, it is difficult to map it to any known voting system. x^{**} is arbitrarily small for δ close to zero. Therefore, this mechanism approaches unanimity as δ approaches zero. As mentioned above, some international organizations do follow unanimity, such as the WTO and NATO. One could argue that in those organizations, sometimes x^{**} is implemented and labeled as $x = 1$. Moreover, countries could 'nominally' all agree and vote in favor of x^{**} with the implicit threat that any action $x > x^{**}$ will be rejected.

The next step is to better understand the optimal equilibrium when δ has intermediate values. In what follows, let us study the cases in which $\delta^{**} < \delta < \delta^*$. The type of equilibria within this region, which preserves veto power, can be divided in two parts, separated by a threshold $\hat{\delta}$ such that $\delta^{**} \leq \hat{\delta} \leq \delta^*$. Let $\mathbf{Y_k}$ be the subset of states such all of the first N_A members and exactly k of the remaining N_B members agree. Each of those subsets has $\binom{N_B}{k} = \frac{N_B!}{(N_B-k)!k!}$ elements.

Let k'_B be the minimum of:

i) The largest number k such that $k\overline{y}_B + (N_B - k)\underline{y}_B < 0$.

ii) The argument that solves:[29]

$$\min_{k_B} \left\{ \frac{-\underline{y_B}}{-\underline{y_B} + p^{N_A} \sum_{k=k_B+1}^{N_B-1} \mathbb{P}(k) E[y_B|k]} \right\}$$

[29]Note that the summation goes until $N_B - 1$. Indeed, if the summation were to go all the way to N_B, then part (i) in this definition would be sufficient for this second part.

s.t.

$$\frac{-\underline{y_B}}{-\underline{y_B} + p^{N_A}\sum_{k=k_B+1}^{N_B-1}\mathbb{P}(k)E[y_B|k]} \geq 0$$

where $\mathbb{P}(k) = \binom{N_B}{k}p^k(1-p)^{N_B-k}$ and $E[y_B|k] = (k\overline{y}_B + (N_B - k)\underline{y}_B)/N_B$.

Proposition 6. *Assume that the conditions in lemma 5 hold. Then, there are thresholds* $\delta^{**} = \underline{d}_0 \leq \underline{d}_1 \leq \cdots \leq \underline{d}_{k'_B} \leq \underline{d}_{k'_B+1} = \hat{\delta}$ *such that when* $\delta \in (\underline{d}_{k_B-1}, \underline{d}_{k_B})$*:*

i) In the state where everyone agrees, $x = 1$*.*

ii) All subsets $\mathbf{Y_k}$ *such that* $0 \leq k < k_B$ *will satisfy* $x(\mathbf{y}) = 0$ *for* $\mathbf{y} \in \mathbf{Y_k}$*, and the participation constraints will be slack in those states.*

iii) All subsets $\mathbf{Y_k}$ *such that* $k_B \leq k < N_B - 1$ *will satisfy* $x(\mathbf{y}) = \hat{x} > 0$ *for* $\mathbf{y} \in \mathbf{Y_k}$*, and* $\hat{x}$ *adjusts in such a way that the participation constraints will be binding in those states.*

iv) $\hat{x}$ *increases in* δ*.*

Proof. See Appendix E. □

In words, as δ increases above δ^{**}, the state where all type-A countries agree and all type-B countries disagree has a discontinuous fall to $x = 0$. Moreover, as δ continues increasing, the states where all type-A countries agree and only one type-B country disagrees will also have a discontinuous fall to $x = 0$. These 'falls' continue occurring as δ increases for all subsets $\mathbf{Y_k}$ such that $k \leq k'_B$. The next result describes the outcome when δ continues increasing above $\hat{\delta}$.

Proposition 7. *There are thresholds* $\hat{\delta} = \overline{d}_{k'_B+1} \leq \overline{d}_{k'_B} \ldots \overline{d}_1 \leq \overline{d}_0 = \delta^*$ *such that when* $\delta \in (\overline{d}_{k_B+1}, \overline{d}_{k_B})$*:*

- *All subsets* $\mathbf{Y_k}$ *such that* $k < k_B$ *will satisfy* $x(\mathbf{y}) = 0$ *for* $\mathbf{y} \in \mathbf{Y_k}$*, and the participation constraints will be slack in those states.*

- *The subset* $\mathbf{Y_k}$ *such that* $k = k_B$ *will satisfy* $x(\mathbf{y}) = \hat{x} > 0$ *for* $\mathbf{y} \in \mathbf{Y_k}$*, and the participation constraints will be slack in those states.*

- *All subsets* $\mathbf{Y_k}$ *such that* $k'_B + 1 \leq k \leq N_B$ *will satisfy* $x(\mathbf{y}) = 1$ *for* $\mathbf{y} \in \mathbf{Y_k}$*, and the participation constraints will be binding in those states, except at* $k = N_B$*.*

- $\hat{x}$ *adjusts in such a way that the participation constraints will be binding for states in* $\mathbf{Y_k}$ *such that* $k'_B + 1 \leq k \leq N_B - 1$*.*

Proof. See Appendix E. □

The previous result states that it is not optimal to set $x = 0$ in states such that such that $k_B\overline{y}_B + (N_B - k_B)\underline{y}_B > 0$. Instead, $x(\mathbf{y})$ will jump to 1, on such states, and $x(\mathbf{y})$ will jump to some $\hat{x} > 0$ in the states where k'_B type-B members agree. Moreover $\hat{x}$ is increasing in δ, and once it has reached 1 in those states k'_B, $\hat{x} > 0$ will be implemented in states where all type-A and $k'_B - 1$ type-B members agree, and so forth. Moreover, note that it is possible to have some number k_B such that $\overline{d}_{k_B} = \overline{d}_{k'_B+1}$ or $\underline{d}_{k_B} = \underline{d}_{k'_B+1}$. Example 3 describes a simple case in which this situation occurs. Finally, Appendix E provides a table with a graphical summary of propositions 6 and 7.

Furthermore, as $\delta \to \hat{\delta}$, the states with $k'_B + 1$ type-B members partially implement some $\hat{x} > 0$. However, as δ increases, it is no longer optimal to reduce the support (number of the states) such that $x > 0$.[30] Indeed, what determines $\underline{d}_{k'_B+1} = \hat{\delta}$ is the restriction $x \leq 1$. Therefore, as $\delta \to \hat{\delta}$, $x \to 1$. When $\delta > \hat{\delta}$, all states such that all type-A members agree and at least $k'_B + 1$ type-B members also agree will implement an action $x = 1$. Moreover, there is a set of states characterized by k_B type-B countries such that $x = \hat{x} < 1$; the threshold k_B decreases with δ, meaning that the action is

[30] That is why the expression for $\underline{d}_{k'_B+1} = \hat{\delta}$ (see equation (22) in Appendix E) is different from the rest of the thresholds $\underline{d}_{k_B}$ (see equation (21) in Appendix E)

implemented in more states.

Although the previous two results provide a very detailed characterization, they lack a useful interpretation. Fortunately, there are equivalent mechanisms that seem more natural and can implement the same allocations. Indeed, the next corollary provides an example of such mechanisms. Namely, the mechanism proposed will be a council; that is, a subset of members are chosen as 'representatives,' and they vote in the hope of implementing the optimal allocation.

As mentioned above, veto power can be implemented using voting weights. Therefore, there is an implicit way to obtain veto power. One immediate example would be to (i) set a council size equal to N_A; (ii) only $N_A - 1$ out of the first N_A members can vote; (iii) one out of the remaining N_B members can vote; (iv) the (council's) voting rule is unanimity; (v) in states where veto power would be used, one of the $N_A - 1$ type-A members that would vote 'no' is elected as part of the council; (vi) in states where veto power would not be used, but it is optimal not to implement the action, one type-B member that would vote 'no' is elected as part of the council; and finally, (vii) in states where it is optimal to implement $x > 0$, elect any $N_A - 1$ type-A members and one type-B member that would vote 'yes.'

On the other hand, since the United Nations uses explicit veto power, it is worth exploring in greater detail mechanisms that explicitly use veto power. The following corollary characterizes a large class of such alternative mechanisms:

Corollary 1. *Consider the same conditions and notation as in propositions 6 and 7:*

- *The same optimal mechanism from either of those two propositions can be implemented by a council-like voting system, where the first N_A members always have the right to vote and have veto power, a subset $\tilde{N}$ of the remaining N_B members will also have the right to vote, and a* *number $N^* \leq N_A + \tilde{N}$ of council members must vote in favor to implement an action.*

- *The $\tilde{N}$ members rotate among the type-B countries. Although rotation is ex ante stochastic, it is state dependent.*

- *For a given discount factor, let $k(\delta)$ be the number that indicates the minimum number of 'yes' votes of type-B countries required to implement some $x > 0$. Then, the range of pairs $(\tilde{N}, N^*)$ that implement the optimal equilibrium satisfies:*

$$\max\{0, k(\delta) - 1 + \tilde{N} - N_B\} < N^* - N_A \leq \min\{k(\delta), \tilde{N}\} \tag{11}$$

Proof. See Appendix E. □

Remark 2. *Note that for N_B larger than $3N_A$, equation (11) is satisfied by setting $N^* = 2N_A - 1$ and $\tilde{N} = 2N_A$. This particular mechanism mimics the case of the United Nations Security Council.*

The last point from the previous corollary, which is illustrated in remark 2, shows how the optimal equilibrium can be remarkably similar to the United Nations Security Council. Although the United Nations votes on several issues each year, this mechanism provides a good approximation that rationalizes the rotation of the non-permanent members of the Security Council for one issue (or one 'aggregate' issue). Namely, the first N_A members are equivalent to the P5, as they always vote and have veto power. The action is implemented if all $N_A = 5$ members and at least $N^* - N_A = N_A - 1 = 4$ of the remaining N_B members agree. Moreover, only 4 type-B members with a positive shock are required.

Note that it is perfectly valid to elect to the council members that have a negative payoff. The fact that in reality most of the resolutions are passed with unanimity does not necessarily exclude the possibility that some members disagree. Indeed, those members know that their vote will not be pivotal, precisely because the council was carefully elected to implement certain actions. As a consequence, they might as well vote in favor.

Moreover, the United Nations' charter states that *"...The General Assembly shall elect ten (...) Members of the United Nations to be non-permanent members of the Security Council, due regard being specially paid, in the first instance to the contribution of Members of the United Nations to the maintenance of international peace and security and to the other purposes of the Organization..."*[31] This statement relates election to the Security Council seats to compliance with United Nations objectives.

Another relevant stylized fact of the Security Council is that there is heterogeneous rotation. That is, countries such as Japan, India and Brazil are part of the Security Council far more often than other members. This behavior can be explained via heterogeneous payoffs among the non-permanent members. There is no reason to study such heterogeneity in a formal fashion, as the characterization of the equilibrium with two types is already computationally complex. However, it would be insightful to verify numerically that the optimal mechanism can actually mimic such equilibrium. The next section will explore that case.

3.3 Numerical Examples

Finally, I will provide two numerical examples. The first example will depict the equilibrium in propositions 5, 6, and 7. The second example is intended to show heterogeneous rotation. To do so, we will add one more type of country: *type-C* that will have a different support of payoff shocks.

Example 3. The simplest case to consider is three countries and one veto power holder: $N = 3$,

[31] (United Nations, 2015, ch. V, art. 23, p. 1.).

$N_A = 1$. We will numerically solve four scenarios. The first three have the same payoff supports: For the type-A country, payoffs are in $\{-2, 1\}$; for type-B countries, payoffs are in $\{-1, 3/2\}$. The difference among the first three cases will be the probability distributions: $p = 0.9$, $p = 0.6$ and $p = 0.5$. The fourth case has a payoff support on $\{-1, 3\}$ for all countries, and $p = 0.4$. Figure 4 shows the optimal mechanism x for each of these four cases, for $\delta \in (0, 1)$.

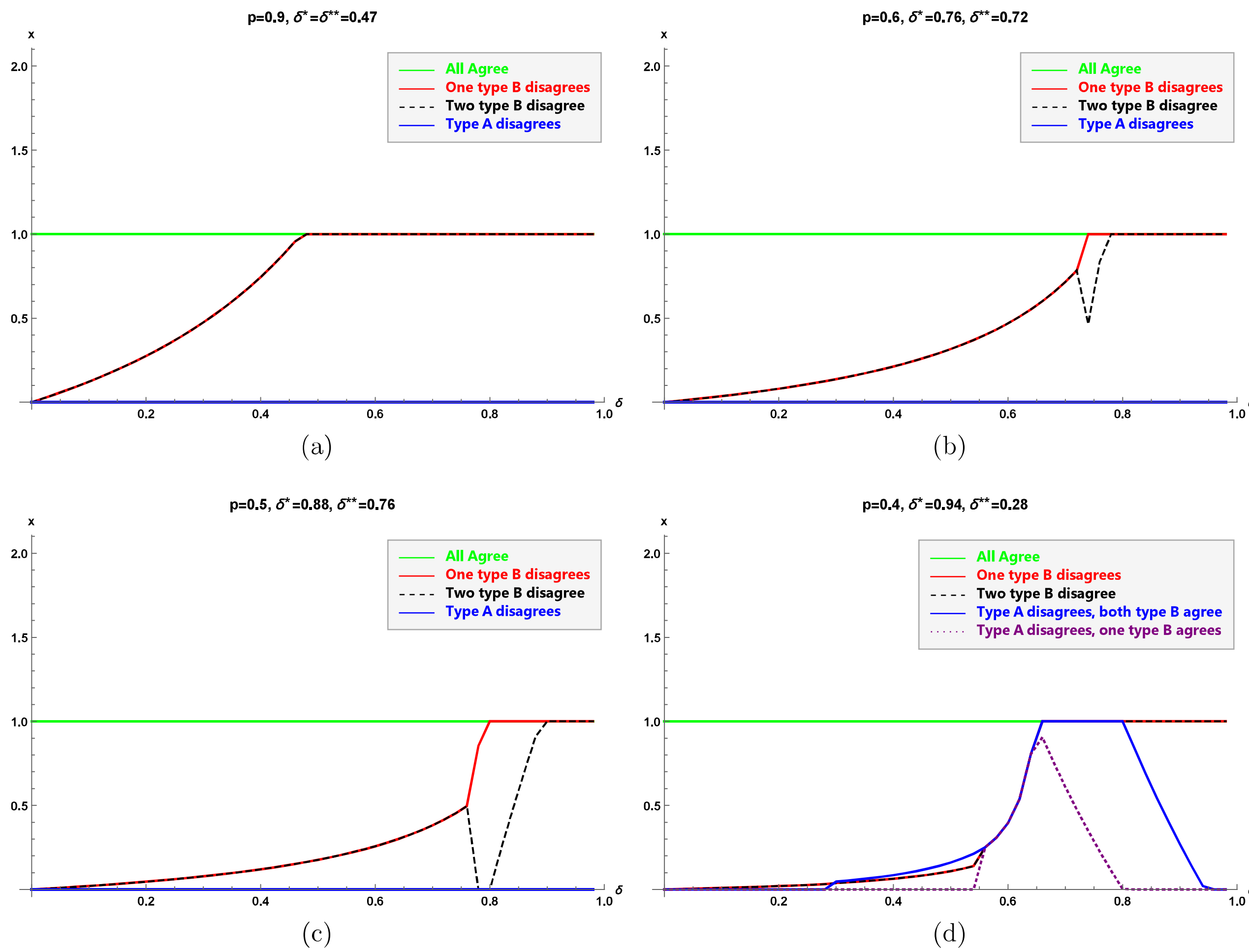
Figure 4.a shows the case when $\delta^* = \delta^{**}$. In the Pareto-efficient allocation, the unique type-A country is a dictator. Thus when type-A disagrees, the optimal allocation is $x = 0$. As the discount factor falls below $\delta^* = 0.47$, since $\delta^* = \delta^{**}$, the optimal allocation equals $x = x^{**}x^P + (1 - x^{**})x^U$, and therefore, $0 < x < 1$ in any state in which type-A agrees and at least one type-B disagrees.

In figures 4.b and 4.c, the conditions of propositions 6 and 7 hold, so type-A still has veto power (the blue line stays at zero), but the optimal allocation differs across states $(\overline{y}_A, \overline{y}_B, \underline{y}_B)$ and $(\overline{y}_A, \underline{y}_B, \underline{y}_B)$ for $\delta \in (\delta^{**}, \delta^*)$. Finally, when $\delta < \delta^{**}$, the optimal equilibrium is again characterized by $x = x^{**}x^P + (1 - x^{**})x^U$ (the red and black lines merge).

Finally, in figure 4.d, the unique type-A country does not hold veto power for $\delta < \delta^{**}$ since the conditions of propositions 6 and 7 do not hold (the blue and purple lines are strictly above zero). However, for a small enough discount factor δ, the type-A country regains veto power, and $x = x^{**}x^P + (1 - x^{**})x^U$ is once again optimal (the red and black lines merge at $x = x^{**}$, and the blue and purple lines drop to zero).

Example 4. For the heterogeneous rotation example, let us consider six countries: one 'mechanism designer' , two 'costly' non-veto power holder countries and three 'cheap' non-veto power holder countries . That is, $N = 6$, $N_A = 1$, $N_B = 2$, and $N_C = 3$. The payoffs are: $y_A \in \{-7, 1\}$, $y_B \in \{-5, 3\}$, and $y_C \in \{-3, 2\}$. Moreover, $\lambda_A = 1$, $\lambda_B = \lambda_C = 0$, and the probability of a positive shock is $p = 0.7$.

Figure 4: Optimal x as a function of δ



Figures (a), (b) and (c) have preference shocks in $\{-2, 1\}$ for type-A and $\{-1, 3/2\}$ for type-B countries. Figure (d) has preference shocks in $\{-1, 3\}$ for all countries.

Numerically, $\delta^{**} = 0.61$ (a variation of proposition 5 still holds), $\delta^{*} = 0.92$, and the optimal mechanism grants type-A countries veto power for all discount factors δ.

I will select one particular discount factor that illustrates heterogeneous rotation. Let us consider $\delta = 0.83$. Denote the number of type-j countries that agree with taking the action in any given state as k_j. Since veto power is optimal, we only need to consider variations in k_B and k_C. Table 1 describes the optimal action as a function of the state.

Table 1: Optimal action for $\delta = 0.83$

		$k_B =$		
		0	1	2
$k_C =$	0	0	0	0
	1	0	0	0.625
	2	0	0.566	0.625
	3	0.566	0.566	1

Similar to corollary 1, there are many different ways to implement this mechanism using a council. In this example, I will describe one of them. Consider a council with four members. The type-A country is a permanent member and holds veto power. Moreover, three out of four votes are required to implement an action $x > 0$. The following table has inputs of the form $(n_B, n_C) = (v_1, v_2, v_3)$. That expression should be read as follows: In a given state, n_B type-B countries are part of the council, n_C type-C countries are part of the council, and their votes are (v_1, v_2, v_3), which can be either 'yes' y or 'no' n. The order of the votes represents the number of each type of voter. For example, an entry $(0, 3) = (n, n, n)$ means that all type-C countries are part of the council and all vote 'no.' An entry $(2, 1) = (y, n, n)$ means that two type-B countries and one type-C are part of the council, and only one of the type-B countries votes 'yes,' while the rest of the council votes 'no.' With this notation, table 2 describes one possible council configuration that implements the optimal mechanism:

Note that only the states in which $x > 0$ have two 'yes' votes, which in addition to the veto power

Table 2: A council with 3 rotating members and equilibrium votes

		k_B = 0	k_B = 1	k_B = 2
k_C	0	$(2,1)=(n,n,n)$	$(1,2)=(n,n,n)$	$(0,3)=(n,n,n)$
	1	$(2,1)=(n,n,y)$	$(1,2)=(n,n,n)$	$(2,1)=(y,y,n)$
	2	$(2,1)=(n,n,y)$	$(1,2)=(n,y,y)$	$(2,1)=(y,y,n)$
	3	$(1,2)=(n,y,y)$	$(1,2)=(n,y,y)$	$(2,1)=(y,y,y)$

holder makes it possible to optimally implement the actions. Moreover, it is easy to check that when $x = 0.625$, two type-B members vote in favor; however, when $x = 0.566$, two type-C members vote in favor. With this distribution of power, we can calculate the probability that each type will be on the council, conditional on the event that A does not exert its veto power. Table 3 shows the probability of each event and the probability that each type will secure a seat in each state:

Table 3: Probabilities of securing a seat on the council

			k_B = 0	k_B = 1	k_B = 2
Prob of event	k_C	0	0.002	0.011	0.013
		1	0.017	0.079	0.093
		2	0.040	0.185	0.216
		3	0.031	0.144	0.168
Prob of B to get a seat	k_C	0	0.002	0.006	0.000
		1	0.017	0.040	0.093
		2	0.040	0.093	0.216
		3	0.015	0.072	0.168
Prob of C to get a seat	k_C	0	0.001	0.008	0.013
		1	0.006	0.053	0.031
		2	0.013	0.123	0.072
		3	0.021	0.096	0.056

By summing the probabilities for all states, we can see that each type-B country has an ex ante probability of securing a seat of 0.761, while each type-C country has an ex ante probability of securing a seat of 0.492. Despite that B is a smaller group, the total voting power of type-B countries is $0.761\times$

$2 = 1.522$ seats on average; however, for type-C countries, the total voting power is $0.492 \times 3 = 1.477$ seats on average. Finally, it is easy to check that, as expected, there is a total of three $(1.522 + 1.477)$ rotating seats in this configuration.

4 Conclusions

To finalize this study, I wish to (i) summarize the results, (ii) provide detailed interpretations that relate the findings more closely to what we observe in reality, and (iii) discuss further extensions. We have studied the optimal decision-making rule for an international organization under stationary perfect public equilibria. I showed how this optimal mechanism can be mapped onto a weighted voting system. As a consequence, the voting system is endogenous to the model. Moreover, I discussed how voting weights are fairly general in the sense that they can implement a large set of well-known voting systems, such as one-country-one vote, dictatorship, oligarchy, and veto power. Under the optimal mechanism, when the members are patient, the Pareto-efficient allocation can be implemented, and therefore, the voting weights are constant over time. When the members are not very patient, the decision-making power is state dependent and therefore ex ante stochastic.

Moreover, the study of 'impatient players' has several interpretations. It can be regarded as countries literally valuing the present more than the future. On the other hand, for the thresholds in most of the results, decreasing the discount factor has similar implications as decreasing the ratio of the good payoff to the bad payoff $(-\overline{y}/\underline{y})$. This means that the type of equilibrium being played does not only depend on the discount factor, which is probably uniform across organizations for each country. Indeed, the type of equilibrium will vary from organization to organization because the payoff structures are different. Thus, comparative statics on δ can be mapped to comparative statics on the payoffs.

We also studied the case in which only a subset of the members have a positive Pareto weight and showed the conditions under which they will have veto power under the optimal mechanism. Moreover, for a moderate discount factor, the optimal equilibrium can be implemented by a council-like voting system that resembles how the members of the United Nations rotate on the Security Council. Specifically, I related the randomness in voting power to how the United Nations assigns power via the Security Council. In reality, this process follows a more complex protocol, so it is worth discussing how the actual Security Council's elections are held and how this mechanism can be mapped onto my model. Every year, at the United Nations, a few members are nominated (by other members or by themselves) as candidates for a seat on the Security Council. Then, all members vote on each of the candidates. Finally, a candidate country is elected if it has the support of at least two-thirds of all members.

Now, let me explain how this protocol relates to my model. There are three points to note: First, it is possible that the preferences are revealed 'informally' prior to the election of Security Council members. Moreover, following corollary 1, once the profile of preferences has been revealed, all members already know what countries will be part of the Security Council and what action x will be implemented. Second, as the preferences are known, the voting protocol for elections of Security Council members is not the actual voting described in my model. Indeed, what my model would say is that given the current state $\mathbf{y}$, the mechanism requires country i to be on the council. Therefore, the rest of the members will nominate i, give it a nominal/sympathy vote to elect it, and then implement the desired action. Third, once elections have taken place, only the Security Council members will participate in 'formal' voting on whether to take an action x, but from corollary 1, we already know that the members of the Security Council were elected in such a way that their vote will implement the desired action x. Moreover, recall that an action partially implemented, $0 < x < 1$, has the interpretation that members compromise on a Pareto-dominated action to satisfy all incentive constraints. This 'second best' action x is the resolution voted on by the Security Council; however, it is perfectly acceptable to nominally 'discuss' $x = 1$ and pretend some verbal bargaining.

I also presented a few extensions to the model, two of which are particularly challenging and relevant for future research, namely, imperfect monitoring and unrestricted payoffs. Both of them have the property that the optimal voting weights will not only be random but also change over time. Moreover, both extensions have a testable implication for future research: The rotation of Security Council seats depends positively on countries' performance on past resolutions. In addition to the testable implication that associates past performance with Security Council seats, country characteristics also affect voting weights. For example, military power might be an important determinant of decision-making power at the United Nations but perhaps not very important at the World Bank.

Finally, although I focused on the rotation of Security Council members, the model could be applied to other institutions. A perhaps cynical example would be that oligarchs in some countries collide, and the political parties rotate power depending on stochastic preference shocks. The model could also fit public good provision in environments where the good is repeatedly delivered, such as how often the road to a village should be repaired. In general, the mechanism could be applied to settings in which monetary transfers are not used and the players have different opinions on the issue.

Appendix

Please contact the author if you are interested in the proofs.

References

Abreu, Dilip, David Pearce, and Ennio Stacchetti (1990) "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, Vol. 58, pp. 1041–1063.

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin (2012) "Dynamics and stability of constitutions, coalitions, and clubs," *American Economic Review*, Vol. 102, pp. 1446–76.

Aghion, Philippe, Alberto Alesina, and Francesco Trebbi (2004) "Endogenous political institutions," *The Quarterly Journal of Economics*, Vol. 119, pp. 565–611.

Ansolabehere, Stephen, James M Snyder, Aaron B Strauss, and Michael M Ting (2005) "Voting weights and formateur advantages in the formation of coalition governments," *American Journal of Political Science*, Vol. 49, pp. 550–563.

Athey, Susan and Kyle Bagwell (2001) "Optimal Collusion with Private Information," *The RAND Journal of Economics*, Vol. 32, pp. 428–465.

Azrieli, Yaron and Semin Kim (2014) "Pareto efficiency and weighted majority rules," *International Economic Review*, Vol. 55, pp. 1067–1088.

Barbera, Salvador and Matthew O. Jackson (2004) "Choosing How to Choose: Self-Stable Majority Rules and Constitutions," *The Quarterly Journal of Economics*, Vol. 119, pp. 1011–1048.

Barbera, Salvador and Matthew O Jackson (2006) "On the weights of nations: Assigning voting weights in a heterogeneous union," *Journal of Political Economy*, Vol. 114, pp. 317–339.

Bourantonis, D. (2005) *The History and Politics of UN Security Council Reform*, Routledge Advances in International Rela Series: Routledge.

Dreher, A. and J.R. Vreeland (2014) *The Political Economy of the United Nations Security Council: Money and Influence*: Cambridge University Press.

Dreher, Axel, Matthew Gould, Matthew D. Rablen, and James R. Vreeland (2014) "The determinants of election to the United Nations Security Council," *Public Choice*, Vol. 158, pp. 51–83.

Eguia, Jon X (2011) "Voting blocs, party discipline and party formation," *Games and Economic Behavior*, Vol. 73, pp. 111–135.

Harstad, Bård (2010) "Strategic delegation and voting rules," *Journal of Public Economics*, Vol. 94, pp. 102–113.

Jackson, Matthew O and Leeat Yariv (2015) "Collective dynamic choice: the necessity of time inconsistency," *American Economic Journal: Microeconomics*, Vol. 7, pp. 150–78.

Koray, Semih (2000) "Self-Selective Social Choice Functions Verify Arrow and Gibbard-Satterthwaite Theorems," *Econometrica*, Vol. 68, pp. 981–996.

Kuziemko, Ilyana and Eric Werker (2006) "How Much Is a Seat on the Security Council Worth? Foreign Aid and Bribery at the United Nations," *Journal of Political Economy*, Vol. 114, pp. 905–930.

Lagunoff, Roger (2009) "Dynamic stability and reform of political institutions," *Games and Economic Behavior*, Vol. 67, pp. 569–583.

Maggi, Giovanni and Massimo Morelli (2006) "Self-Enforcing Voting in International Organizations," *American Economic Review*, Vol. 96, pp. 1137–1158.

Schmitz, Patrick W and Thomas Tröger (2012) "The (sub-) optimality of the majority rule," *Games and Economic Behavior*, Vol. 74, pp. 651–665.

Ticchi, Davide and Andrea Vindigni (2010) "Endogenous constitutions," *The Economic Journal*, Vol. 120, pp. 1–39.

United Nations (2015) *Charter of the United Nations and Statute of the International Court of Justice*: United Nations Publications.

Voeten, Erik (2001) "Outside Options and the Logic of Security Council Action," *The American Political Science Review*, Vol. 95, pp. 845–858.