IDEC DP^2 Series Vol. 10 No. 8

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

Johann Caro-Burnett
NERPS
Hiroshima University

Department of Development Policy
Division of Development Science
Graduate School for International
Development and Cooperation (IDEC)
Hiroshima University
1-5-1 Kagamiyama, Higashi-hiroshima
7398529 Japan

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

Johann Caro-Burnett†

December 10, 2020

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. Thus, if the organization wants to take the action, incentives must be provided. I show that the optimal stationary equilibrium is equivalent to a mechanism characterized by voting weights. I study how this optimal mechanism depends on the discount factor. In particular, I show that within a class of parameter cases, the optimal mechanism mimics the way voting power is distributed among UN members via Security Council seats.

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]

I study the design of the decision-making process for an international organization.[2] I propose a repeated game with three elements that international organizations typically 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.[3] 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.

[1] See the literature review for examples of studies on endogenous voting systems.

[2] Although, the model could fit a broader class of organizations with similar characteristics, I 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. Second, as I will show below, under certain parameter conditions, the results of the model greatly resemble 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.

[3] 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 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, 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.

As a preview of the results, I will first find the optimal equilibrium. This optimal equilibrium can be implemented via different equivalent mechanisms, one of them is voting weights. Then, I find the differences in the optimal equilibria under the entire spectrum of discount factors. Although a folk theorem is valuable, I pay especial attention to parameter conditions that do not allow for the efficient outcome to be implemented. Moreover, the heterogeneity of the members' preferences allows for differences in the voting weights.[4] 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 implement the 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 with costly demands.[5] The second group has zero Pareto weight and a relatively lower cost of complying.[6]

Then, I show that for a large range of discount factors, the optimal mechanism assigns voting power only to a subset of countries (composed of both, the mechanism designers and the zero Pareto weight countries). This distribution of power depends on the profile of preference shocks and resembles 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.[7] This mechanism is remarkably similar to how the United

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

[5] See lemma 3.

[6] Zero Pareto weight is just a benchmark. The results in section 3 would still hold for 'small' Pareto weights in the second group.

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

Nations rotates decision-making power among the non-permanent members of the Security Council.[8]

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 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.

There are a few studies on endogenous decision-making rules. Broadly speaking, some of them (including the present paper) 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). One key difference from their paper is that here the members of the organization are heterogeneous. With this difference, 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 (implemented via rotation) in their decision-making power, such as the United Nations Security Council. Another key difference is that Maggi and Morelli (2006) restrict attention to symmetric equilibria.[9] In the present paper, it would not make sense to assume symmetric payoffs, since I allow for heterogeneous countries. Instead, I restrict attention to stationary equilibrium payoffs. Although in principle this restriction may seem strong, the reader should note that homogeneous players plus symmetric equilibria implies stationary equilibrium payoffs when looking at optimal equilibria.[10] A final key 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

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

[9] The same simplifying assumption is used in Athey and Bagwell (2001).

[10] More specifically, any optimal equilibrium is payoff equivalent to a stationary payoffs equilibrium, as shown in lemma 1 of Maggi and Morelli (2006).

countries. Moreover, this compromise resembles 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.

Aghion et al. (2004), Harstad (2005), and Ticchi and Vindigni (2010) are other examples of studies on efficiency and endogenous voting. Harstad (2005) has a result that captures one of the comparisons between Maggi and Morelli (2006) and the present paper. Harstad (2005) proposes a static model in which the decision-making rule is chosen prior to investment and the realization of preference shocks. The needed supermajority rule ('winning coalition'), broadly speaking, becomes smaller with the heterogeneity of the players. This means that less 'yes' votes are needed, or it is easier to implement an efficient outcome. In the present paper, I show in proposition 4 that even looking at discrete choice mechanisms, there are discount factors where heterogeneity helps implementing an outcome strictly better than the best static equilibrium.

Barbera and Jackson (2004), Koray (2000), and Lagunoff (2009) are examples of studies that investigate endogenous mechanisms by focusing 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 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.

Jackson and Yariv (2015), Schmitz and Tröger (2012), and Harstad (2010) study endogenous

decision-making rules and focus on large preference spaces. Jackson and Yariv (2015) show that when a set of heterogeneous consumers must agree on a common consumption stream, the only way to aggregate preferences that satisfies efficiency and does not have a dictator must be time-inconsistent. Schmitz and Tröger (2012) study optimality in a one-period game, where actions are enforceable, and the payoffs can take more than just two values. Harstad (2010) studies the choice of representatives. He shows how a larger majority threshold (supermajority) can influence the outcome towards a more conservative outcome.

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. Azrieli and Kim (2014) and Schmitz and Tröger (2012) study a one-period game, with preferences belonging to a large set of possible values. They focus on incentive compatibility because the payoffs (from participating in the organization's decision) can take more than just two values; however, they do not consider the problem of enforceability. Those two studies, show that the optimal decision-making rule is a weighted voting system that can be state dependent. This resembles my result for moderate discount factors. However, in their case, the reason relies on incentive compatibility, which is not a surprising result (in the mechanism design literature, when incentive compatibility is binding, the outcome typically depends on the preference shock). On the other hand, in my work, incentive compatibility is not binding; instead, stochastic voting power is a consequence of enforceability.

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[11] and find the mechanism that maximizes the weighted sum of payoffs. As it 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).[12] 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

[11] 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.

[12] 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).

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 x. 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. At the beginning of the stage 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 independent across countries and periods.[13] 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 $\overline{y}_i > 0$, 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 (sequential) equilibrium payoffs. 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. The other equilibrium payoff can be attained by never taking 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

[13] 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.

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 define the Pareto-efficient rule for this model. 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:[14]

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

where $x=1$ means to implement the action, $y_i(\mathbf{y})$ 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. Moreover, the solution is to implement whenever:

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

Note that 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.

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

[14] The reader may have noted that redefining a new support for payoffs equal to $(\lambda_i\underline{y}_i,\lambda_i\overline{y}_i)$ could make the usage of Pareto weights redundant. However, the previous transformation fails when $\lambda_i=0$. This special case has the interpretation that although a country has non-degenerate preferences, the organization does take that country into account at the maximization problem. Moreover, since the country's participation is necessary, the country should still have some decision-making power in equilibrium. This is precisely the case studied in section 3, where only the P5 UN members have a positive Pareto weight.

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.[15] 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.[16] 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, on the equilibrium path, the status quo payoff is strictly smaller than the payoff from having zero decision-making power.[17] 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

[15] 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.

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

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

mutual best responses to one another.

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.[18] 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:

$$\max_{x}(1-\delta)\sum_{t=0}^{\infty}\delta^{t}E\left[x_{t}(\mathbf{y}_{t},\mathbf{h}_{t})\sum_{i\in N}\lambda_{i}y_{i,t}\right] \tag{2}$$

[18] This would not hold under imperfect monitoring.

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.[19] Equation (5) is the *truth-telling* condition: Members 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

[19]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.

organization and award zero payoffs to all countries.

Lemma 1. *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 1, 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.[20]

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:

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

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

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 2. *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 2 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:

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.[21]

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 and study what the equilibrium will look like on the interval $(0, \delta^*)$.

Let us propose an alternative and less abstract way of implementing an allocation in this environment. 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. Moreover, 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.

[21] See proposition 4 in section 2.5.

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:

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

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

Proof. See Appendix C. □

Moreover, note that if $\delta \geq \delta^*$, the above mechanism can be implemented using voting weights that are independent of the current state of the world. The previous set of 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 studies 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[22] 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.

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

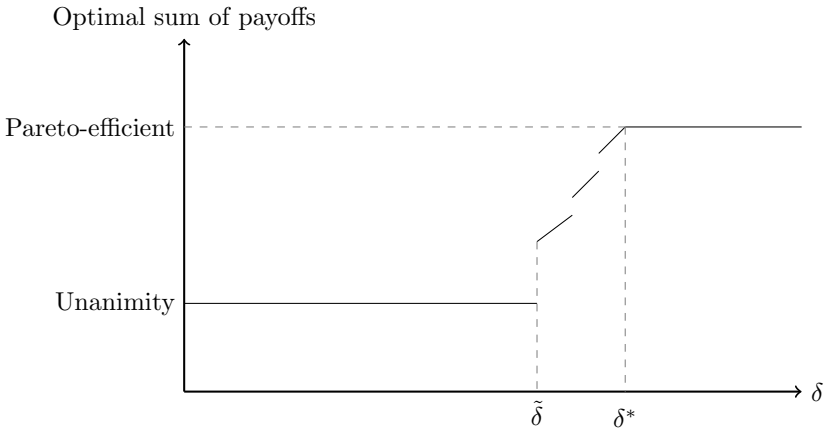
Proposition 3 will still be valid, but the voting weights and target will depend on one additional term that captures how binding the new restrictions are due to the introduction of the outside option.

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 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 1.

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

[22] With the addition of an outside option, part (i) of lemma 1 is no longer true.

Figure 1: Optimal stationary mechanism payoff as a function of δ



(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. There are different ways to model imperfect monitoring. A simple configuration would be that the final outcome x is not perfectly observed. Instead, a signal $s \in [0,1]$ is publicly observed, and it induces a probability distribution $Q(\cdot)$ over the vector of country participations $\mathbf{d}$. More precisely, if two signals satisfy $s < s'$ and two vectors of country participations satisfy $\mathbf{d} < \mathbf{d}'$, then $Q(\mathbf{d}|s) > Q(\mathbf{d}'|s)$ and $Q(\mathbf{d}|s') < Q(\mathbf{d}'|s')$.

Then, assuming that truth-telling is satisfied, a low signal would hint that perhaps countries who declared $y_i < 0$ may have defected. In principle, the decision made by the organization x should depend on the history of observed signals. From Abreu et al. (1990), there is a recursive solution to the equilibrium. In such recursive solution, the future payoff of a member will depend on some combination of the most recent signal and the reported preference shock.

A more interesting configuration could allow for noisy public signals of each country's compliance. Let us consider two such signals. One signal can be labeled as the *good signal*: $\overline{s}_{i,t}$ and be highly correlated with the country's participation, and the other signal can be labeled as the *bad signal*: $\underline{s}_{i,t}$ and be highly correlated with the country's defection. For instance, the probability that the signal is $\overline{s}_{i,t}$ when country i participated takes a value $q > 1/2$; and the probability that the signal is $\overline{s}_{i,t}$ when the country has not participated takes a value of $1 - q$. Again, this model can be solved using a recursive approach where a country's future payoff will be increasing on its own compliance signal. In terms of a voting mechanism, this means that the ex-post voting weight of a country will also be weakly increasing in the country's own compliance 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, it is reasonable to 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 stronger conjecture would be that after two histories of the same length $\mathbf{h}_t$ and $\hat{\mathbf{h}}_t$ such that for all $\tau \leq t$ satisfying that whenever $y_{i,t} < 0$, $\hat{y}_{i,t} < 0$ and $y_{j,t} = \hat{y}_{j,t}$ for $j \neq i$, then:

$$\sum_{t'=t+1}^{\infty} \delta^{t'-t} \left(E\left[x_{t'}(\mathbf{y}_{t'}, \mathbf{h}_{t'}) y_{i,t'} | \mathbf{h}_t\right] - E\left[x_{t'}(\mathbf{y}_{t'}, \mathbf{h}_{t'}) y_{i,t'} | \hat{\mathbf{h}}_t\right] \right) \leq 0$$

Or roughly speaking, the longer a country that has received negative shocks and was requested to participate by taking an action, the higher will have to be its promised future payoff. In terms of the voting mechanism, this would mean that such countries will have accumulated a 'stock of promised decision-making power' over time; thus, in the future, this country will be less likely to keep being requested to participate on states with negative shocks. Moreover, as we will see in the next section, the random voting weights can be implemented by fully delegating decision making power to a rotating council, at least in the case of stationary payoffs. If this mechanism of power delegation also holds in the unrestricted payoffs case, the previously mentioned stock of decision making power would imply a positive relation between the number of years outside the council and the probability to get membership at the council. This property is in line with empirical findings in Dreher et al. (2014).

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 to propose and vote on resolutions that are compulsory to all members of the United Nations.[23] 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 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.

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

3.1 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.[24] 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 other countries to join the United Nations. Therefore, it is reasonable to assume that they had all the Pareto 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

[24] 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."*

the optimal mechanism.

Lemma 3. *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 3 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:[25]

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

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 3 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 $\hat{\delta}$.

[25] 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.

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 $x > 0$.[26] Indeed, what determines $\underline{d}_{k'_B+1} = \hat{\delta}$ is the restriction $x \leq 1$. Therefore, as $\delta \to \hat{\delta}$,

[26] 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)

$x \to 1$. 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 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.[27]

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..."*[28] 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 is 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 1. The simplest case to consider is three countries and one veto power holder: $N = 3$,

[27] This also relates to Downs et al. (1996), where they argue that there is a self-selection bias in the resolutions that are voted. International organizations only discuss publicly the issues that they know will reach an agreement, and omit issues that will never pass the voting process.

[28] (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 2 shows the optimal mechanism x for each of these four cases, for $\delta \in (0, 1)$.

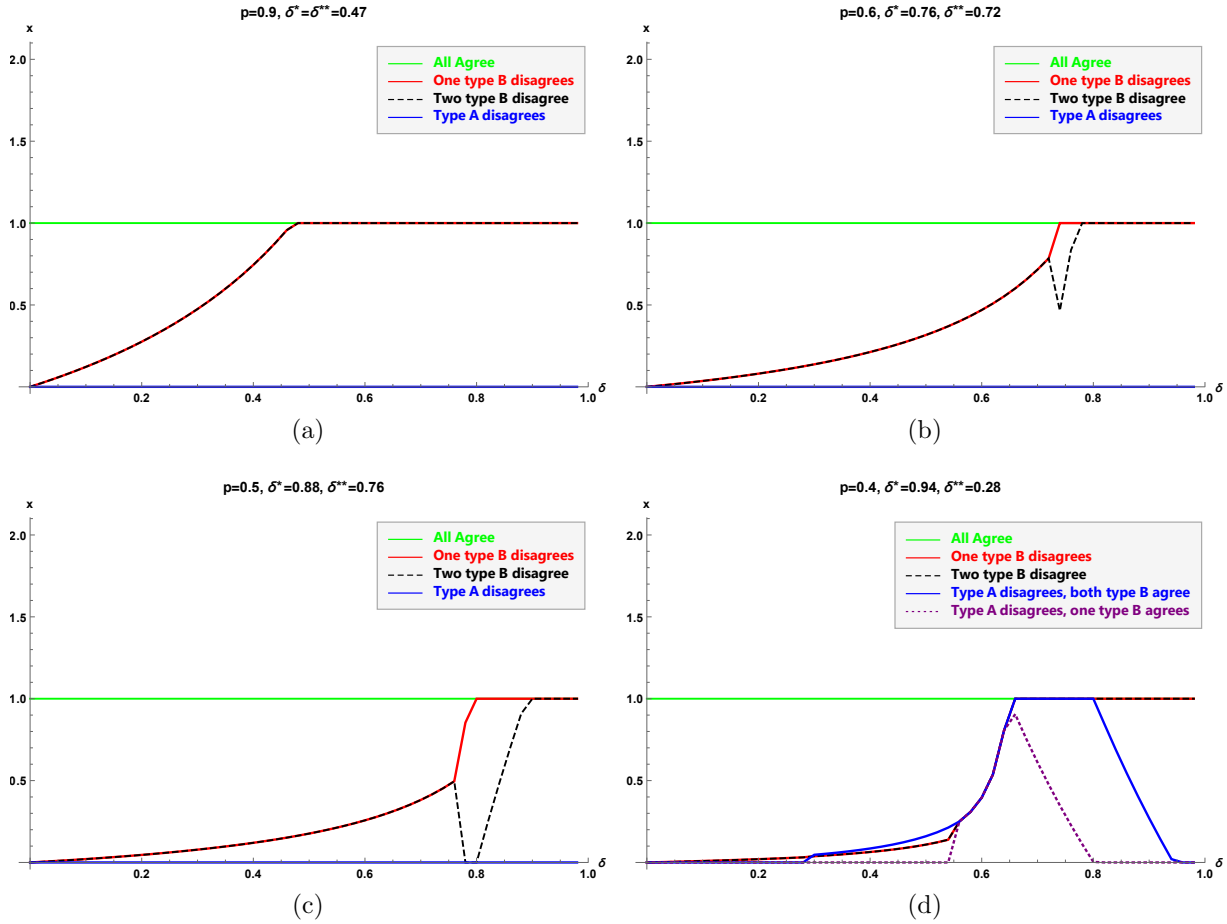
Figure 4.a shows the case when $\delta^* = \delta^{**}$. In the Pareto-efficient allocation, the unique type-A country is a dictator. Thus when type-A disagrees, the optimal allocation is $x = 0$. As the discount factor falls below $\delta^* = 0.47$, since $\delta^* = \delta^{**}$, the optimal allocation 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 2. 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 $p = 0.7$. 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.

Figure 2: 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.

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 $\delta = 0.83$

$k_C =$ \ $k_B =$	0	1	2
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 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

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	$(1,2) = (n,y,y)$	$(1,2) = (n,y,y)$	$(2,1) = (y,y,y)$

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

A Proofs of Lemma 1

Proof of Lemma 1. (i) If at the optimum, a member has veto power, then the action is only implemented in states in which that country has a positive payoff (because that member is reporting truthfully at the optimum). Therefore, the expected payoff of that member is a weighted sum of positive numbers and zeros. Thus, it cannot be negative. Moreover, since the payoffs are stationary, all future expected payoffs are the same and positive.

If a member does not have veto power, then there is at least one state in which the action is implemented and the member has a negative payoff. Therefore, from equation (4), the future expected payoff has to be positive. Moreover, since the payoffs are stationary, all future expected payoffs are the same and positive. Similar to the previous case, this has to hold for every expected payoff after every history.

(ii) x increasing means that $x(\mathbf{y_{-i}}, \overline{y}_i) \geq x(\mathbf{y_{-i}}, \underline{y}_i)$ for all $\mathbf{y_{-i}}$ and i. Therefore, taking expectations over $\mathbf{y_{-i}}$, the inequality still holds. This implies that 5 holds.

(iii) Since the state space $\mathbf{Y}$ is finite, problem (2) has a solution. Moreover, restrictions (3), (4), and (5) imply that any feasible mechanism is an equilibrium of the repeated game that has grim trigger strategies. It only remains to show that there is no equilibrium of the repeated game that has a payoff larger than the solution to (2). Let us assume that there is such equilibrium: $\hat{x}$. This candidate for the optimal equilibrium cannot have grim trigger strategies; otherwise, it could be characterized as a feasible mechanism. Then, let us define a mechanism $\hat{\hat{x}}$ that is identical to $\hat{x}$ on the equilibrium path and has grim trigger strategies off the equilibrium path. Since they differ only off the equilibrium path, the payoffs are unchanged. Moreover, $\hat{\hat{x}}$ will satisfy (3), (4), and (5), and therefore, it cannot have a payoff greater than the solution to (2).

□

B Proof of Lemma 2

Proof. As it is standard in the literature, the equilibrium to the infinitely repeated game can be restated as the mutual best-response of the two-period game with payoffs: $(1-\delta)x(\mathbf{y})y_i+\delta v_i(\mathbf{y})$.[29] The difficulty is that for any state $\mathbf{y}$, the profile of payoffs $(v_i(\mathbf{y}))_{i=1}^N$ has to belong to the set of PPE payoffs V. Moreover, the stationary payoffs restriction implies that for any $\mathbf{y} \neq \hat{\mathbf{y}}$, $(v_i(\mathbf{y}))_{i=1}^N = (v_i(\hat{\mathbf{y}}))_{i=1}^N$. Thus, the maximization problem becomes:

$$\max_{\tilde{x},v} \sum_{\mathbf{y}\in\mathbf{Y}} \mathbb{P}(\mathbf{y}) \sum_i \lambda_i \left[(1-\delta)\tilde{x}(\mathbf{y})y_i + \delta v_i(\mathbf{y})\right] \tag{12}$$

subject to

$$(1-\delta)\tilde{x}(\mathbf{y})y_i + \delta v_i(\mathbf{y}) \geq 0, \forall i, \mathbf{y}$$

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

$$v_i(\mathbf{y}) = v_i(\hat{\mathbf{y}}), \forall i, \mathbf{y} \neq \hat{\mathbf{y}} \tag{13}$$

$$(v_i(\mathbf{y}))_{i=1}^N \in V \tag{14}$$

where restriction (13) makes it explicit that the payoffs are stationary and restriction (14) states that the 'promised payoffs' are credible. In principle, we do not know what the set V is, but we can further simplify problem (12) by noticing that its solution will satisfy $\sum_{\mathbf{y}} \mathbb{P}(\mathbf{y})\tilde{x}(\mathbf{y})y_i = v_i$. Thus problem (12) becomes problem (6). The claim that $x_t(\cdot, \mathbf{h_t}) = \tilde{x}(\cdot)$ is a solution to the repeated game follows immediately, given that problem (12) is equivalent to problem (2).

[29] Abreu et al. (1990).

□

C Proofs of Propositions 1, 2, 3, and 4

Proof of Proposition 1. Let us first focus on the action-taking incentive constraint; then, we will show that the other incentive constraints are satisfied. Note that we only need to provide incentives to those countries that disagreed on taking the action whenever $x = 1$:

$$(1-\delta)y_i + \delta v_i^P \geq 0, \forall i, \mathbf{y} \text{ such that } x(\mathbf{y}) = 1$$

where $v_i^P = E[x^P y_i]$ is the Pareto-efficient expected payoff, which is positive because we assumed that $p\overline{y}_i + (1-p)\underline{y}_i > 0$. For each country that is not a dictator under the Pareto-efficient allocation, we define $\underline{\delta}_i = \frac{-\underline{y}_i}{-\underline{y}_i + v_i^P}$. Finally, the desired discount factor will be the largest of each country's minimal requirements: $\delta^* = \max\{\underline{\delta}_1, \underline{\delta}_2, \ldots, \underline{\delta}_N\}$. Whenever the discount factor is at least δ^*, the participation constraints are satisfied for all members.

The Pareto-efficient allocation is increasing in the preference shock: If $\mathbf{y} \geq \mathbf{y}'$, then $\mathbf{y} \geq \mathbf{y}'$; therefore, from lemma 1, it satisfies the truth-telling condition. Finally, $v_i^P \in V_i$, as the Pareto-efficient mechanism x^P can be sustained as an equilibrium with those same payoffs.

□

Proof of Proposition 2. To show that the optimal equilibrium delivers a payoff strictly higher than unanimity, it suffices to show that there is a feasible action $\hat{x}$ that is strictly better than unanimity. To do so, first consider a given $0 < \delta < \delta^*$. Let v_i^U be the unanimity expected payoff of country i. Then, define $\hat{x}(\mathbf{y}; \delta) = 1$ for the unanimity state, $\hat{x}(\mathbf{y}; \delta) = 0$ if the Pareto optimum allocation at that state is zero, and $\hat{x}(\mathbf{y}; \delta) = \frac{\delta \min_i\{v_i^U\}}{(1-\delta)\max_i\{-y_i\}}$ otherwise. Note that by construction, $E[x^U y_i] < E[\hat{x} y_i] < E[x^P y_i]$, where x^P is the Pareto-efficient action. Moreover, by construction $\hat{x}$ satisfies the participation constraint, and $0 < E[x^U y_i] < E[\hat{x} y_i]$ implies that it satisfies the voluntary membership constraint. It only remains to show

that it satisfies the truth-telling condition, but $x^P > 0$ if and only if $\hat{x} > 0$ implies that $\hat{x}$ is increasing in $\mathbf{y}$, and therefore, truth-telling is satisfied.

□

Proof of Proposition 3. Consider a relaxed version of problem (6), without truth-telling and ignoring that the actions $\tilde{x}$ are constrained to belong to a set of equilibrium actions:

$$\max_{x,\tilde{x}} \sum_{\mathbf{y}\in\mathbf{Y}} \mathbb{P}(\mathbf{y}) \sum_{i\in N} \lambda_i \left((1-\delta)x(\mathbf{y})y_i + \delta \sum_{\hat{\mathbf{y}}\in\mathbf{Y}} \mathbb{P}(\hat{\mathbf{y}})\tilde{x}(\hat{\mathbf{y}}|\mathbf{y})\hat{y}_i \right)$$

subject to

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

Let us multiply the constraint by its respective Lagrange multiplier, normalized by the (strictly positive) probabilities $\mathbb{P}$, and add the sum over all states to the objective function:

$$\max_{x,\tilde{x}} \sum_{\mathbf{y}\in\mathbf{Y}} \mathbb{P}(\mathbf{y}) \sum_{i\in N} (\lambda_i + \gamma_i(\mathbf{y})) \left((1-\delta)x(\mathbf{y})y_i + \delta \sum_{\hat{\mathbf{y}}\in\mathbf{Y}} \mathbb{P}(\hat{\mathbf{y}})\tilde{x}(\hat{\mathbf{y}}|\mathbf{y})\hat{y}_i \right)$$

Since x and $\tilde{x}$ are additive separable, we can relabel $\mathbf{y}$ as $\hat{\mathbf{y}}$ and vice versa on all the terms multiplying $\tilde{x}$ to obtain:

$$\max_{x,\tilde{x}} \left\{ (1-\delta) \sum_{\mathbf{y}\in\mathbf{Y}} \mathbb{P}(\mathbf{y}) \sum_{i\in N} (\lambda_i + \gamma_i(\mathbf{y}))\, x(\mathbf{y})y_i + \delta \sum_{\mathbf{y}\in\mathbf{Y}} \sum_{\hat{\mathbf{y}}\in\mathbf{Y}} \mathbb{P}(\mathbf{y})\mathbb{P}(\hat{\mathbf{y}}) \sum_{i\in N} (\lambda_i + \gamma_i(\hat{\mathbf{y}}))\, \tilde{x}(\mathbf{y}|\hat{\mathbf{y}})y_i \right\}$$

Finally, we use the fact that under stationary payoffs $x(\mathbf{y}) = \tilde{x}(\mathbf{y}|\hat{\mathbf{y}})$ regardless of the previous period $\hat{\mathbf{y}}$. With this, we obtain the following expression:

$$\max_{x} \sum_{\mathbf{y}\in\mathbf{Y}} \mathbb{P}(\mathbf{y})x(\mathbf{y}) \sum_{i\in N} \left(\lambda_i + (1-\delta)\gamma_i(\mathbf{y}) + \delta \sum_{\hat{\mathbf{y}}\in\mathbf{Y}} \mathbb{P}(\hat{\mathbf{y}})\gamma_i(\hat{\mathbf{y}}) \right) y_i \tag{15}$$

Note that a preference shock can be rewritten as $y_i = a_i(\mathbf{y})(\overline{y}_i - \underline{y}_i) + \underline{y}_i$, where a_i is an indicator function that equals one whenever the state $\mathbf{y}$ gives country i a positive shock. After some simplifications, we obtain the desired expressions for the weights m_i and the target M.[30]

Finally, the mechanism implemented by voting weights is increasing, and $x(\mathbf{y}) = \tilde{x}(\mathbf{y}|\hat{\mathbf{y}})$ implies that $\tilde{x}$ belongs to the set of equilibrium actions.

□

Proof of Proposition 4. (i) When $\delta \geq \delta^*$, the optimal mechanism is the Pareto-efficient allocation, which is characterized by x being either 0 or 1.

(ii) By contradiction, assume that for every arbitrarily small δ, there is a feasible decision $x \in \{0, 1\}$ that delivers a payoff strictly higher than unanimity. Then, there is at least one country i and a state $\mathbf{y}$ such that $x(\mathbf{y}) = 1$, $x(\mathbf{y})y_i > 0$ and $\mathbf{y}$ is not the unanimity state. Therefore, there is one country j such that $x(\mathbf{y})y_j = \underline{y}_j < 0$. For j, the participation constraint in that state requires $(1-\delta)x(\mathbf{y})y_j + \delta E[xy_j] \geq 0$. An upper bound for the expected payoff is that of dictatorship $E[xy_j] \leq p\overline{y}_j$. Finally, since $x(\mathbf{y}) = 1$, for δ small enough $(1-\delta)\underline{y}_j + \delta p\overline{y}_j < 0$; therefore, the allocation is not feasible.

(iii) If $\tilde{\delta} < \delta^*$, then for any discount factor in the middle $\tilde{\delta} < \delta < \delta^*$, the optimal allocation has a payoff strictly better than unanimity but strictly worse than the Pareto allocation. Assume that the individual thresholds are all equal, $\underline{\delta}_i = \delta^*$ for all i. Then, $\delta < \delta^*$ implies that all countries have their participation constraints violated. Therefore, any feasible x will implement the action in fewer states than would be Pareto-efficient. This means that not only $\sum \lambda_i E[xy_i] < \sum \lambda_i E[x^P y_i]$ but also $E[xy_i] < E[x^P y_i]$ for every country. Since $x \in \{0, 1\}$, $(1-\delta)x^P(\mathbf{y})y_j + \delta E[x^P y_j] < 0$ implies that $(1-\delta)x(\mathbf{y})y_j + \delta E[xy_j] < 0$ as long as $x > 0$. Therefore, the only state that would be feasible to implement is unanimity. However, we considered a discount factor $\tilde{\delta} < \delta < \delta^*$, thus ruling out unanimity. Thus, a necessary condition for this to hold is $\underline{\delta}_i \neq \underline{\delta}_j$ for at least two countries i and j.

□

[30] Note that the γ_i values are pinned down by $\gamma_i(\mathbf{y}) \sum_{\hat{\mathbf{y}} \in \mathbf{Y}} \mathbb{P}(\hat{\mathbf{y}}) \tilde{x}(\hat{\mathbf{y}}|\mathbf{y}) \hat{y}_i = 0$.

D Proof of Proposition 5

Proof. Using Kuhn-Tucker's sufficiency conditions, we need to find multipliers γ_i such that the proposed equilibrium is a solution to (15):

$$\max_{x\in[0,1]} \sum_{\mathbf{y}\in\mathbf{Y}} \mathbb{P}(\mathbf{y})x(\mathbf{y}) \sum_{i\in N} \left(\lambda_i + (1-\delta)\gamma_i(\mathbf{y}) + \delta \sum_{\hat{\mathbf{y}}\in\mathbf{Y}} \mathbb{P}(\hat{\mathbf{y}})\gamma_i(\hat{\mathbf{y}}) \right) y_i$$

The candidate to solution is characterized by:

$$x(\mathbf{y}) = \begin{cases} 1, & \sum_{i\in N} \left(\lambda_i + (1-\delta)\gamma_i(\mathbf{y}) + \delta \sum_{\hat{\mathbf{y}}\in\mathbf{Y}} \mathbb{P}(\hat{\mathbf{y}})\gamma_i(\hat{\mathbf{y}}) \right) y_i > 0 \\ x^{**}, & \sum_{i\in N} \left(\lambda_i + (1-\delta)\gamma_i(\mathbf{y}) + \delta \sum_{\hat{\mathbf{y}}\in\mathbf{Y}} \mathbb{P}(\hat{\mathbf{y}})\gamma_i(\hat{\mathbf{y}}) \right) y_i = 0 \\ 0, & \sum_{i\in N} \left(\lambda_i + (1-\delta)\gamma_i(\mathbf{y}) + \delta \sum_{\hat{\mathbf{y}}\in\mathbf{Y}} \mathbb{P}(\hat{\mathbf{y}})\gamma_i(\hat{\mathbf{y}}) \right) y_i < 0 \end{cases}$$

Moreover, to simplify notation, we will redefine the set of states. Let us denote any preference profile as a partition of two smaller vectors: $\mathbf{y} = (\mathbf{y}_A, \mathbf{y}_B)$. By concavity,[31] the relevant aspect of any vector $\mathbf{y}_i$ is the number of type-i countries with a positive shock. Let us denote that number as k_i. Replacing the Pareto weights, payoffs, and the new notation of states, the solution to the maximization problem becomes:

$$x = \begin{cases} 1, & N_A\overline{y}_A + \delta N_B E[\gamma_B]E[y_B|N_B] > 0 \\ x^{**}, & N_A\overline{y}_A + (1-\delta)(N_B - k_B)\gamma_B(k_B)\underline{y}_B + \delta N_B E[\gamma_B]E[y_B|k_B] = 0,\ 0 \le k_B \le N_B - 1 \\ 0, & N_A E[y_A|k_A] + \delta N_B E[\gamma_B]E[y_B|k_B] < 0,\ 0 \le k_A \le N_A - 1, 0 \le k_B \le N_B \end{cases} \tag{16}$$

Where $\gamma_B(k_B) = \gamma_i(\mathbf{y}) > 0$ (i) for a country i that is type-B, (ii) in a state $\mathbf{y}$ such that k_B type-B countries had a negative shock, and (iii) all type-A countries had a positive shock (and therefore it does not depend on k_A). $E[\gamma_B]$ is the expectation of $\gamma_B(k_B)$. Finally, $E[y_i|k_i] = (k_i\overline{y}_i + (N_i - k_i)\underline{y}_i)/N_i$. Note from (16) that the incentive constraints are binding only in states where all type-A countries agree and at least

[31] This is also the case since there is homogeneity within each type.

one type-B country disagrees.

Recall that the binomial coefficient equals $\binom{N_B}{k_B} = \frac{N_B!}{(N_B-k_B)!k_B!}$, which we will use to compute the probability of the subset of all the states where k_B type-B and all type-A countries agree: $\mathbb{P}(k_B) = p^{N_A}\binom{N_B}{k_B}p^{k_B}(1-p)^{N_B-k_B}$. Now, we can use the definition of $E[\gamma_B]$, to compute its value:

$$\begin{aligned} E[\gamma_B] &= p^{N_A}\sum_{k_B=0}^{N_B}\mathbb{P}(k_B)E[\gamma_B|k_B] \\ &= p^{N_A}\sum_{k_B=0}^{N_B}\binom{N_B}{k_B}p^{k_B}(1-p)^{N_B-k_B}\frac{(N_B-k_B)\gamma_B(k_B)}{N_B} \end{aligned}$$

Let us define ϵ such that $N_A\overline{y}_A + \delta N_B E[\gamma_B]E[y_B|N_B] + \epsilon = 0$. Next, we add this equation with the rest of the first-order conditions where all type-A countries agree: $N_A\overline{y}_A + (1-\delta)(N_B-k_B)\gamma_B(k_B)\underline{y}_B + \delta N_B E[\gamma_B]E[y_B|k_B] = 0$, but each weighted by $p^{N_A}\binom{N_B}{k_B}p^{k_B}(1-p)^{N_B-k_B}/N_B$:

$$-\frac{p^N}{N_B}\epsilon = p^{N_A}\sum_{k_B=0}^{N_B}\frac{\mathbb{P}(k_B)}{N_B}\left(N_A\overline{y}_A + (1-\delta)(N_B-k_B)\gamma_B(k_B)\underline{y}_B + \delta N_B E[\gamma_B]E[y_B|k_B]\right)$$

where $\gamma_B(k_B)$ is zero only when $k_B = N_B$. From the last two expressions above, we can solve $E[\gamma_B]$:

$$E[\gamma_B] = \frac{N_A\overline{y}_A p^{N_A}(1-p^{N_B})}{N_B\left(\delta p^N\overline{y_B} - (1-\delta)\underline{y_B} - \delta p^{N_A}E[y_B]\right)} \tag{17}$$

and it is easy to verify that the denominator is positive when $\delta < \delta^*$. Now, it remains to check two more conditions: (i) $\gamma_B(k_B) > 0$ for $0 \leq \gamma_B(k_B) \leq N_B - 1$ and (ii) $N_A E[y_A|k_A] + \delta N_B E[\gamma_B]E[y_B|k_B] < 0$ for $0 \leq k_A \leq N_A - 1$ and $0 \leq k_B \leq N_B$. To verify the first condition, we note from the middle expression in (16) that given $E[\gamma_B]$, it suffices to verify the case in which $k_B = 0$. That is:

$$\gamma_B(0) = \frac{N_A\overline{y}_A + \delta E[\gamma_B]N_B\underline{y}_B}{-(1-\delta)N_B\underline{y_B}} > 0$$

we replace $E[\gamma_B]$ from (17) to obtain the desired condition:

$$\delta < \frac{-\underline{y_B}}{-\underline{y_B} + p^{N_A}E[y_B] - p^N\overline{y_B} - P^{N_A}(1-p^{N_B})\underline{y_B}} \tag{18}$$

and again, it is easy to verify that the denominator is positive. Finally, to verify the second condition, it suffices to check that $(N_A-1)\overline{y}_A + \underline{y}_A + \delta N_B E[\gamma_B]\overline{y}_B < 0$. Replacing the solution for $E[\gamma_B]$ from (17), we obtain an upper bound for δ:

$$\delta < \frac{-\underline{y_B}}{-\underline{y_B} + p^{N_A}E[y_B] - p^N\overline{y_B} - \frac{P^{N_A}(1-p^{N_B})N_A\overline{y_A}\ \overline{y_B}}{(N_A-1)\overline{y_A}+\underline{y_A}}}$$

However, from point (ii) in lemma 3, equation (18) is sufficient to satisfy the previous condition. Therefore:

$$\delta^{**} = \min\left\{\frac{-\underline{y_B}}{-\underline{y_B} + p^{N_A}E[y_B] - p^N\overline{y_B} - P^{N_A}(1-p^{N_B})\underline{y_B}}, \delta^*\right\}$$

□

E Proofs of Propositions 6 and 7, Lemma 3, and Corollary 1

Recall that k_A and k_B are the (state-dependent) number of type-A and type-B countries, respectively, in favor of taking the action. With this notation, the following table summarizes the results of section 3:

Proof of Proposition 6. This proof is indeed extremely similar to the proof of proposition 5. It is clear that when $\delta < \delta^{**}$, the solution follows proposition 5. Therefore, we begin by setting $\delta^{**} = \underline{d}_0$. First, we assume

Table 4: Summary of propositions 5, 6 and 7. Optimal x as a function of δ, k_A and k_B

			k_B								
	min δ	max δ	N_B	N_B-1	...	k'_B+1	k'_B	k'_B-1	...	1	0
$k_A = N_A$	0	δ^{**}	1	x^{**}	...	x^{**}	x^{**}	x^{**}	...	x^{**}	x^{**}
	δ^{**}	$\underline{d}_1$	1	$\hat{x}$	...	$\hat{x}$	$\hat{x}$	$\hat{x}$	...	$\hat{x}$	0
					...				...		
	$\underline{d}_{k'_B}$	$\hat{d}$	1	$\hat{x}$	...	$\hat{x}$	0	0	...	0	0
				$\downarrow$	...	$\downarrow$			...		
	$\hat{d}$	$\overline{d}_1$	1	1	...	1	$\hat{x}$	0	...	0	0
	$\overline{d}_{k'_B-1}$	$\overline{d}_{k'_B}$	1	1	...	1	1	1	...	$\hat{x}$	0
	$\overline{d}_{k'_B}$	δ^*	1	1	...	1	1	1	...	1	$\hat{x} \to 1$
	δ^*	1	1	1	...	1	1	1	...	1	1
$k_A < N_A$	0	1	0	0	0	0	0	0	0	0	0

that $k'_B > 0$; otherwise, $\delta^* = \delta^{**}$, and there is nothing else to prove. Then, we need to find a threshold $\underline{d}_1$ satisfying the conditions in proposition 6. Namely, we want to find the maximum discount factor such that the solution to:

$$\max_{x\in[0,1]} \sum_{\mathbf{y}\in\mathbf{Y}} \mathbb{P}(\mathbf{y})x(\mathbf{y}) \sum_{i\in N} \left(\lambda_i + (1-\delta)\gamma_i(\mathbf{y}) + \delta \sum_{\hat{\mathbf{y}}\in\mathbf{Y}} \mathbb{P}(\hat{\mathbf{y}})\gamma_i(\hat{\mathbf{y}}) \right) y_i$$

is characterized by:

$$x = \begin{cases} 1, & N_A\overline{y}_A + \delta N_B E[\gamma_B]\overline{y}_B > 0 \\ \hat{x}, & N_A\overline{y}_A + (1-\delta)(N_B - k_B)\gamma_B(k_B)\underline{y}_B + \delta N_B E[\gamma_B]E[y_B|k_B] = 0,\ 1 \le k_B \le N_B - 1 \\ 0, & N_A\overline{y}_A + \delta N_B E[\gamma_B]\underline{y}_B < 0 \\ 0, & N_A E[y_A|k_A] + \delta N_B E[\gamma_B]E[y_B|k_B] < 0,\ 0 \le k_A \le N_A - 1, 0 \le k_B \le N_B \end{cases}$$

Similar to the proof of proposition 5, we add the previous first-order conditions, weighted by the probability of all events with k_B type-B members agreeing and dive by N_B: $\frac{\mathbb{P}(k_B)}{N_B}$ to obtain an expression for $E[\gamma_B]$:

$$E[\gamma_B] = \frac{N_A \overline{y}_A p^{N_A} \sum_{k_B=1}^{N_B-1} \mathbb{P}(k_B)}{N_B \left(-(1-\delta)\underline{y_B} - \delta p^{N_A} \sum_{k_B=1}^{N_B-1} \mathbb{P}(k_B) E[y_B|k_B]\right)}$$

The denominator is positive, which follows from point (i) of lemma 3 and the fact that $\delta < \delta^*$. Finally, similar to the proof of proposition 5, we need to show that indeed each $\gamma_B(k_B)$ is positive for $1 \leq k_B \leq N_B - 1$, and it suffices to show that it holds for $k_B = 1$. Therefore, from $N_A \overline{y}_A + (1-\delta)(N_B - k_B)\gamma_B(k_B)\underline{y}_B + \delta N_B E[\gamma_B] E[y_B|k_B] = 0$ evaluated at $k_B = 1$, we can solve for $\gamma_B(1)$ and find the maximum δ that ensures it is positive:

$$\delta < \frac{-\underline{y_B}}{-\underline{y_B} + p^{N_A} \sum_{k_B=1}^{N_B-1} \mathbb{P}(k_B) \left(E[y_B|k_B] - E[y_B|1]\right)} \tag{19}$$

The denominator is positive. This follows from point (i) of lemma 3 and the fact that $k'_B > 0$. The next step is to show that it is optimal to give veto power to the first N_A members. It suffices to show that $(N_A - 1)\overline{y}_A + \underline{y}_A + \delta N_B E[\gamma_B]\overline{y}_B < 0$. Replacing the expression for $E[\gamma_B]$, this condition holds when:

$$\delta < \frac{-\underline{y_B}}{-\underline{y_B} + p^{N_A} \sum_{k_B=1}^{N_B-1} \mathbb{P}(k_B) \left(E[y_B|k_B] - \frac{N_A \overline{y}_A}{(N_A-1)\overline{y}_A + \underline{y}_A} \overline{y}_B\right)} \tag{20}$$

where the expression $-\frac{N_A \overline{y}_A}{(N_A-1)\overline{y}_A + \underline{y}_A}\overline{y}_B$ is positive since $(N_A - 1)\overline{y}_A + \underline{y}_A$ is negative. It follows from point (i) of lemma 3 that the denominator is positive. In addition, similar to the proof of proposition 5, we also require that $\delta < \delta^*$. Therefore, if $\delta^* < \frac{-\underline{y_B}}{-\underline{y_B} + p^{N_A} \sum_{k_B=1}^{N_B-1} \mathbb{P}(k_B)(E[y_B|k_B] - E[y_B|1])}$, then $\underline{d}_0 = \underline{d}_1 = \cdots = \hat{\delta}$

Finally, we need to show that $0 \leq \hat{x} \leq 1$. This will indeed give us a non-binding lower bound for δ. From the action-taking constraint, $(1-\delta)\hat{x}\underline{y}_B + \delta p^{N_A}\left(p^{N_B}\overline{y}_B + \sum_{k_B=1}^{N_B-1} \mathbb{P}(k_B)\hat{x} E[y_B|k_B]\right) = 0$. It follows from part (ii) in the definition of k'_B that this condition is satisfied.

The last step is to generalize the previous analysis to all thresholds $\underline{d}_{k_B}$. To do this, note that (19) will become:

$$\delta < \underline{d}_{k_B} = \frac{-\underline{y_B}}{-\underline{y_B} + p^{N_A} \sum_{k=k_B}^{N_B-1} \mathbb{P}(k) \left(E[y_B|k] - E[y_B|k_B]\right)} \tag{21}$$

and it is increasing in $k_B \leq k'_B$. However, note that we need to define thresholds up to k'_B+1. However, at k'_B+1 the summation $\sum_{k=k'_B+1}^{N_B-1}\mathbb{P}(k)E[y_B|k]$ actually decreased from k'_B, since $E[y_B|k'_B+1]>0$. The last threshold $\hat{\delta}$ comes from another restriction, which is $x \leq 1$. From the definition of k'_B:

$$\underline{d}_{k'_B+1} = \hat{\delta} = \frac{-\underline{y_B}}{-\underline{y_B} + p^{N_A}\sum_{k=k'_B}^{N_B}\mathbb{P}(k)E[y_B|k]} \tag{22}$$

Indeed, the previous equation ensures that $\hat{x}$ converges to 1 at $\hat{\delta}$. Finally, the threshold (20) is not binding, as $-\sum_{k=k_B}^{N_B-1}\mathbb{P}(k)\frac{N_A\overline{y}_A}{(N_A-1)\overline{y}_A+\underline{y}_A}\overline{y}_B$ is decreasing in k_B, and point (ii) of lemma 3 was assumed.

□

Proof of Proposition 7. Similar to the previous two proofs, let us define $\hat{\delta} = \overline{d}_{k'_B+1}$, and we need to find $\overline{d}_{k'_B}$ such that $\delta \in (\overline{d}_{k'_B+1}, \overline{d}_{k'_B})$ implies that the solution to:

$$\max_{x\in[0,1]}\sum_{\mathbf{y}\in\mathbf{Y}}\mathbb{P}(\mathbf{y})x(\mathbf{y})\sum_{i\in N}\left(\lambda_i + (1-\delta)\gamma_i(\mathbf{y}) + \delta\sum_{\hat{\mathbf{y}}\in\mathbf{Y}}\mathbb{P}(\hat{\mathbf{y}})\gamma_i(\hat{\mathbf{y}})\right)y_i$$

is characterized by:

$$x = \begin{cases} 1, & N_A\overline{y}_A + \delta N_B E[\gamma_B]\overline{y}_B > 0 \\ 1, & N_A\overline{y}_A + (1-\delta)(N_B-k_B)\gamma_B(k_B)\underline{y}_B + \delta N_B E[\gamma_B]E[y_B|k_B] \geq 0,\ k'_B+1 \leq k_B \leq N_B-1 \\ \hat{x}, & N_A\overline{y}_A + \delta N_B E[\gamma_B]E[y_B|k'_B] = 0 \\ 0, & N_A\overline{y}_A + \delta N_B E[\gamma_B]E[y_B|k_B] \geq 0,\ 0 \leq k_B \leq k'_B-1 \\ 0, & N_A E[y_A|k_A] + \delta N_B E[\gamma_B]E[y_B|k_B] < 0,\ 0 \leq k_A \leq N_A-1, 0 \leq k_B \leq N_B \end{cases} \tag{23}$$

In this case, it is more convenient to start by finding conditions such that $0<\hat{x}<1$. The participation constraints in states when $x=1$ are $(1-\delta)\underline{y}_B + \delta E[xy_B] = 0$. Therefore, $E[xy_B] = -\frac{1-\delta}{\delta}\underline{y}_B$. Now, we use the previous identity and the definition of $E[xy_B]$ to obtain an expression for $\hat{x}$:

$$
\begin{aligned}
-\frac{1-\delta}{\delta}\underline{y}_B &= E[xy_B] \\
&= p^{N_A}\sum_{k_B=0}^{N_B}\mathbb{P}(k_B)x(k_B)E[y_B|k_B] \\
&= p^{N_A}\sum_{k_B=k'_B+1}^{N_B}\mathbb{P}(k_B)E[y_B|k_B] + p^{N_A}\mathbb{P}(k'_B)\hat{x}E[y_B|k'_B]
\end{aligned}
$$

From the previous equation, we can solve for $\hat{x}$:

$$
\hat{x} = \frac{-\frac{1-\delta}{\delta}\underline{y}_B - p^{N_A}\sum_{k_B=k'_B+1}^{N_B}\mathbb{P}(k_B)E[y_B|k_B]}{p^{N_A}\mathbb{P}(k'_B)E[y_B|k'_B]}
$$

Finally, from the expression above, $\hat{x}$ lies on $(0,1)$ as long as δ lies on:

$$
\frac{-\underline{y}_B}{-\underline{y}_B + p^{N_A}\sum_{k_B=k'_B+1}^{N_B}\mathbb{P}(k_B)E[y_B|k_B]} < \delta < \frac{-\underline{y}_B}{-\underline{y}_B + p^{N_A}\sum_{k_B=k'_B}^{N_B}\mathbb{P}(k_B)E[y_B|k_B]} \tag{24}
$$

and the thresholds are well defined and follow the right order since $E[y_B|k'_B] < 0$ from the definition of k'_B. The next condition we need to check is that $\gamma_B \geq 0$. Similar to the proof of proposition 6, the third condition in (23) gives an exact expression for $E[\gamma_B]$:

$$
E[\gamma_B] = \frac{N_A\overline{y}_A}{\delta N_B E[y_B|k'_B]}
$$

Additionally, we can add up all expressions with $\gamma_B(k_B)$, weighted by $\frac{\mathbb{P}(k_B)}{N_B}$; but now we do so to obtain an upper bound for $E[\gamma_B]$:

$$
\begin{aligned}
E[\gamma_B] &= p^{N_A} \sum_{k_B=k'_B+1}^{N_B-1} \frac{\mathbb{P}(k_B)}{N_B} \gamma_B(k_B)(N_B - k_B) \\
&\leq p^{N_A} \sum_{k_B=k'_B+1}^{N_B-1} \frac{\mathbb{P}(k_B)}{N_B} \frac{N_A \overline{y}_A}{(1-\delta)(-\underline{y}_B)} \left(1 - \frac{E[y_B|k_B]}{E[y_B|k'_B]}\right)
\end{aligned}
$$

Combining the previous two equations, we obtain an additional lower bound for δ:

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

However, it is easy to check that (24) is a more restrictive condition. Moreover, note that the lower bound in (24) is precisely $\hat{\delta}$. Finally, we want to ensure that veto power still holds. As before, it suffices to show that the last condition in equation (23) holds for the case $k_A = N_A - 1$ and $k_B = N_B$. That is, $(N_A - 1)\overline{y}_A + \underline{y}_A + \delta E[\gamma_B] N_B \overline{y}_B < 0$. This inequality holds when:

$$
N_A \overline{y}_A \overline{y}_B < \left((N_A - 1)\overline{y}_A + \underline{y}_A\right) E[y_B|k'_B] \tag{25}
$$

However, the second condition in lemma 3 is sufficient for (25).

In general, the expression for the thresholds is:

$$
\overline{d}_{k_B} = \frac{-\underline{y}_B}{-\underline{y}_B + p^{N_A} \sum_{k=k_B}^{N_B} \mathbb{P}(k) E[y_B|k]} \tag{26}
$$

and $N_A \overline{y}_A \overline{y}_B < \left((N_A - 1)\overline{y}_A + \underline{y}_A\right) E[y_B|k_B]$ holds for any $k_B \leq k'_B$; therefore, veto power is optimal. Note that $\overline{d}_0 = \delta^*$, since it is the threshold needed to implement the action $x = 1$ in every state in which all type-A countries agree.

□

Proof of Lemma 3. All conditions stated in lemma 3 have been used at least once during the proofs of propositions 6 and 7; moreover, no additional assumptions were made. Therefore, these are sufficient

conditions. □

Proof of Corollary 1. It suffices to show the source of equation (11). Note that if there is a council of size $N_A + \tilde{N}$, the threshold to implement an action cannot exceed the total number of votes. Therefore, $N^* \leq N_A + \tilde{N}$. Moreover, $k(\delta)$ is the minimum number of 'yes' votes that can be selected in a state such that it is optimal to implement some action $x > 0$. As a consequence, $N^* \leq N_A + k(\delta)$. From these two conditions, we obtain that:

$$N^* - N_A \leq \min\{k(\delta), \tilde{N}\}$$

On the other hand, some voting power must be given to type-B countries; otherwise, type-A countries would always implement the Pareto-efficient allocation. Therefore, $N^* > N_A$. Finally, we wish to ensure that we can find enough 'no' votes whenever it is optimal to preserve the status quo. In particular, we seek to satisfy that conditions in states $k(\delta) - 1$. In such states, there number of 'yes' votes coming from type-B countries cannot exceed $N^* - N_A$. If there are $\tilde{N}$ type-B council members, and we can find at most $N_B - (k(\delta) - 1)$ 'no' votes, we need that $\tilde{N} - (N_B - (k(\delta) - 1)) < N^* - N_A$. From the previous two conditions:

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

□

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.

Downs, George W, David M Rocke, and Peter N Barsoom (1996) "Is the good news about compliance good news about cooperation?" *International Organization*, Vol. 50, pp. 379–406.

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 (2005) "Majority rules and incentives," *The Quarterly Journal of Economics*, Vol. 120, pp. 1535–1568.

——— (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.