



Inducing Cooperation through Weighted Voting and Veto Power

Antonin Macé, Rafael Treibich

► To cite this version:

Antonin Macé, Rafael Treibich. Inducing Cooperation through Weighted Voting and Veto Power. 2018. halshs-01630090v4

HAL Id: halshs-01630090

<https://shs.hal.science/halshs-01630090v4>

Preprint submitted on 19 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Inducing Cooperation through Weighted Voting and Veto Power*

Antonin Macé¹ and Rafael Treibich²

¹Paris School of Economics, CNRS and ENS

²University of Southern Denmark

June 19, 2019

Abstract

We study the design of voting rules for committees representing heterogeneous groups (countries, states, districts) when cooperation among groups is voluntary. While efficiency recommends weighting groups proportionally to their stakes, we show that accounting for participation constraints entails overweighting some groups, those for which the incentive to cooperate is the lowest. When collective decisions are not enforceable, cooperation induces more stringent constraints that may require granting veto power to certain groups. In the benchmark case where groups differ only in their population size (i.e, the apportionment problem), the model provides a rationale for setting a minimum representation for smaller groups.

JEL: F53, D02, C61, C73.

1 Introduction

In 1787, when the founding fathers met in Philadelphia to discuss the creation of a new constitution, the most contentious issue revolved around the composition of the future legislature.

*This paper was previously circulated under the title “On the Weights of Sovereign Nations”. We would like to thank the associate editor and three anonymous referees for very valuable advice. We would also like to thank Alessandra Casella, Ernesto Dal Bó, Andreas Kleiner, Jean-François Laslier, Michel Lebreton, Aniol Llorente-Saguer, Sharun Mukand and Stefan Napel for useful comments, as well as audiences from Barcelona, Berkeley, Copenhagen, Glasgow, Hanover, Louvain-la-Neuve, Odense, Paris, Saint-Étienne, and Toulouse. Financial support from the ANR-17-CE26-0003 (CHOp) project and the ANR-17-EURE-001 (a French government subsidy managed by the Agence Nationale de la Recherche under the framework of the Investissements d’avenir programme) is gratefully acknowledged.

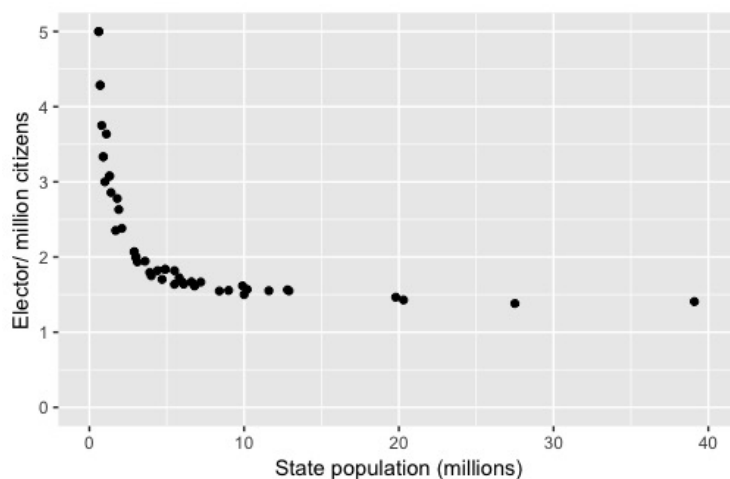
Larger states, led by Virginia, argued in favor of a bicameral legislature, under which states would receive a number of seats proportional to their population in both houses. Smaller states rejected the Virginia Plan, proposing instead the creation of a single house, under which states would receive an equal number of seats, independently of their population. The conflict was so serious that smaller states threatened to leave the union if larger states insisted on the idea of a purely proportional representation:

“The small ones would find some foreign ally of more honor and good faith, who will take them by the hand and do them justice.”

Gunning Bedford Jr., representative for Delaware, 1787.

The issue was resolved by the so-called *Connecticut Compromise*, and the creation of a bicameral legislature, under which states received a weight proportional to their population in the lower house (House of Representatives), but an equal weight in the upper house (Senate). The resulting distribution of seats in the Electoral College¹ appears as a compromise between the principle of “one man, one vote”, leading to efficient and democratic decision-making, and the principle of “one state, one vote”, ensuring the voluntary participation of all states. As illustrated in [Figure 1](#), this distribution is such that the number of electors per capita in each state decreases with the state’s population.

Figure 1: Electors per million citizens in the U.S. Electoral College, by state population (2017).



The tension between the efficiency of a multi-party institution and its acceptability by all parties is not limited to the episode of the Constitutional Convention. In fact, such a tension

¹The Electoral College is a representative committee designed to elect the U.S. president. The number of electors obtained by each state corresponds to the sum of its number of senators (2) and of its number of representatives in the House. In most states, electors are appointed through a winner-takes-all system.

is also inherently present for many international organizations and confederations, when a set of *sovereign* states voluntarily commits to collectively decide on one or several policy areas. One prominent example is the UN Security Council, in which the five permanent members can veto any resolution, thus benefiting from a disproportionate power. The veto power has often been criticized for severely reducing the efficiency of the UN and taking away much of its relevance on the international scene. However, when the Charter of the UN was ratified in San Francisco in 1945, “the issue was made crystal clear by the leaders of the Big Five: it was either the Charter with the veto or no Charter at all” (Wilcox, 1945). In the UN, as in many international organizations, the need to accommodate countries’ voluntary participation is further aggravated by the lack of external enforcement, since sovereign countries cannot be forced to respect the collective decision ex-post.

The importance of both ex-ante and ex-post participation constraints may vary with the nature of collective decisions, sometimes leading to different voting rules within the same organization. This is the case for example at the Council of the European Union, one of the EU’s main decision-making bodies, where the most sensitive decisions are taken at the unanimity, while other type of reforms are taken at a double qualified majority.² At the UN, the General Assembly - who has a softer, more deliberative role than the Security Council - gives an equal representation to all countries and makes decisions at a two third majority rule.

As these examples show, participation constraints play a critical role in shaping collective decision rules. Although the agreed rules vary significantly across institutions, they all appear to depart from efficiency so as to ensure the voluntary participation of all parties. In this article, we propose to study the design of voting rules for committees representing heterogeneous groups (countries, states, districts) when cooperation among groups is voluntary, and the enforceability of collective decisions may not be guaranteed. When should the voting rule depart from the efficient benchmark, and if so, which group should be overweighted? When should some groups benefit from veto power? To address these questions, we take a second-best approach to institutional design by looking for the most efficient rules among those that are politically feasible.³

²Decisions regarding common foreign and security policy, citizenship (the granting of new rights to EU citizens), EU membership, harmonisation of national legislation on indirect taxation, EU finances (own resources, the multiannual financial framework), certain provisions in the field of justice and home affairs (the European prosecutor, family law, operational police cooperation, etc.), the harmonization of national legislation in the field of social security and social protection are taken at the unanimity. Other (less sensitive) decisions are taken according to the following rule: a reform is adopted if approved by 55% of the Member States, representing at least 65% of the EU population. Additionally, a proposal cannot be blocked by less than four Member States.

³An alternative approach would be to explore which decision rules are likely to arise from a bargaining phase among countries in the presence of such constraints. We provide a discussion of this perspective in the Conclusion and in [Section A.9](#) in the Appendix.

Our model features a fixed set of countries choosing whether to delegate some of their competences to a supranational entity.⁴ The choice to transfer a competence is made unanimously *ex ante*, before countries learn about their preferences over future decisions.⁵ If cooperation is agreed upon, decisions are made collectively according to a predetermined voting rule. If cooperation is rejected, countries remain sovereign and make their own decisions.

Our core assumption is that the choice to delegate reflects a trade-off between the efficiency gains from cooperation and a reduced control over decisions. Making collective decisions is profitable for many reasons: it generates coordination gains from harmonized decisions (Loeper, 2011), allows for economies of scale (Alesina et al., 2005), increases bargaining power (Moravcsik, 1998), strengthens commitment (Bown, 2004), etc. However, by forfeiting the right to make their own decisions, countries also lose some decision power. As a result, countries may reject cooperation if they expect to disagree too frequently with the collective decision. The voting rule, which determines how much influence each country exerts on the collective decision, thus plays a critical role in generating cooperation.

We consider in turn the cases of enforceable and non-enforceable collective decisions. When decisions are enforceable, we show that a voting rule induces cooperation if it satisfies a set of *participation constraints*. We then characterize the *optimal rules*, defined as solutions of a welfare maximization problem under these participation constraints. Optimal rules involve weighted voting, but may depart from efficiency to make some countries willing to cooperate.

If collective decisions cannot be enforced, as is often the case in international organizations (Maggi and Morelli, 2006), countries may decide not to comply with collective decisions *ex post*. In that case, compliance incentives arise from a dynamic interaction, as countries may abide by an unwanted decision to obtain future gains from cooperation. We model these incentives in a repeated game, from which we derive a set of *compliance constraints* that a voting rule needs satisfy in order to be *self-enforcing*. We solve for the *optimal self-enforcing rules*, and show that they are weighted and possibly grant veto power to some countries. The result provides a new rationale for the use of veto power:⁶ compliance can sometimes be

⁴For instance, the European Union has exclusive competence over customs unions, competition policy, monetary policy (for countries in the Eurozone), common fisheries policy, and common commercial policy. The EU also holds shared competence (Member States cannot exercise competence in areas where the EU has done so) over various other domains, such as the internal market, agricultural policy, environmental policy, and consumer protection. See Treaty of Lisbon (2007b).

⁵The fact that all EU competences must be voluntarily transferred by its Member States is known as the principle of *conferral* (Treaty of Lisbon, 2007a).

⁶Note that the argument put forth here is conceptually different from the one proposed by Bouton et al. (2018), as the rules under consideration differ. In the present article, we consider dichotomous voting (yes/no), and a country has veto power under a weighted rule if its weight and/or the threshold is high enough. In Bouton et al. (2018), voting is trichotomous (yes/no/veto), so that veto power can be granted independently from the majority rule. Our focus on dichotomous voting is reasonable here as we solely deal with preference aggregation, and we provide a rationale for the veto in this setting. By contrast, Bouton et al. (2018) emphasize the advantage of trichotomous voting (with veto) when both preferences and information are aggregated.

best achieved by giving some “negative power” to a country (i.e., veto power) rather than by compensating it with too much additional “positive power” (i.e., overly large weight).

Finally, we consider a simpler model in which utilities are binary and countries differ only in their population size, but are otherwise (ex ante) identical. This model allows us to address the classic problem of *apportionment*: how should countries’ populations be translated into voting weights of representatives in an international committee? We obtain sharper results in that model. Countries must receive weights proportional to their populations, except for the smallest ones, which must all be weighted equally. The result thus offers a rationale for a minimum representation of smaller countries, as required explicitly in the Treaty of Lisbon ([Treaty of Lisbon, 2007a](#)). It also echoes the distribution of weights in the U.S. Electoral College, where each state is allocated a baseline of two seats plus a number of seats proportional to its population. We extend the characterization of the optimal weights to allow for heterogeneity in efficiency gains across countries and show that smaller countries ought to be overweighted at the optimum even when they gain relatively more from cooperation than larger countries. We further discuss how double majority rules can similarly depart from efficiency to help smaller countries satisfy participation constraints, although they are not optimal in our model. Finally, we focus on optimal self-enforcing rules, which take the form of weighted majorities or unanimity, and derive comparative statics with respect to parameters of the model.

1.1 Related Literature

Our article combines both a normative and a positive approach to voting rules in representative committees. On the normative side, we follow the literature on *apportionment*, which studies the allocation of weights to nations (states) of different sizes in international unions (federations). A first branch of the literature focuses on how to best approximate proportionality when weights are constrained to be integers, such as for the allocation of seats in a parliament ([Balinski and Young, 1982](#)). A second branch of the literature questions the desirability of proportionality, arguing instead in favor of a principle of *degressive proportionality*, which requires weights to increase less than proportionally to states’ populations.⁷ Our article

⁷The literature on degressive proportionality has focused in particular on the *square-root law*, which recommends weights that are proportional to the square-root of each state’s population. Arguments in favor of the *square-root law* are developed by [Penrose \(1946\)](#), [Felsenthal and Machover \(1999\)](#), and [Barberà and Jackson \(2006\)](#), on the grounds of (respectively) equalizing each citizen’s influence, minimizing the mean *majority deficit* (extent of disagreement with the federation-wise majority rule), and following the utilitarian principle. These works are extended by [Beisbart and Bovens \(2007\)](#) and [Kurz et al. \(2017\)](#), who show the fragility of the law to the introduction of a small degree of correlation in citizen’s preferences. Finally, [Koriyama et al. \(2013\)](#) offer a different rationale for degressive proportionality based on the utilitarian principle when citizens exhibit decreasing marginal utility. See [Laslier \(2012\)](#) for a survey.

follows this second strand, building in particular on the utilitarian approach⁸ proposed by Barberà and Jackson (2006) to study voting rules in two-tier democracies, where citizens elect representatives that vote on their behalf. They show in a general framework that an efficient voting rule must weight each state proportionally to its *stake* in the collective decisions, a result we refer to as the *efficient benchmark*.⁹ Similar to our model of apportionment, they also consider two simpler models of preference formation, for which they derive a closed-form solution of the efficient weights. In their *fixed-size block model*, preferences are independent across citizens and the efficient weight of a country is proportional to the square root of its population. In their *fixed-number-of-blocks model*, citizens' preferences are correlated within each country and the efficient weights are proportional to countries' populations. We follow the latter model in our section on apportionment, as it appears to be more consistent with empirical studies (Gelman et al., 2004).

We depart from this literature by adding political feasibility constraints. The premise is that countries' decision to cooperate is voluntary. Starting with the same assumption, but inspired by the formation of monetary unions, Casella (1992) shows that a two-country partnership may require overweighting (in the welfare function of the partnership's decision-maker) the country most tempted to remain sovereign. Our first result on enforceable decisions generalizes her argument to committees with more than two countries by analyzing this trade-off in a voting game. Barberà and Jackson (2004) also follow a positive approach to the design of voting rules, but their focus is on the stability of the voting rule with respect to a constitutional change: a voting rule is *self-stable* if it cannot be overthrown by another voting rule. In contrast, we study the stability of a rule with respect to the composition of the union and require that a rule induces the cooperation of all of its members. Note that the optimal rules and optimal self-enforcing rules that we identify are self-stable¹⁰ among those satisfying the same feasibility constraints, since they are obtained from a welfare-maximization program. The assumption of enforceable decisions is relaxed in a pioneering article by Maggi and Morelli (2006), who consider a union of homogeneous countries engaging in repeated collective decisions. They prove that the optimal self-enforcing rule is either the (efficient) qualified majority rule, or the unanimity rule if the discount factor falls below a critical threshold. Our section on self-enforcing voting rules extends their analysis to the case of a heterogeneous union. In particular, we show that the optimal self-enforcing rule may give veto power to a strict subset of countries. In the apportionment model, veto power is given

⁸The ex-ante utilitarian approach to binary voting rules was initiated by Rae (1969) to provide an argument for the majority rule.

⁹A similar result is provided by Azrieli and Kim (2014) in a mechanism design context. See also Brighouse and Fleurbaey (2010) for a discussion of this idea at the level of political philosophy.

¹⁰With respect to the unanimity rule, taken as the benchmark constitutional rule.

to all countries or none, but the optimal self-enforcing rule may be neither the (efficient) qualified majority rule nor the unanimity rule, for intermediate values of the discount factor.

Finally, a central assumption in our article is that a country's decision to cooperate results from a trade-off between the efficiency of collective decisions and the loss of power in the union.¹¹ Following the seminal article of [Alesina and Spolaore \(1997\)](#) on the (endogenous) size of nations, several articles have explored this rationale for cooperation between countries.¹² [Alesina et al. \(2005\)](#) study the composition and size of international unions when efficiency gains stem from externalities in public good provisions. [Renou \(2011\)](#) studies the effect of the stringency of the supermajority rule on the endogenous composition of the union. Similar to [Renou \(2011\)](#), our article emphasizes the importance of the voting rule on the stability of the union, but differs in that we take into account the heterogeneity of countries.

1.2 Outline

[Section 2](#) introduces the model and the decision game. [Section 3](#) derives the optimal voting rule when collective decisions are enforceable. [Section 4](#) introduces an infinitely repeated version of the decision game and derives the optimal self-enforcing rule. [Section 5](#) illustrates the results of [Section 3](#) and [Section 4](#) in a simple example with five countries. Finally, the model is applied in [Section 6](#) to a simple environment in which utilities are binary and countries differ only in their populations. [Section 7](#) concludes. All the proofs are gathered in [Section A](#).

2 Model

An international union N is made of n countries. Each country has a representative who takes decisions on behalf of its citizens. Representatives must decide whether to remain sovereign or to cooperate, and in the latter case, whether to implement a reform or to stick with the status quo. This is modeled as a game with four stages.

2.1 The Decision Game

In the first stage, each country $i \in N$ decides to remain sovereign, $d_i = 0$, or to cooperate, $d_i = 1$. If at least one country wants to remain sovereign, cooperation is aborted (the game ends), and each country i derives a stand-alone utility $U_i^\emptyset \in \mathbb{R}$. If all countries decide to cooperate, the game continues, and countries have to make a collective decision on the

¹¹See [Demange \(2017\)](#) for a survey of theoretical models on the general tension between the efficiency of large groups and the associated preference heterogeneity.

¹²Note that some authors provide other rationales for international cooperation, such as information aggregation ([Penn, 2016](#)), or even pure preference aggregation ([Cr  mer and Palfrey, 1996](#)).

adoption of a proposed reform.¹³

In the second stage, countries learn the realization of their preferences for the proposed reform. A vector of utilities $u = (u_i)_{i \in N}$ is drawn from a distribution μ . The number u_i measures country i 's aggregate utility if the reform is adopted by all countries. The utilities are drawn independently across countries,¹⁴ and such that for all $i \in N$, $\mathbb{P}_\mu[u_i > 0] > 0$, $\mathbb{P}_\mu[u_i = 0] = 0$ and $\mathbb{P}_\mu[u_i < 0] > 0$. Each country i privately observes its own utility u_i , and the prior μ is common knowledge. If the reform is not adopted by all countries, each country derives a utility of 0.¹⁵

The third stage is a voting stage. Each country reports a message $m_i \in \{0, 1\}$, where $m_i = 1$ is interpreted as a vote in favor of the reform, and $m_i = 0$ is interpreted as a vote against the reform. The collective decision to adopt the reform is made according to a predetermined voting rule v . To keep the model flexible, we define a voting rule as a non-decreasing function $v : \{0, 1\}^N \rightarrow [0, 1]$, where $v(\mathbf{m})$ denotes the probability of accepting the reform, given the vector of messages $\mathbf{m}$.¹⁶ We denote by $\mathcal{V}$ the set of all such voting rules and by $\hat{v}(\mathbf{m}) \in \{0, 1\}$ the realized collective decision; i.e., a random variable $\hat{v}(\mathbf{m})$ such that $\mathbb{P}[\hat{v}(\mathbf{m}) = 1] = v(\mathbf{m})$. For a given profile of votes $\mathbf{m}$, $\hat{v}(\mathbf{m}) = 0$ indicates that countries must keep the status quo and $\hat{v}(\mathbf{m}) = 1$ means that countries must implement the reform.

In the fourth stage, each country i takes an action $a_i \in \{0, 1\}$, taking the value 1 if country i implements the reform, and 0 otherwise. If collective decisions are enforceable, each country must abide by the collective decision, $a_i = \hat{v}(\mathbf{m})$ for all $i \in N$. If collective decisions are not enforceable, then countries may choose to go against the collective decision.

The game thus defined is denoted by $\Gamma_E(v)$ if decisions are enforceable and by $\Gamma_{NE}(v)$ if decisions are not enforceable. In the game $\Gamma_E(v)$, a strategy for $i \in N$ is a vector $s_i = (d_i, m_i)$, with $d_i \in \{0, 1\}$ and $m_i : \mathbb{R} \rightarrow \{0, 1\}; u_i \mapsto m_i(u_i)$. In the game $\Gamma_{NE}(v)$, a strategy for $i \in N$ is a vector $s_i = (d_i, m_i, a_i)$, with $d_i \in \{0, 1\}$, $m_i : \mathbb{R} \rightarrow \{0, 1\}; u_i \mapsto m_i(u_i)$ and $a_i : \mathbb{R} \times \{0, 1\}^N \times \{0, 1\} \rightarrow \{0, 1\}; (u_i, \mathbf{m}, \hat{v}(\mathbf{m})) \mapsto a_i(u_i, \mathbf{m}, \hat{v}(\mathbf{m}))$.

In this article, we particularly focus on the *cooperative profile* of the game; i.e., the profile of strategies such that, for all $i \in N$, $d_i = 1$, $m_i = \mathbb{1}_{u_i > 0}$ and $a_i = \hat{v}(\mathbf{m})$. The expected

¹³Equivalently, one could assume that countries have to make repeated *independent* collective decisions. We assume a single decision for ease of exposition.

¹⁴The independence assumption is ubiquitous in the literature. It emphasizes the conflict of preferences across countries that is central to the model, and it allows for a tractable framework. Note that, if arbitrary patterns of correlation are allowed, the efficient rule may not be weighted.

¹⁵The model does not assume specific population sizes. Each country is characterized by its marginal probability distribution μ_i and its stand-alone utility U_i^0 . Both may reflect the country's population size (as well as the degree of preference homogeneity among its citizens, the quality of its democratic representation, how much it gains from cooperation, etc.), but this need not be explicit.

¹⁶This expression allows for probabilistic decisions, in order to break possible ties. See Koriyama et al. (2013) for an introduction of this class of voting rules, labeled *probabilistic simple games*.

aggregate utility of country i associated with this profile is given by:

$$U_i(v) = \mathbb{E}_\mu \left[v((\mathbf{1}_{u_j > 0})_{j \in N}) u_i \right].$$

A central theme of the article is to identify conditions for which this cooperative profile can be implemented as an equilibrium, as these conditions reflect the constraints that are relevant ex ante, at the constitutional stage where the voting rule is chosen. [Section 3](#) tackles this question when decisions are enforceable, and [Section 4](#) studies the non-enforceable case. Before incorporating such strategic constraints, we introduce the notions of weighted rules, vetoes, welfare, and (first-best) efficient voting rules.

2.2 Weighted Majority Rules and Vetoes

In practice, decision rules used by international committees often take the form of a weighted majority whereby each country is assigned a fixed voting weight and a reform is approved if the total weight of countries in favor exceeds a given threshold (e.g., IMF or Council of the EU before 2014). Formally, a rule v is a *weighted majority rule* if there exist a vector of weights $\mathbf{w} = (w_i)_{i \in N} \in \mathbb{R}^N$, and a threshold $t \in [0, 1]$ such that, for any profile of votes $\mathbf{m} = (m_i)_{i \in N} \in \{0, 1\}^N$,

$$\begin{cases} \sum_{i|m_i=1} w_i > t \sum_{i \in N} w_i & \Rightarrow v(\mathbf{m}) = 1 \\ \sum_{i|m_i=1} w_i < t \sum_{i \in N} w_i & \Rightarrow v(\mathbf{m}) = 0. \end{cases}$$

We say that rule v is weighted and can be represented by the system $[\mathbf{w}; t]$.¹⁷ Whether weighted or not, some rules grant veto power to certain countries (e.g., UN Security Council). Formally, we say that a country $i \in N$ has veto power under rule v if $v(\mathbf{m}) = 0$ whenever $m_i = 0$. We denote by $VE(v) \subseteq N$ the set of countries having veto power under the rule v :

$$VE(v) = \left\{ i \in N \mid m_i = 0 \Rightarrow v(\mathbf{m}) = 0 \right\}.$$

2.3 Welfare and Efficient Voting Rule

For any voting rule v , we define the welfare associated with the cooperative profile under v as:

$$W(v) = \mathbb{E}_\mu \left[v((\mathbf{1}_{u_j > 0})_{j \in N}) \sum_{i \in N} u_i \right] = \sum_{i \in N} U_i(v).$$

¹⁷Note that the definition is agnostic with respect to the tie-breaking rule. Note also that the representation of v may not be unique, even after re-scaling the weights $\mathbf{w}$ by a common factor.

We say that a rule is *efficient* if it achieves the maximum welfare at the cooperative profile; that is, absent any incentive constraint. Following the analysis of [Barberà and Jackson \(2006\)](#), it is useful to define country i 's expected utility from a favorable reform, $w_i^+ = \mathbb{E}_\mu[u_i | u_i > 0]$, and its expected disutility from an unfavorable reform, $w_i^- = -\mathbb{E}_\mu[u_i | u_i < 0]$. From these two numbers, we define country i 's *stake* in the decision as $w_i^e = w_i^+ + w_i^-$, and its *efficient threshold* as $t_i^e = w_i^- / w_i^e$.

Theorem 1. ([Barberà and Jackson, 2006](#); [Azrieli and Kim, 2014](#)) Any efficient voting rule v^e is a weighted majority rule. It is represented by $[\mathbf{w}^e; t^e]$, where the threshold t^e is defined by:

$$t^e = \frac{\sum_{i \in N} w_i^e t_i^e}{\sum_{i \in N} w_i^e}.$$

The result asserts that the efficient rule is essentially unique, in the sense that any efficient rule is represented by the same system of weights $[\mathbf{w}^e; t^e]$, although the tie-breaking rule may differ between two efficient rules.¹⁸ Therefore, we will refer to w_i^e as country i 's *efficient weight*, and the threshold t_i^e is efficient in the sense that it is the threshold of an efficient rule if all countries have the same “efficient threshold”. Note that the result focuses on first-best efficiency and that the associated cooperative profile may not be an equilibrium of the decision game. Incorporating such constraints is the main goal of our article and is the object of the following two sections. In what follows, we assume, without substantial loss of generality, that no country has veto power under the efficient rule (*Assumption NEV*, for no efficient veto).¹⁹

3 Enforceable Decisions

3.1 General Case

We start the analysis by considering the case where decisions prescribed by the voting rule v are enforceable: each country $i \in N$ commits to follow the action plan $a_i = \hat{v}(\mathbf{m})$ for any realization of the messages $\mathbf{m}$. We say that a voting rule v *induces cooperation* if the cooperative profile is a Perfect Bayesian Equilibrium of the game $\Gamma_E(v)$. We denote by $\mathcal{V}_1 \subseteq \mathcal{V}$ the corresponding set of voting rules.

¹⁸Note some subtleties associated with the weight representation. First, there may be other systems of weights $[\mathbf{w}'; t']$ such that efficient rules are represented by $[\mathbf{w}'; t']$. Second, there may be weights $[\mathbf{w}'; t']$ such that some efficient rules are represented by these weights and some other are not.

¹⁹Formally, we require that $\forall i \in N, w_i^- < \sum_{j \neq i} w_j^+$ (Assumption NEV).

Proposition 1. *A voting rule v induces cooperation if and only if each country satisfies the participation constraint: $U_i(v) \geq U_i^\emptyset$ for all $i \in N$.*

We say that a voting rule is *optimal* if it maximizes social welfare in $\mathcal{V}_1$, i.e. if it is a solution of the maximization problem $\max_{v \in \mathcal{V}_1} W(v)$.²⁰ The following theorem describes optimal voting rules.

Theorem 2. *There exists a system of weights $[\mathbf{w}^*; t^*]$ such that any optimal voting rule v^* is weighted and represented by $[\mathbf{w}^*; t^*]$. Countries for which the participation constraint is binding are overweighted relative to their efficient weight, while countries for which the participation constraint is not binding receive their efficient weight.*

Similar to [Theorem 1](#), the result asserts that the optimal rule v^* is essentially unique, in the sense that any optimal rule is represented by the same system of weights $[\mathbf{w}^*; t^*]$. However, contrary to the efficient rule, the optimal rule is such that countries that do not strictly benefit from cooperation may receive more than their efficient weight. We say that these countries are *overweighted*.²¹ In contrast, countries that get strictly more than their stand-alone utility receive their efficient weight.²² Formally, there exists a system $[\mathbf{w}^*; t^*]$, such that any optimal rule v^* is represented by $[\mathbf{w}^*; t^*]$, with, for each country $i \in N$:

$$\begin{cases} U_i(v^*) = U_i^\emptyset & \Rightarrow w_i^* \geq w_i^e \\ U_i(v^*) > U_i^\emptyset & \Rightarrow w_i^* = w_i^e, \end{cases}$$

and where the threshold t^* is the associated weighted average of countries' efficient thresholds:

$$t^* = \frac{\sum_{i \in N} w_i^* t_i^e}{\sum_{i \in N} w_i^*}.$$

At one extremity, if stand-alone utilities are low enough, all countries are willing to cooperate under the efficient voting rule. In that case, the constraints are inoperative, and the efficient rule coincides with the optimal rule. However, as stand-alone utilities become larger, the constraint starts to bind for some countries. The result asserts that, in comparison to the efficient benchmark, these countries should be overweighted, and that the threshold t^* should be closer to their efficient thresholds. This is illustrated in the example of [Section 5](#), where the optimal voting rule, represented by $[(3, 1, 1, 1, 1); 1/2]$, is such that country 1 is

²⁰The existence of a solution is guaranteed when $\mathcal{V}_1$ is non-empty, as the objective function is linear and the set of voting rules $\mathcal{V}_1$ is a closed subset of $[0, 1]^{2^N}$.

²¹We refer here to absolute weights that are higher than the efficient absolute weights. In the paper, we retain this framing, as optimal rules are more easily described in terms of absolute rather than relative weights.

²²Note that the conditions given in [Theorem 2](#) are endogenous. Identifying overweighted countries from exogenous conditions is possible under more specific assumptions, as in [Section 6](#).

overweighted, while countries 2 to 5 get their efficient weight. Country 1's utility $16/3^5$ is equal to its stand-alone utility, while countries 2 to 5's utility $146/3^5$ is larger than their stand-alone utility $32/3^5$.

In contrast with the efficient voting weights, which can be computed independently for each country, the optimal voting weights cannot be obtained separately since they each depend on the complete probability distribution μ and on the vector of stand-alone utilities $(U_i^\emptyset)_{i \in N}$. A country may be overweighted at the optimum if it gains relatively little from cooperation or if it often disagrees with the (efficient) collective decision (as in the example of [Section 5](#)). The level of heterogeneity across countries, both in stakes and preferences, thus plays a crucial role in determining the optimal rule.

Inducing all countries to cooperate may turn out costly if some countries do not benefit enough from cooperation or if they disagree too often with the (endogenous) collective decision. Mechanically, the *cost of participation*, the loss of welfare from having to satisfy the participation constraints,²³ increases with each country's stand-alone utility: decreasing a country's stand-alone utility means relaxing its participation constraint, and thus improving the welfare reached at the optimal rule. However, understanding the effect of other aspects of the model (such as the probability distribution μ) on the cost of participation is more difficult due to the simultaneous effect on the participation constraints and on the efficient decision rule. This ambiguous interplay may lead to counter-intuitive effects. For example, an increase in the efficiency of cooperation may actually increase the cost of cooperation. Consider, for instance, a situation where the efficient decision rule is optimal, and assume that the stake of one country increases (thus increasing the overall efficiency of cooperation). As the new efficient rule weights this country more, other countries whose (ex ante) preferences are opposite to the first country's may end up with a reduced utility. Such countries may then require some additional voting power to cooperate, thus leading to an increase in the cost of participation (from zero to positive). Similarly, an increase in the degree of preference homogeneity may actually increase the cost of participation. Again, starting from a situation where the efficient rule satisfies the participation constraints, raising the homogeneity of preferences may change the efficient voting rule, leading one country's participation constraint to be violated.²⁴ A more homogeneous union may thus induce a larger cost of participation.

²³That is, the difference in welfare between the efficient and the optimal rules.

²⁴Consider, for example, a union of three countries, and assume that the simple majority rule is both efficient and optimal. The probability of favoring the reform is $1/2$ for country 1, $q \in (1/2, 1)$ for country 2, and 1 for country 3. As q increases, the union is more homogeneous, as the probability of any two (or three) countries agreeing is either constant or increasing. However, as U_1 decreases with q (the efficient rule is independent of q , and q only affects the probability of approving the reform when 1 is unfavorable), country 1 may require to be overweighted for high q , and this leads to a positive cost of participation.

4 Non-Enforceable Decisions

We have assumed so far that collective decisions were fully enforceable under cooperation. In fact, enforceability is a major concern for most international organizations, as countries always retain some form of sovereignty and full enforceability is never really achieved. Following Maggi and Morelli (2006), we thus relax the assumption of enforceability and consider an infinitely repeated version of our decision game where countries must repeatedly decide whether to cooperate and, if so, whether to respect the collective decision. In that framework, we show that inducing self-enforcing cooperation is more difficult than inducing cooperation under enforceability. Then, we characterize the optimal self-enforcing rule, which occasionally entails giving veto power to some countries, but not necessarily all.

4.1 Repeated Game

When decisions are not enforceable, considering the one-shot game $\Gamma_{NE}(v)$ is not sufficient, since countries have no incentive to abide by collective decisions in the fourth stage of the game if the game ends right away. A notion of self-enforcing cooperation can instead be introduced if we repeat the decision game. We thus consider the δ -discounted infinitely repeated game $\Gamma_{NE}^\delta(v)$. At each stage $T \in \mathbb{N}$, each country $i \in N$ decides whether to participate or not, $d_i^T \in \{0, 1\}$. Preferences for the reform proposed at stage T , $\mathbf{u}^T$, are drawn from μ , independently of the previous stages. Each country $i \in N$ reports a message $m_i^T \in \{0, 1\}$, observes the action plan $\hat{v}^T(\mathbf{m}^T)$, and takes an action $a_i^T \in \{0, 1\}$, which can differ from $\hat{v}(\mathbf{m}^T)$. At each stage, $\mathbf{d}^T$, $\mathbf{m}^T$, $\hat{v}^T(\mathbf{m}^T)$, and $\mathbf{a}^T$ are publicly observed. All countries are characterized by the same discount factor $\delta \in (0, 1]$.

For a given value of the discount factor δ , we say that a voting rule v is *self-enforcing* if there exists a perfect public equilibrium²⁵ of $\Gamma_{NE}^\delta(v)$ such that the cooperating profile is played at each stage of the game on the equilibrium path. We denote by $\mathcal{V}_\delta \subseteq \mathcal{V}$ the set of self-enforcing rules.

To construct such an equilibrium, we consider the profile of strategies for which each country follows the cooperative strategy absent any deviation and ceases to cooperate forever after any (publicly observed) deviation by a single country i , of the form $d_i^T = 0$ or $a_i^T \neq \hat{v}^T(\mathbf{m}^T)$ for some T .

We observe that, under such a profile, a deviation is profitable for a country when it is unfavorable to a reform approved by the committee. In that case, a deviation yields a short-term benefit for not complying at the current stage in addition to the stand-alone utility at

²⁵The notion of *public perfect equilibrium* is a generalization of subgame perfection for games of incomplete information, commonly employed to analyze games of the type of $\Gamma_{NE}^\delta(v)$ as, for instance, in Athey and Bagwell (2001) or Maggi and Morelli (2006).

the subsequent stages. Compared to the one-shot game, the repeated game thus creates an extra incentive to leave the union, which can only be mitigated by giving veto power to the country tempted to exit. To measure this new temptation to deviate, we define the maximal disutility that country i may suffer from a collective decision:

$$w_i^D = -\min \left\{ w \in \mathbb{R} \mid \mathbb{P}_\mu(u_i = w) > 0 \right\}.$$

Note that $w_i^D \geq w_i^- > 0$. We say that a country $i \in N$ satisfies the *compliance constraint* if

$$U_i(v) \geq U_i^\emptyset + \frac{1-\delta}{\delta} w_i^D.$$

Proposition 2. *A voting rule v is self-enforcing if and only if each country either has veto power and satisfies the participation constraint, or does not have veto power and satisfies the compliance constraint.*

The result establishes the equivalence between the notion of self-enforceability and a set of *endogenous* constraints. Indeed, the constraint that country i should satisfy under rule v is contingent on i having veto power under v . Moreover, we observe that the compliance constraints are more stringent than the participation constraints. As a result, if a voting rule is self-enforcing then it also satisfies the participation constraints. Note that the extreme case $\delta = 1$ coincides with the model of enforceable decisions.

4.2 Optimal Self-Enforcing Rules

We say that the voting rule v is *optimal self-enforcing* if it maximizes social welfare among self-enforcing rules; i.e., if it is a solution of $\max_{v \in \mathcal{V}_\delta} W(v)$.²⁶ From [Proposition 2](#), we immediately get that social welfare is lower under an optimal self-enforcing rule than under an optimal voting rule since $\mathcal{V}_\delta \subseteq \mathcal{V}_1$. The following theorem describes optimal self-enforcing rules.

Theorem 3. *For any optimal self-enforcing voting rule v^{**} there exists a system of weights $[\mathbf{w}^{**}; t^{**}]$ such that v^{**} is weighted and represented by $[\mathbf{w}^{**}; t^{**}]$. Countries for which the compliance constraint is not satisfied are strictly overweighted and have veto power. Countries for which the compliance constraint is binding are weakly overweighted. Countries for which the compliance constraint is satisfied but not binding receive their efficient weight and do not have veto power.*

Formally, for any optimal self-enforcing rule v^{**} , there exists a system $[\mathbf{w}^{**}; t^{**}]$ represent-

²⁶The existence of a solution is guaranteed when $\mathcal{V}_\delta$ is non-empty, as it can be checked that $\mathcal{V}_\delta$ is a closed subset of $[0, 1]^{2^N}$.

ing v^{**} , such that for all $i \in N$:

$$\begin{cases} U_i(v^{**}) < U_i^\emptyset + \frac{1-\delta}{\delta} w_i^D & \Rightarrow w_i^{**} > w_i^e \text{ and } i \in VE(v^{**}), \\ U_i(v^{**}) = U_i^\emptyset + \frac{1-\delta}{\delta} w_i^D & \Rightarrow w_i^{**} \geq w_i^e, \\ U_i(v^{**}) > U_i^\emptyset + \frac{1-\delta}{\delta} w_i^D & \Rightarrow w_i^{**} = w_i^e \text{ and } i \notin VE(v^{**}), \end{cases}$$

and where the threshold t^{**} satisfies:

$$t^{**} \geq \frac{\sum_{i \in N} w_i^{**} t_i^e}{\sum_{i \in N} w_i^{**}},$$

with an equality if no country has veto power, and a strict inequality otherwise.²⁷

Theorem 3 differs from **Theorem 2** in two main respects. First, the benchmark level of utility $U_i^\emptyset$ that separates overweighted countries from non-overweighted countries is increased by an additional $(1-\delta)w_i^D/\delta$. Countries that fall strictly below this augmented utility threshold are strictly overweighted, while countries that fall strictly above receive their efficient weight. Second, in contrast to **Theorem 2**, the benchmark utility also separates countries that benefit from veto power from countries that do not. This is illustrated in the example of **Section 5**, where the optimal self-enforcing rule grants veto power to country 1, but not to countries 2 to 5. Country 1's utility, $72/3^5 \approx 0.30$, falls below its augmented utility threshold of $16/3^5 + 2/5 \approx 0.47$, while countries 2 to 5's utility, $84/3^5 \approx 0.35$, falls above their augmented utility threshold of $32/3^5 + 1/5 \approx 0.33$. The fact that optimal self-enforcing rules may grant veto power to only a strict subset of countries is a major difference to **Maggi and Morelli (2006)**, in which either all countries have veto power or no country has veto power, and this stems from the generality of our model, which allows for heterogeneous countries.²⁸

5 An example

We now develop a simple example to illustrate the results of **Sections 3 and 4**. Consider a union of five countries which must decide, repeatedly, whether to impose embargoes on

²⁷Note again that the conditions given in **Theorem 3** are endogenous. However, it is possible to identify veto countries from exogenous conditions under more specific assumptions. Assume for instance that the utility distribution μ has uniform marginals $(\mu_i)_{i \in N}$, and that these marginals can be ranked, once normalized by $(U_i^\emptyset)_{i \in N}$, according to first-order stochastic dominance. Then, it can be shown that veto countries must have smaller ranks than non-veto countries.

²⁸In the conclusion of their paper, **Maggi and Morelli (2006)** allude to the possibility of such a result when countries are heterogeneous: "If nonegalitarian voting rules are allowed, however, there may be systems that do better than unanimity. One possibility would be to adopt a nonunanimous voting rule but give the "problematic" countries veto power.", page 1153.

tax havens. A sanction is only effective if implemented by all countries. Countries are uncertain about whether to support the embargoes. Country 1 is generally unfavorable and has a probability $1/3$ of supporting a sanction, while countries 2 to 5 are generally favorable and have a probability $2/3$ of supporting the sanctions. Preferences are independent across countries and across decisions. Ex post, if the embargo is effective, country 1 gets a utility of 1 if it is favorable and a disutility of 2 if it is unfavorable. In contrast, countries 2 to 5 get a utility of 2 if they are favorable and a disutility of 1 if they are unfavorable. If the embargo is not effective, all countries get a utility of 0. Countries' preferences are summarized in [Table 1](#).

	Probability of support	Utility if Favorable + Embargo Effective	Utility if Unfavorable + Embargo Effective	Utility if Embargo not Effective
Country 1	$1/3$	1	-2	0
Countries 2, 3, 4, 5	$2/3$	2	-1	0

Table 1: Countries' Preferences.

Under sovereignty, the embargo is implemented effectively only when all countries are favorable, which happens with a small probability of $16/3^5 \approx 0.07$. Ex ante, country 1 gets a utility of $U_1^\emptyset = 16/3^5$, while all other countries get a utility of $U_{2,3,4,5}^\emptyset = 32/3^5$ (from any single decision). Social welfare is equal to $144/3^5$.

In contrast, under cooperation, the embargo may be implemented effectively even if some countries are unfavorable since they must all accept the collective decision. Because all countries have the same stake in the collective decision, the efficient voting rule consists in adopting the embargo by simple majority ([Theorem 1](#)).²⁹ Ex ante, countries 2 to 5 get a utility of:

$$U_{2,3,4,5}^e = \frac{2}{3} \times 2 \times \mathbb{P}(\text{emb. adopted} \mid \text{fav.}) - \frac{1}{3} \times 1 \times \mathbb{P}(\text{emb. adopted} \mid \text{unfav.}) = \frac{228}{3^5} > \frac{32}{3^5},$$

and are thus much better off than under sovereignty. In contrast, country 1, which tends to disagree with the four other countries, is now much worse off:

$$U_1^e = \frac{1}{3} \times 1 \times \mathbb{P}(\text{emb. adopted} \mid \text{fav.}) - \frac{2}{3} \times 2 \times \mathbb{P}(\text{emb. adopted} \mid \text{unfav.}) = -\frac{120}{3^5} < \frac{16}{3^5},$$

which means it would not agree to cooperate ex ante. The only way to ensure cooperation is to give some additional voting power to country 1. The optimal voting rule³⁰ ([Theorem 2](#))

²⁹[Theorem 1](#) asserts that any efficient rule is represented by $[\mathbf{w}^e; t^e]$, where $\mathbf{w}^e = (1, 1, 1, 1, 1)$ and $t^e = 6/15$. When two countries are favorable to the embargo, there is a tie, and it is easy to see that the tie-breaking rule can be chosen arbitrarily (see, for instance, the proof of Theorem 1 in [Barberà and Jackson \(2006\)](#)). Therefore, the simple majority rule is efficient (ties are resolved by blocking a proposal supported by only two countries).

³⁰All optimal rules and optimal self-enforcing rules mentioned in the article have been checked by an algorithm.

consists in *overweighting* country 1 just enough, so that its participation constraint becomes binding: the embargo is adopted either if country 1 and at least one other country are in favor or if all but country 1 are in favor. This voting rule can be represented as a weighted majority rule with weights $\mathbf{w}^* = (3, 1, 1, 1, 1)$ and threshold $t^* = 1/2$.³¹ Country 1 gets exactly its stand-alone utility, $U_1^* = 16/3^5 = U_1^\emptyset$, while countries 2 to 5 now get a reduced utility of $U_{2,3,4,5}^* = 146/3^5$. Social welfare is reduced from $792/3^5$ (under the efficient decision rule) to $600/3^5$, but still much larger than under sovereignty.

If collective decisions cannot be enforced, countries may choose not to implement the embargo even if it has been approved collectively. As defined in [Section 4](#), a voting rule is self-enforcing if there exists a perfect public equilibrium of the associated infinitely repeated game such that the cooperating profile is played at each stage of the game on the equilibrium path. In order for a voting rule to be self-enforcing, the benefit of not implementing the embargo for unfavorable countries must be outweighed by the long-term cost of not sustaining cooperation, which yields an additional compliance constraint ([Proposition 2](#)) that is more stringent than the original participation constraint. In this example, the optimal voting rule (with enforcement) cannot be self-enforcing since country 1's participation constraint is already binding. Here, self-enforcement can only be achieved by granting veto power to country 1 ([Theorem 3](#)). For $\delta = 5/6$, the optimal self-enforcing rule is such that the embargo is adopted if and only if country 1 and at least two other countries are in favor. This voting rule can be represented as a weighted majority rule with weights $\mathbf{w}^{**} = (3, 1, 1, 1, 1)$ and threshold $t^{**} = 2/3$. Country 1 gets a utility of $U_1^{**} = 72/3^5 > U_1^\emptyset$, while countries 2 to 5 now get an even more reduced utility of $U_{2,3,4,5}^{**} = 84/3^5$. Social welfare is reduced from $600/3^5$ (under the optimal rule) to $408/3^5$. Note that even though collective decisions cannot be enforced, social welfare is still much larger under the optimal self-enforcing rule than under sovereignty. [Table 2](#) summarizes the rules and utilities obtained in each of the four considered benchmarks (to simplify the expressions, utilities are multiplied by a factor 3^5).

When the union is divided in two groups of identical countries, the algorithm finds an optimal (resp. optimal self-enforcing) rule among rules that are symmetric within groups. This is without loss of generality, by linearity of the problem. The algorithm is available upon request.

³¹According to [Theorem 2](#), the optimal rule is represented by $[\mathbf{w}^*; t^*]$ with $t^* = 10/21$. We note that the rule is equally well represented by $[\mathbf{w}^*; 1/2]$.

Benchmark	Sovereignty		Efficient	Optimal	Optimal Self-Enforcing
Symbol	$\emptyset$		e	*	**
Voting rule			Simple Majority	1 overweighted	1 overweighted + veto
w_1			1	3	3
$w_{2,3,4,5}$			1	1	1
t			1/2	1/2	2/3
U_1	$\propto$	16	−120	16	72
$U_{2,3,4,5}$	$\propto$	32	228	146	84
Welfare	$\propto$	144	792	600	408

Table 2: Summary of the example.

6 A Model of Apportionment

In this section, we apply our theory to a model where countries vary along a single dimension: their population size. At the heart of this more specific model lies a simple process of preference formation, where citizens' preferences are binary and unbiased ex-ante. We assume that these preferences are correlated within a given country, but independent across countries, and that each country's representative follows the national majority. As in [Koriyama et al. \(2013\)](#), we formalize these assumptions in a parsimonious manner: representatives' preferences are first drawn independently across countries, and citizens' preferences are then derived from their representative's preferences by assuming that a fraction $q > 1/2$ of the country's population shares the same preference as the representative. Our focus on preferences that are correlated within countries is akin to the *fixed-number of blocks* model of [Barberà and Jackson \(2006\)](#), and is in line with empirical studies on real elections ([Gelman et al., 2004](#)). As we explain below, the main insights of this section would also go through if we assumed that citizens' preferences were independent within each country.³²

³²In that case, it would not be meaningful to start from the preferences of the representatives as in [Koriyama et al. \(2013\)](#). Instead, we would start from citizens' preferences, assumed to be independent and identically distributed, and define the opinion of the country's representative as the preference of the majority. This model corresponds to the original framework proposed by [Penrose \(1946\)](#), and to the *fixed-size block model* of [Barberà and Jackson \(2006\)](#). In such a model, the fraction of the population agreeing with the country's representative, $\tilde{q}_i$, is a random variable. As p_i grows large, we obtain $\tilde{q}_i \approx 1/2$ by the law of large numbers, and $\tilde{q}_i - 1/2$ is approximately proportional to $1/\sqrt{p_i}$, by the central limit theorem.

6.1 Model

Under cooperation, proposals are determined exogenously. Ex ante, each country's representative has a probability $1/2$ of agreeing with any of the proposed reforms, independently of other countries. In each country, for any given reform, a (randomly chosen) fraction $q > 1/2$ of citizens agrees with the preference of its country's representative. If the reform ends up being implemented effectively (by all countries), favorable citizens get a utility of 1, while unfavorable citizens get a disutility of 1. If the reform is not adopted effectively, all citizens get a utility of 0. The probability distribution μ associated with this model is such that:

$$\forall i \in N, \quad \mathbb{P}_\mu(u_i = (2q - 1)p_i) = \mathbb{P}_\mu(u_i = -(2q - 1)p_i) = \frac{1}{2}.$$

The stake of country $i \in N$ is thus given by $w_i^+ = w_i^- = w_i^e/2 = (2q - 1)p_i$, and its efficient threshold is $t_i^e = 1/2$. By normalization, the efficient rule is thus represented by $[\mathbf{p}; 1/2]$.³³

Under sovereignty, each country now chooses independently which reforms to implement. In each country, for any given reform, a (randomly chosen) fraction $q > 1/2$ of citizens agrees with the reform. Citizens who are favorable get a reduced utility of $\frac{1}{e} < 1$, while citizens who are unfavorable get a disutility of $\frac{1}{e}$. The parameter $e > 1$ reflects the efficiency gain from cooperating.³⁴ Ex ante, the aggregate (stand-alone) utility of country i with population p_i is thus equal to:

$$U_i^\emptyset = qp_i \frac{1}{e} - (1 - q)p_i \frac{1}{e} = \frac{(2q - 1)p_i}{e}.$$

In this more specific model of apportionment, countries thus vary in their stakes w_i^e , which are proportional to their populations p_i , but are otherwise identical. Note that Assumption NEV boils down to $p_i < (\sum_{j \in N} p_j)/2$ for all $i \in N$, which means that any country accounts for less than half of the total population.

6.2 Optimal Voting Rules

We now obtain sharper predictions for the optimal voting rule: first, overweighted countries are those with the smallest populations; and second, these countries must be given the same voting weight.

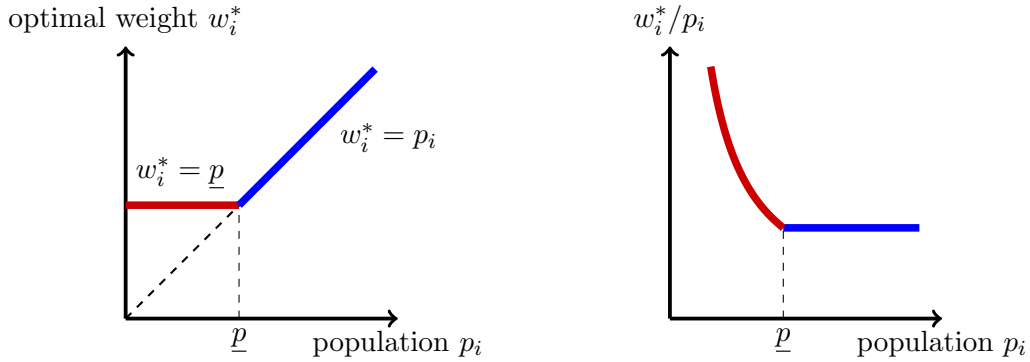
³³By contrast, under independent preferences (i.e. $\tilde{q}_i \approx 1/2$), the stake of country i is $w_i^+ = w_i^- = \mathbb{E}[2(\tilde{q}_i - 1)p_i]$, which is approximately proportional to $\sqrt{p_i}$, by the central limit theorem. In that case, the efficient rule is represented by $[(\sqrt{p_i})_{i \in N}; 1/2]$, this is the original insight of [Penrose \(1946\)](#).

³⁴Note that cooperation is assumed to increase the utility from a favorable reform *and* the disutility from an unfavorable one, by the same factor e , consistent with the view that the collective action goes further in the desired/undesired direction. In a previous version of the article, it was assumed that the disutility of an unfavorable reform was multiplied by a factor e^- , below or above 1. With that alternative (and more general) assumption, the subsequent [Theorem 4](#) remains valid, with a suitable adaptation of the threshold of the optimal rule.

Theorem 4. *In the model of apportionment, there exists $\underline{p} \in \mathbb{R}$ such that any optimal voting rule is a weighted majority rule represented by $[\mathbf{w}^*; 1/2]$, with $w_i^* = \max(p_i, \underline{p})$ for all $i \in N$.*

The intuition behind the result of [Theorem 4](#) is as follows. Start from the efficient weights, which are here proportional to the populations. Smaller countries have a smaller weight and therefore enjoy a smaller *probability of success*: their probability of agreeing with the collective decision is lower. In turn their utilities are lower, relative to their stand-alone utilities. If the efficient rule does not satisfy the participation constraints, then it can be adjusted by increasing the weights of smaller countries, up to the level at which their constraints bind.³⁵

Figure 2: Optimal weights (absolute and per capita) in the model of apportionment



The optimal apportionment rule is illustrated in [Figure 3](#), in absolute and per-capita terms. We first note that the distribution of weights is *degressively proportional*: weights increase with countries' populations (left panel's curve is increasing), but less than proportionally (right panel's curve is decreasing). A sizeable literature on apportionment has already argued in favor of this property,³⁶ but on different grounds than the one we put forth here. In particular, our argument focuses on the bottom of the distribution and supports overweighting small countries that may otherwise have almost no say in the collective decisions. By contrast, previous models recommend degressively proportional rules that have noticeable implications for medium to larger states, often with weights in the order of p^α with $1/2 \leq \alpha \leq 1$.³⁷

The requirement that smaller countries shall be given a minimal and equal representation is actually found explicitly in the Treaty of Lisbon, which specifies a set of constraints for the

³⁵The argument similarly applies in the case of independent preferences (i.e. $\tilde{q}_i \approx 1/2$). In that case, the optimal weights are equal to: $w_i^* = \max(\sqrt{p_i}, \underline{w})$. The main insights thus remain: smaller countries are overweighted, and they receive the same weight.

³⁶[Laslier \(2012\)](#) offers a review of the different arguments in favor of such rules.

³⁷For instance, in the model of [Barberà and Jackson \(2006\)](#), the optimal α is approximately equal to $1/2$ in the *fixed-size block model*, and equal to 1 in the *fixed-number-of-blocks model*. See also [Beisbart and Bovens \(2007\)](#).

composition of the European Parliament.³⁸ Indeed, article 14.2 states that “representation of citizens [in the European Parliament] shall be degressively proportional, with a minimum threshold of six members per Member State” ([Treaty of Lisbon, 2007a](#)). Our article thus offers a theoretical rationale for such a minimal representation threshold.

Finally, the apportionment formula proposed here combines in a simple manner the notions of proportionality and equality, which is reminiscent of several prominent examples. In particular, the overweighting of smaller states echoes the distribution of seats in the U.S. Electoral College,³⁹ and the optimal weights per capita observed in the right panel of [Figure 3](#) mirror the actual ones exhibited in [Figure 1](#). The eight smaller states are allocated the same number of three seats,⁴⁰ representing 4.5% of the seats for only 1.9% of the total population. The same type of apportionment formula has also been proposed for the allocation of seats in the European Parliament, under the name of *Cambridge Compromise*.⁴¹

6.3 Heterogeneous Gains

Our basic model of apportionment assumes that citizens in all countries benefit from the same level of efficiency gain from cooperation. However, it might be sensible in some applications to assume that efficiency gains decrease with a country’s population, larger countries being usually less dependent on international cooperation than smaller countries. In order to investigate the implications of this more general assumption, we introduce a *model of apportionment with heterogeneous gains*, where each country is characterized by a possibly different efficiency gain $e_i \in (1, +\infty)$.

In this more complex model, the efficient rule remains the same (the weighted rule $[\mathbf{p}; 1/2]$) and the stand-alone utilities become $U_i^\emptyset = \frac{(2q-1)p_i}{e_i}$. The reasoning conducted in the proof of [Theorem 4](#) does not extend directly, but we obtain a similar characterization of the optimal weights by resorting to an approximation. The difficulty lies in precisely estimating the expected utility $U_i(v)$, which is proportional to the product of country i ’s population p_i to its *Banzhaf voting power* under rule v , that we denote $BZ_i(v)$.⁴² We use the Penrose limit approximation ([Lindner and Machover, 2004](#)), stating that for n large enough, the ratio of any two countries’ voting power under a weighted voting rule represented by $[\mathbf{w}; 1/2]$ can be

³⁸For a discussion of the application to the allocation of seats in a federal parliament, rather than voting weights in a federal council, see [Koriyama et al. \(2013\)](#).

³⁹Each state is allocated a baseline of two electors plus a number of electors proportional to its population.

⁴⁰Alaska, Delaware, District of Columbia, Montana, North Dakota, South Dakota, Vermont, and Wyoming.

⁴¹The Cambridge Compromise was the result of an academic initiative by the European Parliament, which aimed at formulating a transparent and fair allocation of the seats in the European Parliament. The proposed allocation is based on a similar base + prop formula as in the U.S. Electoral College, whereby each country is allocated a base of six seats plus a number of seats proportional to its population. See [Grimmett \(2012\)](#).

⁴²The Banzhaf index ([Banzhaf, 1964](#)) is a standard measure of ex ante voting power in committees, given by the following formula: $BZ_i(v) = \frac{1}{2^{n-1}} \sum_{S \subseteq N, i \in S} [v(S) - v(S \setminus \{i\})]$.

approximated by the ratio of their respective voting weights.⁴³

Proposition 3. *In the model of apportionment with heterogeneous gains, under the Penrose limit approximation, there exists $a \in \mathbb{R}_+$ such that any optimal voting rule is a weighted majority rule represented by $[\mathbf{w}^*; 1/2]$, with $w_i^* = \max\left(p_i, \frac{a}{e_i}\right)$ for all $i \in N$.*

Under the Penrose approximation, the optimal system of weights $[\mathbf{w}^*; 1/2]$ thus satisfies:

$$w_i^* = \begin{cases} \frac{a}{e_i} & \text{if } p_i e_i < a \\ p_i & \text{if } p_i e_i \geq a. \end{cases}$$

The optimal voting rule is now such that overweighted countries have a smaller population and/or efficiency gain. Furthermore, the voting weights of overweighted countries are inversely proportional to their efficiency gain e_i . When all countries share the same gain, i.e. $e_i = e$ for all i in N , [Proposition 3](#) coincides with the (exact) characterization of the optimal voting weights given in [Theorem 4](#). By contrast, when all countries have the same population, i.e. $p_i = p$ for all i in N , overweighted countries are those with the smallest efficiency gains.

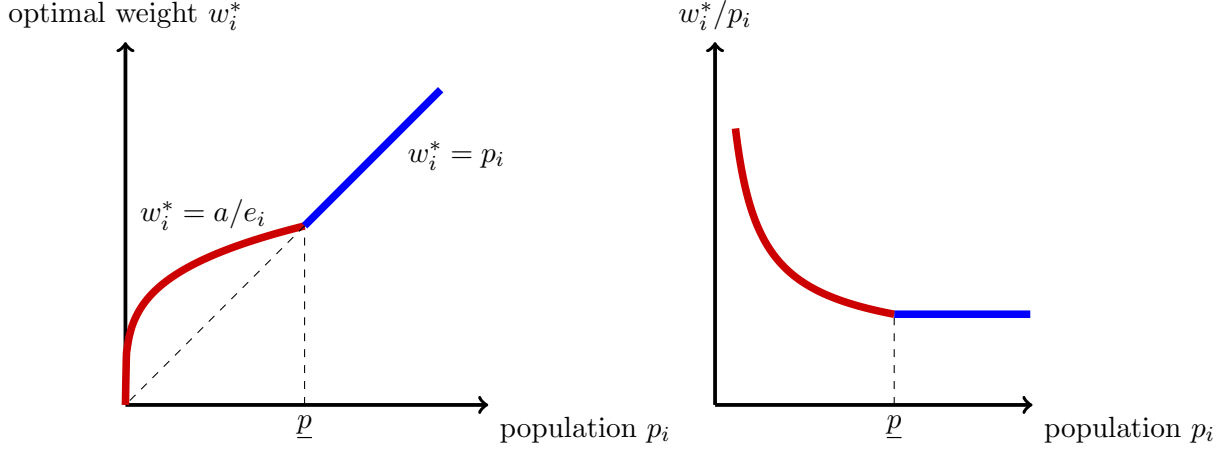
As an application of [Proposition 3](#), consider the case where citizens in smaller countries gain more from cooperation than citizens in larger countries: there exists a decreasing function $e : \mathbb{R}_+ \mapsto (1, +\infty)$ such that $e_i = e(p_i)$ for each $i \in N$. Then, as long as efficiency gains are decreasing less than linearly in p , i.e. $pe(p)$ is increasing in p , [Proposition 3](#) implies that overweighted countries still have a smaller population size than non overweighted countries under the optimal voting rule. The main conclusion of [Theorem 4](#), i.e. smaller countries being overweighted at the optimum, remains true even though efficiency gains are no longer constant across countries. Intuitively, smaller countries still need to be overweighted even though their citizens benefit more from cooperation because larger countries receive a larger voting weight under the efficient voting rule. Larger countries may only be overweighted at the optimum under the more extreme assumption of efficiency gains decreasing more than linearly with the population. We illustrate the shape of the optimal weights when $e_i = e(p_i) \propto p_i^{-1/4}$ for each $i \in N$ in [Figure 3](#).

6.4 Double majority and Bicameralism

As we show in [Theorem 4](#) and [Proposition 3](#), meeting the participation constraints require departing from proportional apportionment by overweighting a group of (generally) smaller countries. In practice, several prominent organizations keep proportional weights, but add a

⁴³The approximation applies well when the number of countries exceeds 15 ([Chang et al., 2006](#)), as in the empirical examples we mention in the introduction (the EU and the US).

Figure 3: Optimal weights (absolute and per capita) in the model of apportionment with heterogeneous gains for $e_i \propto p_i^{-1/4}$



requirement on the number of favorable countries for a proposal to be accepted. This is the case for example at the Council of the European Union, where a proposal needs the support of at least 55% of countries, representing at least 65% of the total population of the EU. The US bicameral legislative system, although of a different nature, can also be interpreted as a form of double majority if one assumes that representatives from the same state vote identically. A successful proposal then requires the support of at least 26 states (through the Senate), representing at least 50% of the population (through the House of Representatives).

Formally, a rule v is an *efficient rule with size threshold* if there exist an efficient rule v^e and a *size threshold* $k \in \{0, \dots, n\}$ such that, for any profile of votes $\mathbf{m} = (m_i)_{i \in N} \in \{0, 1\}^N$,

$$\begin{cases} \#\{i \mid m_i = 1\} \geq k & \Rightarrow & v(\mathbf{m}) = v^e(\mathbf{m}) \\ \#\{i \mid m_i = 1\} < k & \Rightarrow & v(\mathbf{m}) = 0. \end{cases}$$

We write $v = v^e \oplus k$. Note that such a rule v is not a weighted rule in general, and thus cannot be optimal.⁴⁴ As we show in the following proposition, even though these rules are not optimal, they are more likely to satisfy the participation constraints than the efficient voting rule itself.

Proposition 4. *In the apportionment model, assuming that no two countries have the same population and that one country is more populated than a group of two other countries, there*

⁴⁴Whenever the vector of populations is far from being collinear to the constant vector, one may apply the characterization of weighted voting in Taylor and Zwicker (1992) to show that an efficient rule with (positive) size threshold is not weighted. For instance, if $\mathbf{p} = (1, 10, 10, 10, 10, 100, 100)$, the rule $v = v^e \oplus 4$ is not weighted as $v(1236) = v(1457) = 1$ but $v(12345) = v(167) = 0$, violating the property of *trade robustness*. A similar argument implies that the rule of the Council of the EU is not weighted.

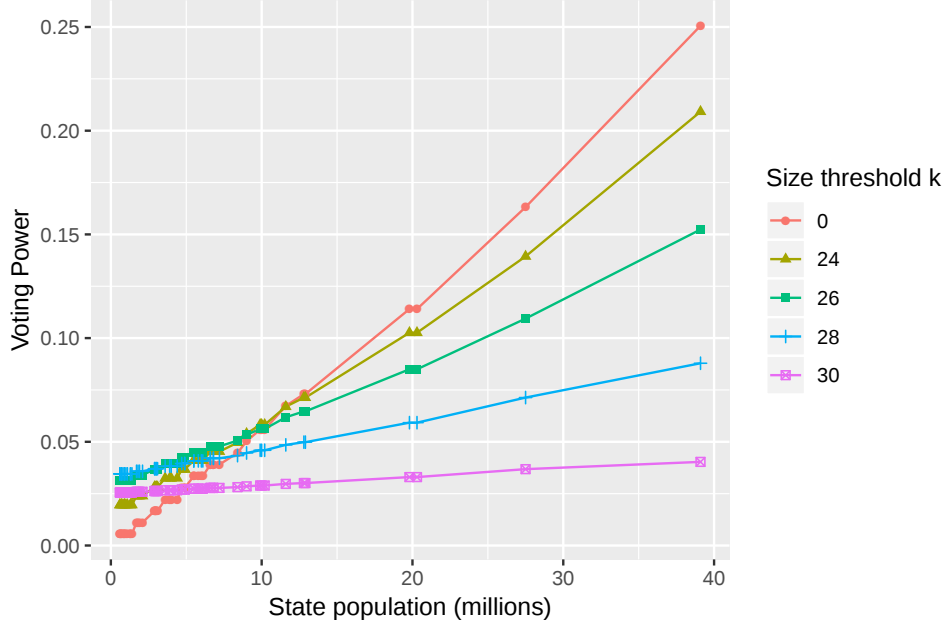
exist two thresholds $\underline{e}, \bar{e} \in (1, +\infty)$, with $\underline{e} < \bar{e}$, such that:

- (i) if $e \geq \bar{e}$, then the efficient voting rule satisfies the participation constraints.
- (ii) if $\underline{e} \leq e < \bar{e}$, then the efficient voting rule does not satisfy the participation constraints but there exists $k^* > 0$ such that the efficient rule with size threshold k^* does.
- (iii) if $e < \underline{e}$, then none of the efficient rules with size threshold satisfy the participation constraints.

The main message of [Proposition 4](#) is contained in statement (ii): for intermediate efficiency gains, adding a certain size threshold k^* to the efficient rule allows to satisfy the participation constraints. Note that the assumption under which the statement holds is very mild: it simply requires that countries' populations are not too similar from each other.⁴⁵

To illustrate the effect of the size threshold on states' ability to satisfy their participation constraints, we plot in [Figure 4](#) the voting power of a U.S. state's representative as a function of the state's population. Note that the voting power is here proportional to $U_i(v)/U_i^\emptyset$, the ratio that determines whether a state satisfies its participation constraint. We plot the corresponding graphs for several rules: the efficient rule v^e represented by $[\mathbf{p}; 1/2]$, and the efficient rules with size threshold k , $v^e \oplus k$, for $k \in \{24, 26, 28, 30\}$.

Figure 4: Voting power of a state's representative, as a function of the state's population, for efficient rules with varying size thresholds, for the U.S. (2017 population figures)



We observe in [Figure 4](#) the two effects associated with the introduction of a size threshold.

⁴⁵For instance, it is satisfied in the EU as Germany is more populated than Malta and Luxembourg taken together, and in the U.S. as California is more populated than Alabama and Arkansas taken together.

Starting from the efficient voting rule ($k = 0$), introducing a small size threshold induces first a distributional effect: smaller countries gain with respect to the efficient benchmark while larger countries lose.⁴⁶ This effect is at the heart of [Proposition 4](#) because increasing the utility of the smallest country makes all countries more likely to satisfy the participation constraints. As the size threshold increases, a second effect comes into play: the rule becomes more stringent, making proposals more difficult to pass, which may ultimately decrease the utility of all countries. This is for example the case in the U.S. legislature when the size threshold goes from 26 to 30.

We conclude that, while not as efficient as the optimal voting rules, double majority rules similarly depart from efficiency by accommodating smaller countries.⁴⁷

6.5 Optimal Self-Enforcing Rules

Finally, we investigate the shape of optimal self-enforcing rules in the model of apportionment. We also obtain sharper predictions: either no country has veto power or all countries have it, and we can map these two cases on a graph parametrized by the efficiency gain e and the discount factor δ .

Theorem 5. *In the model of apportionment, any optimal self-enforcing rule is either the unanimity rule or a weighted majority rule for which no country has veto power. There exist a threshold $\hat{e} > 1$, and two non-increasing functions $\delta^c, \delta^{eff} : (1, +\infty) \rightarrow \mathbb{R}_+$, such that $\delta^c(e) \leq \delta^{eff}(e)$ for all $e > 1$, $\lim_{e \rightarrow \infty} \delta^{eff}(e) < 1$, and:*

- (i) *if $\delta \geq \delta^{eff}(e)$, any optimal self-enforcing rule is an efficient weighted majority rule,*
- (ii) *if $\delta^c(e) \leq \delta < \delta^{eff}(e)$, any optimal self-enforcing rule is a weighted majority rule, with overweighting of small countries,*
- (iii) *if $\delta < \delta^c(e)$ and $e \geq \hat{e}$, the optimal self-enforcing rule is the unanimity rule,*
- (iv) *if $\delta < \delta^c(e)$ and $e < \hat{e}$, there is no self-enforcing rule.*

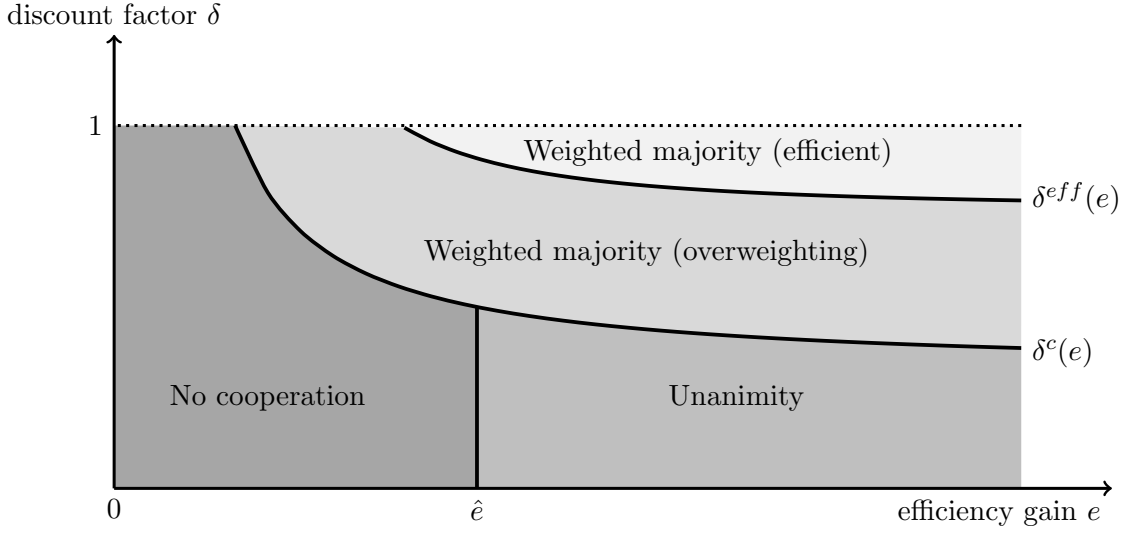
Moreover, for $\delta^c(e) \leq \delta < \delta^{eff}(e)$, there exists a minimal weight $\underline{p}(e, \delta)$, non-increasing in both e and δ , such that any optimal self-enforcing rule is represented by $[\mathbf{w}^{**}; 1/2]$, defined by $w_i^{**} = \max(p_i, \underline{p}(e, \delta))$ for all $i \in N$.

[Theorem 5](#) defines four regions in the space (e, δ) that yield different (or no) optimal self-enforcing rules, as represented in [Figure 5](#) below.

⁴⁶Formally, we can show that the difference in utility associated to the introduction of the size threshold, relative to the stand-alone utility, is decreasing with a country's population.

⁴⁷One may wonder what is the loss of welfare associated to an efficient rule with size threshold, compared to the optimal rule. As an exercise, we computed for each size threshold $k > 10$ the minimal level of efficiency gain e for which $v^e \oplus k$ satisfies the participation constraints and the optimal rule associated to e . Numerically, we observe that the optimal rule always fills at least 66.8% of the welfare gap between $v^e \oplus k$ and the efficient rule v^e .

Figure 5: Optimal self-enforcing rule in the model of apportionment



The figure can be interpreted either horizontally or vertically. First, the line $\delta = 1$ depicts the results we obtain for enforceable decisions. If the efficiency gain e is too small, there is no rule inducing cooperation. If the efficiency gain e is large enough, the efficient voting rule induces cooperation and is therefore optimal. However, for intermediate values of e , the optimal rule involves overweighting small countries, and the extent to which small countries are overweighted decreases with e .

Reading [Figure 5](#) vertically reveals how [Theorem 5](#) extends the main result of [Maggi and Morelli \(2006\)](#). In that article, countries are homogeneous, and there exists a threshold $\bar{\delta}$, below which the optimal self-enforcing rule is the unanimity, and above which the optimal self-enforcing rule is the (efficient) majority rule. In our model, for $e \geq \hat{e}$, there are two thresholds: $\delta^{eff}(e)$ and $\delta^c(e)$. As in the homogeneous model, the optimal self-enforcing rule is the efficient rule when the discount factor is high ($\delta \geq \delta^{eff}(e)$), and it is the unanimity rule when the discount factor is low ($\delta < \delta^c(e)$). What is new here is that we obtain a region of intermediate values of the discount factor ($\delta^c(e) \leq \delta < \delta^{eff}(e)$), for which the optimal self-enforcing rule is an inefficient weighted majority rule, with overweighting of small countries. Moreover, the extent to which small countries are overweighted decreases with the efficiency gain e and with the discount factor δ .

7 Conclusion

We have studied the design of voting rules in representative committees when decisions are binary and cooperation is voluntary. In contrast to the efficient voting rule, which assigns to each

country a voting weight proportional to its stake, the optimal voting rule sometimes involves overweighting certain countries, namely those that have the lowest *endogenous* incentive to cooperate. In the apportionment case, where the heterogeneity reduces to the population size and countries have identical ex-ante preferences, the optimal voting rule assigns an equal (and larger than proportional) weight to smaller countries, while larger countries keep their efficient weight. The theory thus provides a new rationale for the use of degressive proportional apportionment rules. When collective decisions are not enforceable, and cooperation must be agreed on repeatedly, voting rules must satisfy stronger compliance constraints, thus reducing social welfare. At the optimum, countries that do not satisfy compliance constraints benefit from veto power. In the apportionment case, the optimal self-enforcing rule is either unanimous (thus giving veto power to all countries), or such that no country enjoys veto power.

We have assumed throughout the article that cooperation was only beneficial if all countries participated, the “pure” collective action case (Maggi and Morelli, 2006).⁴⁸ Note that all of our results on enforceable decisions remain valid for more general forms of cooperation gains. Indeed, as participation is decided at the unanimity in our model, either all countries cooperate or all remain sovereign. Nevertheless, inducing the cooperation of all countries may prove costly, as in our main example, in which the welfare drops by 24% from the efficient to the optimal rule. When the cooperation gains are “impure”, this cost may be alleviated by allowing a strict subset of countries to cooperate while the others remain sovereign.⁴⁹ Extending the model to such flexible forms of cooperation seems a promising avenue for further research.

We have taken a “constrained” normative approach where a benevolent social planner looks for the welfare-maximizing voting rule while accounting for the effect of the rule on countries’ participation and compliance. A natural alternative would be to consider a fully positive approach where countries bargain over the voting rule. The first participation stage in our game (as described in Section 2) could then be replaced by a bargaining stage where the voting rule (to be used in the subsequent stages) is determined by the following Nash program:

$$\max_{v \in \mathcal{V}} \prod_{i \in N} (U_i(v) - U_i^\emptyset).$$

⁴⁸The “pure” collective action case applies whenever the failure to cooperate of even one country can jeopardize the success of the considered policy. Examples include climate change agreements, immigration policy, or agreements not to harbor terrorists.

⁴⁹This flexible form of cooperation, known as *enhanced cooperation*, was first introduced by the European Union in the Treaty of Amsterdam (1997). It allows a subset of nine or more countries to move forward with cooperation without the consent of the other countries.

The rest of the game would follow in the same fashion. In contrast to the original model, the voting rule is not chosen exogenously so as to maximize social welfare but is obtained endogenously as the outcome of a Nash bargaining between countries in the initial stage. As we explain in [Section A.9](#) in the Appendix, one can show that under enforceable decisions the outcome is a weighted voting rule such that countries that are closer to their stand-alone utility under the bargained voting rule are relatively more overweighted. The result captures a similar intuition as [Theorem 2](#). Countries that gain less from cooperation (relative to their outside option) under the bargained voting rule are more overweighted. For instance, in the example of [Section 5](#), the bargained rule can be represented as a weighted majority rule with weights (4, 1, 1, 1, 1) and a threshold of $1/2$.⁵⁰ Country 1's utility increases from $16/3^5$ (under the optimal voting rule) to $38.8/3^5$, while the utility of countries 2 to 5 decreases from $146/3^5$ to $123.2/3^5$. Social welfare decreases from $600/3^5$ to $531.6/3^5$. More generally, overweighted countries under the optimal voting rule are necessarily better off under the bargained voting rule since their utility under the optimal voting rule is equal to their stand-alone utility. However, social welfare is necessarily lower since the bargained voting rule satisfies all participation constraints. Finally, note that the bargained voting rule is more difficult to compute algebraically because it is defined as the solution of a non-linear maximization problem.

References

- Alesina, A., Angeloni, I., and Etro, F. (2005). International unions. *American Economic Review*, 95(3):602–615.
- Alesina, A. and Spolaore, E. (1997). On the number and size of nations. *The Quarterly Journal of Economics*, 112(4):1027–1056.
- Athey, S. and Bagwell, K. (2001). Optimal collusion with private information. *RAND Journal of Economics*, pages 428–465.
- Azrieli, Y. and Kim, S. (2014). Pareto efficiency and weighted majority rules. *International Economic Review*, 55(4):1067–1088.
- Balinski, M. L. and Young, H. P. (1982). *Fair representation: Meeting the ideal of one man, one vote*. Yale University Press.
- Banzhaf, J. F. (1964). Weighted voting doesn't work: A mathematical analysis. *Rutgers L. Rev.*, 19:317.

⁵⁰ And such that the embargo is approved with probability 0.65 if all but country 1 are in favor and with probability 0.4 if only country 1 is in favor.

- Barberà, S. and Jackson, M. O. (2004). Choosing how to choose: Self-stable majority rules and constitutions. *The Quarterly Journal of Economics*, 119(3):1011–1048.
- Barberà, S. and Jackson, M. O. (2006). On the weights of nations: Assigning voting power to heterogeneous voters. *Journal of Political Economy*, 114(2):317–339.
- Beisbart, C. and Bovens, L. (2007). Welfarist evaluations of decision rules for boards of representatives. *Social Choice and Welfare*, 29(4):581–608.
- Bouton, L., Llorente-Saguer, A., and Malherbe, F. (2018). Get rid of unanimity rule: The superiority of majority rules with veto power. *Journal of Political Economy*, 126(1):107–149.
- Bown, C. (2004). On the economic success of gatt/wto dispute settlement. *The Review of Economics and Statistics*, 86:811–823.
- Brighouse, H. and Fleurbaey, M. (2010). Democracy and proportionality. *Journal of Political Philosophy*, 18(2):137–155.
- Casella, A. (1992). Participation in a currency union. *American Economic Review*, 82(4):847–63.
- Chang, P.-L., Chua, V. C., and Machover, M. (2006). Ls penrose’s limit theorem: tests by simulation. *Mathematical Social Sciences*, 51(1):90–106.
- Crémer, J. and Palfrey, T. R. (1996). In or out? centralization by majority vote. *European Economic Review*, 40(1):43–60.
- Demange, G. (2017). The stability of group formation. *Revue d’économie politique*, 127(4):495–516.
- Felsenthal, D. S. and Machover, M. (1999). Minimizing the mean majority deficit: The second square-root rule. *Mathematical Social Sciences*, 37(1):25–37.
- Gelman, A., Katz, J. N., and Bafumi, J. (2004). Standard voting power indexes do not work: an empirical analysis. *British Journal of Political Science*, 34(04):657–674.
- Grimmett, G. R. (2012). European apportionment via the cambridge compromise. *Mathematical Social Sciences*, 63(2):68–73.
- Koriyama, Y., Laslier, J.-F., Macé, A., and Treibich, R. (2013). Optimal apportionment. *Journal of Political Economy*, 121(3):584–608.

- Kurz, S., Maaser, N., and Napel, S. (2017). On the democratic weights of nations. *Journal of Political Economy*, 125(5):1599–1634.
- Laslier, J.-F. (2012). Why not proportional? *Mathematical Social Sciences*, 63(2):90–93.
- Lindner, I. and Machover, M. (2004). Ls penrose’s limit theorem: proof of some special cases. *Mathematical Social Sciences*, 47(1):37–49.
- Loeper, A. (2011). Coordination in heterogeneous federal systems. *Journal of Public Economics*, 95(7):900–912.
- Maggi, G. and Morelli, M. (2006). Self-enforcing voting in international organizations. *The American Economic Review*, 96(4):1137–1158.
- Moravcsik, A. (1998). *The choice for Europe: social purpose and state power from Messina to Maastricht*. Cornell University Press.
- Penn, E. M. (2016). Engagement, disengagement, or exit: A theory of equilibrium associations. *American Journal of Political Science*, 60(2):322–336.
- Penrose, L. S. (1946). The elementary statistics of majority voting. *Journal of the Royal Statistical Society*, 109(1):53–57.
- Rae, D. W. (1969). Decision-rules and individual values in constitutional choice. *American Political Science Review*, 63(1):40–56.
- Renou, L. (2011). Group formation and governance. *Journal of Public Economic Theory*, 13(4):595–630.
- Taylor, A. and Zwicker, W. (1992). A characterization of weighted voting. *Proceedings of the American Mathematical Society*, 115(4):1089–1094.
- Treaty of Lisbon (2007a). Consolidated version of the treaty on european union - title i: Common provisions.
- Treaty of Lisbon (2007b). Consolidated version of the treaty on the functioning of the european union - part one: Principles.
- Wilcox, F. O. (1945). The yalta voting formula. *American Political Science Review*, 39(5):943–956.

A Appendix: Proofs

A.1 Proof of Proposition 1

In the game $\Gamma_E(v)$, the only relevant beliefs are given by the prior μ at the first stage, and by $\mu(u_i, \cdot)$ for a player i at the third stage. Thus, the only condition to check to see if the cooperating profile is a perfect Bayesian equilibrium is that of sequential rationality.

First, it is clear that sending $m_i = \mathbb{1}_{u_i > 0}$ is always rational in the third stage, as v is non-decreasing. Second, playing $d_i = 1$ at the first stage is rational if and only if the expected outcome of the game under cooperation is no worse than under sovereignty (obtained if $d_i = 0$). The result follows.

A.2 Proof of Theorem 2

In this proof, and in the subsequent proofs, we abuse notation and write $v(M)$ for $v(\mathbf{m})$, where $M \subseteq N$ denotes the coalition of countries that vote in favor of the proposal: $M = \{i \in N | m_i = 1\}$.

Let $\mathcal{V}^* = \arg \max_{v \in \mathcal{V}_1} W(v)$ be the set of optimal rules. The set $\mathcal{V}^*$ is convex: for any $v, v' \in \mathcal{V}^*$ and $\alpha \in (0, 1)$, it is clear that $\alpha v + (1 - \alpha)v' \in \mathcal{V}^*$, by linearity of the objective function and of the constraints that define $\mathcal{V}^*$. Let v^* be an optimal rule such that:

$$v^* \in \arg \max \#\{M \subseteq N | v(M) \in (0, 1)\}.$$

In words, the rule v^* maximizes the recourse to probabilistic tie-breaking among optimal rules. We claim that, if v^* is represented by $[\mathbf{w}; t]$, then any optimal rule $v' \in \mathcal{V}^*$ is also represented by $[\mathbf{w}; t]$. To see that, assume that v^* is represented by $[\mathbf{w}; t]$, but that $v' \in \mathcal{V}^*$ is not. Then, without loss of generality, we can assume that there exists $M \subseteq N$, such that $\sum_{i \in M} w_i > t \sum_{i \in N} w_i$, but $v'(M) < 1$. We have by assumption $v^*(M) = 1$. We obtain a contradiction by considering the rule $v'' = (v^* + v')/2$, as $v'' \in \mathcal{V}^*$, but $\#\{M \subseteq N | v''(M) \in (0, 1)\} > \#\{M \subseteq N | v^*(M) \in (0, 1)\}$. In the remainder of the proof, we solely focus on v^* , as we know that any representation of v^* will be a representation of any optimal rule.

We claim that v^* is also a solution of the following maximization problem (note that v is not constrained to be non-decreasing in that problem) :

$$(\mathcal{P}) : \left| \begin{array}{l} \max \sum_{i \in N} U_i(v) \\ \text{s.t. } \forall i \in N, \quad U_i(v) \geq U_i^\emptyset. \end{array} \right|$$

It suffices to show that any solution of $(\mathcal{P})$ is non-decreasing. For that, let v be a solution of $(\mathcal{P})$ such that $v(M) > v(M')$ with $M \subset M'$. It is straightforward that the rule v' , obtained from v by permuting M and M' , will increase the expected utility of countries in $M' \setminus M$, while

decreasing the expected utility of no country. Hence, v' improves the welfare and satisfies the constraints, a contradiction.

Let P be the probability distribution on the coalitions M of countries favoring the reform (under truthful voting), induced by the probability distribution μ . Formally:

$$\forall M \subseteq N, \quad P(M) = \mathbb{P}_\mu(\{i | u_i > 0\} = M).$$

By assumption, we have that for all $M \subseteq N$, $P(M) > 0$. As countries' utilities are independent, the expected utility of a country $i \in N$ under a rule v writes

$$\begin{aligned} U_i(v) &= \mathbb{E}_\mu \left[v((\mathbb{1}_{u_j > 0})_{j \in N}) u_i \right] \\ &= \sum_{M, i \in M} P(M) v(M) \mathbb{E}_\mu [u_i | u_i > 0] + \sum_{M, i \notin M} P(M) v(M) \mathbb{E}_\mu [u_i | u_i < 0] \\ &= \sum_{M, i \in M} P(M) v(M) w_i^+ - \sum_{M, i \notin M} P(M) v(M) w_i^-. \end{aligned}$$

The Lagrangian of the problem $(\mathcal{P})$ writes

$$\mathcal{L}(v) = \sum_{i \in N} U_i(v) + \sum_{i \in N} \lambda^i [U_i(v) - U_i^\emptyset] + \sum_{M \subseteq N} [\eta^M v(M) + \nu^M (1 - v(M))].$$

Its partial derivative with respect to $v(M)$ (one of the 2^n variables) is

$$\frac{\partial \mathcal{L}}{\partial v(M)}(v) = P(M) \left(\sum_{i \in M} (1 + \lambda^i) w_i^+ - \sum_{i \notin M} (1 + \lambda^i) w_i^- \right) + \eta^M - \nu^M.$$

As v^* is a solution of $(\mathcal{P})$, we can apply the first-order conditions of the Kuhn-Tucker theorem (the constraints are affine functions). There exist non-negative coefficients $(\lambda^i, \eta^M, \nu^M)_{i \in N, M \subseteq N}$ such that

$$\left\{ \begin{array}{ll} (i) & \forall M \subseteq N, \quad \frac{\partial \mathcal{L}}{\partial v(M)}(v^*) = 0 \\ (ii) & \forall i \in N, \quad \lambda^i [U_i(v^*) - U_i^\emptyset] = 0 \\ (iii) & \forall M \subseteq N, \quad \eta^M v^*(M) = 0 \\ (iv) & \forall M \subseteq N, \quad \nu^M (1 - v^*(M)) = 0. \end{array} \right.$$

By the last two lines, η^M and ν^M cannot be simultaneously positive. Therefore, we have

$$\left\{ \begin{array}{lll} \eta^M - \nu^M < 0 & \Rightarrow \nu^M > 0 & \Rightarrow v^*(M) = 1 \\ \eta^M - \nu^M > 0 & \Rightarrow \eta^M > 0 & \Rightarrow v^*(M) = 0. \end{array} \right.$$

By (i), and the formula for the derivative of the Lagrangian, we have that

$$\begin{aligned}
\eta^M - \nu^M < 0 &\Leftrightarrow \sum_{i \in M} (1 + \lambda^i) w_i^+ > \sum_{i \notin M} (1 + \lambda^i) w_i^- \\
&\Leftrightarrow \sum_{i \in M} (1 + \lambda^i) (w_i^+ + w_i^-) > \sum_{i \in N} (1 + \lambda^i) w_i^- \\
&\Leftrightarrow \sum_{i \in M} (1 + \lambda^i) w_i^e > \frac{\sum_{i \in N} (1 + \lambda^i) w_i^-}{\sum_{i \in N} (1 + \lambda^i) w_i^e} \sum_{i \in N} (1 + \lambda^i) w_i^e.
\end{aligned}$$

Setting $w_i^* = (1 + \lambda^i) w_i^e$ and

$$t^* = \frac{\sum_{i \in N} (1 + \lambda^i) w_i^-}{\sum_{i \in N} (1 + \lambda^i) w_i^e} = \frac{\sum_{i \in N} w_i^* t_i^e}{\sum_{i \in N} w_i^*},$$

we get that:

$$\forall M \subseteq N, \quad \begin{cases} \sum_{i \in M} w_i^* > t^* \sum_{i \in N} w_i^* & \Rightarrow v^*(M) = 1 \\ \sum_{i \in M} w_i^* < t^* \sum_{i \in N} w_i^* & \Rightarrow v^*(M) = 0. \end{cases}$$

Therefore we conclude that v^* , and thus any optimal rule in $\mathcal{V}^*$, is weighted and can be represented by $[\mathbf{w}^*; t^*]$. Moreover, by application of the Kuhn-Tucker conditions, we have that $U_i(v^*) = U_i^\emptyset \Rightarrow \lambda^i \geq 0 \Rightarrow w_i^* \geq w_i^e$ and $U_i(v^*) > U_i^\emptyset \Rightarrow \lambda^i = 0 \Rightarrow w_i^* = w_i^e$.

A.3 Proof of **Proposition 2**

The proof is divided in two steps. First, we construct a profile of the repeated game, with cooperation at each stage on the equilibrium path, and we show that it is a perfect public equilibrium if the constraints of the proposition are satisfied. Second, we show that if one constraint is not satisfied, no perfect public equilibrium can sustain cooperation on the equilibrium path.

Step 1: Let v be a voting rule that satisfies the constraints of **Proposition 2**. We consider the following profile of strategies:

- At each stage T , play the cooperative profile of the game $\Gamma_{NE}(v)$.
- If exactly one country is observed to deviate at time T (either $d_i^T = 0$ or $a_i^T \neq \hat{v}^T(\mathbf{m}^T)$ for some $i \in N$), then all countries choose to remain sovereign in any subsequent stage $T' > T$.

Consider a potential deviation from the previous profile, for some country i , and assume it is a (strict) best reply. We note (d_i^0, m_i^0, a_i^0) the first stage of this deviation.

If $d_i^0 = 0$, the deviation yields a stage payoff $U_i^\emptyset$, and a future payoff $U_i^\emptyset$ (given the trigger strategies employed by other players). As $U_i(v) \geq U_i^\emptyset$ (each country satisfies at least the participation constraint), this deviation is not profitable.

If $d_i^0 = 1$, and the deviation is such that $\exists u_i \in \mathbb{R}, m_i^0(u_i) \neq \mathbb{1}_{u_i > 0}$, then it is (weakly) dominated by the strategy $(d_i^0, \mathbb{1}_{u_i > 0}, a_i^0)$. Indeed, as the rule v is non-decreasing, lying only makes it more likely for the action plan $\hat{v}(\mathbf{m})$ to go against the country's will, which is never beneficial, and it doesn't change what happens at subsequent stages as it cannot be detected. We may thus assume that $m_i^0 = \mathbb{1}_{u_i > 0}$.

Let $u_i \in \mathbb{R}$ and $\mathbf{m}_{-i} \in \{0, 1\}^{N \setminus \{i\}}$ be such that $\mu(u_i) > 0$ and $a_i^0 \neq \hat{v}(\mathbb{1}_{u_i > 0}, \mathbf{m}_{-i})$. As the deviation will be detected, it must yield a stage-payoff of more than

$$\hat{v}(\mathbb{1}_{u_i > 0}, \mathbf{m}_{-i})u_i + \frac{\delta}{1 - \delta}(U_i(v) - U_i^\emptyset)$$

(because of the trigger strategies). This deviation can only be profitable if $u_i < 0$. We then distinguish two cases:

- If i has veto power, since $\mathbb{1}_{u_i > 0} = 0$, we have $\hat{v}(\mathbb{1}_{u_i > 0}, \mathbf{m}_{-i}) = 0$. Therefore, the deviation is not profitable.
- If i has no veto, we have:

$$\hat{v}(\mathbb{1}_{u_i > 0}, \mathbf{m}_{-i})u_i + \frac{\delta}{1 - \delta}(U_i(v) - U_i^\emptyset) \geq -w_i^D + \frac{\delta}{1 - \delta}(U_i(v) - U_i^\emptyset) \geq 0$$

because i satisfies the compliance constraint. Therefore, the deviation is not profitable.

Finally, we conclude that the proposed profile of strategies is a perfect public equilibrium.

Step 2: Now suppose that there exists $i \in VE(v)$ such that $U_i(v) < U_i^\emptyset$. Consider a profile such that cooperation is chosen at any stage, and $(\hat{v}^T(\mathbb{1}_{u_i^T > 0})_{i \in N})$ is always implemented. Country i 's expected utility is $U_i(v)$. Therefore, playing $d_i^T = 0$ for all T is a profitable deviation.

Alternatively, suppose that there exists $i \notin VE(v)$ such that $U_i(v) < U_i^\emptyset + \frac{1 - \delta}{\delta}w_i^D$. Consider a profile such that cooperation is chosen at any stage, and $(\hat{v}^T(\mathbb{1}_{u_i^T > 0})_{i \in N})$ is always implemented. Consider the following deviation of player i : follow the first profile, and if there exists some T such that $u_i = -w_i^D$, then play $a_i^T = 0$, and $d_i^{T'} = 0$ for all $T' > T$. The event $\{u_i^T = -w_i^D\}$ occurs almost surely in finite time (for $T < +\infty$), and yields a superior payoff when it occurs. This is thus a profitable deviation.

A.4 Proof of Theorem 3

Let v^{**} be an optimal self-enforcing rule, and let $V^{**} = VE(v^{**})$ be its set of veto countries. By definition, v^{**} is solution of the problem $\max_{v \in \mathcal{V}_\delta} W(v)$, which is equivalent to $\max_{V \subseteq N} \max_{v \in \mathcal{V}_\delta, VE(v)=V} W(v)$. Therefore, v^{**} is solution of $\max_{v \in \mathcal{V}_\delta, VE(v)=V^{**}} W(v)$, and by the argument made at the beginning of the proof of Theorem 2, the rule v^{**} is also a solution of:

$$(\mathcal{P}^{V^{**}}) : \begin{cases} \max_{\{v(M)\}_{V^{**} \subseteq M \subseteq N}} \sum_{i \in N} U_i(v) \\ s.t. \quad \forall i \in V^{**}, \quad U_i(v) \geq U_i^\emptyset \\ s.t. \quad \forall i \in N \setminus V^{**}, \quad U_i(v) \geq U_i^\emptyset + \frac{1-\delta}{\delta} w_i^D \\ s.t. \quad \forall M, \quad V^{**} \not\subseteq M \Rightarrow v(M) = 0. \end{cases}$$

Now, by the arguments made in the proof of Theorem 2, if λ^i denotes the Lagrange multiplier associated to country i 's constraint in $(\mathcal{P}^{V^{**}})$, and if we note $w_i^0 = (1 + \lambda^i)w_i^e$ and

$$t^0 = \frac{\sum_{i \in N} w_i^0 t_i^e}{\sum_{i \in N} w_i^0},$$

we obtain:

$$\forall M \subseteq N, V^{**} \subseteq M, \quad \begin{cases} \sum_{i \in M} w_i^0 > t^0 \sum_{i \in N} w_i^0 \Rightarrow v^{**}(M) = 1 \\ \sum_{i \in M} w_i^0 < t^0 \sum_{i \in N} w_i^0 \Rightarrow v^{**}(M) = 0. \end{cases}$$

Moreover, we know that $V^{**} \not\subseteq M \Rightarrow v^{**}(M) = 0$. Now, we define $w_i^{**} = w_i^0 + K \mathbf{1}_{i \in V^{**}}$, where K is defined as a sufficiently large number, for instance $K = 1 + \sum_{i \in N} w_i^0$. We obtain that if $\sum_{i \in M} w_i^{**} > t^0 \sum_{i \in N} w_i^0 + K \#V^{**}$, then $V^{**} \subseteq M$ and $\sum_{i \in M} w_i^0 > t^0 \sum_{i \in N} w_i^0$, and therefore $v^{**}(M) = 1$. We also have that if $\sum_{i \in M} w_i^{**} < t^0 \sum_{i \in N} w_i^0 + K \#V^{**}$, then $V^{**} \not\subseteq M$ or $\sum_{i \in M} w_i^0 < t^0 \sum_{i \in N} w_i^0$, and therefore $v^{**}(M) = 0$. Finally, denoting

$$t^{**} = \frac{1}{\sum_{i \in N} w_i^{**}} \left(t^0 \sum_{i \in N} w_i^0 + K \#V^{**} \right)$$

We obtain that:

$$\forall M \subseteq N, \quad \begin{cases} \sum_{i \in M} w_i^{**} > t^{**} \sum_{i \in N} w_i^{**} \Rightarrow v^{**}(M) = 1 \\ \sum_{i \in M} w_i^{**} < t^{**} \sum_{i \in N} w_i^{**} \Rightarrow v^{**}(M) = 0. \end{cases}$$

This means that v^{**} is represented by $[\mathbf{w}^{**}; t^{**}]$. Furthermore, note that:

$$\begin{aligned} t^{**} &= \frac{1}{\sum_{i \in N} w_i^{**}} \left(\sum_{i \in N} w_i^0 t_i^e + K \# V^{**} \right) \\ &= \frac{1}{\sum_{i \in N} w_i^{**}} \left(\sum_{i \in N} w_i t_i^e + K \sum_{i \in V^{**}} (1 - t_i^e) \right) \geq \frac{\sum_{i \in N} w_i^{**} t_i^e}{\sum_{i \in N} w_i^{**}}. \end{aligned}$$

Finally, let $i \in N$ be a country such that $U_i(v^{**}) < U_i^\emptyset + \frac{1-\delta}{\delta} w_i^D$. As we have assumed that v^{**} is self-enforcing, it must be that $i \in VE(v^{**})$. Then, by construction, we obtain that $w_i^{**} > w_i^0 \geq w_i^e$.

Conversely, let $i \in N$ be a country such that $U_i(v^{**}) > U_i^\emptyset + \frac{1-\delta}{\delta} w_i^D$. We then show that i has no veto power under v^{**} . By contradiction, suppose that i has veto power. Then, $v^{**}(N \setminus \{i\}) = 0$. For $\varepsilon > 0$, consider v^ε defined by:

$$\begin{cases} v^\varepsilon(N \setminus \{i\}) = \varepsilon \\ \forall M \neq N \setminus \{i\}, \quad v^\varepsilon(M) = v^{**}(M). \end{cases}$$

We have $U_i(v^\varepsilon) = U_i(v^{**}) - \varepsilon P(N \setminus \{i\}) w_i^-$ and $\forall j \neq i, U_j(v^\varepsilon) = U_j(v^{**}) + \varepsilon P(N \setminus \{i\}) w_j^+$. By Assumption NEV, we get $W(v^\varepsilon) > W(v^{**})$. Moreover, v^ε is self-enforcing for ε small enough, hence a contradiction. We obtain that $i \notin VE(v^{**})$. It follows that $w_i^{**} = w_i^0 = w_i^e$.

A.5 Proof of **Theorem 4**

We now have:

$$U_i(v) = \frac{(2q-1)p_i}{2^n} \left(\sum_{M, i \in M} v(M) - \sum_{M, i \notin M} v(M) \right).$$

and $U_i^\emptyset = (2q-1)p_i/e$. Let v^* be an optimal rule maximizing the recourse to probabilistic tie-breaking (i.e. chosen as in the proof of **Theorem 2**, as we know that any representation of v^* will be a representation of any optimal rule). We know from **Theorem 2** that v^* is a weighted majority rule with threshold $t = \frac{1}{2}$, and with weights $(w_i)_{i \in N}$ satisfying⁵¹

$$\begin{cases} \forall i \in N, \quad w_i = (1 + \lambda^i) p_i \geq p_i \\ \forall i \in N, \quad \lambda^i [U_i(v^*) - U_i^\emptyset] = 0. \end{cases}$$

⁵¹For simplicity, the weights obtained in the proof of **Theorem 2** are re-scaled by a factor p_i/w_i^e , independent of i .

Let $S^1 = \{i \in N | U_i(v^*) > U_i^\emptyset\}$ and $S^0 = \{i \in N | U_i(v^*) = U_i^\emptyset\}$, so that $N = S^0 \cup S^1$ is a partition. From the previous system of equations, we have $w_i = p_i$ for all $i \in S^1$. If all parameters λ^i are null, we have $w_i = p_i$ for each country $i \in N$ and the result is obtained. If not, there are some countries in S^0 for which $w_i > p_i$, and the following equation has a unique solution $\underline{p} \in \mathbb{R}$:⁵²

$$\sum_{i \in S^0} \max(p_i, \underline{p}) = \sum_{i \in S^0} w_i.$$

Let us show that v^* is represented by the modified system of weights $[\mathbf{w}^*; 1/2]$ defined by:

$$\begin{cases} \forall i \in S^0, & w_i^* = \max(p_i, \underline{p}) \\ \forall i \in S^1, & w_i^* = p_i. \end{cases}$$

Note that the vector $\mathbf{w}^*$ can be obtained from $\mathbf{w}$ by a finite sequence of (Pigou-Dalton) transfers of the form $(w_i \rightarrow w_i + \alpha, w_j \rightarrow w_j - \alpha)$ with $i, j \in S^0$ and $w_i < w_i + \alpha \leq w_j - \alpha < w_j$. Let us show that if v^* is represented by a vector $[\mathbf{w}; t]$, it is represented by the vector $[\mathbf{w}'; t]$, when $\mathbf{w}'$ has been obtained from $\mathbf{w}$ by such a transfer.

Note $Q = t \sum_{i \in N} w_i = t \sum_{i \in N} w'_i$ and let R be a coalition such that $\sum_{k \in R} w'_k > Q$. We have two cases:

- In all cases but $(i \in R, j \notin R)$, we have

$$\sum_{k \in R} w_k \geq \sum_{k \in R} w'_k > Q = t \sum_{k \in N} w_k,$$

and we get $v^*(R) = 1$, as v^* is represented by $[\mathbf{w}; t]$.

- If $i \in R$ and $j \notin R$, we may have $\sum_{k \in R} w_k \leq Q$. Let us show that $v^*(R) = 1$. Assume by contradiction that $v^*(R) < 1$. Let $\sigma : N \rightarrow N$ be the transposition between i and j . We have $v^*(\sigma(R)) = 1$ (by the previous argument, since $j \in \sigma(R)$). Moreover, since v^* is represented by the system of weights $[\mathbf{w}; t]$ with $w_i < w_j$, we have for any coalition M :

$$\begin{cases} i, j \in M & \Rightarrow & v^*(\sigma(M)) = v^*(M) & (\text{since } \sigma(M) = M) \\ i, j \notin M & \Rightarrow & v^*(\sigma(M)) = v^*(M) & (\text{since } \sigma(M) = M) \\ i \in M, j \notin M & \Rightarrow & v^*(\sigma(M)) \geq v^*(M) \\ i \notin M, j \in M & \Rightarrow & v^*(\sigma(M)) \leq v^*(M). \end{cases}$$

⁵²Note that the function $f : x \mapsto f(x) = \sum_{i \in S^0} \max(p_i, x)$ is increasing and continuous, with $f(0) < \sum_{i \in S^0} w_i$ and $\lim_{x \rightarrow \infty} f(x) = +\infty$.

We obtain:

$$\begin{aligned}
\frac{U_j(v^*)}{U_j^\emptyset} &= \frac{e}{2^n} \left(v^*(\sigma(R)) + \sum_{M, j \in M, M \neq \sigma(R)} v^*(M) - \sum_{M, j \notin M} v^*(M) \right) \\
&\geq \frac{e}{2^n} \left(v^*(\sigma(R)) + \sum_{M, j \in M, M \neq \sigma(R)} v^*(\sigma(M)) - \sum_{M, j \notin M} v^*(\sigma(M)) \right) \\
&> \frac{e}{2^n} \left(v^*(R) + \sum_{M, i \in M, M \neq R} v^*(M) - \sum_{M, i \notin M} v^*(M) \right) \\
&> \frac{U_i(v^*)}{U_i^\emptyset}.
\end{aligned}$$

We get a contradiction with the assumption that $i, j \in S^0$. Finally, it must be that $v^*(R) = 1$.

Similarly, one can show that $\sum_{k \in R} w'_k < Q$ implies $v^*(R) = 0$. Finally, v^* is represented by $[\mathbf{w}'; t]$. By induction, v^* (and thus any optimal rule) is represented by $[\mathbf{w}^*; 1/2]$.

Finally, let us show that $w_i^* = \max(p_i, \underline{p})$ for any $i \in S^1$. Let $i \in S^1$ and $j \in S^0$. As $\frac{U_i(v^*)}{U_i^\emptyset} > \frac{U_j(v^*)}{U_j^\emptyset}$, and v^* is represented by $[\mathbf{w}^*; 1/2]$, it must be that $w_i^* \geq w_j^*$ (by an argument similar to the previous computation). We have $w_j^* = \max(p_j, \underline{p}) \geq \underline{p}$, and thus $w_i^* \geq \underline{p}$. As we already know that $w_i^* = p_i$, we can write $w_i^* = \max(p_i, \underline{p})$.

A.6 Proof of Proposition 3

Proof. We first note that a country's utility is proportional to its Banzhaf voting power:

$$U_i(v) = \frac{(2q-1)p_i}{2^n} \left(\sum_{M, i \in M} v(M) - \sum_{M, i \notin M} v(M) \right) = \frac{e_i U_i^\emptyset}{2} BZ_i(v).$$

By Theorem 2 and the unbiased assumption, there exists $\mathbf{w}^*$ such that any optimal voting rule v^* is weighted and represented by $[\mathbf{w}^*; 1/2]$, since $t_i^e = 1/2$ for all countries. Denote by $O(v^*) \subset N$ the subset of overweighted countries under the optimal voting rule,

$$O(v^*) = \{i \in N \mid w_i^* > p_i\}$$

Let $i, j \in O(v^*)$. Then by Theorem 2, $U_i(v^*) = U_i^\emptyset$ and $U_j(v^*) = U_j^\emptyset$. By the Penrose limit approximation,

$$1 = \frac{U_i(v^*)/U_i^\emptyset}{U_j(v^*)/U_j^\emptyset} = \frac{e_i BZ_i(v^*)}{e_j BZ_j(v^*)} \approx \frac{e_i w_i^*}{e_j w_j^*}.$$

We obtain that $\forall i, j \in O(v^*)$, $w_i^*/w_j^* = e_j/e_i$, hence there exists $a > 0$ such that $\forall i \in O(v^*)$, $w_i^* = a/e_i$. Moreover, for each such country $i \in O(v^*)$, we have $p_i e_i < w_i^* e_i = a$.

Now, let $i \notin O(v^*)$ and $j \in O(v^*)$. By application of [Theorem 2](#) and using Penrose limit approximation, we obtain:

$$1 \leq \frac{U_i(v^*)/U_i^\emptyset}{U_j(v^*)/U_j^\emptyset} = \frac{e_i B Z_i(v^*)}{e_j B Z_j(v^*)} \approx \frac{e_i w_i^*}{e_j w_j^*} = \frac{e_i p_i}{e_j w_j^*}.$$

Hence, for any $i \notin O(v^*)$, we have both $w_i^* = p_i$ and $p_i \geq w_j^* e_j / e_i = a / e_i$. This concludes the proof. $\square$

A.7 Proof of [Proposition 4](#)

Proof. The proof is divided in two parts. First, we show that if a rule of the form $v^e \oplus k$ satisfies the participation constraint for the smallest country, then all participation constraints are satisfied. Then, we show that there exists $k^* \in \{1, \dots, n\}$ such that the smallest country is strictly better off under $v^e \oplus k^*$ than under v^e . Throughout the proof, we assume that the population vector $\mathbf{p}$ is strictly ordered, so that $v^e(M)$ is non-decreasing as function of the population of M .

1. Let v^e be an efficient rule, it is represented by $[\mathbf{p}; 1/2]$. We order countries by increasing order of population, so that $p_1 < p_2 < \dots < p_n$. For a size threshold $k \in \{0, \dots, n\}$ and a country $i \in N \setminus \{1\}$, we write:

$$\begin{aligned} \frac{U_i(v^e \oplus k)}{U_i^\emptyset} - \frac{U_1(v^e \oplus k)}{U_1^\emptyset} &= \frac{e}{2^n} \left(\sum_{M, i \in M, \#M \geq k} v^e(M) - \sum_{M, i \notin M, \#M \geq k} v^e(M) \right. \\ &\quad \left. - \sum_{M, 1 \in M, \#M \geq k} v^e(M) + \sum_{M, 1 \notin M, \#M \geq k} v^e(M) \right) \\ &= \frac{e}{2^n} \left(\sum_{M, i \in M, 1 \notin M, \#M \geq k} v^e(M) - \sum_{M, i \notin M, 1 \in M, \#M \geq k} v^e(M) \right) \\ &= \frac{e}{2^n} \sum_{M, i \notin M, 1 \notin M, \#M \geq k-1} [v^e(M \cup \{i\}) - v^e(M \cup \{1\})] \\ &\geq 0. \end{aligned}$$

The last inequality comes from the assumption that $p_1 \neq p_i$, implying that for any coalition M , the population of $M \cup \{i\}$ is strictly higher than that of $M \cup \{1\}$. Hence, if country 1 satisfies the participation constraint for $v^e \oplus k$, then all countries also do.

2. For the second part, we use the assumption that $p_n > p_1 + p_2$. It follows that by adding countries to $\{n\}$ and $\{1, 2\}$ respectively, we may find a partition of the set of countries $N = S \cup T$ such that $\#S \leq \#T$ and $\sum_{i \in S} p_i > \sum_{i \in T} p_i$. We conclude that there is a coalition S of size at most $n/2$ such that $v^e(S) = 1$.

Let k^* be the minimal size of a non-losing coalition under the efficient rule: $k^* = \min\{\#M \mid v^e(M) > 0\}$. We know that $k^* \leq n/2$, and we can write the relative gain of country 1 when a size threshold $(k^* + 1)$ is added as

$$\begin{aligned} \frac{U_1(v^e \oplus (k^* + 1)) - U_1(v^e)}{U_1^\emptyset} &= \frac{e}{2^n} \left(\sum_{M, 1 \notin M, \#M < k^* + 1} v^e(M) - \sum_{M, 1 \in M, \#M < k^* + 1} v^e(M) \right) \\ &= \frac{e}{2^n} \left(\sum_{M, 1 \notin M, \#M = k^*} v^e(M) - \sum_{M, 1 \in M, \#M = k^*} v^e(M) \right). \end{aligned}$$

Writing M_{+j-1} for $M \cup \{j\} \setminus \{1\}$, we have for any coalition M with $1 \in M$ and $j \notin M$, $v^e(M) \leq v^e(M_{+j-1})$, since the population of coalition M_{+j-1} is strictly larger than that of M . So, we write

$$\begin{aligned} \sum_{M, 1 \in M, \#M = k^*} v^e(M) &\leq \sum_{M, 1 \in M, \#M = k^*} \left(\frac{1}{n - k^*} \sum_{j \notin M} v^e(M_{+j-1}) \right) \\ &\leq \sum_{M, 1 \notin M, \#M = k^*} v^e(M). \end{aligned}$$

The remaining point is to show that this inequality is strict. For that, it suffices to show that there exists a coalition M of size k^* , with $1 \in M$, $n \notin M$ and for which $v^e(M) < v^e(M_{+n-1})$. As v^e is represented by $[\mathbf{p}; 1/2]$ and by definition of k^* , we must have $v^e(M^{max}) > 0$, where $M^{max} = \{n - k^* + 1, \dots, n\}$ is the coalition of size k^* with a maximal population. As $k^* \leq n/2$, we have that $v^e(M^{min}) = 0$, where $M^{min} = \{1, \dots, k^*\}$ is the coalition of size k^* with a minimal population (since the population of M^{min} is strictly lower than that of $N \setminus M^{min}$).

To find the desired coalition, we construct a sequence of coalitions as follows. Let $(S^p)_{p=1 \dots P}$ be a sequence of coalitions such that $S^1 = \{2, \dots, k^*\}$, $S^P = \{n - k^* + 1, \dots, n - 1\}$, each coalition S^p is of size $k^* - 1$, and the difference in population between two consecutive coalitions is positive but bounded by $p_n - p_1$: $0 < \sum_{i \in S^{p+1}} p_i - \sum_{i \in S^p} p_i < p_n - p_1$. As p_1 and p_n are the smallest and largest populations (respectively), such a sequence of coalition

exists. Now, we write

$$\begin{aligned}
0 < v^e(M^{max}) - v^e(M^{min}) &= v^e(S_{+n}^P) - v^e(S_{+1}^1) \\
&= \sum_{p=1}^P [v^e(S_{+n}^p) - v^e(S_{+1}^p)] + \sum_{p=1}^{P-1} [v^e(S_{+1}^{p+1}) - v^e(S_{+n}^p)] \\
&\leq \sum_{p=1}^P [v^e(S_{+n}^p) - v^e(S_{+1}^p)],
\end{aligned}$$

where the inequality arises from the observation that the population of S_{+n}^p is larger than that of S_{+1}^{p+1} , by construction of the sequence (S^p) . Thus, there must be a p such that $v^e(S_{+n}^p) > v^e(S_{+1}^p)$, the strict inequality that we were looking for. Therefore, we obtained that $U_1(v^e \oplus (k^* + 1)) > U_1(v^e)$.

To conclude, we define $\bar{e}$ as the lowest parameter e such that v^e satisfies the participation constraints, and we define $\underline{e}$ as the lowest parameter e such that there exists $k \in \{1, \dots, n\}$ for which $v^e \oplus k$ satisfies the participation constraints. We have shown that $\underline{e} < \bar{e}$, and this concludes the proof. $\square$

A.8 Proof of Theorem 5

We introduce the notion of relative utility of a country under a rule v as the ratio between its utility under v and the utility it would get if it was a dictator:

$$\forall i \in N, \quad u_i(v) = \frac{U_i(v)}{w_i^e/2} = \frac{U_i(v)}{(2q-1)ep_i} = \frac{U_i(v)}{eU_i^\emptyset}.$$

With this notation, i 's compliance constraint can be written as: $u_i(v) \geq \frac{1}{e} + \frac{1-\delta}{\delta}$.

Claim 1 : The optimal rule is either unanimous or a weighted majority rule with threshold $1/2$.

By application of Theorem 3, it suffices to show that the optimal self-enforcing rule v cannot have a set of veto players $V = VE(v)$ such that $\emptyset \subsetneq V \subsetneq N$. Assume by contradiction that it is the case, and take $i \in V$ and $j \notin V$. We have:

$$\begin{aligned}
u_i(v) &= \frac{1}{2^n} \sum_{M, V \subseteq M} v(M) \\
u_j(v) &= \frac{1}{2^n} \left(\sum_{M, V \subseteq M, j \in M} v(M) - \sum_{M, V \subseteq M, j \notin M} v(M) \right).
\end{aligned}$$

Since $j \notin V$, there exists a coalition M with $V \subseteq M$, $j \notin M$ and $v(M) > 0$. Therefore,

$u_i(v) > u_j(v)$. As v is self-enforcing, we have $u_i(v) > u_j(v) \geq \frac{1}{e} + \frac{1-\delta}{\delta}$: i 's constraint is not binding. For $\varepsilon > 0$, consider now v^ε defined by:

$$\begin{cases} v^\varepsilon(N \setminus \{i\}) = \varepsilon \\ \forall M \neq N \setminus \{i\}, \quad v^\varepsilon(M) = v(M). \end{cases}$$

We have $u_i(v^\varepsilon) = u_i(v) - \frac{\varepsilon}{2^n}$ and $\forall j \neq i, u_j(v^\varepsilon) = u_j(v) + \frac{\varepsilon}{2^n}$. As $p_i < \sum_{j \neq i} p_j$, we have $W(v^\varepsilon) > W(v)$. Moreover, v^ε is self-enforcing for ε small enough, hence a contradiction.

Claim 2 : If simple majority is self-enforcing, the optimal self-enforcing rule is a weighted majority rule with threshold $1/2$. If unanimity is self-enforcing, but simple majority is not, then unanimity is the optimal self-enforcing rule. If neither unanimity nor simple majority is self-enforcing, there is no self-enforcing rule.

Let us note v^m the simple majority rule. If v^m is self-enforcing, as unanimity is strictly welfare-dominated by v^m ,⁵³ we get from Claim 1 that the optimal self-enforcing rule is a weighted majority rule with threshold $1/2$.

If v^m is not self-enforcing, note that no country satisfies the compliance constraint under simple majority (simple majority yields the same relative utility for all countries, and they all face the same constraint). We show that no weighted majority v can then be self-enforcing. Indeed, we have:

$$\begin{aligned} \sum_{i \in N} u_i(v) &= \frac{1}{2^n} \sum_{i \in N} \left(\sum_{M, i \in M} v(M) - \sum_{M, i \notin M} v(M) \right) \\ &= \frac{1}{2^n} \sum_{M \subseteq N} \left(\sum_{i \in M} v(M) - \sum_{i \notin M} v(M) \right) \\ &= \frac{1}{2^n} \sum_{M \subseteq N} (2\#M - n)v(M) \\ &\leq \sum_{i \in N} u_i(v^m). \end{aligned}$$

At least one country has a (weakly) lower relative utility under v than under v^m , therefore v cannot be self-enforcing (as v does not grant veto to any country, the endogenous constraints are the same for v and v^m). To conclude, if simple majority is not self-enforcing, the only possible optimal self-enforcing rule is unanimity, and it can be optimal self-enforcing only when it is self-enforcing.

⁵³The relative utility of any country is $\frac{1}{2^{n-1}} \binom{n-1}{\lfloor \frac{n-1}{2} \rfloor}$ under simple majority and $1/2^n$ under unanimity.

Let u^m be the relative utility of any country under simple majority. It is easy to see that: $u^m = \frac{1}{2^{n-1}} \binom{n-1}{\lfloor \frac{n-1}{2} \rfloor}$. Simple majority is self-enforcing if and only if

$$\begin{aligned} u^m \geq \frac{1}{e} + \frac{1-\delta}{\delta} &\Leftrightarrow \frac{1}{\delta} \leq 1 + u^m - \frac{1}{e} \\ &\Leftrightarrow \delta \geq \frac{1}{1 + u^m - \frac{1}{e}} := \delta^c(e). \end{aligned}$$

Let $u^{eff} > 0$ be the relative utility of the smallest country under the efficient voting rule. The (efficient) population-weighted majority rule is self-enforcing, and is therefore the optimal self-enforcing rule, if and only if

$$\delta \geq \frac{1}{1 + u^{eff} - \frac{1}{e}} := \delta^{eff}(e).$$

By the proof of claim 2, it is easy to see that $u^{eff} \leq u^m$, thus $\delta^{eff}(e) \geq \delta^c(e)$. Moreover, as $u^{eff} > 0$, we have $\lim_{e \rightarrow \infty} \delta^{eff}(e) < 1$.

Finally, unanimity is self-enforcing if and only if $\frac{1}{2^n} \geq \frac{1}{e}$, that is if and only if $e \geq 2^n := \underline{e}$.

Claim 3 : For $\delta \geq \delta^c(e)$, there exists a minimal weight function $\underline{p}(e, \delta)$ non-increasing in both e and δ , such that, for each e and δ , any optimal self-enforcing rule is represented by $[\mathbf{w}; 1/2]$, with for all $i \in N$, $w_i = \max(p_i, \underline{p}(e, \delta))$.⁵⁴

We note $k(e, \delta) = \frac{1}{e} + \frac{1-\delta}{\delta}$, this function is decreasing in both e and δ . Consider the following problem:

$$(\mathcal{P}^k) : \begin{cases} \max_{\{v(M)\}_{M \subseteq N}} \sum_{i \in N} p_i u_i(v) \\ s.t. \quad \forall i \in N, \quad u_i(v) \geq k \\ s.t. \quad \forall M \subseteq N, \quad 0 \leq v(M) \leq 1. \end{cases}$$

Following the proofs of [Theorem 2](#) and [Theorem 4](#), we know that any solution v of $(\mathcal{P}^k)$ is a weighted rule represented by a certain vector $[\mathbf{w}; 1/2]$ with:

- for all $i \in N$, $w_i = \max(p_i, \underline{p})$
- $\underline{p}$ is the solution⁵⁵ of $\sum_{i \in N} \max(p_i, \underline{p}) = \sum_{i \in N} p_i(1 + \lambda_*^i)$

⁵⁴The claim may seem obvious, as increasing e and/or δ relaxes the self-enforcing constraints. Note however that the welfare attached to a rule weighted by $[\mathbf{w}; 1/2]$, with $w_i = \max(p_i, \underline{p})$, may be non-monotonic as a function of $\underline{p}$. One can construct such an example with $\mathbf{p} = (2, 4, 4, 5)$ and $\underline{p} = 2$ or 4 or 5.

⁵⁵To be precise, $\underline{p}$ is defined in the proof of [Theorem 5](#) as the solution of $\sum_{i \in S^0} \max(p_i, \underline{p}) = \sum_{i \in S^0} p_i(1 + \lambda_*^i)$, and it is shown at the end of the proof that for all $i \in S^1 = N \setminus S^0$, $\max(p_i, \underline{p}) = p_i = p_i(1 + \lambda_*^i)$. Therefore, the above definition is equivalent.

- for all $i \in N$, λ_*^i is the Lagrangian coefficient associated to country i 's constraint, related to a solution v^* of the problem $(\mathcal{P}^k)$.

With this definition, it is clear that $\underline{p}$ increases with $\sum_{i \in N} p_i \lambda_*^i$. Let us show that this last quantity is non-decreasing as a function of k . The linear program $(\mathcal{P}^k)$ can be re-written as follows (we multiplied each country i 's constraint by a factor $-p_i$):

$$(\mathcal{P}^k) : \begin{cases} \max_{\{v(M)\}_{M \subseteq N}} \frac{1}{2^n} \sum_{M \subseteq N} \left(\sum_{i \in M} p_i - \sum_{i \notin M} p_i \right) v(M) \\ s.t. \quad \forall i \in N, \quad \frac{1}{2^n} \left(\sum_{M, i \notin M} p_i v(M) - \sum_{M, i \in M} p_i v(M) \right) \leq -k p_i \\ s.t. \quad \forall M \subseteq N, \quad v(M) \leq 1 \\ s.t. \quad \forall M \subseteq N, \quad v(M) \geq 0. \end{cases}$$

The dual of $(\mathcal{P}^k)$, of which the variables $(\lambda_*^i)_{i \in N}$ are solution, is the following linear program:

$$(\mathcal{D}^k) : \begin{cases} \min_{\{\lambda^i\}_{i \in N}, \{\nu^M\}_{M \subseteq N}} \sum_{M \subseteq N} \nu^M - k \sum_{i \in N} p_i \lambda^i \\ s.t. \quad \forall M \subseteq N, \quad \frac{1}{2^n} \left(\sum_{i \notin M} p_i \lambda^i - \sum_{i \in M} p_i \lambda^i \right) + \nu^M \geq \frac{1}{2^n} \left(\sum_{i \in M} p_i - \sum_{i \notin M} p_i \right) \\ s.t. \quad \forall i \in N, \quad \lambda^i \geq 0 \\ s.t. \quad \forall M \subseteq N, \quad \nu^M \geq 0. \end{cases}$$

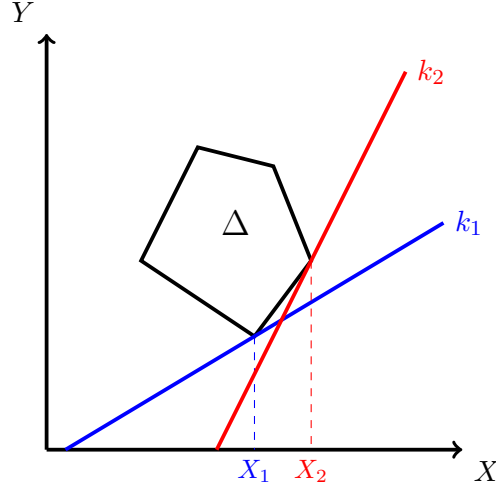
Now, consider the mapping $\Phi : ((\lambda^i), (\nu^M)) \mapsto (X = \sum_{i \in N} p_i \lambda^i, Y = \sum_{M \subseteq N} \nu^M)$. It is a linear mapping from $\mathbb{R}^{2^n} \times \mathbb{R}^N$ into $\mathbb{R}^2$, which transforms any convex polyhedron into a convex polyhedron. Therefore, for any solution $((\nu^M), (\lambda^i))$ of the program $(\mathcal{D}^k)$, there is a corresponding solution (X, Y) of a the 2-dimensional program $(\mathcal{D}^k)'$ defined by:

$$(\mathcal{D}^k)' : \begin{cases} \min_{X, Y} Y - kX \\ s.t. \quad (X, Y) \in \Delta \\ s.t. \quad X \geq 0 \\ s.t. \quad Y \geq 0, \end{cases}$$

where $\Delta \subset \mathbb{R}^2$ is a convex polyhedron. It is clear (see [Figure 6](#)) that X is non-decreasing as a function of k , in the following strong sense: for any $k_2 > k_1$ and any solutions (X_1, Y_1) of $(\mathcal{D}^{k_1})'$ and (X_2, Y_2) of $(\mathcal{D}^{k_2})'$, we have that $X_2 \geq X_1$.

To conclude, there exists a non-decreasing function $\underline{p}(k)$, such that for each k , any solution of $(\mathcal{P}^k)$ is represented by $[\mathbf{w}; 1/2]$, with for all $i \in N$, $w_i = \max(p_i, \underline{p}(k))$.

Figure 6: Solutions of $(\mathcal{D}^{k_1})'$ and $(\mathcal{D}^{k_2})'$ for $k_2 > k_1$.



A.9 Bargained Voting Rule

The voting rule agreed upon via the bargaining process described in the conclusion is solution to the following maximization problem:

$$(\mathcal{P}^B) : \begin{cases} \max_{\{v(M)\}_{M \subseteq N} \in [0,1]^{(2^N)}} \sum_{i \in N} \ln(U_i(v) - U_i^\emptyset) \\ \text{s.t. } \forall i \in N, \quad U_i(v) > U_i^\emptyset. \end{cases}$$

We show the following proposition.

Proposition 5. *If some decision rules strictly satisfy all participation constraints, then the bargained voting rule v^B exists and is unique. There exists a system of weights $[\mathbf{w}^B; t^B]$ such that it is weighted and represented by $[\mathbf{w}^B; t^B]$. Countries that are closer to their stand-alone utility under the bargained voting rule are relatively more overweighted.*

Formally, for any two countries $i, j \in N$: $U_i(v^B) - U_i^\emptyset > U_j(v^B) - U_j^\emptyset$ implies $w_i^B/w_i^e < w_j^B/w_j^e$.

Proof. The function $f : v \mapsto \sum_{i \in N} \ln(U_i(v) - U_i^\emptyset)$ is strictly concave on the non-empty, open and convex set $\{v \mid \forall i \in N, U_i(v) > U_i^\emptyset\}$ and $\lim_{U_i(v) \rightarrow U_i^\emptyset} f(v) = -\infty$. Thus, the problem $(\mathcal{P}^B)$ admits a unique solution. The Lagrangian writes

$$\mathcal{L}(v) = \sum_{i \in N} \ln(U_i(v) - U_i^\emptyset) + \sum_{M \subseteq N} [\eta^M v(M) + \nu^M (1 - v(M))].$$

Its partial derivative with respect to $v(M)$ (one of the 2^n variables) is

$$\frac{\partial \mathcal{L}}{\partial v(M)}(v) = P(M) \left(\sum_{i \in M} \frac{w_i^+}{U_i(v) - U_i^\emptyset} - \sum_{i \notin M} \frac{w_i^-}{U_i(v) - U_i^\emptyset} \right) + \eta^M - \nu^M.$$

Denoting by v^B the solution of $(\mathcal{P}^B)$, we obtain (as in the proof of [Theorem 2](#)):

$$\forall M \subseteq N, \quad \begin{cases} \sum_{i \in M} w_i^B > t^B \sum_{i \in N} w_i^B & \Rightarrow v^B(M) = 1 \\ \sum_{i \in M} w_i^B < t^B \sum_{i \in N} w_i^B & \Rightarrow v^B(M) = 0. \end{cases}$$

where $w_i^B = \left(\frac{1}{U_i(v^B) - U_i^\emptyset} \right) w_i^e$ and $t^B = \frac{\sum_{i \in N} w_i^B t_i^e}{\sum_{i \in N} w_i^B}$.

Therefore we conclude that v^B is weighted and can be represented by $[\mathbf{w}^B; t^B]$. Furthermore, by definition of $\mathbf{w}^B$, $U_i(v^B) - U_i^\emptyset > U_j(v^B) - U_j^\emptyset$ implies $w_i^B/w_i^e < w_j^B/w_j^e$.

□