amse
école d'économie d'aix-marseille
aix-marseille school of economics

Working Papers / Documents de travail

Optimal Voting Rules under Participation Constraints

Antonin Macé
Rafael Treibich

WP 2017 - Nr 42

Optimal Voting Rules under Participation Constraints*

Antonin Macé[1] and Rafael Treibich[2]

[1] Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE
[2]University of Southern Denmark

November 6, 2017

Abstract

We study the design of voting rules for international unions when countries' participation is voluntary. While efficiency recommends weighting countries proportionally to their *stakes*, we show that accounting for participation constraints entails *overweighting* some countries, those for which the incentive to participate is the lowest. When decisions are not enforceable, cooperation requires the satisfaction of more stringent constraints, that may be mitigated by granting a veto power to some countries. The model has important implications for the problem of apportionment, the allocation of voting weights to countries of differing populations, where it provides a rationale for setting a minimum representation for small countries. (JEL: F53, D02, C61, C73)

"You may if you wish go home from this Conference and say that you have defeated the veto. But what will be your answer when you are asked: Where is the Charter?" —U.S. Senator Tom Connally at the 1945 San Francisco Conference.

1 Introduction

In 1787, the founding fathers of the US Constitution faced a contentious challenge: how to accommodate a fair representation of states at the federal level, while preserving a say to small states in the new institutions? This issue was resolved by the so-called *Connecticut Compromise*, under which states received a weight proportional to their populations in the House,

*We would like to thank Alessandra Casella, Ernesto Dal Bó, Andreas Kleiner, Jean-François Laslier, Michel Lebreton, Aniol Llorente-Saguer and Sharun Mukand for useful comments, as well as audiences from Berkeley, Copenhagen, Glasgow, Odense, Paris, Saint-Étienne and Toulouse.

but an equal weight in the Senate. The resulting distribution of seats at the Electoral College appears as a compromise between the notions of proportionality, conducing to efficiency, and equality, conducing to acceptability. This distinctive feature of the US Constitution was decisive in its adoption and has remained unchanged until now.

The tension between the efficiency of a multi-party institution and its acceptability by all parties is not limited to the anecdotal episode of the Constitutional Convention. In fact, such a tension is also inherently present for many intergovernmental organizations and confederations, when a group of sovereign states voluntarily commits to collectively decide on one or several policy areas. One prominent example is the Council of the European Union (EU), one of its main decision-making bodies, whose decision rules have often been changed after fierce debates. The Treaty of Lisbon established the latest rule: a reform is adopted if it is approved by at least 55% of the Member States representing at least 65% of the EU population.[1] This rule exhibits again a compromise between proportionality and equality, which ensures its acceptability by the smallest Member States. Another important example is the UN Security Council, in which the five permanent members can veto any resolution, and consequently benefit from a disproportionate power. When the Charter of the UN was ratified at 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).[2] In that case, the acceptability of the rule entailed not only overweighting some countries, but ensuring them a veto power.

The goal of the present paper is to shed light on the tension between the efficiency of a voting rule and its acceptability, in the context of international organizations. Should countries be weighted according to their populations, or should small countries be overweighted? Should some important countries benefit from a veto power? To address these questions, we take a second-best approach to institutional design, by looking for normatively appealing rules among those that are politically feasible when countries' participation is voluntary.

Our model features a group of countries choosing whether to delegate some of their competences to a supranational entity.[3] The choice to transfer a competence is made unanimously[4] ex-ante, before a decision is taken. If cooperation is agreed upon, the decision is made collectively according to a predetermined voting rule. If cooperation is rejected, countries remain

[1]The rule applies for most policy areas. Additionally, a proposal cannot be blocked by less than four Member States.

[2]The quote is also cited in Bouton et al. (2014).

[3]For instance, the European Union has exclusive competence over custom 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).

[4]The fact that all EU competences must be voluntary transferred by its member states is known as the principle of *conferral* (Treaty of Lisbon, 2007a).

sovereign and they retain an associated stand-alone utility.

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 externalities (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 loose 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, countries commit to accepting the outcome of these decisions even if they end up disagreeing. In that case, we show that cooperation can be established if the voting rule satisfies a set of participation constraints. This leads us to study a constrained optimization problem: determining which rule delivers the highest (ex-ante, utilitarian) welfare, subject to the participation constraint of all countries. Our first result asserts that the optimal rule is a weighted majority rule, whereby each country is assigned a fixed voting weight and a reform is adopted if the total weight of favorable countries exceeds a certain threshold. Furthermore, we show that the weight of a country should be equal to its efficient weight, the weight it would receive absent any participation constraint, unless its participation constraint is binding, in which case it should be larger. This result thus offers a justification for overweighting countries that have the lowest (endogenous) incentive to participate.

We then relax the assumption of enforceability of collective decisions, as this property is not likely to be satisfied in the context of intergovernmental organizations (Maggi and Morelli, 2006). We define a notion of self-enforcing cooperation, by considering a repeated version of the previous decision game: it requires that countries choose to cooperate at each stage, and also comply with the collective decision even when they disagree. We show that a rule is self-enforcing if it satisfies a set of endogenous constraints: countries with a veto must satisfy their participation constraint, while countries without a veto must satisfy a more stringent "compliance" constraint. Building on this result, we show that self-enforcing rules take the form of weighted majority rules. Similar to the enforceable case, the weight of a country should be equal to its efficient weight, unless its utility falls below a specific benchmark level of utility, in which case it should be larger. Furthermore, we show that some of the overweighted countries may also benefit from a veto power. The result thus provides a rationale for the use of veto power: compliance can sometimes be best achieved by giving some "negative power" to a country (i.e. a 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 the heterogeneity between countries is governed by a single parameter, such as the country's population. This model can be interpreted as addressing the *apportionment problem*: 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 at the US Electoral College, where each state is allocated a baseline of 2 seats plus a number of seats proportional to its population. Then, we investigate the implication of self-enforceability in this simplified model, and show that it never leads to recommending veto power for a subset of countries: either the rule must be a weighted majority rule with no veto or it must be the unanimity rule. We further show that the former case prevails when the efficiency gain is high and/or the discount factor is high. We also show that, in that case, the minimum representation threshold decreases with the efficiency gain and the discount factor.

1.1 Relation to the literature

Our paper combines both a normative and a positive approach to voting rules in international unions. 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, 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 for a principle of *degressively proportionality*, which requires weights to increase less than proportionally with states' populations.[5] Our paper follows this second strand, building in particular on the utilitarian approach[6] 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

[5]The 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); 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), or following the utilitarian principle. These works are extended by Beisbart and Bovens (2007) and Napel et al. (2016), 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) offers a different rationale for degressive proportionality, based on the utilitarian principle, when citizens exhibit decreasing marginal utility. See Laslier (2012) for a survey.

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

weight each state proportionally to its *stake* in the decisions.[7] Depending on the assumption made on the correlation of preferences within states, the stake of a state coincides either with its population or with the square root of its population.

We depart from this literature by adding political feasibility constraints. Such a positive approach to voting rules has been introduced in a couple of important papers. Barberà and Jackson (2004) consider the stability of a voting rule with respect to a constitutional rule: a voting rule is stable if it is not overthrown by another voting rule. In contrast, we study the stability of a rule with respect 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-stable[8] among those satisfying the same feasibility constraints, as they are obtained from a welfare-maximization program. Our focus on self-enforceability of collective decisions is inspired by the pioneering paper of Maggi and Morelli (2006). They consider a union of homogeneous countries and prove that the optimal self-enforcing rule is either a (qualified) majority or the unanimity. Our section on self-enforceability extends their analysis to the case of an heterogeneous union: we show that the optimal rule may give a veto power to a strict subset of countries in general, as hinted in the conclusion of their paper. However, we prove that if the heterogeneity is controlled by a single parameter, like the population of a country, then the veto power should either be given to all countries or none.

The premise of our paper is that countries' participation to an international union 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. As mentioned in the conclusion of that paper, generalizing the argument requires analyzing this trade-off in a voting game, and that is what our first result on overweighting achieves.

Finally, our main assumption in this work 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. Following the seminal paper of Alesina and Spolaore (1997) on the (endogenous) size of nations, several papers have explored this rationale for cooperation between countries. Alesina et al. (2005) explores 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 paper emphasizes the importance of the voting rule on the stability of

[7] A 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 on the level of political philosophy.

[8] With respect to the unanimity rule, taken as the benchmark constitutional rule.

the union, but differs in that we take into account the heterogeneity of countries. Finally, let us note that some authors have provided other rationales for international cooperation, such as information aggregation (Penn, 2015), or even pure preference aggregation (Crémer and Palfrey, 1996).

1.2 Example

Consider a union of 5 countries which must decide, repeatedly, whether to impose embargoes on 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 favorable and a disutility of 1 if unfavorable. If the embargo is not effective, all countries get a utility of 0.

Before preferences over future decisions are realized, countries must decide whether to sign a cooperation treaty, i.e. agree to take all embargo decisions collectively with all other countries in the union, or remain sovereign, i.e. take all embargo decisions independently of other countries. The treaty is only effective if signed by all countries in the union and is assumed enforceable.

Under sovereignty, the embargo is implemented effectively only when all countries are favorable, which happens with a very small probability $16/3^5$. Ex-ante, country 1 gets a utility $U = 16/3^5$, while all other countries get utility $U = 32/3^5$ from any 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. Ex-ante, the utility of each country depends both on the preferences of other countries and on the decision rule used to make the collective decision after preferences have realized.

Here, because all countries have the same stake in the collective decision, the efficient voting rule consists in adopting the embargo at the simple majority (Theorem 1). Ex-ante, countries 2 to 5 get a utility:

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

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:

$$U_1 = \frac{1}{3}\mathbb{P}(\text{emb. adopted} \mid \text{fav.}) - \frac{4}{3}\mathbb{P}(\text{emb. adopted} \mid \text{unf.}) = -\frac{120}{3^5} < \frac{16}{3^5},$$

which means it would not agree to cooperate ex-ante. Here, the only way to ensure cooperation is to give some additional voting power to country 1. The optimal decision rule (Theorem 2) 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 $(9, 3, 3, 3, 3)$ and threshold $1/2$. Country 1 gets exactly its stand-alone utility $U = 16/3^5$, while countries 2 to 5 now get a reduced utility $U = 146/3^5$. 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 adopt the embargo even if this has been decided collectively. In that case, looking at the previous one-shot game is not sufficient, as countries have no incentive to abide by collective decisions if the game ends right away. We thus consider the infinitely repeated version of that game where countries decide at each stage whether to cooperate, and in case of cooperation, whether to implement the collective decision. Let $\delta = 5/6$ be the common discount factor. In order for the voting rule to be self-enforcing (i.e. induce cooperation and compliance on the equilibrium path), the benefit of not implementing the embargo for unfavorable countries must be outweighed by the long-term cost of not sustaining cooperation. The associated compliance constraints turns out to be more stringent than the participation constraints (Proposition 2). As a result, the previous optimal rule cannot be self-enforcing since country 1's participation constraint was already binding. Self-enforcement can only be achieved by granting a veto power to country 1 (Theorem 3). The optimal self-enforcing voting rule is such that the embargo is adopted if and only if country 1 and and at least two other countries are in favor. This voting rule can again be represented as a weighted majority rule, and we observe that country 1 benefits from a veto power under this rule, as it can block any proposal. Country 1 gets utility $U = 72/3^5 > U_1^0$, while countries 2 to 5 now get an even reduced utility of $U = 84/3^5$. 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. The following table 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).

Benchmark		Sovereignty	Efficient	Optimal	Optimal Self-Enforcing
Voting rule			Simple Majority	1 overweighted	1 has veto power
U_1	$\propto$	16	-120	16	72
$U_{2,3,4,5}$	$\propto$	32	228	146	84
Welfare	$\propto$	144	792	600	408

Table 1: Summary of the example.

1.3 Outline

The model is introduced in Section 2. The main result, deriving the *optimal* voting rule, is provided in Section 3 and is followed by a discussion on the social cost of participation constraints. Section 4 explores the condition of self-enforceability, and derives the optimal self-enforcing rule. Then, the model is applied in Section 5 to a simple environment in which states differ only in their populations.

2 Model

An international union N is made of n countries. Each country has one representative who takes decisions on the behalf of its citizens. Representatives must decide whether to remain sovereign or to cooperate, and if so, 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, and each country i derives a stand-alone utility $U_i^{\emptyset} \geq 0$. If all countries decide to cooperate, the game continues and countries have to make a decision on the adoption of a proposed reform.

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 (u_i) are independent across countries, 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.

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 countries follow a pre-determined voting rule v. To keep the model flexible, we define a voting rule as a non-decreasing function $v : \{0,1\}^N \to [0,1]$, where $v(\mathbf{m})$ denotes the probability of accepting the reform, given the vector of messages $\mathbf{m}$.[9] We denote 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 value 1 if country i implements the reform, and value 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 this paper, we particularly focus on the cooperative profile of the game, i.e. the profile 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 aggregate utility to country i associated to this profile is given by:

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

A common theme of the article will be to identify conditions for which this cooperative profile can be implemented as an equilibrium. 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

[9] This expression allows for probabilistic decisions, in order to break possible ties.

$\mathbf{m} \in \{0,1\}^N$,

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

We say that the rule v is weighted and can be represented by the vector of weights $[\mathbf{w}; t]$.[10] Among the class of weighted voting rules, some rules mechanically grant a veto power to some countries (e.g. UN Security Council). Formally, we say that a country $i \in N$ has a veto power under a rule v if $v(\mathbf{m}) = 0$ whenever $m_i = 0$. We denote by $VE(v) \subseteq N$ the set of countries having a veto power under the rule v:

$$VE(v) = \{ i \in N \quad | \quad m_i = 0 \;\Rightarrow\; v(\mathbf{m}) = 0 \}.$$

2.3 Welfare and efficient voting rule

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

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

We say that a rule is *efficient* if it achieves the maximum welfare at the cooperative profile, this 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:*

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

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 cooperative

[10] Note 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.

profile may not be an equilibrium of the decision game. Incorporating such constraints is the main goal of our paper, it is the object of the following two sections.

As we want to understand how these constraints shape optimal rules, we will assume, without substantial loss of generality, that efficient rules are well-behaved. The following assumption guarantees that no country has a veto power under an efficient rule.[11]

Assumption 1. *For any country* $i \in N$, $w_i^- < \sum_{j \neq i} w_j^+$.

3 Enforceable decisions

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 $\mathbf{m}$, whenever they cooperate. We are interested in voting rules that induce cooperation (in the first stage) at equilibrium.

Proposition 1. *The cooperative profile is a perfect Bayesian equilibrium of the game* $\Gamma_e(v)$ *if and only if each country satisfies the participation constraint:* $U_i(v) \geq U_i^{\emptyset}$ *for all* $i \in N$.

We denote by $\mathcal{PC}$ the set of voting rules satisfying the participation constraints.

3.1 Optimal Voting Rules

We look for voting rules maximizing the social welfare when participation is voluntary. We say that a voting rule is *optimal* if it is a solution[12] of the maximization problem $\max_{v \in \mathcal{PC}} W(v)$. The following theorem describes optimal voting rules.

Theorem 2. *Any optimal voting rule is a weighted majority rule. Countries for which the participation constraint is binding are overweighted relative to their efficient weight, while remaining countries receive their efficient weight.*

Any optimal voting rule v^* is such that a country i which gets strictly more than its stand-alone utility receives its efficient weight w_i^e, while countries which do not strictly benefit from cooperation may receive more than their efficient weight. We say that these countries are

[11] Indeed, under Assumption 1, we have:

$$\sum_{j \in N} w_j^- < \sum_{j \neq i} w_j^e, \text{ thus } \sum_{j \neq i} w_j^e > t^e \sum_{j \in N} w_j^e.$$

It follows from Theorem 1 that for any efficient rule v^e, $i \notin VE(v^e)$.

[12] Note that the existence of a solution is guaranteed as the objective function is linear, and the set of voting rules $\mathcal{PC}$ is a closed subset of $[0,1]^{2^N}$.

overweighted. Formally, v^* can be represented by a vector of weights $[\mathbf{w}; t]$ such that for each country $i \in N$:

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

and the threshold t is a 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. As stand-alone utilities become larger however, 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 Section 1.2, where the optimal voting rule $[(9, 3, 3, 3, 3), 1/2]$ is such that country 1 is 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 efficient weights, which can be computed independently for each country, optimal voting weights may not 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 disagrees often with the (efficient) collective decision (as in the example of Section 1.2). 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,[13] increases with each country's stand-alone utility: decreasing a country 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.

[13] That is, the difference in welfare between the efficient and the optimal rules.

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.[14] A more homogeneous union may thus induce a larger 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 decisions. In that framework, we show that inducing self-enforcing cooperation is harder than inducing cooperation under enforceability. Then, we characterize the optimal self-enforcing rule, which occasionally entails giving a veto power to some countries, but not necessarily all countries.

4.1 Repeated Game

When decisions are not enforceable, considering the one-shot game $\Gamma_{ne}(v)$ is not sufficient, as 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 infinite repeated game $\Gamma^{\delta}_{ne}(v)$. At each stage $t \in \mathbb{N}$, each country $i \in N$ decides whether to participate $d_i^t \in \{0, 1\}$. 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 $a_i^t \in \{0, 1\}$, that can differ from $\hat{v}(\mathbf{m}^t)$. At each stage, $\mathbf{d}^t$, $\mathbf{m}^t$,

[14] Consider 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 are $1/2$ 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.

$\hat{v}^t(\mathbf{m}^t)$ 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 equilibrium[15] of $\Gamma^{\delta}_{ne}(v)$ such that the cooperating profile is played at each stage of the game on the equilibrium path. 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 $d^t_i = 0$ or $a^t_i \neq \hat{v}^t(\mathbf{m}^t)$ for some t.

We observe that, under such a profile, a deviation is most profitable for a country when the committee approved a reform the country wanted to block. 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 the subsequent stages. Compared to the one-shot game, the repeated game thus creates an extra-incentive to leave the union, that can only be mitigated by giving a veto power to the country tempted to exit.

We define the maximal disutility that country i may suffer from a decision by

$$w^D_i = -\min\ \{\, w \in \mathbb{R} \mid \mathbb{P}_\mu(u_i = w) > 0 \,\}\,.$$

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

$$U_i(v) \geq U^{\emptyset}_i + \frac{1-\delta}{\delta} w^D_i.$$

Proposition 2. *A voting rule v is self-enforcing if and only if for all $i \in N$ either i has a veto power and satisfies the participation constraint, or i does not have a veto power and satisfies the compliance constraint.*

We denote by $\mathcal{SE}$ the set of self-enforcing rules. The result establishes the equivalence between the notion of self-enforceability and a set of *endogenous* constraints. Indeed, the constraint that a country i should satisfy under a rule v is contingent on i having a 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.

[15] The 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 $\Gamma^{\delta}_{ne}(v)$, as for instance in Athey and Bagwell (2001) or Maggi and Morelli (2006).

4.2 Optimal self-enforcing rules

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

Theorem 3. *Any optimal self-enforcing rule is a weighted majority rule. Countries for which the compliance constraint is not satisfied are strictly overweighted and have a 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 a veto power.*

Formally, an optimal self-enforcing rule v^* can be represented by a vector $[\mathbf{w}; t]$, such that for all $i \in N$:

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

and

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

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

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 $(1-\delta)w_i^D/\delta$.[16] Countries who fall strictly below this augmented utility threshold are strictly overweighted, while countries who fall strictly above receive their efficient weight. Second, in contrast with Theorem 2, the benchmark utility also separates countries who benefit from a veto power from countries who do not. This is illustrated in the example of section Section 1.2, where the optimal self-enforcing rule grants a 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 $16/3^5 + 2/5 \approx 0.47$, while countries 2 to 5's utility $84/3^5 \approx 0.35$ falls above their augmented utility threshold $32/3^5 + 1/5 \approx 0.33$. The fact that the optimal self-enforcing rule may grant a veto power to only a strict subset of countries is a major difference with Maggi

[16] When $\delta = 1$, this additional term equals 0 so that $\mathcal{SE} = \mathcal{PC}$.

and Morelli (2006), in which either all countries have a veto or no country has a veto, and this stems from the generality of our model which allows for heterogeneous countries.[17]

5 A Model of Apportionment

In this section, we consider a more specific model of apportionment where utilities are binary and countries differ only in their population size, which allows for more tractable results.

5.1 Model

Under sovereignty, the representative of each country chooses which reform(s) to implement. In each country, for any given reform, a (randomly chosen) fixed fraction $q > 1/2$ of citizens agrees with the representative and get a utility of 1, while the rest of the population disagrees and get a disutility of -1. Ex-ante, the utility of country i with population p_i is thus equal to:

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

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 each other. In each country, for any given reform, a fixed fraction $q > 1/2$ of citizens agrees with the opinion of its country's representative. If the reform ends up being implemented effectively (by all countries), favorable citizens get a utility of e, while unfavorable citizens get a disutility of $-e$. The parameter e can be interpreted as a per-capita efficiency gain from cooperation, and we assume that $e > 1$.[18] If the reform is not adopted effectively, all citizens get a utility of 0. The probability distribution μ associated to this model is such that:

$$\forall i \in N, \qquad \mathbb{P}_\mu\left(u_i = (2q-1)ep_i\right) = \mathbb{P}_\mu\left(u_i = -(2q-1)ep_i\right) = \frac{1}{2}.$$

The efficient weight 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 $t_i^e = 1/2$.

[17] Note that the possibility of having only a strict subset of veto countries at the optimal self-enforcing rule does not hinge on countries having biased preferences (as assumed in the example of Section 1.2). For example consider $N = \{1, 2, 3, 4, 5\}$, μ such that $\mathbb{P}_\mu(u_1 = 2) = \mathbb{P}_\mu(u_1 = -2) = \mathbb{P}_\mu(u_{2-5} = 1) = \mathbb{P}_\mu(u_{2-5} = -1) = 1/2$, $U_1^{\emptyset} = 0.8$ and $U_{2-5}^{\emptyset} = 0$. Then for $\delta = 0.95$, an optimal self-enforcing rule is such that a proposal is accepted (i) with probability 1 whenever country 1 and at least 2 of the remaining countries are in favor and (ii) with probability 0.45 whenever country 1 and one of the remaining countries are in favor. That voting rule gives country 1 a veto power.

[18] Note 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 paper, 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.

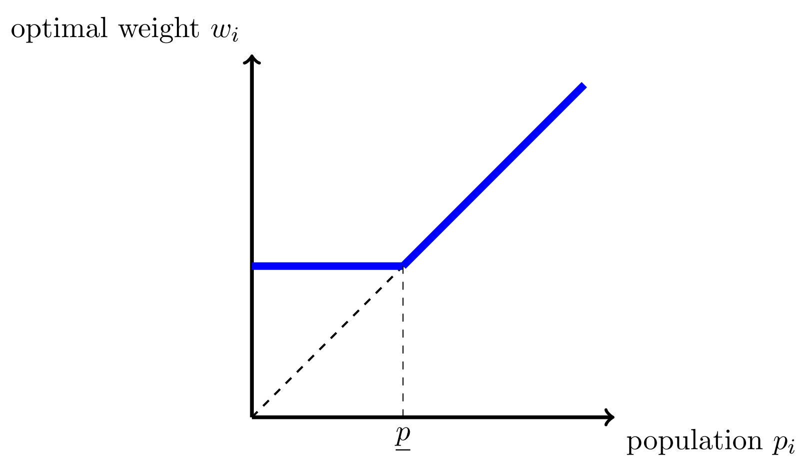
In that more specific model, countries thus vary in their stakes w_i^e, which are proportional to their population p_i, but are otherwise identical.[19] In particular, they have the same ex-ante probability of agreeing with a given reform, equal to 1/2. Note that Assumption 1 boils down to $p_i < (\sum_{j \in N} p_j)/2$ for all $i \in N$, which means that any country accounts for less than half of the total population.

5.2 Optimal Voting Rules

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

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

Figure 1: Optimal voting weights in the model of apportionment



The optimal apportionment rule is illustrated in Figure 1. We first note that the distribution of weights is *degressively proportional*: weights increase with country's populations, but less than proportionally. A sizable literature on apportionment has already argued in favor of this property,[20] 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,

[19] Note that we have assumed that the per-capita efficiency gain from cooperation, e, is the same for all countries. This assumption is maintained throughout the section to keep the interpretation of the results simple. However, with varying gains e_i, all the results remain valid by replacing p_i by $e_i p_i$ in the subsequent Theorem 4 and Theorem 5.

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

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 $1/2 \leq \alpha \leq 1$.[21]

The requirement that smaller countries shall be given a minimal and equal representation is actually found explicitly in the Treaty Lisbon, which specifies a set of constraints for the composition of the European Parliament.[22] Indeed, article 14.2 states that "Representation of citizens [at the European Parliament] shall be degressively proportional, with a minimum threshold of six members per Member State" (Treaty of Lisbon, 2007a). Our paper thus offers a theoretical rationale for such 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 US Electoral College where each state is allocated a baseline of 2 seats plus a number of seats proportional to its population. The 8 smaller states are allocated the same number of 3 seats,[23] 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 at the European Parliament, under the name of *Cambridge Compromise*.[24]

5.3 Optimal Self-Enforcing Rules

We also obtain sharper predictions for the optimal self-enforcing rule: either no country has a veto power or all countries have it, and we can map these two cases on a graph parametrized by the per-capita 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 a veto power. There exists a threshold $\underline{e} > 0$ and two non-increasing functions $\delta^c, \delta^{eff} : \mathbb{R}_+ \to \mathbb{R}_+$, with for all $e \in \mathbb{R}_+$, $\delta^c(e) \leq \delta^{eff}(e)$ and $\lim_{e\to\infty} \delta^{eff}(e) < 1$, such that:*

(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,

[21] For 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).

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

[23] Alaska, Delaware, District of Columbia, Montana, North Dakota, South Dakota, Vermont, Wyoming.

[24] The 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 at the European Parliament. The proposed allocation is based on a similar base + prop formula as in the US Electoral College, whereby each country is allocated a base of 6 seats plus a number of seats proportional to its population. See Grimmett (2012).

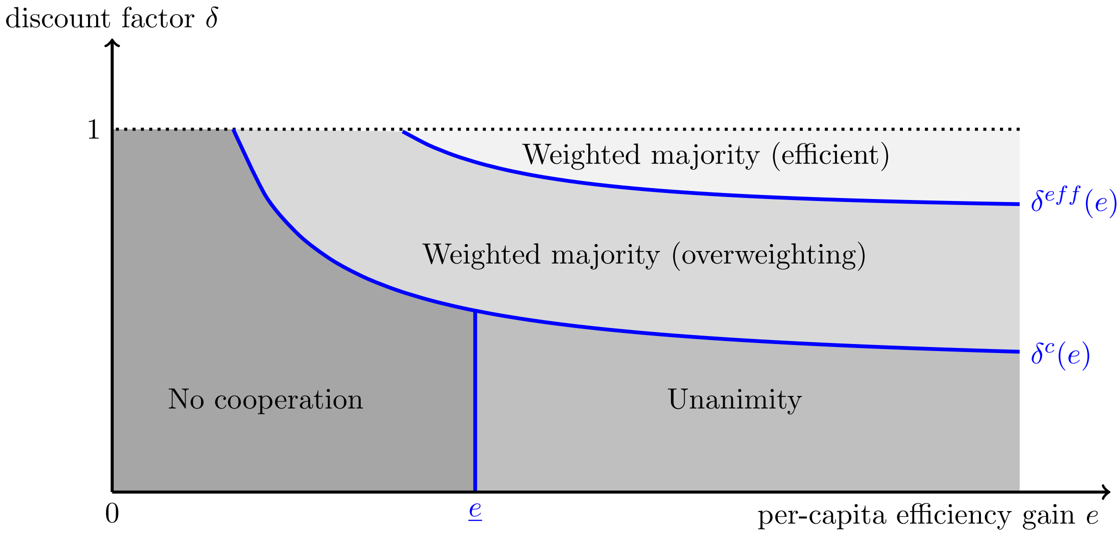
(iii) if $\delta < \delta^c(e)$ and $e \geq \underline{e}$, the optimal self-enforcing rule the is unanimity rule,

(iv) if $\delta < \delta^c(e)$ and $e < \underline{e}$, 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 an optimal self-enforcing rule is represented by $[\mathbf{w}; 1/2]$, defined by: for all $i \in N$, $\quad w_i = \max\Big(p_i, \underline{p}(e, \delta)\Big)$.

Theorem 5 defines four regions in the space (e, δ), that yield different (or no) optimal self-enforcing rules, as represented in Figure 2 below.

Figure 2: 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 per-capita efficiency gain e is too small, there is no rule inducing cooperation. If the per-capita 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 2 vertically reveals how Theorem 5 extends the main result of Maggi and Morelli (2006). In that paper, countries are homogeneous, and there exists a threshold $\overline{\delta}$, 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 $e \geq \underline{e}$, there are two thresholds: $\delta^{eff}(e)$ and $\delta^c(e)$. As in the homogeneous model, the efficient rule is the optimal self-enforcing rule when the discount factor is high ($\delta \geq \delta^{eff}(e)$), and the unanimity rule is the optimal self-enforcing rule when the low 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 a weighted majority rule with overweighting of small countries. Moreover, the extent to which small countries are overweighted decreases with the per-capita efficiency gain e and with the discount factor δ.

References

Alesina, A., I. Angeloni, and F. Etro (2005): "International Unions," *American Economic Review*, 95, 602–615.

Alesina, A. and E. Spolaore (1997): "On the number and size of nations," *The Quarterly Journal of Economics*, 112, 1027–1056.

Athey, S. and K. Bagwell (2001): "Optimal collusion with private information," *RAND Journal of Economics*, 428–465.

Azrieli, Y. and S. Kim (2014): "Pareto efficiency and weighted majority rules," *International Economic Review*, 55, 1067–1088.

Balinski, M. L. and H. P. Young (1982): *Fair representation: meeting the ideal of one man, one vote*, Yale University Press.

Barberà, S. and M. O. Jackson (2004): "Choosing how to choose: Self-stable majority rules and constitutions," *The Quarterly Journal of Economics*, 1011–1048.

——— (2006): "On the weights of nations: assigning voting power to heterogeneous voters," *Journal of Political Economy*, 114, 317–339.

Beisbart, C. and L. Bovens (2007): "Welfarist evaluations of decision rules for boards of representatives," *Social Choice and Welfare*, 29, 581–608.

Bouton, L., A. Llorente-Saguer, and F. Malherbe (2014): "Get rid of unanimity: The superiority of majority rule with veto power," Tech. rep., National Bureau of Economic Research.

Bown, C. (2004): "On the Economic Success of GATT/WTO Dispute Settlement," *The Review of Economics and Statistics*, 86, 811–823.

Brighouse, H. and M. Fleurbaey (2010): "Democracy and proportionality," *Journal of Political Philosophy*, 18, 137–155.

Casella, A. (1992): "Participation in a Currency Union," *American Economic Review*, 82, 847–63.

Crémer, J. and T. R. Palfrey (1996): "In or out?: Centralization by majority vote," *European Economic Review*, 40, 43–60.

Felsenthal, D. S. and M. Machover (1999): "Minimizing the mean majority deficit: The second square-root rule," *Mathematical Social Sciences*, 37, 25–37.

Grimmett, G. R. (2012): "European apportionment via the Cambridge Compromise," *Mathematical Social Sciences*, 63, 68–73.

Koriyama, Y., J.-F. Laslier, A. Macé, and R. Treibich (2013): "Optimal apportionment," *Journal of Political Economy*, 121, 584–608.

Laslier, J.-F. (2012): "Why not proportional?" *Mathematical Social Sciences*, 63, 90–93.

Loeper, A. (2011): "Coordination in heterogeneous federal systems," *Journal of Public Economics*, 95, 900–912.

Maggi, G. and M. Morelli (2006): "Self-enforcing voting in international organizations," *The American Economic Review*, 96, 1137–1158.

Moravcsik, A. (1998): *The choice for Europe: social purpose and state power from Messina to Maastricht*, Cornell University Press.

Napel, S., N. Maaser, and S. Kurz (2016): "On the Democratic Weights of Nations," *Journal of Political Economy*.

Penn, E. M. (2015): "Engagement, Disengagement, or Exit: A Theory of Equilibrium Associations," *American Journal of Political Science*.

Penrose, L. S. (1946): "The elementary statistics of majority voting," *Journal of the Royal Statistical Society*, 109, 53–57.

Rae, D. W. (1969): "Decision-rules and individual values in constitutional choice," *American Political Science Review*, 63, 40–56.

Renou, L. (2011): "Group formation and governance," *Journal of Public Economic Theory*, 13, 595–630.

Treaty of Lisbon (2007a): "Consolidated version of the Treaty on European Union - Title I: Common Provisions," .

——— (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, 943–956.

6 Appendix: proofs

6.1 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.

6.2 Proof of Theorem 2

In this proof, and in the subsequent proofs, we abuse notation and we 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 v^* be an optimal voting rule. We claim that v^* is a solution of the following maximization problem (note that in the problem, the function v is not assumed to be non-decreasing):

$$(\mathcal{P}) : \left| \begin{array}{l} \max\limits_{\{v(M)\}_{M \subseteq N} \in [0,1]^{(2^N)}} \sum\limits_{i \in N} U_i(v) \\ \text{s.t.} \quad \forall i \in N, \quad U_i(v) \geq U_i^{\emptyset}. \end{array} \right.$$

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

Note that the probability distribution μ defines a probability distribution P on the coalition M of countries favoring the reform (under truthful voting), formally:

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

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

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

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 $(\lambda^i, \eta^M, \nu^M)$ such that

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

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

$$\begin{cases} \eta^M - \nu^M < 0 & \Rightarrow \nu^M > 0 \quad \Rightarrow v^*(M) = 1 \\ \eta^M - \nu^M > 0 & \Rightarrow \eta^M > 0 \quad \Rightarrow v^*(M) = 0. \end{cases}$$

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

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

We conclude by setting $w_i = (1+\lambda^i)w_i^e$ and

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

6.3 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.

Assume first that the constraints are satisfied under the rule v. We consider the following profile of strategies:

- play at each step t the cooperating profile of the game $\Gamma_{ne}(v)$
- if exactly one country is observed to deviate at time t (either $d_i^t = 0$ or $a_i^t \neq \hat{v}^t(\mathbf{m}^t)$), then choose sovereignty at each 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 $d_i^0 = 1$, if 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 changes what happens at subsequent stages at 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 $\hat{v}(\mathbb{1}_{u_i>0}, \mathbf{m}_{-i})u_i + \dfrac{\delta}{1-\delta}(U_i(v) - U_i^{\emptyset})$ (because of the trigger strategies). This deviation can only be profitable if $u_i < 0$, and we distinguish two cases:

- if i has a veto, as $\mathbb{1}_{u_i>0} = 0$, 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 $\hat{v}(\mathbb{1}_{u_i>0}, \mathbf{m}_{-i})u_i + \dfrac{\delta}{1-\delta}(U_i(v) - U_i^{\emptyset}) \geq -w_i^D + \dfrac{\delta}{1-\delta}(U_i(v) - U_i^{\emptyset}) \geq 0$, because i satisfies the compliance constraint. Therefore, the deviation is not profitable.

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

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

6.4 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 $\max_{v \in \mathcal{SE}} W(v)$, which is equivalent to $\max_{V \subseteq N} \max_{v \in \mathcal{SE}, VE(v) = V} W(v)$. Therefore, v^* is solution of
$\max_{v \in \mathcal{SE}, VE(v) = V^*} W(v)$, and by the argument made at the beginning of the proof of Theorem 2, this problem is equivalent to:

$$(\mathcal{P}^{V^*}) : \left| \begin{array}{l} \max_{\{v(M)\}_{V^* \subseteq M \subseteq N}} \sum_{i \in N} U_i(v) \\ s.t. \quad \forall i \in V^*, \quad U_i(v) \geq U_i^{\emptyset} \\ s.t. \quad \forall i \in N \backslash V^*, \quad U_i(v) \geq U_i^{\emptyset} + \dfrac{1-\delta}{\delta} w_i^D. \\ s.t. \quad \forall M, \quad V^* \nsubseteq M \quad \Rightarrow \quad v(M) = 0. \end{array} \right.$$

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 $t^0 = \dfrac{\sum_{i \in N} w_i^0 t_i^e}{\sum_{i \in N} w_i^0}$, we obtain:

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

Moreover, we know that $V^* \nsubseteq M \;\; \Rightarrow \;\; v^*(M) = 0$. Now, we define $w_i = w_i^0 + K\mathbb{1}_{i \in V^*}$, 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 $V^* \nsubseteq M$ or

$\sum_{i\in M} w_i^0 < t^0 \sum_{i\in N} w_i^0$, and therefore $v^*(M) = 0$. Finally, we define

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

Finally, we have obtained that:

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

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

Finally, let $i \in N$ be a country such that $U_i(v^*) < U_i^{\emptyset} + \dfrac{1-\delta}{\delta} w_i^D$. 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 $U_i(v^*) > U_i^{\emptyset} + \dfrac{1-\delta}{\delta} w_i^D$, we will show that i has no veto power under v^*. By contradiction, suppose that i has a veto power: we have in particular $v^*(N\backslash\{i\}) = 0$. For $\varepsilon > 0$, consider now v^ε defined by:

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

We have $U_i(v^\varepsilon) = U_i(v^*) - \varepsilon P(N\backslash\{i\}) w_i^-$ and $\forall j \neq i,\ U_j(v^\varepsilon) = U_j(v^*) + \varepsilon P(N\backslash\{i\}) w_j^+$. By Assumption 1, we have $W(v^\varepsilon) > W(v)$. Moreover, v^ε is self-enforcing for ε small enough, hence a contradiction. We obtain that $i \notin VE(v^*)$. It follows that $w_i = w_i^0 = w_i^e$.

6.5 Proof of Theorem 4

We have:

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

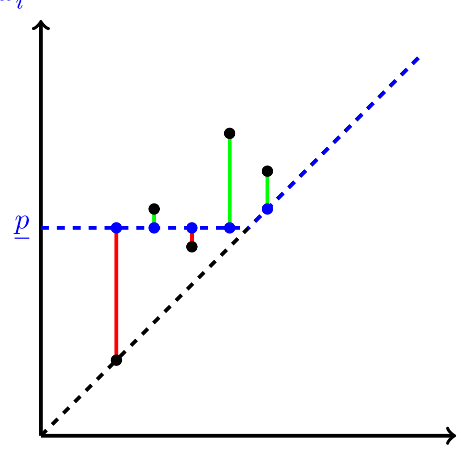
and $U_i^\emptyset = (2q-1)p_i$. Let v^* be an optimal rule. We know from Theorem 2 that v^* is a weighted majority rule with threshold $t = \frac{1}{2}$, and with weights $(w_i)_{i\in N}$ satisfying[25]

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

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 $\underline{p} \in \mathbb{R}$ (as illustrated in Figure 3):

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

Figure 3: Definition of $\underline{p}$ (total red length=total green length)



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 $(w_i \to w_i + \alpha, w_j \to w_j - \alpha)$ 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 $[\mathbf{w}'; t]$,

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

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 T be a coalition such that $\sum_{k\in T} w'_k > Q$. We have two cases:

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

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

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

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

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

  We obtain:

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

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

Similarly, one can show that $\sum_{k\in T} w'_k < Q$ implies $v^*(T) = 0$. Finally, v^* is represented by $[\mathbf{w}'; t]$. By induction, v^* 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 $w_i^* = \max(p_i, \underline{p})$.

6.6 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:

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

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, represented by $[\mathbf{w}; 1/2]$.

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:

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

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:

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

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, represented by $[\mathbf{w}; 1/2]$. 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, we get from Claim 1 that the optimal self-enforcing rule is a weighted majority rule, represented by $[\mathbf{w}; 1/2]$.

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.

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

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

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 $\lim_{e\to\infty} \delta^{eff}(e) < 1$.

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, there exists an optimal self-enforcing rule represented by $[\mathbf{w}; 1/2]$, with for all $i \in N$, $w_i = \max\big(p_i, \underline{p}(e,\delta)\big)$.[26]

[26] The claim may seem obvious, as increasing e and/or δ relaxes the self-enforcing constraints. Note however

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

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

Following the proofs of Theorem 2 and Theorem 5, we can show that any solution v of $(\mathcal{P}^k)$ is a weighted rule represented by $[\mathbf{w}; 1/2]$ with:

- for all $i \in N$, $w_i = \max(p_i, \underline{p})$
- $\underline{p}$ is the solution[27] of $\sum_{i\in N} \max(p_i, \underline{p}) = \sum_{i\in N} p_i(1+\lambda^i)$
- for all $i \in N$, λ^i is the Lagrangian coefficient associated to the i's constraint in 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 increases with k. The linear program $(\mathcal{P}^k)$ can be re-written as follows (we multiplied each i's constraint by a factor $-p_i$):

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

The dual of $(\mathcal{P}^k)$ is the following linear program:

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

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

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.

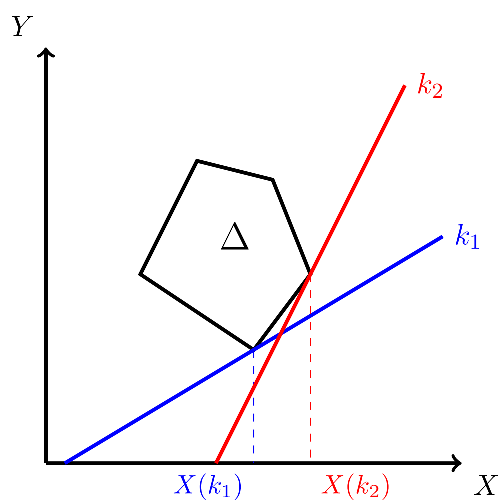
[27] To be precise, $\underline{p}$ is defined in the proof of Theorem 5 as the solution of $\sum_{i\in S^0} \max(p_i, \underline{p}) = \sum_{i\in S^0} p_i(1+\lambda^i)$, and it is shown at the end of the proof that for all $i \in S^1 = N \backslash S^0$, $\max(p_i, \underline{p}) = p_i = p_i(1+\lambda^i)$.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

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 below with $X = \sum_{i\in N} p_i \lambda^i$. The reduced program writes:

$$(\mathcal{D}^k)' : \left| \begin{array}{ll} \min_{X,Y} Y - kX & \\ s.t. & (X,Y) \in \Delta \\ s.t. & X \geq 0 \\ s.t. & Y \geq 0, \end{array} \right.$$

where $\Delta \subset \mathbb{R}^2$ is a convex polyhedron. It is clear (see Figure 4) that X increases as k increases, in the following sense: there exists a non-decreasing function $X(k)$ such that for each k, $X(k)$ is the first coordinate of a solution of $(\mathcal{D}^k)'$.

Figure 4: Solutions of $(\mathcal{D}^{k_1})'$ and $(\mathcal{D}^{k_2})'$ for $k_2 > k_1$.



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