Stable Consitituitions

Daeyoung Jeong* Semin Kim†

November 5, 2018

Abstract

This study identifies a set of stable constitutions. A constitution is a pair of voting rules (f, F) where f is for the choice of final outcome, and F is for the decision on the change of a voting rule from the given rule f. A constitution is stable if any possible alternative rule does not get enough votes to replace the given rule f under the rule F. We fully characterize the set of interim stable constitutions among anonymous voting rules. We also characterize the properties of the interim stable constitutions among general weighted majority rules.

JEL Classification: C72, D02, D72, D82.

Keywords: Weighted majority rules, decision rules, self-stability

*New York University Abu Dhabi; daeyoung.jeong@gmail.com

†School of Economics, Yonsei University, South Korea; seminkim@yonsei.ac.kr

Contents

1 Introduction 3

1.1 Related Literature . . . 6

2 Definitions 8

2.1 Environment . . . 8

2.2 Voting Rules and Constitutions . . . 8

2.3 Interim Stability . . . 9

3 Interim Stable Constitution 14

3.1 Qualified Majority Rules . . . 14

3.2 Weighted Majority Rules . . . 17

3.3 Environment Independence . . . 20

4 Discussion 21

4.1 Fixed weights . . . 21

4.2 Ex ante Stability . . . 23

4.3 Set of alternative rules and constitutions . . . 25

A Appendix 28

A.1 Proofs . . . 28

A.2 Consistency of Definition . . . 35

A.3 Extra Lemmas and Propositions . . . 38

A.3.1 Super Majority Rules . . . 45

1 Introduction

Since different voting rules may result in different voting outcomes, the welfare of an individual voter may depend not only on her preference over possible voting outcomes, but also on a voting rule they use. So, it is natural that a voter with a certain preference over possible voting outcomes also has a preference over different voting rules: For example, a voter who prefers a status quo to a reform may prefer unamimity rule to simple majority rule. When the change on voting rule is possible in a certain way, one individual may want to change the voting rule according to her interest, but another may not. A rule change in a society is, therefore, closely tied up with individuals' preferences over voting outcomes and voting rules.

This study identifies a set of interim stable constitution. A constitution is a pair of voting rules (f, F) where f is for the choice of final outcome, and F is for the decision on the change of a voting rule from the given rule f. A constitution is stable if any possible alternative rule does not get enough votes to replace the given rule f under the rule F. Unlike the previous studies on stable voting rules (Barberà and Jackson, 2004; Azrieli and Kim, 2016), which assume that the decision on the change of a voting rule takes place before individuals' preferences over possible outcomes have been realized, we assume that individuals' preferences have been realized even before making any decision.[1]

On many occasions, voters are aware of their own preferences over economic outcomes even before they decide on a voting rule to use. For example, a legislature, who are aware of the characteristics of an upcoming bill, may have a chance

[1]Our study is an obvious generalization of Holmström and Myerson (1983), which assume individuals with realized preferences over possible-https://www.overleaf.com/project/5b90db18049e9f1591843ebe voting outcomes make decisions on changing the decision rule under unanimity.

to change their decision rule beforehand. More specifically, consider the 'recent' changes in the United States Senate regarding the Rule XXII of the Standing Rules. The Rule XXII states the supports of "three-fifths of the Senators duly chosen and sworn" are required to stop a debate or 'filibuster' over a proposal at hand, so as for the Senate to proceed to a final vote on it.[2] At the same time, the Senate requires "two-thirds of the Senators presenting and voting" to amend its own rules. These rules, which have stood for a long time since 1975, have recently been challenged by another parliamentary procedure called the "Nuclear Option." Once invoked by the majority leader, this option can basically change the rule for amending the Senate rules from two-third supermajority to simple majority. In 2013, the Nuclear Option was invoked, and the U.S. Senate can have successfully changed the rule for stopping a filibuster against all executive branches and judicial nominees other than the Supreme Court of the United States from three-fifth supermajority to simple majority. Once again in 2017, the Nuclear Option was invoked, and the Senate abolished the exception on the Supreme Court nomination. Thus, now in the U.S. Senate, a debate on a proposal regarding any nomination can be ended only with simple majority.

Then, why have the 'long-lived' three-third supermajority rule of the Rule XXII been challenged recently? We take note of two aspects of these historical incidents. First, for the incidents, the voting rule for amending the Senate rules has been changed from two-third supermajority to simple majority: The former is stronger than the given rule, three-third supermajority, but the latter is weaker. That is, the given rule f had stably been survived when the voting rule for a rule

[2] Thus, given this rule, more than two-fifths of senators (40 out of 100 senators) can filibuster against any proposal: A senator or senators can speak and debate indefinitely to prevent a final vote on a proposal.

change F was stronger than itself, but has been replaced by a weaker rule when F had been weaker than itself. That is, the relative strength of rules may play an important role in the concept of stability. Second, the senators had been aware of the their own preferences over the proposals at hand even before they decided on a voting rule to use. In 2013, the U.S. Senate with a Democratic majority had lead the change to approve the nomination to the United States Court of Appeals for the District of Columbia Circuit, and in 2017, the Senate with a Republicatn majority lead the change to approve the nomination to Supreme Court. So, the Senate's preferences on the nominations has been reflected on the rule changes. That is, in this situation, the voting body may make a decision on the rule change strategically based on their preferences. In this study, we discuss the significance of the above mentioned factors in the concept of stability: the relative strength of voting rules and the timing of a rule change decision.

This study helps us to explore the implications by comparing the set of interim stable decision rules with the set of ex-ante stable decision rules, which has been identified by previous studies. A voting rule may be interim stable, but not ex-ante, or possibly vice versa. The comparison, therefore, allows us to examine the impact of the timing of the rule change on the stability of decision rules, and also the feasible set of economic outcomes.

In analysis, we first start with the simplest possible case where all voters have the same voting power under any voting rule. We show that a constitution is interim stable if and only if 1) the voting rule for the ordinary decision is not stronger than the voting rule for the change of the rule, and 2) the combining toughness of those two rules should be higher then a certain level. Intuitively, the first condition prevents liberal voters from forming a coalition to change the given rule to a weaker voting rule: If there are enough number of liberal voters so that they can change the given rule to a weaker one, they don't have to change it since the given rule is weak enough for them to achieve the liberal outcome. On the other hand, the second condition prevents conservative voters from forming a coalition to change the given rule to a stronger voting rule: If there are enough number of conservative voters so that they can change the given rule to a stronger rule, they don't have to change the rule since the given rule is strong enough to prevent the voting body from implementing the liberal outcome.

We generalize our analysis further by considering a set of weighted majority rules where individual voters may have different voting powers. Notwithstanding the asymmetry of voters under a general weighted majority rule makes the analysis considerably complicated, we can still obtain meaningful conditions that characterize the set of stable weighted majority constitutions.

1.1 Related Literature

The two papers, Barberà and Jackson (2004) and Holmström and Myerson (1983) motivate this project. Barberà and Jackson (2004) introduce the ex-ante self-stability of voting rules and focus on the qualified majority rules. Unlike them, we define the interim self-stability of voting rules and study not only the qualified majority rules but also general voting rules. The interim self-stability is similar to the durability of decision rules defined by Holmström and Myerson (1983) in that an agent utilizes the preferences information in the interim stage. While they use the unanimous rule to choose between rules, we start with the given rule itself and try to extend the argument with the various rules. It can show the effects of those variations on the set of stable rules.

In our model, agents' preferences over voting rules are endogenously determined from their assessments regarding their preferences over alternatives. Such a model was first suggested in early papers by Rae (1969), Badger (1972), and Curtis (1972). While these papers only consider anonymous voting rules with the same weight to all agents, we study weighted majority rules which allow the heterogenous weights for agents.

The seminal book of Neumann and Morgenstern (1953, Section 5) theoretically investigates weighted majority rules. The main interest of the book is the measures of the voting power of agents under the rule. A common scenario leading to heterogeneous voting weights is that of a representative democracy with heterogenous district sizes. An early paper on this topic is Penrose (1946). Recently, Barberà and Jackson (2006) and Fleurbaey (2008) point out the advantage of weighted majority rules from a utilitarian point of view. Also, Azrieli and Kim (2014) show that, in a standard mechanism design setup, weighted majority rules naturally arise from considerations of efficiency and incentive compatibility. We investigate another property, the stability of weighted majority rules.

The idea that the same voting rule used to choose between alternatives is also used to choose between voting rules can be found in the social choice literature. Koray (2000) introduces the concept of self-selection for social choice functions. See also Barberà and Beviá (2002) and Koray and Slinko (2008).

2 Definitions

2.1 Environment

A society faces a binary decision whether to implement the Reform (R) or to keep the Status-quo (S), so the set of alternatives is $A = \{R, S\}$. In the society, there are $n \geq 2$ agents (voters), $N = \{1, 2, ..., n\}$. Each agent can either prefer R or S, which indicates the type of the agent, $t_i \in T_i = \{r, s\}$. The probability of agent i being a type t_i is $p_i(t_i)$ and $p_i(t_i = r) + p_i(t_i = s) = 1$. We assume that there is no agent who is indifferent between R and S, and that $p_i(t_i) > 0$ for any $t_i \in T_i$. Let $T = T_1 \times \cdots \times T_n$ be the set of type profiles. We assume that types are independent across agents, so let $P(t) = \underset{i\in N}{\Pi} p_i(t_i)$ be the probability of a type profile $t \in T$. For the technical convenience, we abuse the notation, $P(t_{-i}) = \frac{p(t)}{p_i(t_i)}$ for the probability of a type profile of other agents excluding agent i.

An agent's utility depends on the chosen alternative and on his own type, $u_i : A \times T_i \to \mathbb{R}$. We normalize the utility such that $u(R, r) = a$, $u(R, s) = -1$, and $u(S, r) = u(S, s) = 0$. Thus a society can be characterized by the pair (p_r, a), where $p_r = (p_1(r), ..., p_n(r))$.

2.2 Voting Rules and Constitutions

A voting rule is any mapping $f : T \to [0, 1]$, with the interpretation that, $f(t)$ is the probability that the reform R is chosen when the agents' type profile is $t \in T$. The set of voting rules is $\overline{\mathbf{G}}$. We mainly focus on an important subset $\mathbf{G} \subset \overline{\mathbf{G}}$, the set of weighted majority rules. We refer readers to Azrieli and Kim (2014) for a discussion of importance of these rules based on Pareto efficiency. The formal definition is following.

Definition 1 (Weighted Majority Rule)**.**

The voting rule f is a *Weighted Majority Rule* if there are non-negative weights $w^f = \left(w_1^f, ..., w_n^f\right)$ and a quota $0 \leq q^f < \sum_{i \in N} w_i^f$ such that

$$f(t) = \begin{cases} 1 & if \quad \sum_{\{i:t_i=r\}} w_i^f > q^f \\ 0 & if \quad \sum_{\{i:t_i=r\}} w_i^f \leq q^f. \end{cases}$$

And a weighted majority rule f is denoted by $\left(w^f, q^f\right)$.

We define a constitution as a pair of weighted majority rules.

Definition 2 (Constitution)**.**

A *Constitution* is a pair of weighted majority rules $(f, F) \in \mathbf{G} \times \mathbf{G}$ where f is for the choice of final outcome, S or R and F is for the choice of rules, a given rule f or an alternative rule g.

Note that $F = (w^F, q^F)$ is also the weighted majority rule for the choice of f or g. To economize on notation, we drop the superscripts, so write $F = (w, q)$. Also we consider a given rule f as Status quo and g as Reform for F. The corresponding type is based on the preference over f and g which is endogenously determined in our voting game. The detailed implementation of F in the voting game will be explained in the next section.

2.3 Interim Stability

We now define the concept of interim stability of a constitution (f, F) with a two-stage voting game, Γ. Timing of the game Γ is as follows. In the first stage, individual voters observe their own type t_i. Then under a rule $F = (w, q)$, agents play a simultaneous voting game whether to vote for the incumbent rule f or to vote for the alternative rule g. The alternative rule g would replace the incumbent rule f if $\sum w_i > q$, where the sum of weights is taken over all voters who vote for g, and f would be maintained otherwise. In the second stage, agents make a decision on $A = \{R, S\}$ by the rule chosen in the first stage, either f or g. To restrict our focus on the decision on rule choice, we assume that voters act sincerely in the second stage: r-type votes for R and s-type for S.[3]

Roughly, we say a constitution (f, F) is interim stable if we can find a Nash equilibrium for any alternative voting rule g such that the rule g would never replace the incumbent rule f when the decision is made by the rule F. The rest of this section formally defines the interim stability of a constitution.

Let $\sigma_i(t_i)$ be the probability that Agent i would vote for g in the first stage when her type is t_i. To reject the alternative rule g all the time, the probability that g gets sufficient support should be zero for all $t \in T$. In other words, the alternative g is always rejected if and only if

$$\sum_{\{j:\sigma_j(t_j)>0\}} w_j \leq q, \quad \forall t \in T. \tag{C.1}$$

If Condition (C.1) holds, then the voting strategies in the first stage, $\sigma = (\sigma_i)_i^n$, together with sincere voting under f and g at the second stage, form a Nash equilibrium if and only if

$$\sum_{t_{-i}} P(t_{-i})\, \gamma_i(t_{-i}) u_i(f(t), t_i) \geq \sum_{t_{-i}} P(t_{-i})\, \gamma_i(t_{-i}) u_i(g(t), t_i) \quad \forall i, \;\; \forall t_i \in T_i, \tag{C.2}$$

[3] In our voting game, the sincere voting strategy is a weakly dominant strategy for any agent.

where

$$\Phi_i = \{H_i \subseteq N \setminus \{i\} |\, q - w_i < \sum_{j \in H_i} w_j \leq q\},$$

and

$$\gamma_i(t_{-i}) = \sum_{H_i \in \Phi_i} \left(\prod_{j \in H_i} \sigma_j(t_j) \right) \left(\prod_{j \in N/(H_i \cup \{i\})} (1 - \sigma_j(t_j)) \right).$$

Agent i is pivotal if the agents in $H_i \in \Phi_i$ vote for the alternative rule g and all others $j \notin H_i$ vote for the incumbent rule f. $\gamma_i(t_{-i})$ is the voter i's probability of being pivotal given the other agents' strategies and types, σ_{-i} and t_{-i}. Therefore, Condition (C.2) implies that either Agent i is never pivotal, or she is expected to be weakly better off under f than g.

Note that, in a simultaneous voting game, there generally exists a trivial Nash equilibrium in which no agent votes for g, unless an agent has the dictatorial power under F. In such an equilibrium, where the condition (C.1) and (C.2) are satisfied, f defeats any alternative rule g. So, in order to define a reasonable concept of interim stability, we refine the equilibria of the game Γ by requiring a type of sequential rationality for agents' voting strategy profile σ given that they share a consistent 'posterior' belief.

Similarly to Holmström and Myerson (1983), we assume that agents have some apprehensions for being pivotal due to others' possible mistakes in voting. That is, even though an agent is never pivotal, she may still have some posterior belief over the other agents' types if she happens to be pivotal due to others' mistakes in voting. We characterize a posterior belief given that Agent i is pivotal as follows.

$$\mu_i(t_{-i}) = \lim_{k\to\infty} \frac{P\left(t_{-i}\right)\sum_{H_i\in\Phi_i}\rho(H_i; t_{-i}, \sigma^k)}{\sum_{\hat{t}_{-i}\in T_{-i}} P\left(\hat{t}_{-i}\right)\sum_{H_i\in\Phi_i}\rho(H_i; \hat{t}_{-i}, \sigma^k)} \tag{C.3}$$

$$\forall i, \forall t_i \in T_i, \forall t_{-i} \in T_{-i},$$

where

$$\rho(H_i; t_{-i}, \sigma^k) = \left(\prod_{j\in H_i}\sigma_j^k(t_j)\right)\left(\prod_{j\in N\setminus(H_i\cup\{i\})}(1-\sigma_j^k(t_j))\right)$$

$$\sigma_j^k(t_j) > 0 \quad \forall k, \forall j, \forall t_j \in T_j$$

$$\sigma_j(t_j) = \lim_{k\to\infty}\sigma_j^k(t_j) \quad \forall j, \forall t_j \in T_j$$

Since the limit of denominator of Condition (C.3), which represents i's probability of being pivotal given σ, could be zero, we characterize the distribution in the style of the trembling hand model. Agent i believes that, when she is pivotal, the others' type profile is t_{-i} with the probability $\mu_i(t_{-i})$ given their mistakes σ_{-i}^k.

Given this belief, we require that, for any type t_i of any agent i,

$$\text{if } \ \sigma_i(t_i) = 0, \tag{C.4}$$

$$\text{then } \sum_{t_{-i}}\mu_i(t_{-i})u_i(f(t), t_i) \geq \sum_{t_{-i}}\mu_i(t_{-i})u_i(g(t), t_i).$$

Condition (C.4) imposes that, conditional on that the agent i is pivotal, if she never votes for g, then she is expected to be weakly better off under f than g. In other words, this condition prevents an agent who strictly prefers the alternative g from voting against it.

One may notice that the form of Condition (C.2) is similar with Condition (C.4). We here argue that if we have Condition (C.1) and (C.4) together, Condition (C.2) is redundant. Condition (C.1) implies that, if some agent i is pivotal, then $\sigma_i(s) = \sigma_i(r) = 0$. By construction, if some agent i is pivotal, then the limit of the denominator of Condition (C.3) is positive and the numerator is $P(t_{-i}) \times \gamma(t_{-i})$. Therefore, if some agent i is pivotal, Condition (C.2) is equivalent to Condition (C.4). In addition, if some agent is not pivotal, Condition (C.2) does not matter. Therefore, from now on, we can ignore Condition (C.2) and focus only on Condition (C.1), (C.3) and (C.4).

Now, we define the concept of equilibrium rejection.

Definition 3 (Equilibrium rejection)**.**
Consider a constitution (f, F). A strategy profile and a belief (σ, μ) consists an equilibrium rejection of g under F if Condition (C.1), (C.3) and (C.4) are all satisfied.

Now, we formally define the interim stability of a constitution (f, F) by using the above definition. We consider the competition between rules depending on the size of the set of alternative rules. Generally, lager the set of competing alternative rules is, smaller the set of interim stable constitution is. So, we define the interim stability in a specific set of the alternative rules, denoted by $\mathbf{S} \subseteq \overline{\mathbf{G}}$.

Definition 4 (Interim Stability)**.**
Let $\mathbf{S} \subseteq \overline{\mathbf{G}}$ be given. A constitution (f, F) is interim stable in $\mathbf{S}$ if there exists an equilibrium rejection of any alternative rule $g \in \mathbf{S}$ under F.

Barberà and Jackson (2006) focused on a special type of constitution where the society uses the same voting rule on the decisions of the rule change and the final outcome, so $F = f$. Similarly we can define the interim "self-stability" of a voting rule as follows.

Definition 5 (Interim Self-Stability)**.**

Let $\mathbf{S} \subseteq \overline{\mathbf{G}}$ be given. A voting rule f is interim self-stable in $\mathbf{S}$ if the constitution (f, f) is interim stable in $\mathbf{S}$.

3 Interim Stable Constitution

3.1 Qualified Majority Rules

In this subsection, as in Barberà and Jackson (2004), we focus on the special type of weighted majority rules where all voters have the same voting power $w = (1, ..., 1)$ and $q \in \{0, 1, ..., n-1\}$. We name such a rule a *Qualified majority rule*, which treats each voter equally. They are classified according to the quota: a standard majority rule ($q = q^s \equiv \frac{n}{2}$ if n is even and $\frac{n-1}{2}$ if n is odd), a sub majority rule ($q < q^s$), and a super majority rule ($q > q^s$). We denote by $\mathbf{G}(\textsc{i})$ the set of qualified majority rules, $w = \textsc{i} \equiv (1, ..., 1)$.

We first consider a constitution composed of qualified majority rules $(f, F) \in \mathbf{G}(\textsc{i}) \times \mathbf{G}(\textsc{i})$ and characterize the set of interim stable constitutions in the set of qualified majority rules $\mathbf{G}(\textsc{i})$.

Proposition 1. *A constitution $(f, F) \in \mathbf{G}(\textsc{i}) \times \mathbf{G}(\textsc{i})$ is interim stable in $\mathbf{G}(\textsc{i})$ if and only if*

1. $q \geq q^f$ *and*

2. $q + 1 \geq n - q^f$.

To discuss the conditions more effectively, we define a minimal winning coalition of a voting rule. A minimal winning coalition C^f of a weighted majority rule f is just the right size of a group of agents who can make change under the rule. So, the set of minimal winning coalitions of a weighted majority rule f, denoted by Ψ^f, is the set of minimal winning coalitions C^f's such that 1) $\sum_{i \in C^f} > q^f$ and 2) $\sum_{i \in \left(C^f \setminus \{j\}\right)} \leq q^f$ for any $j \in C^f$. Note that, when f is a qualified majority rule, $|C^f| = q^f + 1$ for any $C^f \in \Psi^f$. Similarly, we define a minimal veto coalition of a voting rule such that a coalition of agents who can together veto any change but cannot if they lose any one of the agents. For a qualified majority rule f, the size of a minimal veto coalition is $N - q^f$. We also denote $N_r(t) = \{i : t_i = r\}$ and $N_s(t) = \{i : t_i = s\}$ for $t \in T$.

Now, let's get back to two simple conditions in Proposition 1; $q \geq q^f$ and $q + 1 \geq n - q^f$. The first condition prevents r-types from forming a coalition to support the change to an alternative rule, and the second condition prevents s-types from doing that. More specifically, the former says the size of a minimal winning coalition of F, $q+1$, is not less than the size of a minimal winning coalition of f, $q^f + 1$. If there are enough number of r-type agents so that they can replace the given rule with a weaker one they all prefer, $|N_r(t)| > q$, they don't have to change it since the given rule is weak enough for them to get the Reform in the second stage voting game, $|N_r(t)| > q^f$. The latter condition says the size of minimal winning coalition of F, $q + 1$, is not less than the size of minimal veto coalition of f, $n - q^f$. If there are enough number of s-type voters so that they can change the given rule to a stronger rule they all prefer, $|N_s(t)| > q$, they don't have to change the rule since they can veto any attempt to reform under the given rule f, $|N_s(t)| > n - q^f - 1$.

Those conditions are consistent with observations from the reality. We usually observe a constitution where F is relatively more conservative than f. Moreover, we have few real world examples with a long-lived constitution consisting of two weak voting rules, such as a constitution with two sub-majority rules with $q+q^f < n-1$, which is consistent with the second condition.

Next, we check if the constitutions characterized in Proposition 1.1 are still stable against a set of more general alternatives, the set of any weighted majority rules, G. It is hard to find the real world example where a given qualified majority rule was replaced by a weighted majority rule under which agents' voting powers vary. One may argue that it is because of the social norm: The social norm may require fairness consideration that a voting rule treats all voters equally. Here, in a positive analysis, we examine if a certain qualified majority rule can endure any general weighted majority rules without such a social norm.

Theorem 1. *A constitution $(f, F) \in \mathbf{G}(\textsc{i}) \times \mathbf{G}(\textsc{i})$ is interim stable in $\mathbf{G}$ if and only if it is interim stable in $\mathbf{G}(\textsc{i})$.*

This theorem explains why it is hard for one society to move from an "one-person, one-vote" decision rule to another with asymmetric voting powers. If a society starts with a stable constitution composed of qualified majority rules, it may not need any strong normative arguments to keep the fairness or equality in voting powers.

The following corollaries help readers to recognize easily the different results compared to other papers.

Corollary 1.1. *A qualified majority rule f is interim self-stable in $\mathbf{G}$ if and only if it is a standard or super majority rule.*

This corollary characterizes the set of interim self-stable qualified majority rules. It turns out that a qualified majority rule is interim self-stable if and only if it is not a sub majority rule. We here simply point out that majority rule and most super majority rules cannot survive in Azrieli and Kim (2014) and sub-majority rule can survive in a certain society of Barberà and Jackson (2006). The detailed discussion is in Section 4. On the other hand, in reality, we frequently observe not only standard majority rules but also super majority rules: the unanimity rule under jury conviction systems or super-majority rules under legislatures.

Corollary 1.2. *If a constitution (f, F) with q is interim stable in $\mathbf{G}$, a constitution (f, F') with $q' > q$ is also interim stable in $\mathbf{G}$.*

Corollary 2 implies that the stronger the voting rule for the rule choice is, the bigger the set of interim stable constitutions is. Consequently, even some sub majority rule f can survive when F is more conservative than a standard majority rule even though a sub majority rule is not interim self-stable in $\mathbf{G}$. Note that it is caused by the change of F in our setup while it is by the change of society in the setup of Barberà and Jackson (2006). Both may explain why we rarely observe a sub majority rule in reality, not never in use.

3.2 Weighted Majority Rules

In this section, we relax the anonymous constraints for the given constitution (f, F) as well as a alternative g, so that different agents could have different voting powers under any decision rule considered or used. Even though not as common as anonymous voting rules, non-anonymous weighted majority rules where agents have different voting powers can be easily found in reality. A stockholder meeting is one typical example: stockholders' weights are determined by the amounts of the stocks they possess. Another example is a legislature with a veto player: In a presidential system, the president may have the veto power, so has the power to refuse to approve a bill.

The following proposition discusses the necessary condition of interim stable constitution among weighted majority rules. We denote by $\bar{C}$ the minimal winning coalition which contains agents with highest weights under F.

Proposition 2 (Necessary Condition of Interim Stable Constitution).
A constitution $(f, F) \in \mathbf{G} \times \mathbf{G}$ is interim stable in $\mathbf{G}$ only if

1. *$\exists C^f \in \Psi^f$ such that C^f is not a proper superset of C for any $C \in \Psi^F$ and*

2. *for some $i \in \bar{C}$, $\exists \hat{C} \ni i$ such that $\hat{C} \cap C^f \neq \emptyset$ for any $C^f \in \Psi^f$.*

The conditions in Proposition 2 is a generalized version of those in Proposition 1. The first condition is for r-type voters: If there are a coalition of r-type voters who can change the given rule f to a weaker rule under F, but cannot achieve the Reform by themselves under f, then the constitution (f, F) is not interim stable. The second condition is for s-type voters: If there are a coalition of s-type voters who can change the given rule f to a stronger rule under F, but cannot veto the Reform by themselves under f, then the constitution (f, F) is not interim stable.[4]

[4]Note that this proposition only specifies a necessary condition. Since the set of alternatives is infinite $|G| = \infty$ and there are too many tedious alternatives which have never been considered in reality, it is not easy to pin down a sufficient condition in this setting. For example, for any given constitution (f, F), we can come up with a weird alternative g that gives all powers to one minimal winning coalition of F and assigns zero weights for the others. This alternative may not be interesting to consider, but still in the set of alternative G and makes the given rule f hard to be stable. Here, in order to restrict our attention to realistic situations, we examine if a typical example of a general weighted majority rule is interim stable.

Corollary 2.1. *A constitution* (f, F) *is not interim stable if* F *is a sub majority rule.*

We also provide the sufficient condition of interim stable constitution. Roughly speaking, the condition is the existence of a group which has a great voting power both for F and f.

Proposition 3. *A constitution* $(f, F) \in \mathbf{G} \times \mathbf{G}$ *is interim stable in* $\mathbf{G}$ *if*

1. $\exists V \subseteq N$ *such that* $V \subseteq C$ *for any* $C \in \psi^F$ *and* $V \cap C^f \neq \phi$ *for any* $C^f \in \psi^f$ *and*

2. *for all* $i \in V$ *and all* $C \ni i$, $\exists \bar{C}^f \in \psi^f$ *such that* $((C/V) \cup \{i\}) \supseteq \bar{C}^f$

The following corollaries demonstrate interesting characteristics and examples of interim stable (or non-stable) constitutions.

Corollary 3.1. *A constitution* (f, F) *is interim stable in* G *if* F *is the unanimity rule.*

Now, think about the typical example mentioned above, a presidential system with a veto player. In technical terms, a veto player is an agent who is in any minimal winning coalition. So, if there is a veto agent, without her support, a change cannot be made (or the reform R cannot be achieved). The following corollary shows that a decision rule with a veto agent is interim self-stable.

Corollary 3.2. *A voting rule* f *is interim self-stable in* G *if* f *has a veto player.*

3.3 Environment Independence

In this subsection, we show how robust the concept of interim stability is to our setup of environment, the society. Specifically, we examine if the set of interim stable constitution can be affected by the society (p_r, a).

Remark 1. *For any (p_r, a) and (p'_r, a'), Proposition 1-5 and Theorem hold unchanged.*

The additional proof for the remark is unnecessary because the proofs of Proposition 1-5 and Theorem are not affected by the society (p_r, a). Proposition 1-5 and Theorem contain various results depending on the set of constitutions and the set of alternative rules. That is, Proposition 1 and Theorem show that the set of interim stable constitutions $(f, F) \in \mathbf{G}(\textsc{i}) \times \mathbf{G}(\textsc{i})$ is independent to society. From the others, it is not shown in the case of interim stable constitutions $(f, F) \in \mathbf{G}(w) \times \mathbf{G}(\textsc{w})$ or $\mathbf{G} \times \mathbf{G}$. However, the following proposition argues the independence of environment if we extend the set of alternative rules from the set of weighted majority rules $\mathbf{G}$ to the set of arbitrary voting rules $\overline{\mathbf{G}}$.

Proposition 4 (Environment Independence)**.**
For any (p_r, a) and (p'_r, a'), a constitution $(f, F) \in \mathbf{G} \times \mathbf{G}$ is interim stable in $\overline{\mathbf{G}}$ at (p_r, a) if and only if it is interim stable in $\overline{\mathbf{G}}$ at (p'_r, a').

Even though both the remark and Proposition 6 argue the independence of environment of interim stable constitution, they are dissimilar in the proof ideas. In the proof of Proposition 6, if a rule g defeats f in the society (p_r, a) then we can construct a new rule g' which defeats f in another society (p'_r, a'). Since the way of constructing g' uses randomization of alternatives, the rule g' is not a weighted majority rule. Hence, the idea needs an arbitrary voting rule as an alternative to f. On the other hand, the remark relies on the assumption that we allow mixed strategies for agents, which can equip the agents with the corresponding rejection equilibrium fit to the society.

The remark and Proposition 6 are noteworthy as their implication is significantly different to that of Azrieli and Kim (2016) and Barberà and Jackson (2004) which focus on ex-ante self-stability of voting rules. Specifically, ex-ante self-stable voting rules in Azrieli and Kim (2016) are not affected by p_r since they consider an arbitrary voting rule as an alternative to f. The ex-ante self-stable qualified majority rules in Barberà and Jackson (2004) heavily depend on p_r since they allow only qualified majority rules as an alternative to f and they are interested in the connection between the self-stability and p_r. The reason why the utility of agents a also does not affect our stability is that agents make decision at interim stage where $r\,(s)$ type agent obtains only $a\,(1)$ or 0 in the possible future events.

4 Discussion

4.1 Fixed weights

In reality, agents' voting powers may have never be the object of change. There may be a rule or regulation which fixes agents' voting powers before they make decisions on a rule change and/or an economic outcome. For example, before any stockholder meeting, stockholders' relative voting powers have been determined by the amounts of the stocks they possess, and is seldomly changed by the result of a meeting.

Here, we discuss the interim stability in a set of weighted majority rules in which an agent's weight does not vary over rules. So, in this setting, the rules are differentiated only in quota. So, we restrict our focus on the set of weighted majority rules under which agents' weights are fixed: For any f and g, $(w_i^f)_{i=1}^n = (w_i^g)_{i=1}^n$. Denote $\mathbf{G}(\mathrm{w})$ the set of weighted majority rules with given weights $\mathrm{w} \equiv (w_i)_{i=1}^n$. The following proposition specifies a sufficient condition of interim stable fixed-weight constitutions. Denote by $w(Y) \equiv \sum_{i \in Y} w_i$ the sum of the weights of a coalition $Y \subseteq N$. Note that Ψ^F is the set of minimal winning coalitions of the voting rule F.

Proposition 5 (Fixed-weight Constitution: Sufficient condition)**.**
A constitution $(f, F) \in \mathbf{G}(\mathrm{w}) \times \mathbf{G}(\mathrm{w})$ is interim stable in $\mathbf{G}(\mathrm{w})$ if

1. *$q \geq q^f$ and*

2. *$\forall i$, $\exists \hat{C} \in \Psi^F$ such that $i \in \hat{C}$ and $w(N \setminus \hat{C}) \leq q^f$.*

Proposition 5 shares a similar implication with Proposition 1. The first condition prevents r-types from forming a winning coalition for an equilibrium rejection, and the second condition prevents s-types from doing that. The former says the minimum size of minimal winning coalitions of F is not less than the minimum size of minimal winning coalitions of f, q^f. If there are enough number of r-type agents so that they can change the given rule to a weaker one they all may prefer, $w(N_r(t)) > q$, they don't have to change it since the given rule is weak enough for them to get the Reform, $w(N_r(t)) > q^f$. The latter says if any agent under the given rule is a member of a coalition which is a minimal winning coalition of F as well as a veto coalition of f, then the constitution (f, F) is interim stable. If there are enough number of s-type agents so that they can change the given rule to a stronger rule they all prefer, $w(N_s(t)) > q$, one may believe they don't have to change the rule since they are strong enough, $N_s(t) \supseteq \hat{C}$, so that they can veto any attempt to reform under the given rule f, $q^f \geq w(N \setminus \hat{C}) \geq w(N \setminus N_s(t))$.

So far, we discuss a sufficient condition of interim stable constitutions with fixed weights. The following proposition describes a necessary condition which is also strong enough to cover the cases with the anonymous constraint discussed in Section 3. We denote by $\bar{C}$ the minimal winning coalition which contains agents with highest weights under F.

Proposition 6 (Fixed-weight Constitution: Necessary condition)**.**
A constitution $(f, F) \in \mathbf{G}(\mathrm{w}) \times \mathbf{G}(\mathrm{w})$ *is interim stable in* $\mathbf{G}(\mathrm{w})$ *only if*

1. $q \geq q^f$ *and*

2. *for some* $i \in \bar{C}$, $\exists \hat{C} \ni i$ *such that* $w(N \setminus \hat{C}) \leq q^f$.

The first condition is the same as the corresponding condition of Proposition 1, and the second condition is the generalized version of the second condition of it.

4.2 Ex ante Stability

We discuss the connection and comparison between our Interim stability and Ex ante stability that AKAzrieli and Kim (2016) and BJBarberà and Jackson (2004) investigate.[5] The fundamental difference is on the timing of comparison of voting rules in terms of information. When agents do not know their types (Ex ante stage), they vote for a rule between the incumbent rule f and an alternative rule

[5] AKAzrieli and Kim (2016) and BJBarberà and Jackson (2004) use the term 'self-stability' whey they deal with voting rules as well as constitutions

g. In the Appendix, we show the formal consistency of the definition between ex ante stability and interim stability.

The informational difference is similar to the difference of ex ante and interim Pareto efficiency in HMHolmström and Myerson (1983). Then one may wonder if an ex ante stable constitution is interim stable as HMHolmström and Myerson (1983) shows that an ex ante Pareto efficient rule is interim Pareto efficient. In general, such a relationship does not hold in the concept of stability. The counter example is a sub majority rule which is not interim self-stable in $\mathbf{G}(\text{I})$ can be ex ante self-stable in $\mathbf{G}(\text{I})$ in a certain society (Example 3 in BJBarberà and Jackson (2004)). An intuitive explanation compared to Pareto efficiency is the following. For the Pareto efficiency, if a rule is not interim Pareto efficient then there exists a rule g which makes all agents better off with at least one strictly better off in terms of interim expected utility. The same rule g can prevent f from being ex ante Pareto efficient. This simple logic does not work regarding the stability because the rule g that prevents f from being interim self-stable cannot guarantee the welfare improvement of all agents. Thus the rule f could be ex ante self-stable.

We find, however, if the set of competing alternative rules is $\overline{\mathbf{G}}$ then any ex ante self-stable rule is interim self-stable. The set of ex ante self-stable voting rules characterized in AKAzrieli and Kim (2016) consists of voting rules with at least one veto player and a certain structure of non-veto players and one qualified majority rule with the quota $q = n - 1$. Since the proof of corollary (regarding veto player) holds even in $\overline{\mathbf{G}}$, we need only the argument regarding the qualified majority rule with $q = n - 1$ to show the following proposition. The proof is in the Appendix.

Proposition 7. *If a weighted majority rule f is ex ante self-stable in $\overline{\mathbf{G}}$, then it is interim self-stable in $\overline{\mathbf{G}}$.*

4.3 Set of alternative rules and constitutions

As in the definition of interim stable constitution, we emphasize that the set of interim stable constitutions depends on the set of alternative rules. Here we denote a set of alternative rules by $\mathbf{A} \subseteq \overline{\mathbf{G}}$. The set of constitutions in which we are interested can be various such that $(f, F) \in \mathbf{S} \times \mathbf{S}' \subseteq \overline{\mathbf{G}} \times \overline{\mathbf{G}}$. It is obvious that for any set of constitutions $\mathbf{S} \times \mathbf{S}'$ and sets of alternative rules $\mathbf{A} \subset \mathbf{A}'$, a constitution which is interim stable in $\mathbf{A}'$ is also interim stable in $\mathbf{A}$. Since $\mathbf{G}(\textsc{i}) \subset \mathbf{G}(\mathbf{w}) \subset \mathbf{G}$, we know that a constitution $(f, F) \in \mathbf{G}(\textsc{i}) \times \mathbf{G}(\textsc{i})$ which is interim stable in $\mathbf{G}$ is interim stable in $\mathbf{G}(\textsc{i})$. Our theorem, though, shows the opposite direction, which is not obvious.

The classification of stability depending on the set of alternative rules and constitutions, and on the informational stage may provide a general framework or help us in detail to study the literature of stability. For instance, we can understand that BJBarberà and Jackson (2004) investigate a constitution $(f, F) \in \mathbf{G}(\textsc{i}) \times \mathbf{G}(\textsc{i})$ which is ex ante stable in $\mathbf{G}(\textsc{i})$ and that AKAzrieli and Kim (2016) study a constitution $(f, F) \in \mathbf{G} \times \mathbf{G}$ which is ex ante stable in $\overline{\mathbf{G}}$. In this framework, we can interpret that a durable rule in HMHolmström and Myerson (1983) is a rule f in the constitution $(f, \mathit{unanimity\, rule})$ which is interim stable in $\overline{\mathbf{G}}$.[6]

[6]HMHolmström and Myerson (1983) consider a lager set of decision rules than ours such as the set of decision rules with a general number of alternatives.

References

Azrieli, Y. and Kim, S. (2014). Pareto efficiency and weighted majority rules. *International Economic Review*, 55(4):1067–1088. 1.1, 2.2, 3.1

Azrieli, Y. and Kim, S. (2016). On the self-(in)stability of weighted majority rules. *Games and Economic Behavior*, 100:376–389. 1, 3.3, 4.2, 5, 4.3, A.2, A.2

Badger, W. (1972). Political individualism, positional preferences, and optimal decision rules. In Niemi, R. G. and Weisberg, H. F., editors, *Probability models of collective decision making.* Merrill, Columbus, Ohio. 1.1

Barberà, S. and Beviá, C. (2002). Self-selection consistent functions. *YJETH¡/cja:jid¿ Journal of Economic Theory*, 105(2):263–277. 1.1

Barberà, S. and Jackson, M. (2006). On the weights of nations: Assigning voting weights in a heterogeneous union. *Journal of Political Economy*, 114(2):317–339. 1.1, 2.3, 3.1, 3.1

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. 1, 1.1, 3.1, 3.3, 4.2, 5, 4.3

Curtis, R. (1972). Political individualism, positional preferences, and optimal decision rules. In Niemi, R. G. and Weisberg, H. F., editors, *Probability models of collective decision making.* Merrill, Columbus, Ohio. 1.1

Fleurbaey, M. (2008). Weighted majority and democratic theory. *Mimeo*. 1.1

Holmström, B. and Myerson, R. B. (1983). Efficient and durable decision rules with incomplete information. *Econometrica*, 51(6):pp. 1799–1819. 1, 1.1, 2.3, 4.2, 4.3, 6, A.2

Koray, S. (2000). Self-selective social choice functions verify arrow and gibbard-satterthwaite theorems. *Econometrica*, 68(4):981–995. 1.1

Koray, S. and Slinko, A. (2008). Self-selective social choice functions. *Social Choice and Welfare*, 31(1):129–149. 1.1

Neumann, J. V. and Morgenstern, O. (1953). *Theory of games and economic behavior,*. Princeton University Press, Princeton. 1.1

Penrose, L. S. (1946). The elementary statistics of majority voting. *Journal of the Royal Statistical Society*, 109(1):53–57. 1.1

Rae, D. W. (1969). Decision-rules and individual values in constitutional choice. *American Political Science Review*, 63:40–56. 1.1

A Appendix

A.1 Proofs

Proof of Proposition 1.

(If part) First, consider any alternative rule g with $q^g > q^f$. Set $\sigma_i(s) = \sigma_i(r) = 0$ and $\sigma_i^k(s) > \sigma_i^k(r)$. By construction, $f = R$ if $g = R$. So, $\sigma_i(r) = 0$ is always justified. For $t_i = s$, $\mu_i(t_{-i}) > 0$ only when $f(t) = S$.[7] So, $\sigma_i(s) = 0$ is also justified.

Second, consider any alternative rule g with $q^g < q^f$. Set $\sigma_i(s) = \sigma_i(r) = 0$ and $\sigma_i^k(s) < \sigma_i^k(r)$. By construction, $f = R$ only if $g = R$. So, $\sigma_i(s) = 0$ is always justified. For $t_i = r$, $\mu_i(t_{-i}) > 0$ only when $f(t) = R$.[8] So, $\sigma_i(r) = 0$ is also justified.

(Only if part) Suppose $q < q^f$. Consider an alternative rule g with $q^g = 0$. By construction, $f = R$ only if $g = R$. For any arbitrary i with $t_i = r$, $g = R$. Moreover, if there exists an equilibrium rejection, $\mu_i(t_{-i}) > 0$ for some t_{-i} with $f(r, t_{-i}) = S$. Hence, from Condition (C.4), $\sigma_i(r) > 0$. We have picked an arbitrary i, so there cannot be an equilibrium rejection of g.

Suppose now $q + q^f < n - 1$. Consider the unanimous rule g. By construction, $f = R$ if $g = R$. For any arbitrary i with $t_i = s$, $g = S$. Moreover, if there exists an equilibrium rejection, $\mu_i(t_{-i}) > 0$ for some t_{-i} with $f(s, t_{-i}) = R$. Hence, from Condition (C.4), $\sigma_i(s) > 0$. We have picked an arbitrary i, so there cannot be an equilibrium rejection of g. □

Lemma 1 (Equilibrium property 1)**.**

[7] The agent i believes she is pivotal only when at least $q+1$ agents including her have s-type.
[8] The agent i believes she is pivotal only when at least $q+1$ agents including her have r-type.

Consider a constitution (f, F). If there exists a equilibrium rejection (σ, μ) of an alternative rule g, then, for each i such that $\sigma_j(s) + \sigma_j(s) > 0$ for any $j \geq i$, there exists a set of agents $\tilde{H} \in \Phi_i$ and a type profile $\tilde{t}_{-i} \in T_{-i}$ such that

$$\lim_{k\to\infty} \frac{\rho(\tilde{H}; \tilde{t}_{-i}, \sigma^k)}{\rho(H; t_{-i}, \sigma^k)} > 0 \qquad \forall H \in \Phi_i, \forall t_{-i} \in T_{-i} \tag{A.1}$$

which satisfies $\sigma_j(s) = \sigma_j(r) = 0$ for all $j \in N \setminus (\tilde{H} \cup \{i\})$.

***Proof** of Lemma 1.* Without loss of generality, suppose $w_i \geq w_j$ if and only if $i \geq j$ under F.

First, consider $i = n$. Suppose not. So for any $\tilde{H}$ and a type profile $\tilde{t}_{-i} \in T_{-i}$ which makes Equation (A.1) goes to zero in the slowest speed, there is some agent $j \in N \setminus (\tilde{H} \cup \{i\})$ with $\sigma_j(s) + \sigma_j(r) > 0$. Define $\hat{N}^+ \equiv \{j \in N \setminus (\tilde{H} \cup \{i\}) | \sigma_j(s) + \sigma_j(r) > 0\}$. Also define $\tilde{N}^+ \equiv \{j \in \tilde{H} | \sigma_j(\tilde{t}_j) > 0\}$. We know $w(\hat{N}^+ \cup \tilde{N}^+ \cup \{i\}) \leq q$ ($\because \sigma$ consists an equilibrium rejection) and $w(\tilde{H} \cup \{i\}) > q$. Now, pick $j \in \tilde{H} \setminus (\hat{N}^+ \cup \tilde{N}^+)$ with the lowest weight, and add her into the set $(\hat{N}^+ \cup \tilde{N}^+ \cup \{i\})$. Repeat it until the set turns to a winning coalition. Denote the winning coalition WC' and also $H' \equiv WC' \setminus \{i\}$. By construction, $H' \in \Phi_i$ ($\because i = n$ has the highest weight). Then we find a contradiction, since $\rho(H'; t_{-i}, \sigma)$ goes to zero in a speed that is slower than $\rho(\tilde{H}; \tilde{t}_{-i}, \sigma)$.

Second, consider $i = n - 1$. A similar argument from above works for any $j < i$. So, we only need to show that there exists some $\tilde{H}$ such that $j = n$ with $\sigma_j(s) + \sigma_j(r) > 0$ is (also) not in $N \setminus (\tilde{H} \cup \{i\})$. Suppose $j \in N \setminus (\tilde{H} \cup \{i\})$ with $\sigma_j(s) + \sigma_j(r) > 0$. Follow the same logic from above to find H'. Because $j = n$ is not in $\tilde{H} \setminus (\hat{N}^+ \cup \tilde{N}^+)$, we still have $H' \in \Phi_{i=n-1}$. Then, we can show that $\rho(H'; t_{-i}, \sigma)$ goes to zero in a speed that is slower than $\rho(\tilde{H}; \tilde{t}_{-i}, \sigma)$. Contradiction.

Similar arguments apply for any $i \in N$. $\square$

Proof *of Theorem 1.* **Case 1:** Consider the case where $\forall C^g$, $\exists C^f \subseteq C^g$. Set $\sigma_i(s) = \sigma_i(r) = 0$ and $\sigma_i^k(s) > \sigma_i^k(r)$. By construction, $f = R$ if $g = R$. So, $\sigma_i(r) = 0$ is always justified. For $t_i = s$, $\mu_i(t_{-i}) > 0$ only when $f(t) = S$. ($\because$ The agent i believes she is pivotal only when at least $q+1$ agents including herself are s-type. And we know $q^f \geq n - (q+1)$ by construction.) So, $\sigma_i(s) = 0$ is also justified.

Case 2: Consider the case where $\exists \bar{C}^g$ such that $\bar{C}^g \not\supseteq C^f$ $\forall C^f$.

Define $\bar{C}^f \equiv \arg\max_{C^f} w^g(C^f)$. (More than one?) Then, by construction, for some C^g, $\bar{C}^f \supsetneq C^g$, so $\exists i \in \bar{C}^f$ such that $w^g(\bar{C}^f \setminus \{i\}) > q^g$.

Case 2-1: Consider the case where $\forall i \in \bar{C}^f$, we have $w^g(\bar{C}^f \setminus \{i\}) > q^g$. Set, for any i, $\sigma_i(s) = \sigma_i(r) = 0$, $\sigma_i^k(s) = k^{-\frac{2}{w_i^g}}$ and $\sigma_i^k(r) = k^{-\frac{1}{w_i^g}}$. So, $\sigma_i^k(s) < \sigma_i^k(r)$ for any i, and $\sigma_i^k(r) < \sigma_j^k(r)$ for any i and j such that $w_i^g < w_j^g$. For $t_i = s$, $\mu_i(t_{-i}) > 0$ only when $g(t) = R$. ($\because$ For any i, $\mu_i(t_{-i}) > 0$ only when $\forall j \in (\bar{C}^f \setminus \{i\})$ has $t_j = r$.) So, $\sigma_i(s) = 0$ is okay. For $t_i = r$, $\mu_i(t_{-i}) > 0$ only when $f(t) = R$. ($\because$ The agent i believes she is pivotal only when at least $q+1$ agents including herself are r-type. And we know $q^f \leq q$ by construction.) So, $\sigma_i(r) = 0$ is okay.

Case 2-2: Consider the case where for some $i \in \bar{C}^f$, we have $w^g(\bar{C}^f \setminus \{i\}) \leq q^g$. Define $J \equiv \{i \in \bar{C}^f : w^g(\bar{C}^f \setminus \{i\}) \leq q^g\}$. Pick $\bar{C}^F \supseteq \bar{C}^f$. Set, for any $i \in J$, $\sigma_i(s) = \epsilon$ and $\sigma_i(r) = 1$, and for any $i \in \bar{C}^F \setminus \bar{C}^f$, set $\sigma_i(s) = 1$ and $\sigma_i(r) = \epsilon$. For all others, such that $i \in \bar{C}^f \setminus J$ or $i \in N \setminus \bar{C}^F$, set $\sigma_i(s) = \sigma_i(r) = 0$ and $\sigma_i^k(s) < \sigma_i^k(r)$. We show that this construction can form a equilibrium rejection of g with some small enough positive value of ϵ.

Check if $\sigma_i(r) = 0$ is okay. If she believes $f = R$, then it is okay. If she believes

$f(t) = S$, she knows $g(t) = S$. (Suppose not, so she believes $f = S$ and $g = R$. If all agents in J are r-type, then some $C^f \supset J$ containing her i contains only r-type agents. So, to have $f = S$, she knows at least one agents in J should be s-type. Then, by construction of $\bar{C}^f \supset J$, more than $g^f + 1$ agents should be r-type to make $g = R$. So, $f = R$. Contradiction.) So for any belief, $\sigma_i(r) = 0$ is okay.

Now, check if $\sigma_i(s) = 0$ is okay. Suppose $\epsilon = 0$. Then she always believes $g = R$. Moreover, she believes $f = S$ with positive probability. (Only agents in J and $q^f - |J|$ more agents are guaranteed to be r types, and all others could be s-type with positive probability.) Therefore, $\sum_{t_{-i}} \mu_i(t_{-i}) u_i(f(t), t_i) - \sum_{t_{-i}} \mu_i(t_{-i}) u_i(g(t), t_i) > 0$ when $\epsilon = 0$. If we denote by κ_i the event that makes $g = R$ and $f = S$, then $\lim_{\epsilon \to 0} \mu(\kappa_i) = 1$. Write

$$
\begin{aligned}
&\sum_{t_{-i}} \mu_i(t_{-i}) u_i(f(t), t_i) - \sum_{t_{-i}} \mu_i(t_{-i}) u_i(g(t), t_i) \\
&= \mu(\kappa_i) \left(u_i(f(t) = S, t_i = s) - u_i(g(t) = R, t_i = s) \right) \\
&\quad + (1 - \mu(\kappa_i)) \left(\sum_{t_{-i}} \mu_i(t_{-i} | \kappa_i^c) u_i(f(t), t_i = s) - \sum_{t_{-i}} \mu_i(t_{-i} | \kappa_i^c) u_i(g(t), t_i = s) \right) \\
&= \mu(\kappa_i) \left(0 - (-1) \right) \\
&\quad + (1 - \mu(\kappa_i)) \left(\sum_{t_{-i}} \mu_i(t_{-i} | \kappa_i^c) u_i(f(t), t_i = s) - \sum_{t_{-i}} \mu_i(t_{-i} | \kappa_i^c) u_i(g(t), t_i = s) \right) \\
&\geq \mu(\kappa_i) + (1 - \mu(\kappa_i)) \times (-1) = \left(2\mu(\kappa_i) - 1 \right).
\end{aligned}
$$

Since $\lim_{\epsilon \to 0} \mu(\kappa_i) = 1$, for a small enough ϵ, $(2\mu(\kappa_i) - 1)$ is always positive. Therefore, $\sigma_i(r) = 0$ is okay. $\square$

Proof *of Proposition 5.*

Find H such that $H \in \Phi_i$ for any $i \in N \setminus H$ and $(N \setminus C^*) \subset H$ for some $j \in N \setminus H$.[9]

Consider first the case with $q^g < q^f$. Set $\sigma_i(t_i) = 0$ for all i and t_i and $\sigma_i^k(r) = k^{-\frac{1}{w_i}}$ and $\sigma_i^k(s) = k^{-\frac{2}{w_i}}$. For any s-type agent, $\sigma_i(s) = 0$ is okay, since $g(\cdot) = S$ implies $f(\cdot) = S$. For an r-type agent, $\mu(t_{-i}) > 0$ only when $f(t_i = r, t_{-i}) = R$. So $\sigma_i(r) = 0$ is justified.

Consider now the case with $q^g > q^f$. Set $\sigma_i(t_i) = 0$ for all i and t_i and $\sigma_i^k(r) = k^{-\frac{2}{w_i}}$ and $\sigma_i^k(s) = k^{-\frac{1}{w_i}}$. For any r-type agent, $\sigma_i(r) = 0$ is okay, since $g(\cdot) = R$ implies $f(\cdot) = R$. For an s-type agent i, Equation (A.1) with $\tilde{H} = \hat{C} \setminus \{i\}$ where $\hat{C}$ has the highest weights among the minimal coalitions contain the agent i converges to zero in a speed that is slower than that with any other $\tilde{H}$. By construction, $w(N \setminus \hat{C}) \leq q^f$. So, $\mu(t_{-i}) > 0$ only when $f(t_i = r, t_{-i}) = S$. Therefore, $\sigma_i(s) = 0$ is justified. $\square$

Proof *of Proposition 6.*

We prove by contradiction.

First, suppose $q < q^f$ and there exists some equilibrium rejection (σ, μ) of g with $q^g = 0$. For any i, $\mu_i(t_{-i}) > 0$ for some t_{-i} such that $f(t_i = r, t_{-i}) = S$, while $g(t_i = r, t_{-i}) = R$ always. So, $\sigma_i(r)$ should be positive from Condition (C.4). Contradiction.

Second, suppose $\forall i \in \bar{C}$, $\forall C \ni i$, $w(N \setminus C) > q^f$. And suppose there exists some equilibrium rejection (σ, μ) of the unanimity rule g. Consider the agent $i = n$. From Lemma 1, we know for Equation A.1, $j \in N \setminus (\tilde{H} \cup \{i\})$ should have $\sigma_j(s) = \sigma_j(r) = 0$. By construction, there exist a minimal winning coalition $C \ni i$ such that $C \subset (\tilde{H} \cup \{i\})$, and any $j \in \tilde{H} \setminus C$ should have $\sigma_j(s) + \sigma_j(r) > 0$. (If not,

[9] How? Add i with the smallest weight into the set $N \setminus C^*$. Repeat until the set becomes to be a winning coalition of all i not in the set.

there should exists some slower H' which does not contain j with $\sigma_j(s)+\sigma_j(r)=0$ than $\tilde{H}$.) Also, if $\sigma_j(r)=0$ for some $j \in \tilde{H} \setminus C$, there exists some $\tilde{H}'$ and t'_{-i} where $j \in N \setminus \tilde{H}'$ and $t_j = r$, which gives the same convergence speed for $\rho(\tilde{H}'; t'_{-i}, \sigma)$ with $\rho(\tilde{H}; t_{-i}, \sigma)$. We know C has a mutually exclusive C^f and have shown that all $j \in C^f$ have r-types with positive probability in the sense of posterior belief μ. Therefore, for $i = n$, $\mu_i(t_{-i}) > 0$ for some $f(t_i = s, t_{-i}) = R$, while $g(t_i = s, t_{-i}) = S$ always. So, $\sigma_i(s)$ should be positive from Condition (C.4).

For some $i \in \bar{C}$ such that $i \neq n$, a similar logic can be applied. So, for all $i \in \bar{C}$, $\sigma_i(s) > 0$. Contradiction. $\square$

Proof *of Proposition 2.*

We prove by contradiction.

First, suppose $\forall C^f \in \Psi^f$, $\exists C \in \Psi^F$ such that $C \subseteq C^f$. So, $f = R$ implies $F = R$. Also, suppose there exists some equilibrium rejection (σ, μ) of g with $q^g = 0$. For any i, $\mu_i(t_{-i}) > 0$ for some t_{-i} such that $f(t_i = r, t_{-i}) = S$, while $g(t_i = r, t_{-i}) = R$ always. So, $\sigma_i(r)$ should be positive from Condition (C.4). Contradiction.

The second part of the proof is the same as that of Proof of "Only if part" in Proposition 6. $\square$

Proof *of Proposition 3.*

Consider a strategy profile and a belief system (σ, μ) such that, for all $i \in V$, $\sigma_i(r) = 0$, $\sigma_i(s) = 0$, and for all $j \neq i \in \underset{C \in \psi^F}{\cup} C$, $\sigma_j(r) = 1$, $\sigma_j(s) = 0$ and for other agents l, $\sigma_l(r) = \sigma_l(s) = 1$. And set $\sigma_i^k(r) = k^{-2}$, $\sigma_i^k(s) = k^{-1}$, and $\sigma_j^k(s) = k^{-2}$.

For the agent i, if $t_i = r$, $\sigma_i(r) = 0$ is justified since $f(t_i, t_{-i}) = R$ for any t_{-i} from the condition 2 in the proposition.

If $t_i = s$, $\sigma_i(s) = 0$ is justified since $f(t_i, t_{-i}) = S$ for any t_{-i} from that $V \cap C^f \neq \phi$ for any $C^f \in \Psi^f$. Similarly, $\sigma_j(s) = 0$ is also justified.

Then, all i in V reject any alternative rule g. It implies that F chooses f since $V \subseteq C$ for any $C \in \Psi^F$. □

***Proof** of Proposition 4.* Suppose a constitution (f, F) is not interim stable for a society (p_r, a). We want to show that the (f, F) is not interim stable for any other arbitrary society. Since the constitution is not interim stable for (p_r, a), we can find an alternative g for which no equilibrium rejection exists. This implies that, for any strategy profile and belief system which satisfy Condition (C.1) and (C.3) violates Condition (C.4) for some t_i. Note that Condition (C.3) is all about σ^k for a society. That is, any σ satisfies Condition (C.1) and any corresponding μ generated by Condiion (C.3) with any possible σ^k which converges to σ violates Condition (C.4) for a particular t_i: That is, for a particular t_i

$$\sum_{t_{-i}} \mu_i (t_{-i}) (u_i(g(t), t_i) - u_i(f(t), t_i)) > 0.$$

We argue that, for another society (p'_r, a'), there exists an alternative voting rule g' such that any σ satisfies Condition (C.1) and any corresponding μ' generated by Condiion (C.3) with any possible σ^k which converges to σ violates Condition (C.4) for the same t_i: That is, for a particular t_i

$$\sum_{t_{-i}} \mu'_i (t_{-i}) (u_i(g'(t), t_i) - u_i(f(t), t_i)) > 0.$$

For each type profile t, we define $\alpha_t = \frac{p_i(t_i)\mu_i(t_{-i})}{p'_i(t_i)\mu'_i(t_{-i})}$ if $\mu_i(t_{-i}) > 0$ and $\alpha_t = 0$ otherwise. Let $\alpha = \max_{t \in T} \alpha_t$. Define the voting rule g' by $g'(t) = \frac{\alpha_t}{\alpha} g(t) +$

$\left(1-\frac{\alpha_t}{\alpha}\right) f(t)$. Then, for the t_i,

$$
\begin{aligned}
0 &< \sum_{t_{-i}} \mu_i\left(t_{-i}\right)\left(u_i(g(t), t_i) - u_i(f(t), t_i)\right) \\
&= \sum_{t_{-i}} \mu_i'\left(t_{-i}\right) \frac{p_i'\left(t_i\right)}{p_i\left(t_i\right)} \alpha_t \left(u_i(g(t), t_i) - u_i(f(t), t_i)\right) \\
&= \sum_{t_{-i}} \mu_i'\left(t_{-i}\right) \frac{p_i'\left(t_i\right)}{p_i\left(t_i\right)} \alpha \left[\frac{\alpha_t}{\alpha} u_i(g(t), t_i) + \left(1-\frac{\alpha_t}{\alpha}\right) u_i(f(t), t_i) - u_i(f(t), t_i)\right] \\
&= \frac{p_i'\left(t_i\right)}{p_i\left(t_i\right)} \alpha \left(\sum_{t_{-i}} \mu_i'\left(t_{-i}\right)\left[u_i(\tfrac{\alpha_t}{\alpha} g(t) + \left(1-\tfrac{\alpha_t}{\alpha}\right) f\left(t\right), t_i) - u_i(f(t), t_i)\right]\right) \\
&= \frac{p_i'\left(t_i\right)}{p_i\left(t_i\right)} \alpha \left(\sum_{t_{-i}} \mu_i'\left(t_{-i}\right)\left[u_i(g'\left(t\right), t_i) - u_i(f(t), t_i)\right]\right)
\end{aligned}
$$

Thus, if there is no equilibrium rejection of g in the original society, there is also no equilibrium rejection of g' in the new society. The converse is obvious. □

A.2 Consistency of Definition

Here, we discuss the consistency of our definition of interim self stability with the ex-ante self-stability à la Azrieli and Kim (2016) and the durability à la Holmström and Myerson (1983).

Consider the "ex-ante environment" studied in Azrieli and Kim (2016), where agents vote on rule change before their types are realized. We rewrite our conditions and definition as follows.

To reject the alternative rule g all the time, the probability that g gets sufficient weighted votes should be zero. In other words, the alternative g is always rejected if and only if

$$\sum_{\{j:\sigma_j>0\}} w_j \leq q. \tag{A.2}$$

If Equation (A.2) holds, then honest behavior in f and g (we consider incentive compatible f and g), together with the voting strategies in the first stage, $\sigma = (\sigma_1, \ldots, \sigma_n)$ form a Nash equilibrium if and only if

$$\gamma_i \left(u_i(f) - u_i(g)\right) \geq 0 \quad \forall i, \tag{A.3}$$

where

$$u_i(f) = a \sum_{\{t \in T: t_i = r\}} p(t) f(t) - \sum_{\{t \in T: t_i = s\}} p(t) f(t),$$

$$\Phi_i = \{H_i \subseteq N/\{i\} |\, q - w_i < \sum_{j \in H_i} w_j \leq q\},$$

and

$$\gamma_i = \sum_{H_i \in \Phi_i} \left(\prod_{j \in H_i} \sigma_j\right) \left(\prod_{j \in N/(H_i \cup \{i\})} (1 - \sigma_j)\right).$$

We require that, for any individual i,

$$\text{if } \; u_i(f) < u_i(g), \text{then } \; \sigma_i = 1. \tag{A.4}$$

This condition imposes that, if the expected utility of individual i in the alternative decision rule g would be higher than in the current rule f, then individual i should vote for g.[10]

$w(Y) := \sum_{i \in Y} w_i$ denotes the total weight of coalition Y.

Proposition 8. *For a given weighted majority rule f, $w(\{i : u_i(f) < u_i(g)\}) \leq q$ for any alternative rule g if and only if there exists a strategy profile σ that satisfies conditions (A.2), (A.3), and (A.4).*

Proof of Proposition 8.

(Only if part)

Suppose $w(\{i : u_i(f) < u_i(g)\}) \leq q$. Then, set $\sigma_i = 1$ for any $i \in \{i : u_i(f) < u_i(g)\}$ and $\sigma_i = 0$ for any $i \notin \{i : u_i(f) < u_i(g)\}$. The condition (A.2) and (A.4) are satisfied. For an individual i with $\sigma_i = 1$, $\gamma_i = 0$. For an individual i with $\sigma_i = 0$, $(u_i(f) - u_i(g)) \geq 0$ by construction. Therefore, the condition (A.3) is satisfied.

(If part)

Suppose not. That is, the conditions (A.2), (A.3), and (A.4) are all satisfied, but

$$w(\{i : u_i(f) < u_i(g)\}) > q.$$

Since we suppose the condition (A.4) is satisfied, $\{j : u_j(f) < u_j(g)\} \subseteq \{j : \sigma_j > 0\}$, which implies $w(\{i : u_i(f) < u_i(g)\}) \leq w(\{j : \sigma_j > 0\})$. Then, the condition (A.2) is violated, since $\sum_{\{j : \sigma_j > 0\}} w_j \geq w(\{i : u_i(f) < u_i(g)\}) > q$. Contradiction. $\square$

One may wonder why we don't use the simple condition as $w(\{i : u_i(f) < u_i(g)\}) \leq q$ in Azrieli and Kim (2016) to define interim self-stability. To do that, in our set-

[10] In the second stage, since f and g are incentive compatible, we simply assume that all individuals report their true types.

ting, we need to add up the weights of agents i's who have

$$\sum_{t_{-i}} \mu_i(t_{-i}) u_i(f(t), t_i) < \sum_{t_{-i}} \mu_i(t_{-i}) u_i(g(t), t_i),$$

which is a part of the condition (C.4). But as in the condition (C.3), the posterior belief μ_i can only be calculated with a strategy profile for the first stage voting game σ. That is, to define interim self-stability in a way analogous to Azrieli and Kim (2016), we need a complete characterization of a Nash equilibrium with a sequentially rational strategy profile and a consistent belief system.

A.3 Extra Lemmas and Propositions

We show that if any individual can be in a minimal winning coalition which veto the ordinary decision together, then the weighted majority rule is interim self stable.

Lemma 2 (Fixed-weight Environment: Sufficient condition)**.**
A weighted majority rule $f \in \mathbf{G}(w)$ *is interim self-stable in* $\mathbf{G}(w)$ *if* $\forall i$, $\exists C \ni i$ *such that* $w(N \setminus C) \leq q$.

We omit the proof of this lemma since it is a corollary of Proposition 5 for a constitution (f, f). Instead, we provide an intuition behind it. Lemma 2 says if any individual under the given rule is a member of a minimal winning coalition which is a veto coalition at the same time, then the rule is interim self-stable. If there are enough number of r-type voters so that they together can change the given rule to a weaker rule, it is obvious that individuals does not have to vote for the change since the given rule is already weak enough. If there are enough number of s-type voters so that they can change the given rule to a stronger rule they all prefer, one may believe they don't have to change the rule since they are strong enough to veto any attempt to reform under the given rule.

We show that, if a decision rule is an interim self-stable weighted majority rule, then there exists an individual with a relatively high voting power such that any minimal coalition containing the individual is not mutually exclusive with all other minimal coalitions. Note that this condition is identical to the necessary (and sufficient) condition of interim self-stability from the previous sections.

Proposition 9 (Necessary Condition in the General Environment)**.**
A weighted majority rule $f \in \mathbf{G}$ is interim self stable in $\mathbf{G}$ only if for some $i \in \bar{C}$, $\exists \hat{C} \ni i$ such that $\hat{C} \cap C^f \neq \emptyset$ for any $C^f \in \Psi^f$.

Lemma 3 is an analogy of the sufficient condition of super majority qualified majority rules.

Lemma 3 (Fixed-weight Environment: Sufficient Condition 2)**.**
A weighted majority rule $f \in G(w)$ is interim self-stable among $G(w)$ if there exists a minimal winning coalition C^* such that $w(N \setminus C^*) + w_i \leq q$ for some $i \in C^*$.

Proof of Lemma 3.
Let $\Psi^{\hat{f}}$ denote the set of minimal winning coalitions(MWCs) under a decision rule $\hat{f}$. We also define, for a decision rule $\hat{f}$ and a type t_i, $T_{t_i}^{\hat{f}} \equiv \left\{t_{-i} : \hat{f}(t_i, t_{-i}) = R\right\}$, and for a set of type profile $\tilde{T} \subseteq T_{-i}$, $\mu_i(\tilde{T}) \equiv \sum_{t_{-i} \in \tilde{T}} \mu_i(t_{-i})$. For the convenience of notation, $w_i \geq w_j$ if and only if $i \geq j$.

Find H such that $H \in \Phi_i$ for any $i \in N \setminus H$ and $(N \setminus C^*) \subset H$.[11] By construction, $w(N \setminus (H \cup \{i\})) \leq q$ for any $i \in N \setminus H$ and $C^* \cap H \neq \emptyset$.

First, consider the case where $q_f > q_g$. So, if $g(t) = S$, then $f(t) = S$. And if $f(t) = R$, then $g(t) = R$. Suppose a strategy profile (σ, μ) such that $\sigma_i(s) = 0$ for any i and $\sigma_j(r) = 0$ for any $j \notin H$ and $\sigma_j(r) = 1$ for $j \in H$, and any arbitrary $\sigma_i^k(t_i)$ which converges to $\sigma_i(t_i)$ for any i and t_i. $\sigma_i(s)$ is always justified, since $g(t) = S$ implies $f(t) = S$. For any $i \notin H$, $\gamma_i(t_{-i}) > 0$ only when $t_j = r$ for all $j \in H$. Then, for $t_i = r$, Condition (C.2) is satisfied, and so $\sigma_i(r) = 0$ is justified. For any $i \in H_{\underline{i}}$, $\gamma_i(t_{-i})$ is always zero, $\sigma_j(r) = 1$ it is okay.

Second, consider the case where $q_f < q_g$. So, if $f(t) = S$, then $g(t) = S$. And if $g(t) = R$, then $f(t) = R$. Suppose a strategy profile (σ, μ) such that $\sigma_i(r) = 0$ for any i and $\sigma_j(s) = 0$ for any $j \notin H$ and $\sigma_j(s) = 1$ for $j \in H$, and any arbitrary $\sigma_i^k(t_i)$ which converges to $\sigma_i(t_i)$ for any i and t_i. $\sigma_i(r)$ is always justified, since $g(t) = R$ implies $f(t) = R$. For any $i \notin H$, $\gamma_i(t_{-i}) > 0$ only when $f(t_i, t_{-i}) = S$. Then, for $t_i = s$, Condition (C.2) is satisfied, and so $\sigma_i(s) = 0$ is justified. For any $i \in H_{\underline{i}}$, $\gamma_i(t_{-i})$ is always zero, so $\sigma_j(s) = 1$ is okay. □

Lemma 4. *If $\exists i^*$ such that, $\forall C^* \ni i^*$, $\exists C \cap C^* = \emptyset$, then $w(N) > 2q$.*

Lemma 5. *If, for some i^*, $w_{i^*} < w(N) - 2q$, then $\sigma_{i^*}(s) = 1$ in equilibrium.*

Lemma 6. *If there exists an equilibrium rejection, then $w(\{i | w_i < w(N) - 2q\}) \leq q$.*

Lemma 7. *There exists a minimal winning coalition C^* such that $w(N \setminus C^*) + w_{i^*} \leq q$ for some $i^* \in C^*$ if and only if no pair of minimal winning coalitions is mutually exclusive.*

[11] How? Add i with the smallest weight into the set $N \setminus C^*$. Repeat until the set becomes to be a winning coalition of all i not in the set.

Proof. Consider $A \equiv C^* \setminus \{i^*\}$ and $B \equiv \{i^*\} \cup (N \setminus C^*)$. We know $w(A) \leq q$ and $w(B) \leq q$. Then we cannot find any partition of N with two winning coalitions. □

Lemma 8. *There exists an agent i such that for any $C^* \ni i$ $w(N \setminus C^*) \leq q$ if and only if no pair of minimal winning coalitions is mutually exclusive.*

Proof. The "If" part is straightforward.

(Only if) Suppose not. So, there exists a pair of minimal winning coalitions C and C' such that $C \cap C' = \emptyset$. By construction i is neither in C nor in C'. We know $w(N) \geq w(C)+w(C')+w_i > q+q+w_i$. For any $C^* \ni i$, we have $w(N \setminus C^*) \leq q$ and $q < w(C^*) \leq q+w_i$. So, $w(N) = w(N \setminus C^*)+w(C^*) \leq q+q+w_i$. Contradiction. □

Lemma 9.

There is a minimal winning coalition $\bar{C} \in \Psi^f$ where for any $i \in \bar{C}$, any minimal winning coalition $C_i \ni i$ has a mutually exclusive minimal winning coalition $C' \in \Psi^f$ such that $C_i \cap C' = \emptyset$ if and only if any minimal winning coalition in Ψ^f has a mutually exclusive minimal winning coalition in Ψ^f.

Proof.

(Only if)

Assume a minimal winning coalition $\bar{C} \in \Psi^f$ where for any $i \in \bar{C}$, any minimal winning coalition $C_i \ni i$ has a mutually exclusive minimal winning coalition $C' \in \Psi^f$ such that $C_i \cap C' = \emptyset$. If there exists a minimal winning coalition $\tilde{C}$ which is not mutually exclusive with any other minimal winning coalition, then $\tilde{C}$ should not contain any $i \in \bar{C}$. So, $\tilde{C} \cap \bar{C} = \emptyset$. Contradiction.

(If)

It is obvious. □

Lemma 10.

If any minimal winning coalition has a mutually exclusive minimal winning coalition, there exists a pair of mutually exclusive minimal winning coalitions C and C' such that $w_i \geq w_{i'}$ for any $i \in C$ and $i' \in C'$.

Proof. There always exists a minimal winning coalition $\bar{C} \in \Psi^f$ such that for all $i \in \bar{C}$, $w_i \geq w_j$ for any $j \in N \setminus \bar{C}$. If the minimal winning coalition $\bar{C}$ has a mutually exclusive minimal winning coalition C', then $w_i \geq w_{i'}$ for any $i \in \bar{C}$ and $i' \in C'$. □

Lemma 11 (Necessary Condition 1: Single Agent Minimal Winning Coalition)**.** *If, under f, there exists an agent i who consists a minimal winning coalition by itself $C = \{i\}$ and a mutually exclusive minimal winning coalition $\tilde{C}$ such that $\tilde{C} \cap C = \emptyset$, f is not interim self stable.*

Proof of Lemma 11.

By construction, the agent i is always pivotal, $\gamma(t_{-i}) = 1$ for all t_{-i}.

Consider an alternative rule g such that $w_i^g = w_i^f$ and $w_j^g = 0$ for all $j \neq i$.

Then, for $t_i = s$, Equation (C.2) is violated since

$$\sum_{t_{-i}} P(t_{-i})\, u_i(f(t), t_i) = \sum_{t_{-i} \in T^f_{t_i = s}} P(t_{-i})\, u_i(f(t), t_i) = -\sum_{t_{-i} \in T^f_{t_i = s}} P(t_{-i}) < 0.$$

So, there is no equilibrium rejection of g. □

Lemma 12 (Necessary Condition 2: Small Quota)**.** *If there exists a minimal winning coalition $C \in \Psi^f$ such that for any $i \in C$ and for any minimal winning coalition $C_i \ni i$, $w(N \setminus C_i) > 2q$, f is not interim self-stable.*

Proof of Lemma 12.

Let an alternative rule g be the unanimous rule. By construction, for any agent $i \in C$, $T^f_{t_i=s} \neq \emptyset$ and $T^g_{t_i=s} = \emptyset$.

We prove by contradiction. Let's suppose there exists an equilibrium of g, (σ, μ).

For $t_i = s$, suppose $\sigma_i(s) \neq 1$. From Equation (C.4), $-\mu_i(T^f_s) \geq -\mu_i(T^g_s)$. We know $\mu_i(T^g_r) = 0$, we should have $\mu_i(T^f_s) = 0$.

So, if a type profile $\tilde{t}_{-i} \in T_{-i}$ and some set of agents $\tilde{H} \in \Phi_i$,

$$\lim_{k \to \infty} \left(\prod_{j \in \tilde{H}} \sigma^k_j(\tilde{t}_j) \right) \left(\prod_{j \in N \setminus (\tilde{H} \cup \{i\})} (1 - \sigma^k_j(\tilde{t}_j)) \right)$$

goes to zero in a speed that is no faster than for any other $H \in \Phi_i$ and type profile $t_{-i} \in T_{-i}$, $\tilde{t}_{-i}$ should not be in $T^f_{t_i=s}$. It means that, for a minimal winning coalition $C_i \ni i$ which is a subset of $\tilde{H}$, any $j \in \left(\tilde{H} \setminus C_i\right)$ have either $\sigma_j(s) > 0$ or $\sigma_j(r) > 0$. Also, there should not be any minimal winning coalition with all r types in $\tilde{t}_{-i}$. It means that $w(\left\{j \in N \setminus (\tilde{H} \cup \{i\}) | \tilde{t}_j = r\right\}) \leq q$. Then, we should have enough number of agents $j \in N \setminus (\tilde{H} \cup \{i\})$ such that $\tilde{t}_j = s$ and

$$w(\left\{j \in N \setminus (\tilde{H} \cup \{i\}) | \tilde{t}_j = s\right\}) \geq w(N \setminus (\tilde{H} \cup \{i\})) - q.$$

It implies that for such j with $\tilde{t}_j = s$ we should have $\sigma_j(r) = 1$. But, since $w(N \setminus C_i) > 2q$, $w(N \setminus C_i) = w(\tilde{H} \setminus C_i) + w(N \setminus (\tilde{H} \cup \{i\}))$ and

$$\begin{aligned} & w(N \setminus (\tilde{H} \cup \{i\})) \\ & \quad = w(\left\{j \in N \setminus (\tilde{H} \cup \{i\}) | \tilde{t}_j = s\right\}) + w(\left\{j \in N \setminus (\tilde{H} \cup \{i\}) | \tilde{t}_j = r\right\}), \end{aligned}$$

we have

$$w(\Big\{j \in N \setminus (\tilde{H} \cup \{i\}) | \tilde{t}_j = s\Big\}) + w(\tilde{H} \setminus C_i) > q.$$

The above result violates Equation (C.1). So, σ cannot be an equilibrium rejection. Hence, $\sigma_i(s)$ should be 1.

However, this is true for any $i \in C$. Contradiction. □

Lemma 13.

A weighted majority rule $f \in G(w)$ is interim self-stable among $G(W)$ if, for any individual i, there exists a minimal winning coalition including i which is not mutually exclusive with any other minimal winning coalition.

Proof.

Denote $\hat{C}_i$ a minimal winning coalition which is not mutually exclusive with any other minimal winning coalition. Since we are focusing on the case where $w_f = w_g$ for any g, it is either $q_f < q_g$ or $q_f > q_g$.

First, consider the case where $q_f < q_g$. So, if $f(t) = S$, then $g(t) = S$. And if $f(t) = R$, then $g(t)$ could be either S or R. Suppose a strategy profile (σ, μ) such that $\sigma_i(t_i) = 0$ for all i and $t_i \in T_i$, $\sigma_i^k(s) = k^{-\frac{1}{w_i}}$ and $\sigma_i^k(r) = k^{-\frac{2}{w_i}}$. Pick an agent i. Suppose for a minimal winning coalition C_m and a type profile t_{-i}, the convergence speed of

$$\lim_{k \to \infty} \left(\prod_{j \in C_m \setminus \{i\}} \sigma_j^k(t_j) \right) \left(\prod_{j \in N \setminus C_m} (1 - \sigma_j^k(t_j)) \right)$$

is slower than for any other minimal winning coalition. By construction, $W(C_m) \geq W(\hat{C}_i)$ and $t_j = s$ for $j \in C_m \setminus \{i\}$. For such t_{-i}, $f(t) = S$ if $t_i = s$, and $f(t)$ could be either S or R if $t_i = r$, because C_m has no mutually exclusive minimal winning coalition. Thus the right hand side of Equation (C.4) is weakly less than the left for any type and any agent. This is true for any i. The strategy profile σ and the derived belief system μ is an equilibrium rejection of g.

Second, consider the case where $q_f > q_g$. So, if $f(t) = S$, then $g(t)$ could be either S or R. And if $f(t) = R$, then $g(t) = R$. Suppose a strategy profile (σ, μ) such that $\sigma_i(t_i) = 0$ for all i and $t_i \in T_i$, $\sigma_i^k(s) = k^{-\frac{2}{w_i}}$ and $\sigma_i^k(r) = k^{-\frac{1}{w_i}}$. Pick an agent i. Suppose for a minimal winning coalition C_m and a type profile t_{-i}, the convergence speed of

$$\lim_{k\to\infty} \left(\prod_{j \in C_m \setminus \{i\}} \sigma_j^k(t_j) \right) \left(\prod_{j \in N \setminus C_m} (1 - \sigma_j^k(t_j)) \right)$$

is slower than for any other minimal winning coalition and type profile. By construction, $W(C_m) \geq W(\hat{C}_i)$ and $t_j = r$ for $j \in C_m \setminus \{i\}$. For such t_{-i}, $f(t)$ could be either S or R if $t_i = s$, and $f(t) = R$ if $t_i = r$, because C_m has no mutually exclusive minimal winning coalition. Thus the right hand side of Equation (C.4) is weakly less than the left for any type and any agent. This is true for any i. The strategy profile σ and the derived belief system μ is an equilibrium rejection of g. $\square$

A.3.1 Super Majority Rules

Lemma 14 (Weak altenative)**.**

Consider a qualified majority rule f. If g has a C^g where $n^g \notin C^g$ and $w^f(C^g) \leq q^f$, then there exist an equilibrium rejection of g.

Proof of Lemma 14. Set $\sigma_i(t_i) = 0$, $\sigma_i^k(r) = k^{-\frac{1}{w_i^g}}$ and $\sigma_i^k(s) = k^{-\frac{2}{w_i^g}}$. Condition (C.1) and (C.2) are trivially satisfied. Since $\sigma_i^k(r) > \sigma_i^k(s)$ for any k and i, and $\sigma_i^k(r) > \sigma_j^k(r)$ for any $i > j$, we have $f(t_{-i}, t_i = r) = R$ and $g(t_{-i}, t_i) = R$ for any i and t_{-i} with $\mu_i(t_{-i}) > 0$. Therefore, Condition (C.4) are satisfied for any i and t_i. $\square$

Lemma 15. *Consider a qualified majority rule f with $q^f \geq \frac{n-1}{2}$. If g has a C^f where $w^g(N \setminus C^f) + w_i^g \leq q^g$ for any $i \in C^f$, then there exists an equilibrium rejection of g.*

Proof. Consider C^f with highest weights under g. Set $\sigma_i(t_i) = 0$, $\sigma_i^k(r) = k^{-\frac{2}{w_i^g}}$ and $\sigma_i^k(s) = k^{-\frac{1}{w_i^g}}$. Since $\sigma_i^k(r) < \sigma_i^k(s)$ for any k and i, and $\sigma_i^k(s) > \sigma_j^k(s)$ for any $i > j$, we have $f(t_{-i}, t_i = s) = S$ and $g(t_{-i}, t_i) = S$ for any i and t_{-i} with $\mu_i(t_{-i}) > 0$. Therefore, Condition (C.4) are satisfied for any i and t_i. $\square$

Lemma 16. *Consider a qualified majority rule f with $q^f \geq \frac{n-1}{2}$. The followings are equivalent.*

1. *A rule g has a C^f where $w^g(N \setminus C^f) + w_i^g \leq q^g$ for any $i \in C^f$.*

2. *$\bar{C}^f$ with highest weights under g satisfies $w^g(N \setminus \bar{C}^f) + w_i^g \leq q^g$ for any $i \in \bar{C}^f$.*

3. *Consider $\bar{C}^f$ with highest weights under g. Then, $w^g(N \setminus \bar{C}^f) + w_{n^g}^g \leq q^g$.*