9-20-2021

A Political Reciprocity Mechanism

Roland Pongou

Jean-Baptise Tondji
The University of Texas Rio Grande Valley

A Political Reciprocity Mechanism*

Roland Pongou[1, 2] and Jean-Baptiste Tondji[3]

[1]Department of Economics, University of Ottawa
[2]Harvard T.H. Chan School of Public Health, Harvard University
[3]Department of Economics and Finance, The University of Texas Rio Grande Valley

September 20, 2021

Abstract

We consider the problem of designing legislative mechanisms that guarantee equilibrium existence, Pareto-efficiency, and inclusiveness. To address this question, we propose a finite-horizon voting procedure that embeds *clauses of reciprocity*. These clauses grant voters the right to oppose actions that are not in their interest, retract actions that face opposition, and punish harmful actions. We study voters' strategic behavior under this voting procedure using two classical approaches. Following the blocking approach, we introduce two related solution concepts—the *reciprocity set* and the *sophisticated reciprocity set*—to predict equilibrium policies. We then show that these solution concepts (1) are always non-empty; (2) only select Pareto-efficient policies; (3) strategically protect minority interests; and (4) are compatible with classical notions of fairness and Rawlsian justice in distributive problems. Following the non-cooperative approach, we provide an implementation of each of these solution concepts in subgame perfect equilibrium, which makes them applicable in a wide range of legislative settings. We also extend them to effectivity functions, a large class of games that includes strategic form games. A comparative analysis shows that the *reciprocity mechanism* has other desirable features and properties that distinguish it from other well-known voting mechanisms and solution concepts.

KEYWORDS: Reciprocity Mechanism, Voting, (Sophisticated) Reciprocity Set, Equilibrium Existence, Pareto-efficiency, Strategic Inclusiveness, Fairness, Rawlsian Justice, Minority.
JEL CLASSIFICATION: P16, D72, C7, J15, H41.

---

*Corresponding authors: Roland Pongou (rpongou@uottawa.ca or rop103@harvard.edu) and Jean-Baptiste Tondji (jeanbaptiste.tondji@utrgv.edu). We are grateful to the Editor Tilman Börgers, the Associate Editor, and two anonymous referees for insightful and constructive comments and suggestions that have helped improve the paper. We also acknowledge with gratitude the useful discussions with Victoria Barham, Braz Camargo, Christopher Cotton, Lars Ehlers, Diego Escobari, Anthony Heyes, John List, Filipe Martins-da-Rocha, Daniel Monte, Aggey Semenov, Roberto Serrano, Ken Shepsle, Matt Webb, Philip Reny, and Stanley Winer and valuable comments of seminar participants at the 50th Annual Conference of the Canadian Economics Association, Carleton University, FGV-Sao Paulo School of Economics, Summer School of Institute Research at the African School of Economics, University of Ottawa, and The University of Texas Rio Grande Valley.

1 Introduction

This paper solves the problem of designing legislative mechanisms that guarantee equilibrium existence, Pareto-efficiency, and inclusiveness. Studies show that the pursuit of self-interest generally leads to the emergence and persistence of public policies that are inefficient and non-inclusive.[1] Experimental studies suggest that these issues can be resolved if agents are induced to display reciprocal and pro-social behaviors in decision-making.[2] However, to our knowledge, in the context of voting, there does not currently exist any formal mechanism that is able to incentivize selfish agents to display such behaviors. This study addresses this problem by proposing a simple voting procedure that incorporates *clauses of reciprocity* and therefore provides incentives for agents to display reciprocal and pro-social behaviors. To capture rational behavior under this procedure, we define two related solutions concepts following the *blocking approach* and implement them using a non-cooperative *implementation-theoretic approach*. Our voting procedure therefore leads to a mechanism that we call a *political reciprocity mechanism*. We show that the solution concepts guarantee the existence of an equilibrium, Pareto-efficiency, and inclusiveness in that they *strategically* protect minority interests and are compatible with well-known concepts of fairness and Rawlsian justice. We provide an overview of our analysis and main results below.

1.1 A voting procedure with reciprocity clauses

Let N be a set of agents, A a set of policies, $\mathcal{C}$ a constitution that defines *winning* coalitions, and V a voting procedure consisting of finite or infinite stages. A *political economy* is defined as $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V, R(\theta)\rangle$, where θ is a state of the world, and $R(\theta)$ is a profile of preferences defined over the policy space A. In this paper, we primarily study political economies in which the voting procedure $V \equiv V^{rm}$ comprises the following four stages. **Stage 0**: A status quo policy $x \in A$ is chosen by Nature, a social planner or an agent. **Stage 1**: A *winning* coalition S can object against x by proposing a policy y. If no winning coalition objects, x remains in place and the voting process ends. If a winning coalition objects by proposing y, Nature can end the voting process and in this case, y becomes the new policy. If Nature does not end the process, it continues to the next stage.[3] **Stage 2**: Each agent $i \notin S$ can oppose y. If there is no opposition, y is adopted and the process ends. If there is any opposition, S is given the opportunity to withdraw y in the next stage. **Stage 3**: If S withdraws y, x remains in place and the process ends. If S maintains y, the process moves to the final stage. **Stage 4**: The opposing agents can form a *winning* coalition T to replace y with z. If T succeeds, z is elected and the process ends. Otherwise, y is elected and the process ends.

[1]See, for example, Falkinger et al. (2000), and Acemoglu et al. (2011).

[2]See, for instance, Axelrod (1984), Fehr and Gächter (1998), Fehr and Gächter (2000), Falk and Fischbacher (2006), and Nash et al. (2012). A **reciprocal behavior** consists of returning favor for favor and harm for harm, and opposing harmful actions. A **pro-social behavior** is an action that promotes social *efficiency*.

[3]We also study a version of the voting procedure in which this role of Nature is removed.

We designate the procedure V^{rm}, a reciprocity voting procedure because Stages 2, 3, and 4 contain **reciprocity clauses (or rules)** that make it possible for agents to: (i) oppose actions that are not in their interest (Stage 2); (ii) retract actions that harm others and therefore face opposition (Stage 3); and (iii) punish actions that harm them (Stage 4). Thanks to these clauses, we show that the voting procedure V^{rm} induces rational agents to act in a reciprocal manner towards one another. In addition, because we provide a non-cooperative implementation of this voting procedure (see below), we will call it a *political reciprocity mechanism*.

1.2 The reciprocity set: blocking versus non-cooperative approach

We formalize the rational behavior of agents under the voting procedure V^{rm} following both the *blocking* and the *non-cooperative* approaches. As put by Dutta and Vohra (2017), the *blocking approach* relies on a coalitional game structure to describe the feasible actions of each coalition in the game and abstracts from the details of the negotiation process (see also Serrano (2020)), whereas the *non-cooperative bargaining* approach relies on specifying details including the protocol that describes the possible actions and order of moves of each player. Our formalization of rational behavior under the procedure V^{rm} takes into account the fact that agents are farsighted given the sequential nature of this procedure.

Following the **blocking approach**, a policy is said to be an *equilibrium* if it cannot be "blocked" by rational agents. In other words, an *equilibrium* policy is such that, if proposed as the status quo, no *winning* coalition will have an incentive to deviate from it under the voting procedure V^{rm}. The **reciprocity set** is the set of all equilibrium policies. Following the traditional literature on coalitional bargaining (Aumann, 1959), the reciprocity set assumes that agents are prudent or ambiguity-averse in the face of uncertainty regarding the future of the game. When such uncertainty does not exist (that is, Nature cannot end the voting process in Stage 1), agents can make moves that are more strategic. This consideration leads to the **sophisticated reciprocity set**, a refinement of the reciprocity set.

Following the **non-cooperative approach**, we formalize the voting procedure V^{rm} as an extensive form game. We give an implementation of the sophisticated reciprocity set by showing that a status quo policy is the unique outcome of all subgame perfect equilibria if and only if it belongs to the sophisticated reciprocity set. We provide a similar non-cooperative implementation of the reciprocity set.[4]

Furthermore, like the Nash equilibrium and the core, we show that the reciprocity set (resp. sophisticated reciprocity set) can be defined by a *domination relation*, and exploit this fact to extend these solution concepts to the class of effectivity function games, which is a very large class of games that includes both transferable-utility games and strategic form games.[5]

[4]Various ways of implementing cooperative solution concepts have been proposed in the literature. Our implementation of the (sophisticated) reciprocity set follows a similar approach as the status quo-based implementation provided by Lagunoff (1994) for the core.

[5]We find, for example, that the reciprocity set selects (Cooperate; Cooperate) as the unique equilibrium of the two-agent Prisoners' Dilemma Game.

Quite interestingly, by following these different approaches, our analysis espouses the view of scholars who argue that the boundary between the two main branches of game theory is blurry (see, for instance, Roth and Wilson (2019), and Roth (2020)). As stated by Roth:

> Today, particularly in areas of applied economics such as market design, cooperative and noncooperative game theory are viewed more as models at different levels of detail than as models of different kinds of games. (Roth, 2020, p. xx)

In our study, we show that these approaches are complementary. There are two main differences between our framework and other models of political decision-making that are used in the literature:[6]

(i) Given that the voting procedure V^{rm} is a sequential procedure, our framework differs from one-shot games (von Neumann and Morgenstern, 1944; Nash, 1951; Gillies, 1959; Aumann, 1959; Schwartz, 1976; Miller, 1980). At the same time, V^{rm} comprises only four stages (excluding Stage 0), which can be more appropriate in a legislative context given the fact that it minimizes the time-cost of decision-making. In fact, the voting procedure V^{rm} is a modification of legislative procedures that are generally employed in democratic countries like the United States, Canada, the United Kingdom, and many others. In these countries, legislative decision-making follows a succession of stages. However, we show that unlike our framework, these procedures can lead to non-Pareto-efficient policies.

(ii) By giving agents an opportunity to amend (or retract) their actions after the first move, the reciprocity mechanism also differs from other studies on sequential multilateral bargaining (see, for example, Harsanyi (1974), Baron and Ferejohn (1989), Chwe (1994), Eguia and Shepsle (2015), Fotso et al. (2017), and Diermeier et al. (2017)). We provide additional distinctions in Section 7.

1.3 Outline of results

Reciprocal behavior. The reciprocity mechanism V^{rm} induces *rational reciprocal behavior*. This behavior consists of avoiding actions that harm others and abstaining from free-riding, which is *positive* reciprocity (Theorem 1). On the non-cooperative side, one can find a simple extensive form game whose equilibrium prediction coincides with the sophisticated reciprocity set (Theorem 2) and the reciprocity set (Theorem 3), respectively. The latter finding implies that reciprocal and pro-social behaviors can be induced under a non-cooperative framework among selfish agents. In fact, the reciprocity set selects cooperation as the unique equilibrium in the prisoners' dilemma game.

Equilibrium existence. Examining the existence of equilibria for discrete and continuous policy spaces, we show that the reciprocity set is always nonempty if agents have strict or linear preferences (Theorem 4 and Theorem 5). This finding implies that the reciprocity mechanism resolves the classical paradox of voting that exists in the core. We also find that if preferences are *single-peaked* over a left-right political spectrum, and if the constitution is the majority rule, there exists at least one equilibrium policy and at most two equilibria. In particular, if the number of voters is odd, the equilibrium is unique and it corresponds to the median voter's ideal point (Proposition 3).

[6] See Section 7 and the Supplement Materials for details on such models.

Pareto-Efficiency. One attractive property of the reciprocity set is that it only selects Pareto-efficient policies regardless of the nature of the policy space (Theorem 6). This finding implies that the reciprocity set, in addition to *strategically* preserving minority interests (Proposition 4), resolves the well-known conflict between individual rationality (the pursuit of self-interest) and Pareto-efficiency (or collective rationality) in political decision making. It therefore prevents political failure regardless of the degree to which political opinions are antagonistic.

Strategic protection of minority interests. In our treatment of minority interests, a *minority group* is a losing coalition or a set of agents who, in a particular context, may favor a policy that a majority of agents dislike. A *minority option* is a policy favored by a minority group. A minority option is said to be *strategic* if it is majority-dominated by a policy that is also majority-dominated. We find that a strategic minority option can be selected under the reciprocity mechanism because of agents' *farsighted* behavior (Proposition 4 and Theorem 7). This property is not necessarily explained by the fact that agents display reciprocal behavior. In fact, some of the other farsighted solution concepts also protect strategic minority options. However, we differ in that only the reciprocity set (and the sophisticated reciprocity set) can *always* protect strategic minority options while also preserving policy efficiency. In this sense, a combination of reciprocal and farsighted behaviors leads to the strategic protection of minorities in a way that is efficient. Also, the (sophisticated) reciprocity set never selects *non-strategic* minority options. Our analysis offers a new explanation to the "decisive minority" phenomenon (Campbell (1999)).

Fairness and Rawlsian justice. We show that two classical concepts of distributive justice always belong to the reciprocity set, when the latter solution concept is applied to the classical distribution problem modeled through a transferable-utility voting game. These concepts are the Shapley value (Shapley, 1953) and the nucleolus (Schmeidler, 1969). Their goal is to distribute the surplus generated by agents that get together to cooperate on a project. What these concepts have in common is that they give equal pay to equally productive agents and they are Pareto-efficient. They differ in that the Shapley value is generally viewed as reflecting meritocratic justice (see, for example, Demeze-Jouatsa et al. (2021), and the references therein), whereas the nucleolus is considered as a formalization of Rawlsian justice because it maximizes the payoff of least productive agents (see, for example, Legros (1987), and Iñarra et al. (2020)). By showing that both solution concepts lead to allocations that always belong to the reciprocity set, we show that this latter solution concept is consistent with classical notions of distributive justice. In this respect, we differ from the core in that it may not contain the Shapley value even when it is not empty. The core always contains the nucleolus when it is not empty. We differ in that while the core can be empty under our mild conditions, the reciprocity set is never empty. The sophisticated reciprocity set preserves all the above properties (Theorem 7).

Comparison with other solution concepts. In Section 7 and the Supplemental Materials, we compare the reciprocity mechanism to other well-known procedures (and related solution concepts) that mainly follow the traditional *blocking approach*, namely the one-shot mechanism, a modified version of the reciprocity mechanism, and several farsighted coalitional procedures. These solutions concepts include, among others, the von Neumann-Morgenstern (vNM) stable set (von Neumann and Morgenstern, 1944), the Core (Gillies, 1959), the Harsanyi stable set (Harsanyi, 1974), the Top cycle set (Schwartz, 1976), the Uncovered set (Miller, 1980), and the Largest consistent set (Chwe, 1994). We find that the reciprocity set and the sophisticated reciprocity set have several comparative advantages. For example, as mentioned above, an equilibrium policy might not exist in a one-shot game, and farsighted coalitional procedures generally select non-Pareto-efficient policies. These two pitfalls are avoided by the reciprocity set under mild conditions. Also, although our goal is not to provide a systematic comparison of the different solution concepts we examine, we provide examples in which the *reciprocity set* refines the predictions of some of the solution concepts developed to capture rationality under the other procedures. Following the *non-cooperative bargaining approach*, other methods in modeling strategic interactions in social environments include multilateral bargaining in legislatures. Notable examples include Baron and Ferejohn (1989), Banks and Duggan (2000), and Diermeier et al. (2017). However, our framework differs from these studies and the large literature that they have inspired in at least three important respects that we describe in Section 7. We strongly emphasize that the only purpose of this comparison is to show that the reciprocity set is not a *redundant* solution concept. Our goal is absolutely not to highlight the relative merits of each approach, which is beyond the scope of this paper.

In Section 2, we describe in detail the voting procedure V^{rm}. Section 3 focuses on the equilibrium concepts that capture strategic behavior under the procedure V^{rm}. First, we introduce the reciprocity set, and we show that reciprocal actions are rational. Second, we introduce the sophisticated reciprocity set, a refinement of the reciprocity set. Third, we provide a full implementation of each of these solution concepts, which follows a non-cooperative approach. Finally, we show that the reciprocity set and the sophisticated reciprocity set can be generalized to the domain of effectivity function games, which constitute a large class of games that include strategic form games. Section 4 and Section 5 study the properties of the reciprocity set. Section 6 discusses some inclusive properties of our new equilibrium concepts under the reciprocity mechanism. In Section 7, we summarize the comparative analyses of the reciprocity set and the sophisticated reciprocity set with other solutions that follow the blocking approach. Section 8 provides additional contributions of our study to the literature. In Section 9, we conclude. We relegate additional elaborations on examples and the comparative analyses of Section 7 to the Supplemental Materials.

2 Setup

2.1 Preliminary Definitions and Notations

A society denoted N, consisting of agents $i = 1, ..., n$, is faced with the problem of choosing a policy from a non-empty policy space A by voting, and Θ is the set of possible states of the world. For simplicity, we suppose that the policy space is the same in all states. We denote $\mathcal{N}$ as the set of all non-empty subsets (or coalitions) of N. For any coalition $S \in \mathcal{N}$, we use $-S$ to denote the complement of S.

Agents hold common information, and in each state $\theta \in \Theta$, each agent $i \in N$ has an ordinal preference relation denoted $R_i(\theta)$ on A; the strict part of $R_i(\theta)$ is denoted by $P_i(\theta)$, and indifference is denoted $I_i(\theta)$. The preference profile in state $\theta \in \Theta$ is denoted $R(\theta) = (R_1(\theta), ..., R_n(\theta))$. Our analysis covers *weak order* and *linear order* preferences. A preference relation $R_i(\theta)$ is a *weak ordering* if it is reflexive, complete, and transitive. A *linear ordering* is a reflexive, anti-symmetric, transitive, and complete binary relation. For $x, y \in A$, $yR_i(\theta)x$ indicates that agent i weakly prefers y to x; $yP_i(\theta)x$ indicates that agent i prefers y to x; and $yI_i(\theta)x$ indicates that agent i is indifferent between y and x. Moreover, for $S \in \mathcal{N}$, $yP_S(\theta)x$ indicates that $yP_i(\theta)x$ for each $i \in S$ (we say S prefers y over x); $yR_S(\theta)x$ indicates that $yP_i(\theta)x$ for some $i \in S$ and $yI_j(\theta)x$ for some other $j \in S$ (we say S weakly prefers y over x); and $yI_S(\theta)x$ indicates that $yI_i(\theta)x$ for all $i \in S$ (we say S is indifferent between y and x).

In this study, we consider both discrete and continuous policy spaces. When A is discrete, we assume that it is finite. For a continuous policy space, we assume that A is a compact and convex subset of a multidimensional vector space, and is endowed with the topology of closed convergence.[7] When A is continuous, we also assume that preferences are continuous. Preference continuity means that an agent who prefers a policy x over another policy y prefers any policy that is *close* enough to x over any policy that is *close* enough to y.

A *constitution* is a distribution of decisive power among the different subsets of the society. It is formalized as a function $\mathcal{C}$ that maps each coalition $S \in \mathcal{N}$ into either 1 or 0 (with $\mathcal{C}(\emptyset) = 0$); $\mathcal{C}(S) = 1$ means that S is a *winning coalition*, and $\mathcal{C}(S) = 0$ means that S is a *losing coalition*. Let $\mathbb{C}$ denote the set of all constitutions, and $W(\mathcal{C})$ the set of winning coalitions under the constitution $\mathcal{C} \in \mathbb{C}$. We impose the following usual conditions on $W(\mathcal{C})$: (1) $W(\mathcal{C}) \neq \emptyset$; (2) for any $S, T \in \mathcal{N}$ such that $S \subset T$, if $S \in W(\mathcal{C})$, then $T \in W(\mathcal{C})$; and (3) for any $S \in \mathcal{N}$, if $S \in W(\mathcal{C})$, then $N - S \notin W(\mathcal{C})$. For $x, y \in A$, we denote $yP(\theta)x$ to indicate that there exists $S \in W(\mathcal{C})$ such that $yP_S(\theta)x$.[8]

A *voting procedure*, denoted V, is a description of the process governing the selection of policies. A voting procedure that is encountered in many real-life settings is the "one-shot" procedure (von Neumann and Morgenstern, 1944; Gillies, 1959), where a challenger is pitted against a status quo, and if supported by a winning coalition (as defined by the prevailing constitution), it replaces the status quo, and the election process ends; otherwise, the status quo remains in place and the election

[7]This is a usual assumption that has proven useful in the analysis of the existence of equilibria in continuous economies; see, for example, Arrow and Debreu (1954) and Hildenbrand (1974).

[8]Under a constitution, a winning coalition has the right (or power) to replace any policy x by any other policy y and can therefore be said to detain global power. Our model easily extends to situations in which coalitions have only local power; in such situations, a coalition might have the right to replace a policy a by a different policy b, but might not have the right to replace c by d; an example is the election of state representatives in the United States of America, where the residents of a state are not eligible to vote for the representatives of a different state. Such situations can be modelled as *effectivity functions* (Chwe, 1994), which constitute a large class of games including strategic form games.

process ends.[9] There exist several variants of this election format (see, among others, Schwartz (1976), Miller (1980), Börgers (2000), and Borgers (2004)). Other real-life voting procedures are dynamic. Under these procedures, a challenger is pitted against a status quo; if supported by a winning coalition, it becomes the new status quo, and the process starts afresh. Some dynamic voting procedures are designed to last only two or three rounds of vote, whereas others can go on indefinitely and only stop when a policy is reached that no winning coalition is willing to replace. Early contributions include, among others, Farquharson (1969), Harsanyi (1974), Rubinstein (1980), Baron and Ferejohn (1989), Chwe (1994), and Ray and Vohra (2015).

A **political economy** (or **game**) is an environment which consists of a set of agents, a policy space, a voting procedure, a constitution, and a preference profile over the policy space. It is formalized as a list $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V, R(\theta) \rangle$. We denote by $\mathbb{P}$ the set of all political economies.

2.2 A Political Economy with a Reciprocity Voting Procedure

In this study, we primarily examine political economies $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta) \rangle$, where the voting procedure V^{rm} comprises the following four stages which we describe below and represent in Figure 1. Following the **blocking approach**, we primarily focus on actions that coalitions can take or cannot take, but we do not offer details on the actions of individual agents; this is because the blocking approach assumes that coalitions are the primary units of decision-making. However, in Section 3.3, following the **non-cooperative approach**, we provide a possible description of each agent's action space.

Stage 0: A status quo policy $x \in A$ is chosen by **Nature** (or by a social planner or an agent).

Stage 1 (**Objection**): If a winning coalition S proposes that x be replaced by another policy y, the pair (y, S) is called a *plausible objection* against x. If no objection against x exists, then x remains in place, which ends the process. If a *plausible objection* (y, S) against x exists, y is recognized as a bill, and if **Nature** does not end the process, the voting procedure advances to the second stage.

Stage 2 (**Right to opposition**): Each agent who is not a member of the sponsoring coalition S has the right to oppose the bill y. If there is no opposition, y becomes the new policy, which ends the process. However, if there is any opposition, the process moves to the next stage.

Stage 3 (**Withdrawal**): Assume there is an opposition to y. Then, S has the right to either withdraw or maintain the bill y. If S withdraws the bill y, the status quo x remains in place

[9] It is important to emphasize the fact that it is the prevailing constitution that determines the set of winning coalitions. For example, if the constitution is the majority rule, then a winning coalition is any set that contains more than half of the voting population; if it is the unanimity rule, the unique winning coalition is the entire voting population. Under certain constitutions such as the ones used in the United Nations Security Council, the World Bank and the International Monetary Fund, some voters have more power than others. Our findings hold regardless of how the constitution is defined.

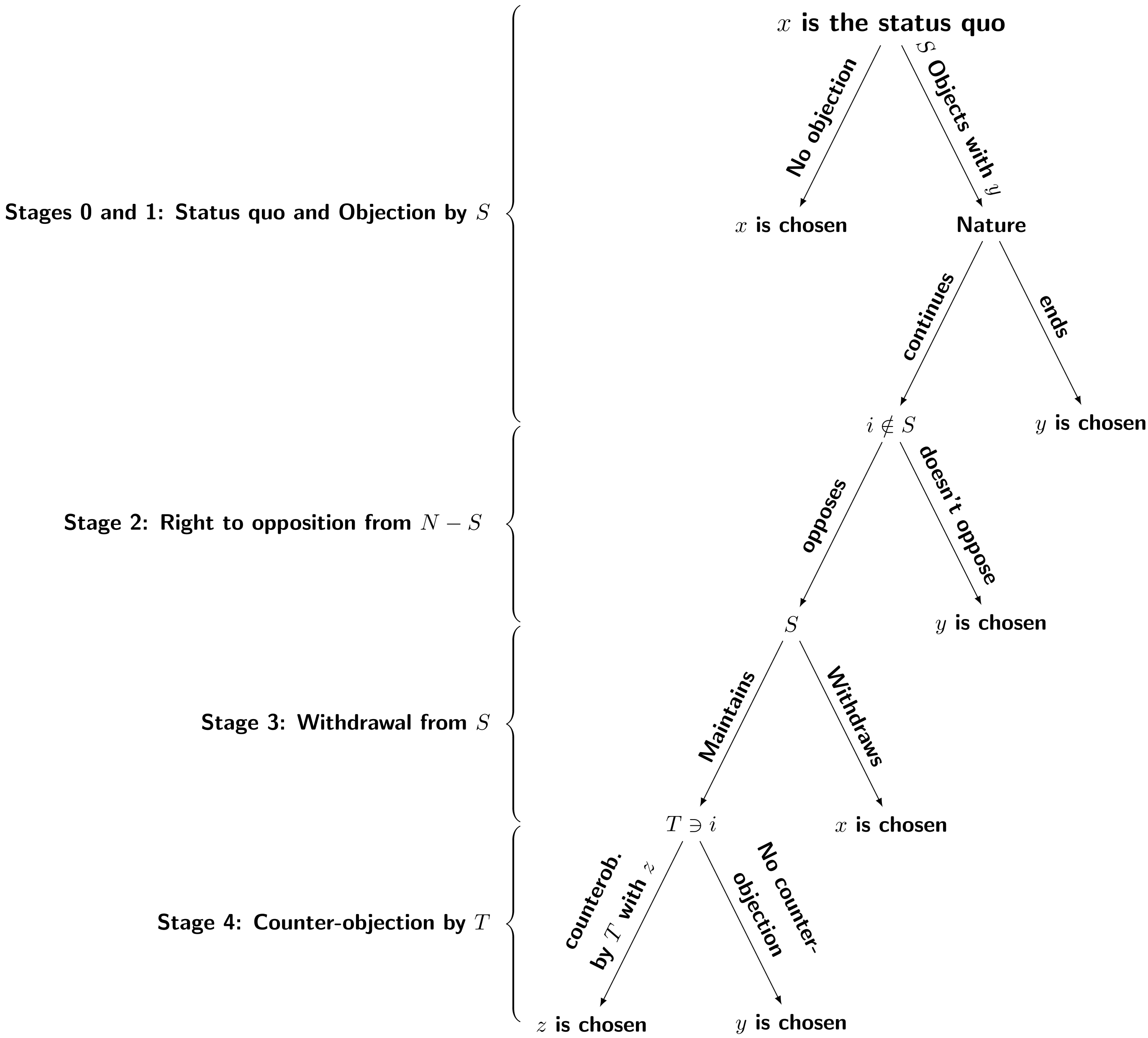

Figure 1: Representation of the voting procedure V^{rm}

and the process ends.[10] However, if S maintains y, then the process moves to the next and final stage, where opposing agents have a right to formulate a counter-objection.

Stage 4 (**Counter-objection**): Suppose that S maintains the bill y. Each opposing agent has the right to invite other agents to form a winning coalition T (obviously the coalition T is different from the coalition S) in order to replace y with another policy z. In this case, z is called an amendment of y and (z, T) is a *plausible counter-objection* to the objection (y, S). If no *plausible counter-objection* to (y, S) exists, then y becomes the new policy, which ends the process. If a *plausible counter-objection* (z, T) exists, z is elected as the new policy and

[10] Recent actions by the U.S. Immigration and Customs Enforcement (ICE) in the Student and Exchange Visitor Program (SEVP) during the peak of COVID-19 pandemic illustrate a decision-making process in which opposition can lead to the retraction of an amendment to the status quo policy (Herpich, 2020).

the process ends.

Three remarks are in order. First, Stages 2, 3 and 4 in the voting procedure V^{rm} incorporate **rules** or **clauses of reciprocity** that make it possible for agents to: (i) oppose actions that are not in their interest (Stage 2); (ii) retract actions that face opposition because they harm others (Stage 3); and (iii) punish or retaliate against actions that are harmful (Stage 4). Indeed, we will show that this voting procedure induces *rational* agents to display *reciprocal actions*—returning favor for favor and harm for harm—and prevents free-riding. For this reason, we will call the mechanism that implements the procedure V^{rm} a *reciprocity mechanism* (see Section 3.3). The reciprocity clauses in the procedure V^{rm} distinguish it from other well-known sequential decision-making procedures in the literature, as we discuss in Section 7 and the Supplemental Materials.

Second, note that there is no agreement that binds the members of a sponsoring coalition S. Furthermore, if an opposition arises at Stage 2, the agents in S who want to withdraw their support of the bill y do not need to cooperate to do so. They can withdraw in a non-cooperative manner, and the bill will be withdrawn only if the remaining members of S who did not withdraw form a *losing coalition*. Following the blocking approach, we assume that agents are self-interested, and they join a coalition only to advance their personal interests. In addition, the voting procedure does not give any special power or veto-right to a winning coalition to initiate the first move. Any winning coalition can initiate the move from the status quo. Agents have full information about existing alternatives and agents' preferences along the process. The advantage of a first mover is its ability to revise its move at the second stage if there is an opposition against its proposal.[11]

Third, the presence of **Nature** in the voting procedure follows the literature (see, for example, Dubey and Shubik (1977), Palfrey and Rosenthal (1985), Osborne and Rubinstein (1994), and Acemoglu et al. (2009)) and is a realistic assumption. However, this assumption is not necessary for our results to hold. In fact, it has been argued that many exogenous factors that are outside the control of voters interfere with decision-making in real-life politics, and that such factors can disrupt a process or even prevent a coalition from forming. It follows that the voting process might end after the move from the status quo x to y initiated by a winning coalition S, even if S anticipates that, under "normal" circumstances, there will be a subsequent move from y to z initiated by another winning coalition T. This is because T might not be able to form due to exogenous circumstances such as sickness of one of the members, an unanticipated change in the political atmosphere, hidden threats directed towards certain members, etcetera. Although the presence of **Nature** is realistic and affects the behavior of the sponsoring coalition S, we will show in Section 6.3 that removing this kind of uncertainty from the process does not affect our main results.

[11]Note, however, that any winning coalition can move first. Therefore, the first-mover advantage does not affect the equilibrium concepts discussed later in the paper.

3 Equilibrium Concepts

3.1 Equilibrium Policies: The Reciprocity Set

We introduce the **reciprocity set** as a solution concept to determine the set of equilibrium policies–policies that will not be replaced if proposed as status quo–under the reciprocity voting procedure V^{rm} described above. This solution concept therefore formalizes the rational behavior of agents by answering the question of when a winning coalition will choose to formulate a *plausible* objection against a status quo under the procedure V^{rm}. This formalization takes into account the fact that agents are necessarily farsighted, given the sequential nature of this voting procedure.

Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$ be a political economy, $x \in A$ be a status quo, and $S \in W(\mathcal{C})$ be a winning coalition that is deciding whether or not to replace x with another policy y. The two following conditions should be satisfied for S to replace x by y:

1. Each member of S should prefer y over x (i.e., $yP_S(\theta)x$; y is said to be majority-preferred to x); and

2. For any winning coalition T different from S that prefers an alternative policy z over y (i.e., $zP_T(\theta)y$), if some member of T prefers the status quo x over the proposed policy y (i.e., $not(yR_T(\theta)x)$), then each member of S should weakly prefer z over x (i.e., $zR_S(\theta)x$). More formally, this condition is expressed as follows:

$$[\forall (z,T) \in A \times W(\mathcal{C}),\ S \neq T,\ zP_T(\theta)y \text{ and } not(yR_T(\theta)x)] \text{ implies } [zR_S(\theta)x].$$

The first condition is a natural requirement that appears in all two-stage bargaining models of rationality in the literature (see, for example, Aumann and Maschler (1964), Mas-Colell (1989), and all of the subsequent studies inspired by this paper). This condition can be viewed as an expression of prudence or ambiguity aversion. It expresses the fact that the sponsoring coalition S considers the presence of **Nature** in the voting procedure and therefore cannot fully predict the future of the game after introducing the bill y. This is because the game might end at y for the reasons mentioned earlier. Therefore, even if each member of S knows that there exists another winning coalition T that will be interested in deviating from the proposed policy y to a more preferred policy z and even if z is not an inferior option for S than x, it is not clear that T will be formed. It follows that y might end up being selected, with the implication that S cannot introduce option y without preferring it over the policy x.[12]

[12]Interestingly, a reader has also mentioned that this assumption can be interpreted as an expression of *bounded rationality*. Such assumptions are often made for realistic reasons. For instance, Demuynck et al. argue that myopic moves are "...natural in complex social environments where the number of possible states is large and agents have little information about the possible actions other agents may take or the incentives of other agents." Demuynck et al. (2019, p. 2). See also Jackson and Wolinsky (1996) for similar arguments in the context of network formation. In Section 6.3, we propose the *sophisticated reciprocity set*—a refinement of the reciprocity set that captures strategic behavior in the absence of Nature and therefore gets rid of the first condition above—, and we show that it preserves all the properties of the reciprocity set.

The second condition says that, if, following a *plausible* objection (y, S) against the status quo x, a winning coalition T formulates a *plausible* counter-objection or an amendment (z, T), then S should weakly prefer z over x provided that certain members of T prefer x over y. In fact, it is straightforward to see that if all of the members of T prefer y over x (i.e., $yR_T(\theta)x$) and if some member of S prefer x over z, then the coalition T will not oppose the move from x to y and so will not formulate the *plausible* counter-objection (z, T) against (y, S). Indeed, if T opposes y, then S will withdraw its *plausible* objection, thus resulting in x remaining in place, an outcome that does not benefit the members of T since they all prefer y over x. This means that an opposition leading to a *plausible* counter-objection (z, T) emerges if and only if T prefers z over y (i.e., $zP_T(\theta)y$) and there is at least one member of T who prefers x over y (i.e., $not(yR_T(\theta)x)$).

The **reciprocity set** is the set of all the *equilibrium* policies, with an *equilibrium* policy being a policy that cannot be replaced if it is a status quo under the voting procedure V^{rm}. We define it formally below.

Definition 1. Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$ be a political economy, $S \in W(\mathcal{C})$ be a winning coalition, and $x, y \in A$ be two policies.

1. **Plausible objection**: (y, S) is said to be a *plausible* objection against x if $yP_S(\theta)x$.

2. **Plausible counter-objection**: Let (y, S) be a plausible objection against x. A pair $(z, T) \in A \times W(\mathcal{C})$ is said to be a *plausible* counter-objection against (y, S) if $zP_T(\theta)y$ and $not(yR_T(\theta)x)$.

3. **Unfriendly plausible counter-objection**: Let (y, S) be a plausible objection against x and $(z, T) \in A \times W(\mathcal{C})$ be a plausible counter-objection against (y, S). The plausible counter-objection (z, T) is said to be *unfriendly* if $not(zR_S(\theta)x)$.[13]

4. **Justified plausible objection**: A plausible objection (y, S) against x is said to be *justified* if there is no unfriendly plausible counter-objection against (y, S). If (y, S) is a justified plausible objection against x, this is denoted by $y \gtrdot_S x$.

5. **Reciprocity set**: The reciprocity set of $\mathcal{PE}(\theta)$, denoted $\mathfrak{E}(\mathcal{PE}(\theta))$, is the set of all the policies in A against which no justified plausible objection exists. Formally, given a state $\theta \in \Theta$,

$$\mathfrak{E}(\mathcal{PE}(\theta)) = \{x \in A : \text{ there does not exist } (y, S) \in A \times W(\mathcal{C}) \text{ such that } y \gtrdot_S x\}.$$

Any policy in the reciprocity set is called a *reciprocity equilibrium*.[14]

[13] The notion of an "unfriendly counter-objection" is inspired by the concept of an "unfriendly amendment" found in the legislative jargon.

[14] Note that, in our framework, plausible objections and plausible counter-objections are formulated in terms of strict domination (yP_Sx if yP_ix for each $i \in S$). All of the results of this paper continue to hold if we use weak domination (yP_Sx if yR_ix for each $i \in S$ and yP_jx for some $j \in S$) instead, though the reciprocity set could be refined.

In a political economy, all agents bargain and settle on an equilibrium policy that emerges from the sequence of threats and counter-threats. The emergence of a policy as equilibrium, in general, reflects the power or influence of different population coalitions as well as the preferences of agents. In our framework, this power is provided by the constitution $\mathcal{C}$. The reciprocity set derives its name from the fact that it captures strategic behavior under the reciprocity voting procedure V^{rm}. As we also show in Section 3.1.1 below, *reciprocal actions*—returning favor for favor and harm for harm–are rational under the procedure V^{rm}.

Interestingly, the reciprocity set defines a *social choice correspondence* $\mathfrak{E}$, which maps any political economy $\mathcal{PE}(\theta) \in \mathbb{P}$ to the subset $\mathfrak{E}(\mathcal{PE}(\theta)) \subseteq A$, the set of equilibrium policies in state $\theta \in \Theta$. One important property of the function $\mathfrak{E}$ is Pareto-efficiency, i.e., for all $\theta \in \Theta$, and all $x \in \mathfrak{E}(\mathcal{PE}(\theta))$, x is **(weakly) Pareto-efficient**: there is no $y \in A$ such that $yP_i(\theta)x$ for all $i \in N$. In Section 5, we show that $\mathfrak{E}$ satisfies Pareto-efficiency when the domain of preferences for each agent consists of strict or weak ordering over discrete or continuous policy spaces.

3.1.1 Reciprocal Actions are Rational under the Voting Procedure V^{rm}

We show that the voting procedure V^{rm} provides incentives for agents to take *reciprocal actions*, regardless of whether agents are selfish or not. The procedure encourages the phenomenon known as *positive reciprocity*—returning favor for favor—, and it prevents the emergence of *negative reciprocity*—returning harm for harm—. Indeed, we prove in Theorem 1 below that the following statements are all true:

1. A first-mover coalition S cannot rationally take any action that harms the interests of any potential second-mover coalition T, unless it is the case that no member of S will be worse off after the subsequent move by T (see item (1) of Theorem 1).

2. Similarly, no second-mover coalition T can rationally take an action that harms the interests of a first-mover coalition S (see item (2) of Theorem 1).

3. As a consequence of 1. and 2., if the move of a first-mover coalition S does not hurt any member of a potential second mover T, then S will effectively initiate the move, even if T would like to initiate a subsequent move to a policy z at which some members of S are worse off relative to the status quo (see item (3) of Theorem 1).

Theorem 1. (Reciprocal actions are rational) *Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$ be a political economy. For all $S, T \in W(\mathcal{C})$ and distinct policies $x, y, z \in A$ such that $yP_S(\theta)x$ and $zP_T(\theta)y$:*

1. *If $not(yR_T(\theta)x)$ (that is, if a move from the status quo x to y will hurt some member of T), then S will not formulate any plausible objection (y, S) against x, unless $zR_S(\theta)x$ (that is, unless no member of S will be hurt at z relative to x after T deviates from y to z following the move from x to y by S). If $not(yR_T(\theta)x)$ and $not(zR_S(\theta)x)$, then S will not formulate a plausible objection (y, S) against x.*

2. *Assume that S formulates a plausible objection (y, S) against x. If the move from x to y does not hurt any member of T ($yR_T(\theta)x$), then T will not formulate an unfriendly plausible counter-objection (z, T) against (y, S) following an opposition.*

3. *As a corollary of items 1. and 2., if $yR_T(\theta)x$, then S will formulate a plausible objection (y, S) against x, even if $not(zR_S(\theta)x)$.*

Proof of Theorem 1. Let $S, T \in W(\mathcal{C})$, and $x, y, z \in A$ such that $yP_S(\theta)x$ and $zP_T(\theta)y$.

1. Suppose that $not(yR_T(\theta)x)$ and $not(zR_S(\theta)x)$. Assume by contradiction that S formulates a plausible objection (y, S) against x. Then, since $zP_T(\theta)y$, a member of T who is not a member of S will oppose the plausible objection. Hence, S will either withdraw or maintain its plausible objection following the prescription in Stage 2. Clearly, S will withdraw its plausible objection since, if it does not this, a plausible counter-objection (z, T) against (y, S) will be formulated by T, thus leading to z being selected as the new policy. This is an inferior outcome for some members of S because $not(zR_S(\theta)x)$. It follows that x will remain in place, which implies that formulating a plausible objection (y, S) against x will not profit S. The coalition S will therefore not formulate such an plausible objection, which is a contradiction.

2. Assume that S formulates a plausible objection (y, S) against x, and suppose that $yR_T(\theta)x$. Assume by contradiction that a member of T who is not a member of S opposes this move. Then, S will either withdraw or maintain its plausible objection. Given the fact that the plausible counter-objection (z, T) is a possibility if S does not withdraw y and is an unfriendly plausible counter-objection by definition (since $not(zR_S(\theta)x)$), S will withdraw its plausible objection, thus causing x to remain in place. But this does not benefit coalition T, since $yR_T(\theta)x$. Consequently, T will not formulate any unfriendly plausible counter-objection (z, T) against (y, S), which is a contradiction.

3. Assume that $yR_T(\theta)x$. If S formulates a plausible objection (y, S) against x, there will be no opposition that could lead to an unfriendly plausible counter-objection (z, T) against (y, S). This occurs because such an opposition will lead to S withdrawing its plausible objection and thus causing the election of x. This response will not benefit the members of coalition T. Then, following the plausible objection, either y will be elected if there is opposition or any opposition and subsequent plausible counter-objection (z, T) will be such that $zR_S(\theta)x$. In either case, S will benefit, thus justifying its objection (y, S) against option x. □

It follows from Theorem 1 that the voting procedure V^{rm} encourages positive reciprocity and prevents negative reciprocity from emerging in that it discourages first movers from taking actions that will trigger retaliation by second movers. It can be seen that reciprocal actions are induced by the possibility for a first-mover coalition to withdraw its move if an opposition arises (Stage 2 in V^{rm}) and by the possibility for any second-mover coalition to retaliate against any first move that is harmful.

3.2 A Refinement: The Sophisticated Reciprocity Set

The reciprocity set assumes that each member of a winning coalition S that formulates a plausible objection (y, S) to a status quo policy x should prefer y to x. As acknowledged earlier, this assumption, which goes back to Aumann and Maschler (1964) and Selten (1981), can be viewed as formalizing a prudent behavior or an aversion to ambiguity when there is uncertainty regarding the continuation of the voting process after the first stage. In this section, we assume that such uncertainty does not exist, implying that agents can act more strategically. Therefore, for $\theta \in \Theta$, we denote a political economy $\mathcal{PE}^s(\theta)$ as an array $\langle N, A, \mathcal{C}, V^s, R(\theta)\rangle$, where V^s is the voting procedure V^{rm} without the intervention of Nature. More precisely, the description of the procedure V^s is exactly as that of the procedure V^{rm}, except for Stage 1 of V^{rm} which is modified as follows:

Stage 1 (**Objection**): If a winning coalition S proposes that x be replaced by another policy y, the pair (y, S) is called a *plausible objection* against x. If no objection against x exists, then x remains in place, which ends the process. If a *plausible objection* (y, S) against x exists, y is recognized as a bill.

Under the voting procedure V^s, we modify the definition of a "plausible objection" in the definition of the reciprocity set. This modification leads to a new solution concept that we call the *sophisticated reciprocity set*. We will show that this new solution concept refines the reciprocity set while also preserving all its properties.

To define the sophisticated reciprocity set, we now assume that a coalition may sponsor a move from a status quo policy x to a policy y even if its members do not prefer y to x. The move would be rational only if the members of the sponsoring coalition know that the voting process will move from y to an alternative z at which they are better off compared to x. More generally, a winning coalition S that wishes to replace a status quo x with a policy y should prefer y to x only if the vote process will not move beyond y, which will happen under two different circumstances. The first is that there is no other winning coalition T that has an incentive to formulate a plausible counter-objection (z, T) (that is, $not(zR_T(\theta)y)$ for any $T \in W(\mathcal{C})$, $T \neq S$, $z \in A$). The second scenario is that if such a coalition T exists, T must not be hurt by the move from x to y (i.e., $yR_T(\theta)x$) and S must be hurt by z relative x (i.e., $not(zR_S(\theta)x)$). In this latter circumstance, no member of T will oppose the move from x to y by S because doing so will cause S to withdraw y under the reciprocity voting procedure V^s, which will hurt T. This reasoning is formalized in the definition of a "strategic objection" in Definition 2 below. The *sophisticated reciprocity set* is defined simply by replacing the notion of a plausible objection in the definition of the reciprocity set (see Definition 1) by the notion of a strategic objection.

Definition 2. Let $\mathcal{PE}^s(\theta) = \langle N, A, \mathcal{C}, V^s, R(\theta)\rangle$ be a political economy, S be a winning coalition, and $x, y \in A$ be two policies.

1. **Strategic objection**: (y, S) is said to be a strategic objection against x if for all $(z, T) \in A \times W(\mathcal{C})$, $S \neq T$, $[yR_T(\theta)x$ and $not(zR_S(\theta)x)]$ or $not(zR_T(\theta)y)$ implies $yP_S(\theta)x$.

2. **Plausible counter-objection**: Let (y, S) be a strategic objection against x. A pair $(z, T) \in A \times W(\mathcal{C})$, $S \neq T$, is said to be a counter-objection against (y, S) if $zP_T(\theta)y$ and $not(yR_T(\theta)x)$.

3. **Unfriendly plausible counter-objection**: Let (y, S) be a strategic objection against x and $(z, T) \in A \times W(\mathcal{C})$ be a counter-objection against (y, S). The counter-objection (z, T) is said to be unfriendly if $not(zR_S(\theta)x)$.

4. **Justified strategic objection**: A strategic objection (y, S) against x is said to be justified if there is no unfriendly counter-objection against (y, S). If (y, S) is a justified strategic objection against x, this is denoted by $y >_S^{sd} x$.

5. **Sophisticated reciprocity set**: The sophisticated reciprocity set of $\mathcal{PE}^s(\theta)$, denoted $SR(\mathcal{PE}^s(\theta))$, is the set of all the policies in A against which no justified strategic objection exists. Formally,

$$SR(\mathcal{PE}^s(\theta)) = \{x \in A : \text{ there does not exist } (y, S) \in A \times W(\mathcal{C}) \text{ such that } y >_S^{sd} x\}.$$

Any policy in $SR(\mathcal{PE}^s(\theta))$ is called a *sophisticated reciprocity equilibrium*.

Example 1 illustrates the difference between the reciprocity set and the sophisticated reciprocity set.

Example 1. Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta) \rangle$ and $\mathcal{PE}^s(\theta) = \langle N, A, \mathcal{C}, V^s, R(\theta) \rangle$ be two political economies, where $N = \{1, 2, 3, 4, 5, 6\}$, $A = \{x, y, z, t\}$, the constitution $\mathcal{C}$ is majority rule, and the preference profile $R(\theta)$ over the set A is defined as follows: $yP_1(\theta)tP_1(\theta)zP_1(\theta)x$; $yP_2(\theta)tP_2(\theta)zP_2(\theta)x$; $zP_3(\theta)yP_3(\theta)xP_3(\theta)t$; $tI_4(\theta)zP_4(\theta)xP_4(\theta)y$; $xP_5(\theta)tP_5(\theta)zP_5(\theta)y$; and $xP_6(\theta)tP_6(\theta)zP_6(\theta)y$. Let $S = \{1, 2, 3, 4\}$, $T = \{3, 4, 5, 6\}$, and $U = \{1, 2, 5, 6\}$.

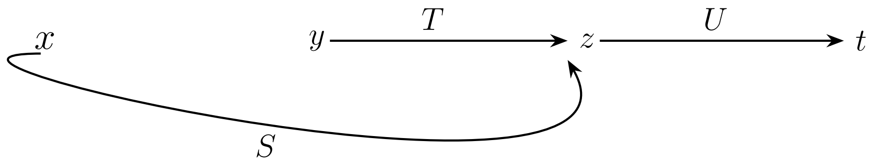

Figure 2: Reciprocity set versus Sophisticated reciprocity set

Figure 2 describes the domination or popularity graph among policies according to agent preferences. The reciprocity set $\mathfrak{E}(\mathcal{PE}(\theta))$ set is larger than the sophisticated reciprocity set $SR(\mathcal{PE}^s(\theta))$: $\mathfrak{E}(\mathcal{PE}(\theta)) = \{x, y, t\}$ and $SR(\mathcal{PE}^s(\theta)) = \{y, t\}$. It is straightforward to see that if x is the status quo, it will not be replaced by any winning coalition if agents behave according to the rationality defining the reciprocity set. However, the winning coalition S will replace x by y if agents behave according to the sophisticated reciprocity set. This is because, despite the fact that S does not prefer y to x, its members know that if they move to y, player 5 will oppose the move, and will subsequently form the winning coalition T to move to z, which is a better outcome for S relative to x.[15]

In Proposition 1, we show that the sophisticated reciprocity set refines the reciprocity set.

Proposition 1. *Let* $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$ *and* $\mathcal{PE}^s(\theta) = \langle N, A, \mathcal{C}, V^s, R(\theta)\rangle$ *be two political economies. Then,* $SR(\mathcal{PE}^s(\theta)) \subseteq \mathfrak{E}(\mathcal{PE}(\theta))$*, and this inclusion may be strict.*

Proof of Proposition 1. Let $x, y \in A$ be two policies, and $S \in W(\mathcal{C})$ be a winning coalition. We note that $y \gtrdot_S^{sd} x$ if and only if: a) $\forall (z,T) \in A \times W(\mathcal{C})$, $S \neq T$, $[yR_T(\theta)x$ and $not(zR_S(\theta)x)]$ or $not(zR_T(\theta)y)$ implies $yP_S(\theta)x$; and, b) $[\forall (z,T) \in A \times W(\mathcal{C}),\ S \neq T,\ zP_T(\theta)y$ and $not(yR_T(\theta)x)]$ implies $[zR_S(\theta)x]$. Let $x \in SR(\mathcal{PE}^s(\theta))$. Assume that $x \notin \mathfrak{E}(\mathcal{PE}(\theta))$. Then, there exists $(y, S) \in A \times W(\mathcal{C})$ such that $y \gtrdot_S x$. It is straightforward to prove that $y \gtrdot_S^{sd} x$, which is a contradiction. Then, $SR(\mathcal{PE}^s(\theta)) \subseteq \mathfrak{E}(\mathcal{PE}(\theta))$. In Example 1, we provide two political economies $\mathcal{PE}^s(\theta)$ and $\mathcal{PE}(\theta)$ such that $SR(\mathcal{PE}^s(\theta)) = \{y, t\} \neq \{x, y, t\} = \mathfrak{E}(\mathcal{PE}(\theta))$. □

3.3 A Non-Cooperative Implementation of the Reciprocity Set and the Sophisticated Reciprocity Set

In this section, we provide a non-cooperative implementation of the reciprocity set and the sophisticated reciprocity. The implementation of the reciprocity set presents a difficulty inherent to the role played by Nature in Stage 1 of the voting procedure V^{rm}. For this reason, we will start with the complete implementation of the sophisticated reciprocity set.[16] A complete implementation of the reciprocity set will follow by assuming that agents solve the ambiguity caused by Nature in Stage 1 of V^{rm} by adopting a "maxmin" behavior. Following this behavior, agents in Stage 1 of V^{rm} vote against a status quo x for a challenger y only if they will not regret this move in the worse case scenario: if Nature ends the game after this move, then y will prevail, otherwise, another policy z will prevail; a "maxmin" behavior implies that an agent who votes for y against x should prefer their least preferred option between y and z to x.

3.3.1 A Non-Cooperative Implementation of the Sophisticated Reciprocity Set

We assume that there is a designer who knows all the components of a political economy except the preference profile $R(\theta)$, given $\theta \in \Theta$. Their goal is to implement the social choice correspondence SR. In the spirit of Lagunoff (1994) who develops a class of sequential voting mechanisms that are relative to a "status quo" outcome for the implementation of the *core* (Edgeworth, 1881; Gillies, 1959), we define a class of (extensive form) mechanisms for the implementation of the sophisticated reciprocity set. Each of these mechanisms can be roughly described as follows. Nature chooses at

[15] Note that following the opposition by agent 5, the coalition S will not withdraw y, giving the opportunity to 5 to form T and move from y to z, ending the game according to the voting procedure V^{rm}.

[16] Remark that given that the sophisticated reciprocity set is included in the reciprocity set, a complete implementation of the former is a partial implementation of the latter.

random a "status quo" policy that will be either adopted or rejected following a sequential voting procedure of perfect information described below. Under this procedure, each agent moves sequentially according to a predetermined and known protocol. In this sequential game, a move will consist of proposing an amendment to the status quo, or opposing or accepting a proposal introduced by another agent, or withdrawing their support from a previous move following the predetermined order. This procedure is a fully described below.

An extensive form game or mechanism is an array $\Gamma = \langle N, A, x, \mathcal{C}, K, \phi, g \rangle$, where N is a set of agents, A the policy space, $x \in A$ a status quo policy, $\mathcal{C}$ the constitution, K the game tree with initial node x (the status quo policy), ϕ a protocol that determines the order of agents' moves in the tree, and $g : Z \times \{\mathcal{C}\} \times \{x\} \longrightarrow A$ is the outcome function, where Z denotes the set of terminal nodes of the tree K. The set of nodes (or steps) of the tree K is denoted $\mathbb{T}$. For all $t \in \mathbb{T} - Z$, agents move according to the protocol ϕ. Let Φ be the set of all protocols ϕ that determine the order in which agents move. Let M_i^t denote the set of actions available to agent i at node t and let M_i denote the set of strategies of agent i. The set of strategy profiles is denoted by $M = \times_{i=1}^n M_i$. For a strategy profile $m \in M$, let $g(m, \mathcal{C}, x, t)$ or simply $g(m, x, t)$ (given that $\mathcal{C}$ is fixed) denote the policy corresponding to m starting at node $t \in \mathbb{T} - Z$; when $t \in Z$, we simply write $g(m, \mathcal{C}, x, t)$ as $g(m, x)$. For a profile $m = (m_i)_{i \in N} \in M$, we denote $m_{-S} = (m_i)_{i \notin S}$, for $S \in \mathcal{N}$. Given a political economy $\mathcal{PE}^s(\theta) = \langle N, A, \mathcal{C}, V^s, R(\theta) \rangle$, the mechanism Γ defines an extensive form game $G = \langle \Gamma, R(\theta) \rangle$, where agents evaluate their payoffs corresponding to each strategy profile m at the policy $g(m, x)$. We denote by $\mathcal{F}(x, \theta)$ the class of all extensive form games attained by varying the protocol ϕ. That is, $\mathcal{F}(x, \theta) = \{G = \langle N, A, x, K, \phi, \mathcal{C}, g, R(\theta) \rangle : \ \phi \in \Phi\}$.

A subgame perfect equilibrium of a game G is a strategy profile $\hat{m} \in M$ such that for all $t \in \mathbb{T}$ and for all $i \in N$,

$$g(\hat{m}, x, t) \ R_i(\theta) \ g(\hat{m}_{-i}, m_i, x, t) \text{ for all } m_i \in M_i.$$

Let SPE(G) denote the set of all policies corresponding to (or selected by) subgame perfect equilibria of the extensive form game G.

We construct a mechanism Γ consisting of six steps illustrated in Figure 4.

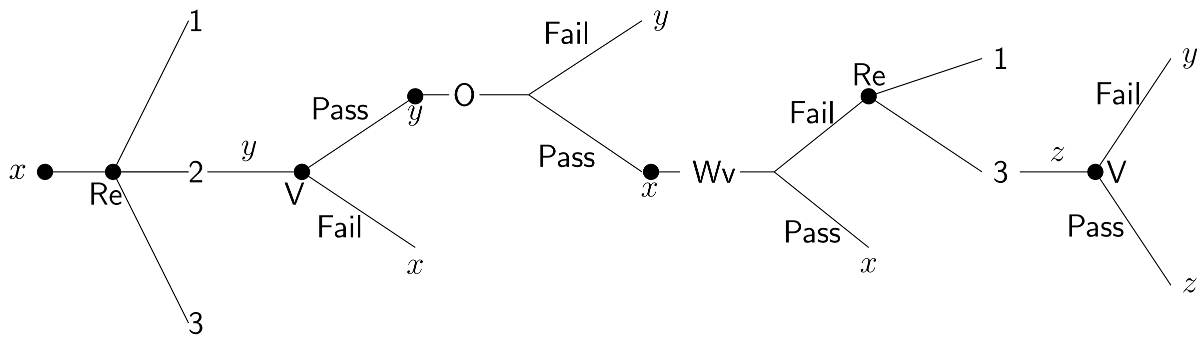

Figure 3: Illustration of the extensive form mechanism. Note: Re = recognition; V = vote; O = opposition; Wv = withdrawal vote.

Step 1. The first agent in the order determined by the protocol $\phi \in \Phi$ makes a proposal y against the status quo x, and each of the other agents chooses the inactive action denoted $\emptyset$. If the first agent does not propose a policy $y \neq x$, the second agent in the protocol ϕ is asked to propose, and so on. If no agent proposes $y \neq x$, the process ends. Otherwise, the game moves to the next step.

Step 2. All agents vote sequentially according to the protocol ϕ between y and x. Each agent has two possible actions: vote "yes" (supports y) or vote "no" (supports x). Let S be the set of agents who support y against x. If S is a winning coalition (i.e., $S \in W(\mathcal{C})$), the vote passes, y is recognized as a bill, and the game advances to Step 3. Otherwise, the status quo x remains in place (that is, S is not a winning coalition or $S \notin W(\mathcal{C})$), and the game ends.

Step 3. The alternative y is on the floor. Each agent i who is not a member of the set S can oppose the bill y. Agent $i \in N - S$ has two actions: vote "yes" (oppose y) or "no" (not oppose y). Implicitly, agents in S do not oppose y; in that case, we assume that each agent in S chooses the inactive action $\emptyset$. All agents in $N - S$ vote, sequentially, according to ϕ. If all agents vote "no" (not oppose y), then y is selected and the game ends. Otherwise, at least an agent in $N - S$ votes "yes" (oppose y) and the vote continues to Step 4.

Step 4. Agents in Step 1 who voted "yes" for y (the coalition S) decide to either withdraw or maintain their support for y. Each agent in S has two actions: vote "yes" (maintains) or "no" (withdraws); and each agent $i \in N - S$ chooses the inactive action $\emptyset$. All agents in S vote, sequentially, according to the protocol ϕ. If the set of agents who vote "no" is such that the remaining agents of S who voted "yes" still form a winning coalition, y is maintained and the game continues to Step 5. Otherwise, the bill y is withdrawn, the status quo x remains in place, and the game ends.

Step 5. The alternative y is on the floor. The first agent (agent i) among the opponents of y in Step 3 ($N - S$), according to the protocol ϕ, proposes an alternative $z_i \neq y$, and each of the other agents chooses the inactive action $\emptyset$.

Step 6. All agents vote, sequentially, according to ϕ. Each agent has two actions: vote "yes" (supports z_i) or "no" (supports y). Let T_i be the set of agents who support z_i against y. If T_i is a winning coalition, then z_i is the chosen policy and the game ends. Otherwise, the first agent (agent $i+1$) among the remaining opponents of y in $N - S - \{i\}$, according to the protocol ϕ, proposes another alternative $z_{i+1} \notin \{y, z_i\}$, and all agents vote according to ϕ. If z_{i+1} is supported by a winning coalition T_{i+1} against y, z_{i+1} is the chosen policy and the game ends. Otherwise, the first agent (agent $i+2$) of y in $N - S - \{i, i+1\}$, according to ϕ proposes $z_{i+2} \notin \{y, z_i, z_{i+1}\}$, and all agents vote according to ϕ. The process continues and ends when either a policy z_j proposed by an opponent $j \in N - S$ is majority-preferred to y or y is not beaten by any of the policies z_j proposed by the opponents of y in $N - S$. In the latter case, y is the chosen policy and the game ends.

When the voting process ends in a given step, agents receive their "utilities" according to the selected policy and their preferences $R(\theta)$. In the mechanism, agents are recognized according to the protocol ϕ, and when it is time to vote, every agent observes the vote as it is cast. Therefore, agents are perfectly informed about the entire history of play whenever they must take an action. Let $\mathbb{T}$ be the set of steps or nodes in our extensive form mechanism. Let h_i^t be the history of moves in the game that contains the identity of proposers, proposals that have been made, and actions taken up to node $t \in \mathbb{T}$, but not including agent i. We can assume that the history at $t = 0$ before the agents move (according to ϕ) in Step 1 is the empty history: $h^i(0) = \varnothing$ for each $i \in N$. At history h_t^i, $t \in \mathbb{T}$, agent i takes action a_i^t such that

$$a_i^t \in \begin{cases} A & \text{if agent } i \text{ is the proposer,} \\ \{\text{yes, no}, \emptyset\} & \text{Otherwise.} \end{cases}$$

We can write that history $h_i^t = (a_1^t, a_2^t, ..., a_{i-1}^t)$. A *strategy* for agent i is a mapping m_i which maps histories h_i^t to actions a_i^t for all $t \in \mathbb{T}$. A strategy profile $m = (m_1, ..., m_n)$ describes a complete sequence of moves in the game given the status quo policy x. This completes the description of the extensive form game. Theorem 2 shows that the sophisticated reciprocity set admits a non-cooperative implementation in the domain $\mathcal{F}$.

Theorem 2. *Let $\mathcal{PE}^s(\theta) = \langle N, A, \mathcal{C}, V^s, R(\theta) \rangle \in \mathbb{P}$ be a political economy and $x \in A$ be a policy,*
(1) if $x \in SR(\mathcal{PE}(\theta))$ then SPE$(G) = \{x\}$ for every $G \in \mathcal{F}(x, \theta)$;
(2) if $x \in$ SPE(G) for some $G \in \mathcal{F}(x, \theta)$, then $x \in SR(\mathcal{PE}(\theta))$.

Theorem 2 shows that equilibrium policies under the reciprocity set are those status quo policies that prevail in all extensive form games in the domain $\mathcal{F}$. By contrast a policy that is not in the reciprocity set can not prevail if proposed as a status quo in any such extensive form game in the domain $\mathcal{F}$.

Proof of Theorem 2. We need some additional notation before proving Theorem 2. Let $x \in A$ be the status quo and $G = \langle N, A, x, K, \phi, \mathcal{C}, g, R(\theta) \rangle \in \mathcal{F}(x, \theta)$ be an extensive form game. Let $\Omega = A \times W(\mathcal{C}) \times N \times \mathcal{N} \times A$, and (y, S, i, T, z) be an element of Ω, with $i \in T - S$, and $y \neq x$. Let $E^x(y, S, i, T, z)$ denote a possible sequence of moves in the game G (given the status quo x; an agent is recognized and proposes y; $S \in W(\mathcal{C})$ consists of agents who vote to support y against x, i.e, $yP_S(\theta)x$; agent $i \notin S$ opposes y; the bill y is not withdrawn; and agents form $T \in \mathcal{N}$ to support z against y, i.e., $zP_T(\theta)y$). A strategy profile m in the game G which corresponds to a sequence $E^x(y, S, i, T, z)$ leads to either y or z as the final policy, i.e., $g(m, x) \in \{y, z\}$; however if $T \in W(\mathcal{C})$, then $g(m, x) = z$.

(1) Let $\mathcal{PE}^s(\theta)$ be a political economy. Consider $G \in \mathcal{F}(x, \theta)$ and assume that $x \in SR(\mathcal{P}^s(\theta))$. We show that $x \in$ SPE(G). Suppose that $x \notin$ SPE(G). Then, any strategy profile $\hat{m} \in M$ which is a subgame perfect equilibrium of the game G does not yield x as the final policy, i.e., $g(\hat{m}, x) \neq x$. Let $\hat{m}$ be such a strategy profile. Then, there exists an element $(y, S, i, T, z) \in \Omega$ such that the

sequence of moves $E^x(y,S,i,T,z)$ generated from $\hat{m}$ leads to either y or z as the final policy, i.e., $g(\hat{m},x) \in \{y,z\}$. We discuss two points. (a) Assume that $[yR_T(\theta)x$ and $not(zR_S(\theta)x)]$ and $T \in W(\mathcal{C})$. Given that $yP_S(\theta)x$, then the pair (y,S) is a strategic objection against x, and $g(\hat{m},x) = z$. Note that (z,T) cannot be a plausible counter-objection against (y,S) since $yR_T(\theta)x$. Let $(z',T') \in A \times W(\mathcal{C})$ be a plausible counter-objection against (y,S), with $z' \neq z$ and $z'R_S(\theta)x$. Then, there exists $j \in T' - S$ such that $xP_j(\theta)y$ and $z'P_{T'}(\theta)y$. Let m be a strategy profile in the game G that yields the sequence of moves $E^x(y,S,j,T',z')$. Given that $z'P_{T'}(\theta)y$ and $T' \in W(\mathcal{C})$, it holds that $g(m,x) = z'$. With $z'R_S(\theta)x$, it follows that the plausible counter-objection (z',T') against (y,S) is friendly, and (y,S) is a justified strategic objection against x. (b) If $not(zR_{T'}(\theta)y)$ for any $T' \in W(\mathcal{C})$, including $T' = T$, then, with $yP_S(\theta)x$, the pair (y,S) is a justified strategic objection against x. Points (a) and (b) imply that $x \notin SR(\mathcal{PE}^s(\theta))$, which is a contradiction. Therefore, $x \in \mathsf{SPE}(G)$.

Now, for a game $G \in \mathcal{F}(x,\theta)$, assume there exists $y \in \mathsf{SPE}(G)$ such that $y \neq x$. Then, there exists a strategy profile $\hat{m} \in M$ which is a subgame perfect equilibrium of the game G with $g(\hat{m},x) = y$. According to the game G, there exists $S \in W(\mathcal{C})$ such that $yP_S(\theta)x$. If there is no $(z,T) \in A \times W(\mathcal{C})$ such that $zP_T(\theta)y$, then (y,S) is a justified plausible objection against x or $x \notin SR(\mathcal{PE}^s(\theta))$, which is a contradiction. Assume there exist $(z,T) \in A - \{y\} \times W(\mathcal{C})$, $T \neq S$, $T \cap (N-S) \neq \emptyset$, and $zP_T(\theta)y$. Let $i \in T \cap (N-S)$ and consider the strategy profile $\overline{m} = (\hat{m}_{-i}, \overline{m}_i)$, where agent i opposes y in Step 3, and if recognized in Step 5 or Step 6, proposes z and votes for z against y, and keeps all other actions according to $\hat{m}_i$. Given that $T \in W(\mathcal{C})$, $g(\overline{m}) = z$ and $z_iP_i(\theta)y$, which is a contradiction, since $y \in \mathsf{SPE}(G)$. Therefore, $y = x$, and $\mathsf{SPE}(G) = \{x\}$ for every $G \in \mathcal{F}(x,\theta)$.

(2) Let $x \notin SR(\mathcal{PE}^s(\theta))$ and $G \in \mathcal{F}(x,\theta)$. There exist $y \in A - \{x\}$ and $S \in W(\mathcal{C})$ such that (y,S) is a justified strategic objection against x. Assume that x is the final policy supported by the subgame perfect equilibrium profile $\hat{m} = (\hat{m}_1, ..., \hat{m}_n) \in M : g(\hat{m},x) = x$.

a) Assume there is no plausible counter-objection against y. Let a_1 be the first agent in S according to the protocol ϕ. By definition, $\hat{m}_{a_1}$ is a best response given any history $h^t_{a_1}$, $t \in \mathbb{T}$. Consider the deviation in which agent a_1 proposes y in Step 1. There are only two scenarios that could yield x as the final policy from this deviation.

Case (a-1). The vote in Step 2 favors x against y.

Case (a-2). The vote in Step 2 favors y against x; an agent $i \in N - S$ opposes policy y; and members in S fail to maintain y.

Consider Case (a-1). Given that (y,S) is a strategic objection against x, it holds that $yP_S(\theta)x$. The vote in Step 2 favors x against y implies that there exists $T \in W(\mathcal{C})$ such that $xP_T(\theta)y$ and $not(yR_T(\theta)x)$. The latter implies that (x,T) is a plausible counter-objection against y, which is absurd by assumption. Additionally, since S and $T \in W(\mathcal{C})$, there exists $i \in S \cap T$ such that $xP_i(\theta)yP_i(\theta)x$ or $xP_i(\theta)x$ by transitivity of $P(\theta)$, which is a contradiction. Consider Case (a-2). Here, as in Case (a-1), the best response for each agent in S is to vote "yes" and not oppose policy y. Therefore, x cannot arise as the final policy of a subgame perfect equilibrium of the game G.

b) Assume that there exists a plausible counter-objection (z, T) against the strategic objection (y, S). Thus, there exists $i \in T - S$ such that $xP_i(\theta)y$, and $zP_T(\theta)y$. Given that the strategic objection (y, S) is justified, then $zR_S(\theta)x$. Following the same reasoning in as in Case a), a subgame perfect equilibrium in the game G can yield only z as the final policy, which is a contradiction. $\square$

3.3.2 A Non-Cooperative Implementation of the Reciprocity Set

In Section 3.3.2, we provide a non-cooperative implementation of the reciprocity set. For simplicity, we will assume that agents have linear preferences. The mechanism that implements this solution concept, denoted as $\Gamma^{\textbf{Na}}$, is described exactly in the same way as the mechanism, Γ, that implements the sophisticated reciprocity set, except for a small modification in Step 2 of the extensive form game $G = \langle \Gamma, R(\theta) \rangle$, which is modified as follows:

Step 2. All agents vote sequentially according to an exogenous protocol ϕ between y and x. Each agent has two possible actions: vote "yes" (supports y) or vote "no" (supports x). Let S be the set of agents who support y against x. If S is a winning coalition, the vote passes, y is recognized as a bill, **and if Nature does not end the voting process**, the game advances to Step 3 below. Otherwise, the status quo x remains in place (that is, S is not a winning coalition), **or Nature ends the process at the policy** y, and the game ends.[17]

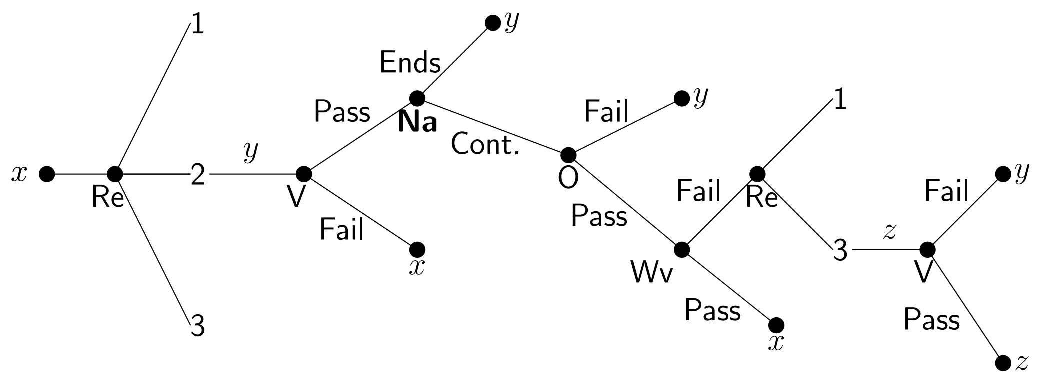

Figure 4: Illustration of the extensive form mechanism. Note: Re = recognition; V = vote; **Na** = Nature; O = opposition; Wv = withdrawal vote.

The game tree resulting from this modification is presented in Figure 4. Given a political economy $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta) \rangle$, the mechanism $\Gamma^{\textbf{Na}}$ defines an extensive form game with Nature, that we denote $G^{\textbf{Na}} = \langle \Gamma^{\textbf{Na}}, R(\theta) \rangle$, where agents evaluate their payoffs corresponding to each strategy profile m at the policy $g(m, x)$ as described in Section 3.3.1. We denote by $\mathcal{F}^{\textbf{Na}}(x, \theta)$ the class of all extensive form games attained by varying the protocol ϕ. That is, $\mathcal{F}^{\textbf{Na}}(x, \theta) = \{G = \langle \Gamma^{\textbf{Na}}, R(\theta) \rangle : \phi \in \Phi\}$. For simplicity, for any game with Nature $G^{\textbf{Na}} = \langle \Gamma^{\textbf{Na}}, R(\theta) \rangle \in \mathcal{F}^{\textbf{Na}}(x, \theta)$, we write $G = \langle \Gamma, R(\theta) \rangle \in \mathcal{F}(x, \theta)$ the corresponding game without Nature.

[17] Nature ends the game with an unknown probability strictly greater than 0.

It follows that Nature creates an uncertainty regarding the continuation of the game process after the move from the status quo. We assume that the social planner perceives ambiguity in voters' preferences and behavior, which is a natural assumption. More precisely, we assume that agents adopt a maxmin behavior which consists of voting against a status quo x for a challenger y only if they will not regret it regardless of whether Nature ends the game process at y (in case those who vote for y form a winning coalition) or not. Assume that agents who vote for y against the status quo x form a winning coalition S. If Nature ends the game after the move from x to y, y will prevail. Otherwise, another policy z will prevail. A maxmin behavior in any equilibrium play in a game $G^{\textbf{Na}}$ implies that for agent i in S, $\min_i \{y, z\} R_i(\theta)x$, where $\min_i \{y, z\}$ is the least preferred policy between y and z from voter i's point of view. Taking this ambiguity aversion into account in an equilibrium play, we define an equilibrium in this uncertain environment as a strategy vector that is a subgame perfect equilibrium of the corresponding extensive form game $G = \langle \Gamma, R(\theta) \rangle \in \mathcal{F}(x, \theta)$, where Nature cannot end the game, and that respects the property according to which $\min_i \{y, z\} R_i(\theta)x$ for each agent i in the sponsoring coalition S. If we take this latter property as a "constraint", then we obtain an equilibrium notion that appropriately captures strategic behavior in our uncertain environment. We will call that notion a **constrained subgame perfect equilibrium**. Formally, a strategy profile $\hat{m} \in M$ is a constrained subgame perfect equilibrium of the game $G^{\textbf{Na}} \in \mathcal{F}^{\textbf{Na}}(x, \theta)$ if:

(a) $\hat{m}$ is a subgame perfect equilibrium of the game $G = \langle \Gamma, R(\theta) \rangle \in \mathcal{F}(x, \theta)$; and

(b) $\min_i \{g(\hat{m}, x), y\} R_i(\theta)x$ for all $i \in S$, where S consists of agents who vote for y against x in $\hat{m}$, and $g(\hat{m}, x) \in A$ is a policy selected by $\hat{m}$.

Let SPE($G^{\textbf{Na}}$) denote the set of all policies selected by constrained subgame perfect equilibria of the extensive form game $G^{\textbf{Na}}$. Theorem 3 shows that the mechanism $\Gamma^{\textbf{Na}}$ implements the reciprocity set the in the domain $\mathcal{F}^{\textbf{Na}}$.

Theorem 3. *Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta) \rangle \in \mathbb{P}$ be a political economy and $x \in A$ be a policy,*
(1) if $x \in \mathfrak{E}(\mathcal{PE}(\theta))$ then SPE($G^{\textbf{Na}}$) $= \{x\}$ for every $G^{\textbf{Na}} \in \mathcal{F}^{\textbf{Na}}(x, \theta)$;
(2) if $x \in$ SPE($G^{\textbf{Na}}$) for some $G^{\textbf{Na}} \in \mathcal{F}^{\textbf{Na}}(x, \theta)$, then $x \in \mathfrak{E}(\mathcal{PE}(\theta))$.

The proof of Theorem 3 uses the same arguments as that of Theorem 2 and is presented in Supplement Materials (Section 1).

The general problem of implementing social choice correspondences through non-cooperative solution concepts such as Nash equilibrium or subgame perfect equilibrium has been extensively analyzed. Canonical mechanisms include, among others, Maskin (1977) and Jackson (1992) and Dutta et al. (1994) for Nash implementation (see also Dasgupta et al. (1979), Maskin (1985), Moore (1992), and Jackson (2001) for excellent surveys); Selten (1981), Moore and Repullo (1988) and Abreu and Sen (1990) for subgame perfect implementation. Some works on the implementation of the core include, among others, Kalai et al. (1979), Lagunoff (1994), Moldovanu and Winter (1994), Perry and Reny (1994), Serrano (1995), Serrano and Vohra (1997), Bergin and Duggan

(1999), Banks and Duggan (2000), and Okada (2012). Gul (1989) and Hart and Mas-Colell (1996) implement the Shapley value; Harsanyi (1974) implements the stable sets; Einy and Wettstein (1999) and Perez-Castrillo and Wettstein (2000) consider the implementation of bargaining sets; and Serrano (1993) implements the Nucleolus. We view the implementation-theoretic approaches in Sections 3.3.1 and 3.3.2 as providing a non-cooperative underpinning of the sophisticated set and the reciprocity set, respectively.

3.3.3 The Predicted (Sophisticated) Reciprocity Set

The non-cooperative implementation of the reciprocity set and the sophisticated reciprocity suggests that one can predict the outcome of vote from any status quo policy. If a policy is in the reciprocity set (resp. sophisticated reciprocity set), by definition, that policy will be selected if proposed as the status quo under the voting procedure V^{rm} (resp. V^s). Otherwise, it will be replaced by another policy. In this section, we introduce the *predicted reciprocity set* (resp. sophisticated reciprocity set) as a solution concept that determines the set of policies that can be selected from a given status quo under the voting procedure V^{rm} (resp. V^s). This solution concept is intuitively similar to that of the subgame perfect equilibrium.

Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$ (resp. $\mathcal{PE}^s(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$) be a political economy, and $x \in A$ be a status quo. If x is not a reciprocity equilibrium (sophisticated reciprocity equilibrium), there exists a justified plausible objection (resp. strategic objection) (y, S) against x. If there is no opposition or plausible counter-objection against y, y will be the predicted outcome at x. If there is any opposition and, in particular, if there is a plausible counter-objection against y that leads to the proposition of amendment z, then z will be the predicted outcome at x. Indeed, given the fact that (y, S) is a justified plausible objection (resp. strategic objection) against x, the latter plausible counter-objection is not unfriendly to the coalition (i.e., S) supporting the plausible objection (resp. strategic objection). The precise description of the *predicted reciprocity set* (resp. *sophisticated reciprocity set*) is provided in Definition 3.

Definition 3. Let $\mathcal{PE}(\theta) = \langle N, A, V^{rm}, \mathcal{C}, R(\theta)\rangle$ (resp. $\mathcal{PE}^s(\theta) = \langle N, A, \mathcal{C}, V^s, R(\theta)\rangle$) be a political economy and $x \in A$ be a status quo. The *predicted reciprocity set* (resp. *predicted sophisticated reciprocity set*) at x, denoted $\mathfrak{E}^x(\mathcal{PE}(\theta))$ (resp. $SR^x(\mathcal{PE}^s(\theta))$), is defined as follows:

$\mathfrak{E}^x(\mathcal{PE}(\theta)) = \{x\}$ if $x \in \mathfrak{E}(\mathcal{P}(\theta))$; otherwise

$\mathfrak{E}^x(\mathcal{PE}(\theta)) = \{y \in A :$ there exist $S \in W(\mathcal{C}),\ y \gtrdot_S x$ and there does not exist $z \in A,\ zPy\}$, or

$\mathfrak{E}^x(\mathcal{PE}(\theta)) = \{z \in A :$ there exist $S \in W(\mathcal{C})$ and $y \in A,\ y \gtrdot_S x$ and $zPy\}$.

$SR^x(\mathcal{PE}^s(\theta))$ is defined similarly as $\mathfrak{E}^x(\mathcal{PE}(\theta))$ by replacing $\mathfrak{E}(\mathcal{PE}(\theta))$ by $SR(\mathcal{PE}^s(\theta)))$ and $y \gtrdot_S x$ by $y \gtrdot_S^{sd} x$ in the definition of $\mathfrak{E}^x(\mathcal{PE}(\theta))$.

From Definition 3, each equilibrium policy is its own prediction, but there might exist predicted policies that are not equilibria. The analysis indicates that the predicted reciprocity set (resp.

predicted sophisticated reciprocity set) from any given status quo policy is non-empty because the voting procedure involves a finite number of sequential actions. In other words, in a political economy $\mathcal{PE}(\theta) = \langle N, A, V^{rm}, \mathcal{C}, R(\theta)\rangle$ (resp. $\mathcal{PE}^s(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$), a reciprocity equilibrium (resp. sophisticated reciprocity equilibrium) is a fixed point of the correspondence $\mathfrak{E}^x(\mathcal{PE}(\theta))$ (resp. $SR^x(\mathcal{PE}^s(\theta))$.

3.4 Applications of The Reciprocity Set to Effectivity Functions and Strategic Form Games

In Section 3.4, we apply the reciprocity set to effectivity functions and strategic form games.[18] To perform these applications, we need additional notations. We use the concept of "effectivity functions" in place of a constitution to distribute power or abilities among coalitions over the policy space.[19] We represent a political economy as an array $\mathcal{PE}(\theta) = \langle N, A, V^{rm}, \{\rightarrow_S\}_{S\in F\in\mathcal{N}}, R(\theta)\rangle$, where N is a finite set of agents, A is the policy space, V^{rm} is the voting procedure with reciprocity clauses, F is a collection of admissible coalitions, $\{\rightarrow_S\}_{S\in F\in\mathcal{N}}$ is an "effectivity function" (Chwe, 1994) describing the distribution of abilities among the admissible coalitions in F—where for each $S \in F$, $\rightarrow_S$ is a binary relation defined over A—, and $R(\theta)$ is a preference profile defined over A. For each element $S \in F$, the effectivity function $\rightarrow_S$ determines the set of actions that S can take in the decision-making process. For instance, for $x, y \in A$, the notation $x \rightarrow_S y$ indicates that S has the ability (or power) to replace x by y if given the opportunity to do so. Additionally, the effectivity function $\rightarrow_S$ is monotonic: for $x, y \in A$ and $S, T \in \mathcal{N}$ such that $S \subset T$, $x \rightarrow_S y$ implies $x \rightarrow_T y$. Monotonicity means that if a coalition S has the power to replace x by y, then a coalition that is larger than S (in the sense of set inclusion) has that power too. This is a plausible assumption in voting games.

We extend the **reciprocity set** to this larger setting. In fact, given a status quo $x \in A$, a pair $(y, S) \in A \times \mathcal{N}$ is a plausible objection against x if $x \rightarrow_S y$ and $yP_S(\theta)x$. A pair $(z, T) \in A \times \mathcal{N}$ is a plausible counter-objection against a plausible objection (y, S) if $y \rightarrow_T z$, $zP_T(\theta)y$ and $not(yR_T(\theta)x)$. The reciprocity set consists of justified plausible objections. Consider a strategic form game $\Lambda = \langle N, (X_i)_{i\in N}, (u_i)_{i\in N}\rangle$, where N is a non-empty set of agents, each agent $i \in N$ has a finite, non-empty, strategy set X_i and a utility or payoff function $u_i : X \times \Theta \rightarrow \mathbb{R}$, where $X = \times_{i\in N} X_i$ is the set of strategy profiles. Given $S \in \mathcal{N}$, let $X_S = \times_{i\in S} X_i$, and given $x \in X$, let $x_S = (x_i)_{i\in S}$. The reciprocity set naturally extends to strategic form games because, as shown in Chwe (1994), they are a subclass of effectivity function games.[20] Indeed, in what follows, we can write a strategic form game using effectivity functions.

[18] For expositional purposes, we will only be working with the reciprocity set. However, all the properties of the reciprocity set that we uncover are also valid for the sophisticated reciprocity set.

[19] See, for example, Chwe (1994) and Fotso et al. (2017), for a brief survey of games that are defined with effectivity functions.

[20] The class of effectivity function games includes other prominent subclasses including transferable-utility games, non-transferable-utility games, and network games. It follows that the domain of the reciprocity set is large.

Our application of the reciprocity set to strategic form games follows Chwe (1994, pp. 313–318) who conducts a similar analysis using the largest consistent set. However, as we explain in Section 7 and in the Supplement Materials, our solution concepts differ significantly in terms of their properties. For example, the largest consistent set may select a non-Pareto-efficient alternative, whereas the reciprocity (sophisticated reciprocity) set may not.

We apply the reciprocity set to make predictions in each of the following situations involving a strategic form game Λ. I. Agents play a simultaneous-move game (Chwe, 1994, p. 313); II. Agents announce "individual contingent treats" (Greenberg, 1990, p. 98), or coalitions of agents announce "coalitional contingent threats" (Greenberg, 1990, p. 102); and III. Coalitional deviations satisfy the "self-enforceability" criterion (see, for example, Bernheim et al. (1987)). We present our applications to the first two situations. One can easily derive the last situation from the analysis below and the contributions of Chwe (1994, pp. 313–318).

I. First, the reciprocity set can predict which of several pure strategy Nash equilibria will emerge as the result of a finite-horizon bargaining procedure with perfect information. Let $A = NE(\Lambda)$ represent the set of pure strategy Nash equilibria of a strategic form game Λ. For $x, y \in A$, let $yR_i(\theta)x$ if $u_i(y,\theta) \geq u_i(x,\theta)$, and let $x \to_S y$ if $x_{N-S} = y_{N-S}$. For example, consider the two-agent game described in Table 1. Agent i's strategy set is $X_i = \{ia, ib, ic\}$, for $i = 1, 2$, and $u(x,\theta) \equiv u(x) = (u_1(x), u_2(x))$ for any $x \in X$. There are three pure strategy Nash equilibria: $NE(\Lambda) = \{(1a, 2a), (1b, 2b), (3a, 3b)\}$. If we consider $A = NE(\Lambda)$, then the reciprocity set is $\mathfrak{E}(\mathcal{PE}(\theta)) = \{(1a, 2a)\}$. The latter consists of the only Pareto-efficient alternative in $NE(\Lambda)$. Given that each option in A is a result of the situation in which agents play a simultaneous-move game, the reciprocity set makes a prediction about which alternative is likely to result from the bargaining process if there is some form of communication or coordination between farsighted agents.

		Agent 2		
		$2a$	$2b$	$2c$
	$1a$	$(3,3)$	$(0,3)$	$(0,0)$
Agent 1	$1b$	$(3,0)$	$(2,2)$	$(0,2)$
	$1c$	$(0,0)$	$(2,0)$	$(1,1)$

Table 1: Of the three Nash equilibria, the reciprocity set selects only one

II. Second, we can apply the reciprocity set to "individual contingent threats situations" or "coalitional contingent threats situations", a setting in which agents in the strategic form game do not play a simultaneous-move game. Instead, each coalition S facing a strategy profile $x \in X$ can announce: "If all you other agents stick to playing x_{N-S}, we will play $y_S \in X_S$ instead of x_S."[21] Agents can announce and update such threats in the decision-making process when they have the opportunity to do so, without any commitment. Let $A = X$. For $x, y \in A$, let $yR_i(\theta)x$ if $u_i(y,\theta) \geq u_i(x,\theta)$, and let $x \to_S y$ if $x_{N-S} = y_{N-S}$. The reciprocity set in the contingent threats situation contains

[21] The quote is adapted from Greenberg (1990, p. 102) and Chwe (1994, p. 315).

the *strong Nash equilibria*. A strategy profile $x \in X$ is a *strong Nash equilibrium* if for all $S \in \mathcal{N}$, and for all $y_S \in X_S - \{x_S\}$, there exists $j \in S$ such that $u_j(x, \theta) \geq u_j((y_S, x_{N-S}), \theta)$. Let $SN(\Lambda)$ be the set of strong (pure strategy) Nash equilibria of the game Λ. A strategy profile $x \in SN(\Lambda)$ implies that no coalition (including the grand coalition, i.e., all agents collectively) can profitably deviate from x. It follows that any $x \in SN(\Lambda)$ is both Pareto-efficient and a Nash equilibrium.

Proposition 2. *Let $\Lambda = \langle N, (X_i)_{i\in N}, (u_i)_{i\in N}\rangle$ be a strategic form game in which N is finite and non-empty and the strategic sets X_i are finite and non-empty. Consider $\mathcal{PE}(\theta) = \langle N, A, V^{rm}, \{\rightarrow_S\}_{S\in F\in\mathcal{N}}, R(\theta)\rangle$ be a political economy corresponding to the contingent threats situation, where $A = \times_{i\in N} X_i$, $yR_i(\theta)x$ if $u_i(y, \theta) > u_i(x, \theta)$, and $x \rightarrow_S y$ if $x_{N-S} = y_{N-S}$. If $x \in SN(\Lambda)$, then $x \in \mathfrak{E}(\mathcal{PE}(\theta))$.*

Proof of Proposition 2. Say $x \in SN(\Lambda)$ and $x \notin \mathfrak{E}(\mathcal{PE}(\theta))$. Then, there exists $(y, S) \in A \times \mathcal{N}$ such that $x \rightarrow_S y$ and $yP_S(\theta)x$. By definition of $\mathcal{PE}(\theta)$, $x_{N-S} = y_{N-S}$. Thus, $yP_S(\theta)x$ implies that $u_i((y_S, x_{N-S}), \theta) > u_i(x, \theta)$, for each $i \in S$. The latter contradicts the fact that x is a strong Nash equilibrium. Therefore, $x \in SN(\Lambda)$ implies $x \in \mathfrak{E}(\mathcal{PE}(\theta))$. □

In the game displayed in Table 1, the profile $(1a, 2a)$ is the only strong Nash equilibrium and the reciprocity set is $\{(1a, 2a)\}$. However, the reciprocity set and the strong Nash equilibrium may differ. For example, consider the Prisoner's Dilemma game: each of the N agents chooses either to cooperate (C) or to defect (D), and D is the best action when the other player plays C. However, if all the players choose D, then the outcome of the game is worse to all of them than the outcome that arises if all choose C. The Prisoner's Dilemma game doesn't admit a strong Nash equilibrium in pure strategies (and the only pure strategy Nash equilibrium is (D,D)), while the profile in which each agent plays C is the unique element in the reciprocity set. Also, in the game described in Table 2 where $N = \{1, 2, 3\}$, and agent i's strategy set is $X_i = \{ia, ib\}$, for $i \in N$. The game admits no strong Nash equilibrium, while the reciprocity set is $\{(1a, 2a, 3a), (1a, 2b, 3b), (1b, 2b, 3a)\}$.

		Agent 2				
		$2a$	$2b$		$2a$	$2b$
Agent 1	$1a$	$1,2,3$	$0,0,0$	$1a$	$4,0,4$	$3,1,2$
	$1b$	$0,4,4$	$2,3,1$	$1b$	$0,0,0$	$4,4,0$
		$3a$	$3a$		$3b$	$3b$
				Agent 3		

Table 2: No strong Nash equilibria exist

These examples highlight some nice properties of the reciprocity set as a solution concept. Indeed, we will show that, unlike the set of pure strategy Nash equilibria, the reciprocity set is never empty (that is, a reciprocity equilibrium always exists), and all reciprocity equilibria are Pareto-efficient.

The next section provides some examples to illustrate the reciprocity mechanism.

3.5 Examples

In the following example, we illustrate the reciprocity set and its non-cooperative implementation to a game consisting of selecting a beverage for a party.

Example 2. (**Juice, wine, or beer?**) Suppose that five friends, Anneli, Barbara, Connie, Diane, and Lucy are selecting a party beverage. For reasons of saving some money from economy of buying in bulk, they decide to choose a common beverage. Let $N = \{1, 2, 3, 4, 5\}$ be the set of friends and $B = \{b_1, b_2, b_3\}$ be the set of beverages, where b_1 is juice, b_2 is wine, and b_3 is beer. The preference profile $R(\theta)$ over beverages is the following: Anneli (friend 1) : $b_3 P_1(\theta) b_2 P_1(\theta) b_1$; Barbara (friend 2) : $b_3 P_2(\theta) b_2 P_2(\theta) b_1$; Connie (friend 3) : $b_2 P_3(\theta) b_1 P_3(\theta) b_3$; Diane (friend 4) : $b_1 P_4(\theta) b_3 P_4(\theta) b_2$; and Lucy (friend 5) : $b_1 P_5(\theta) b_2 P_5(\theta) b_3$. We suppose that the constitution $\mathcal{C}$ is defined as: $S \in W(\mathcal{C})$ if either ($|S| = 3$ and $5 \notin S$) or ($|S| \geq 4$). Figure 5 shows the popularity relationship among the various beverages (for example, beer (b_3) is more popular than wine (b_2) as the former is preferred by three out of the four friends not including friend 5). One can check that

$$b_1 \xrightarrow{\{1,2,3\}} b_2 \xrightarrow{\{1,2,4\}} b_3$$

Figure 5: Popularity relationship among the beverages

the reciprocity set is $\mathfrak{E}(\mathcal{PE}(\theta)) = \{b_1, b_3\}$. To understand why, assume that b_1 is the status quo. Despite the fact that each player in the winning coalition $\{1, 2, 3\}$ prefers b_2 to b_1, this coalition will not form to move to b_2 because player 3 will not join it; in fact, if this coalition forms and moves to b_2 and Nature does not end the game, player 4 will oppose this move in Stage 2 of the game and coalition $\{1, 2, 3\}$ will withdraw b_2 in the Stage 3 because if it does not, players in the winning coalition $\{1, 2, 4\}$ will move to b_3 in Stage 4, which is the worst option for player 3. If b_2 is the status quo, it is clear that $\{1, 2, 4\}$ will move to b_3, which is why b_3 is not a reciprocity equilibrium. If b_3 is the status quo, no winning coalition will form to move to another alternative. It follows that $\mathfrak{E}(\mathcal{PE}(\theta)) = \{b_1, b_3\}$.

Now, let us model this game as an extensive form game and show that the only alternatives that are the outcomes of (subgame perfect) maxmin equilibria if these alternatives are proposed as status quo are b_1 and b_3. Assume the order of moves in the discussion between the friends is given by the protocol $\phi = (12354)$, i.e., Anneli is first, Barbara is second, Connie is third, Lucy is fourth, and Diane is fifth.

1. If b_1 is the status quo, then Anneli has the choice of proposing b_2. If she does that, then each friend in the group $S = \{1, 2, 3\}$ will vote for b_2, and Lucy and Diane will vote for b_1, yielding b_2 the motion on the floor because $S \in W(\mathcal{C})$. Note that Anneli and the other friends make self-interested decisions here (you propose or vote for b_2 against b_1 because you prefer the former to the latter) since there is a possibility that friends might run out of time (Nature), and decide to end the process at b_2. Otherwise, the game continues and Lucy can oppose b_2. Then, each friend in S votes on whether or not to maintain her support for b_2. If the withdrawal fails, Lucy is the first opponent

of b_2 to make a proposal against b_2. According to Lucy's preferences, she will propose b_1, and the vote between b_2 and b_1 will yield b_2 as the winner. Then, Diane, the second opponent against b_2 will propose b_3 against b_2, and b_3 will win because it is more popular than b_2. But b_3 is the worst alternative for Connie (friend 3). It follows from backward induction that Connie will vote for the withdrawal in Step 4 resulting in the remaining coalition of supporters $S - \{3\} = \{1, 2\} \notin W(\mathcal{C})$ becoming a losing coalition, and the status quo b_1 will be the final adopted decision. This conclusion remains if we choose any other order of moves. It follows that all the (subgame perfect) maxmin equilibria lead to the outcome b_1: $\mathsf{SPE}(G^{\mathbf{Na}}) = \{b_1\}$ for every $G \in \mathcal{F}^{\mathbf{Na}}(b_1, \theta)$.

2. If b_2 is the status quo, following the protocol $\phi = (12354)$, Anneli proposes b_3 against b_2, and the friends vote between b_3 and b_2. Each person in $S' = \{1, 2, 4\}$ votes for b_3, and Connie and Lucy vote for b_2. Given that $S' \in W(\mathcal{C})$, the alternative b_3 defeats b_2. If Nature does not end the discussion, then either Connie or Lucy can oppose b_3 in Step 3, and each friend in S' decides to maintain or withdraw her initial support to b_3. Given their preferences, there is no alternative that will defeat b_3, and each person in S' maintains the initial decision in Step 2, and the withdrawal fails. According to the protocol ϕ, Lucy is recognized to make a proposal against b_3, and she can propose either b_1 or b_2. If Lucy proposes b_1, each friend in $\{3, 4, 5\}$ votes for b_1, while Anneli and Barbara vote for b_3. Given that $\{3, 4, 5\} \notin W(\mathcal{C})$, the proposal b_3 remains on the floor. If Lucy proposes b_2, then b_3 will defeat b_2. Given that Connie also proposes either b_2 or b_1 against b_3, it follows that no alternative defeats the latter in Step 6, and from backward induction $\mathsf{SPE}(G^{\mathbf{Na}}) = \{b_3\}$ for every $G^{\mathbf{Na}} \in \mathcal{F}^{\mathbf{Na}}(b_2, \theta)$.

3. If b_3 is the status quo, each friend has no incentive to make a different proposal against b_3, given its popularity. Any other unilateral deviation from that strategy profile is deterred. In fact, according to the protocol $\phi = (12354)$, the first friend who can propose a different choice against b_3 is Connie (friend 3), and she will propose either b_2 or b_1. After Connie's moves, the process that follows is similar to the game described in 2. We have similar situations when either Diane or Lucy moves. Therefore, if b_3 is the status quo, the game ends in Step 1, and b_3 is the final choice. Hence, $\mathsf{SPE}(G^{\mathbf{Na}}) = \{b_3\}$ for every $G^{\mathbf{Na}} \in \mathcal{F}^{\mathbf{Na}}(b_3, \theta)$. Thus, $\mathfrak{E}(\mathcal{PE}(\theta)) = \{b_1, b_3\}$, and the friends can buy either juice or beer for the party.

The second example set out below elucidates the difference between the reciprocity set and the predicted reciprocity set.

Example 3. (**Predicted versus Equilibrium Policies**) Consider a political economy $\mathcal{PE}(\theta)$, where $N = \{1, 2, 3, 4, 5, 6\}$, $A = \{a, b, c, d\}$, $\mathcal{C}$ is the majority rule, and the preference profile $R(\theta)$ over the set A is given as follows: $bP_1(\theta)dP_1(\theta)cP_1(\theta)a$; $dP_2(\theta)bP_2(\theta)cP_2(\theta)a$; $cP_3(\theta)bP_3(\theta)dP_3(\theta)a$; $cP_4(\theta)bP_4(\theta)aP_4(\theta)d$; $aP_5(\theta)dP_5(\theta)cP_5(\theta)b$; and $aP_6(\theta)dP_6(\theta)cP_6(\theta)b$. Let $S = \{1, 2, 3, 4\}$, $T = \{3, 4, 5, 6\}$, and $U = \{1, 2, 5, 6\}$. The domination or popularity graph among policies based on preferences is provided in Figure 6. The reciprocity set is $\mathfrak{E}(\mathcal{PE}(\theta)) = \{b, d\}$, and the predicted reciprocity sets are: $\mathfrak{E}^a(\mathcal{PE}(\theta)) = \{c\}$, $\mathfrak{E}^b(\mathcal{PE}(\theta)) = \{b\}$, $\mathfrak{E}^c(\mathcal{PE}(\theta)) = \{d\}$, and $\mathfrak{E}^d(\mathcal{PE}(\theta)) = \{d\}$. If the policy a is the status quo, then the policy c will be elected, since (c, S) is a justified objection against a. However, c is not a reciprocity equilibrium because if c itself is the status quo, then the winning coalition U will object against c by proposing d. Since there is no other possibility to move from d, the latter will be elected. The alternatives b and d are the only reciprocity equilibria in $\mathcal{PE}(\theta)$.

$$\begin{tikzcd}
a \arrow[r, "S"] \arrow[rr, bend right, "S"'] & b \arrow[r, "T"] & c \arrow[r, "U"] & d
\end{tikzcd}$$

Figure 6: Predicted versus Equilibrium Policies (the arrows indicate the direction of the popularity relationship; for instance b is a more popular policy than a because b is preferred over a by the majority coalition S)

The next example is a modification of the traditional ultimatum game. In this modified version, objections can only be made by winning coalitions with respect to a specific constitution. In the Supplement Materials, we provide more details on how to derive the reciprocity set and the predicted reciprocity set for this example.

Example 4. (Ultimatum game with counter-offer) A population of n agents must share an amount of 100 dollars, with each agent receiving a non-negative portion. The set of feasible allocations is the simplex:

$$A = \{x = (x_1, ..., x_n) \in [0, 100]^n : \sum_{i=1}^{n} x_i \leq 100\}.$$

Assume that the constitution $\mathcal{C}$ is majority rule, and for the state $\theta \in \Theta$, the preference profile $R(\theta) = (R_i(\theta))_{i \in N}$ is defined as follows: for $x, y \in A$, and $i \in N$, $yR_i(\theta)x$ if $u_i(y) \geq u_i(x)$, where $u_i : A \longrightarrow \mathbb{R}$ is agent i's payoff function. The decision-making process follows the voting procedure V^{rm}. An arbitrator proposes a status quo allocation $x_0 = (0, 0, ..., 0)$. The traditional ultimatum game is played between two agents (a proposer and a responder) and, unlike the procedure V^{rm}, it does not allow for further negotiation after the first move. An interesting question is whether a reciprocity equilibrium exists. Assume that each agent i's utility function is strictly increasing in his or her payoff x_i, which is a natural assumption. Then, in this framework, a reciprocity equilibrium always exists, though it might not be unique. Furthermore, if agents have other-regarding preferences (with each agent i's utility depending on both his or her payoff x_i and other agents' payoffs x_{-i}) and if utilities are continuous (but not necessarily differentiable), we show that a reciprocity equilibrium always exists, and that all equilibria are Pareto-efficient. To illustrate, suppose that there are three agents who have the following linear utility functions: $u_1(x_1, x_2, x_3) = x_1 - \frac{1}{3}(x_2 - x_1) - \frac{1}{3}(x_3 - x_1)$; $u_2(x_1, x_2, x_3) = x_2 - \frac{1}{3}(x_1 - x_2) - \frac{1}{3}(x_3 - x_2)$; and $u_3(x_1, x_2, x_3) = x_3 - \frac{1}{3}(x_1 - x_3) - \frac{1}{3}(x_2 - x_3)$. Interpreting these utility functions, agents have fairness considerations or other-regarding preferences in that fixing the payoff of an agent i and increasing the payoff of another agent has the effect of decreasing i's utility. The analysis proves that (see more details in the Supplement Materials) the set of predicted allocations from the status quo x_0 is:

$$\mathfrak{E}^{x_0}(\mathcal{PE}(\theta)) = \left\{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = 100 \text{ and } x_i \geq \frac{100}{6};\ i = 1, 2, 3\right\},$$

and the reciprocity set is (observe that $\mathfrak{E}^{x_0}(\mathcal{P}(\theta)) \subset \mathfrak{E}(\mathcal{P}(\theta))$:

$$\mathfrak{E}(\mathcal{PE}(\theta)) = \left\{\begin{array}{l}(x_1, x_2, x_3):\ x_1 + x_2 + x_3 = 100, \text{ and there do not exist two agents} \\ i, j \in \{1, 2, 3\},\ (i \neq j) \text{ such that } x_i = x_j = 0;\ i, j = 1, 2, 3\end{array}\right\}.$$

The determination of all of the equilibria under the reciprocity mechanism is possible in certain contexts (for instance, Examples 2, 3, and 4). However, in certain other political economies, the process can be tedious. The question then arises as to whether it is feasible to prove the existence and analyze the properties of equilibrium policies under very general conditions. The next section establishes more general results on the existence of reciprocity equilibria in our framework.

4 Existence of Reciprocity Equilibria

The results in Section 4 reveal that a reciprocity equilibrium always exists under very natural assumptions on agents' preferences. Interestingly, when preferences have particular known structures, there might exist only one or two equilibria. However, uniqueness is not to be expected in general, which might explain why structurally identical societies can have very different policies. Each subsection that follows makes a different assumption about the structure of preferences and of the policy space, and analyzes the existence of reciprocity equilibria under this assumption. Generally, the identity of an equilibrium policy depends on three structural factors: (1) the structure of preferences; (2) the nature of the political space; and (3) the size of the voting population.

4.1 Discrete Policy Space

Section 4.1 deals with the existence of equilibria when agents have linear or strict preferences over a finite set of policies. Preference linearity simply means that agents have a strict ordering of all of the available policies. We have the following result.

Theorem 4. *Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$ be a political economy. Assume that A is finite ($|A| < \infty$) and $R_i(\theta)$ is a linear ordering for all $i \in N$. Then, a reciprocity equilibrium exists (that is, the reciprocity set is not empty).*

Proof of Theorem 4. Let $x, y \in A$ be two policies, and $S \in W(\mathcal{C})$ be a winning coalition. Note that (y, S) is justified plausible objection against x or $y \gg_S x$ if and only if a) $yP_S(\theta)x$; and, b) $[\forall(z, T) \in A \times W(\mathcal{C}),\ S \neq T,\ zP_T(\theta)y$ and $not(yR_T(\theta)x)]$ implies $[zR_S(\theta)x]$. Define the real-valued function $f : A \longrightarrow \mathbb{R}$ as follows: for all $x \in A$, define $f(x) = |\{y \in A; yP(\theta)x\}|$ (where

$|X|$ denotes the cardinality of the set X) and let $x_0 \in A$ such that $f(x_0) = \min_{x \in A}\{f(x)\}$. We show that $x_0 \in \mathfrak{E}(\mathcal{PE}(\theta))$. If this assertion is not true, then there exists $y \in A$ and $S \in W(\mathcal{C})$, such that $y \gg_S x_0$. It follows that:

$$\begin{cases} (\alpha)\ yP_S(\theta)x,\ S \in W(\mathcal{C}) \\ (\beta)\ \forall (z,T):\ T \neq S,\ zP_T(\theta)y, T \in W(\mathcal{C}) \text{ and } not(yR_T(\theta)x) \Rightarrow zR_S x \end{cases} \quad (1)$$

By definition of x_0 and given the fact that the binary relation $P(\theta)$ is asymmetric, there exists $c \in A$ such that $cP(\theta)y$ and $not(cP(\theta)x_0)$. Thus, there exists a coalition T, with $T \in W(\mathcal{C})$ such that $cP_T(\theta)y$. If $yP_T(\theta)x_0$, then we have $cP_T(\theta)y$ and $yP_T(\theta)x_0$, so that $cP_T(\theta)x_0$ by transitivity, which contradicts $not(cP(\theta)x_0)$. Suppose that $not(yP_T(\theta)x_0)$. Then according to the assertion (β) in equation (1), we have $cP_S(\theta)x_0$, which contradicts $not(cP(\theta)x_0)$. We conclude that $x_0 \in \mathfrak{E}(\mathcal{PE}(\theta))$. $\square$

Theorem 4 is illustrated in Examples 2 and 3. These examples also reveal that a multiplicity of reciprocity equilibria can emerge in our framework. The empirical implications of this finding are that different policies might prevail in structurally identical economies.

4.2 Continuous Policy Space

We assume that the policy space A is a compact and convex subset of the multidimensional vector space. Without loss of generality, we assume that $A = [0,1]^k$ where k is a natural number. We also denote by LM the Lebesgue measure on the affine manifold spanned by A. Given the nature of the policy space, we analyze the existence of an equilibrium policy when agents display linear and continuous preferences over A.

Theorem 5. *Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$ be a political economy, and assume that agent preferences are continuous, linear, and endowed with the topology of closed convergence. Then, a reciprocity equilibrium exists.*

Proof of Theorem 5. Let $g : A \longrightarrow \mathbb{R}$ be the real-valued function defined over the policy A by $g(x) = LM(\{y \in A : yP(\theta)x\})$. Given the fact that preferences are continuous, and A is compact and convex, there exists $x_1 \in A$ such that $g(x_1) = \min_{x \in A}\{g(x)\}$. Following the same reasoning as in the proof of Theorem 4, we show that $x_1 \in \mathfrak{E}(\mathcal{PE}(\theta))$. $\square$

In Section 4.2.1, we present an application to a classical and widely studied framework.

4.2.1 Left-Right Political Spectrum: Median Voter Theorem

We examine the predictions of the reciprocity mechanism when the policy space can be represented by a left-right political spectrum (see, for example, Castles and Mair (1984), Giddens (1994), Bobbio (1996), and Evans et al. (1996)), agent preferences are *single-peaked*, and the constitution is majority rule. Real-life political situations that can be modeled by this framework are numerous, which is perhaps the reason why it has been widely studied in the literature (see, for example, Black (1948), Inada (1964), Grandmont (1978), Moulin (1980), Sprumont (1991), Thomson (1997), Austen-Smith and Banks (1999), Ehlers et al. (2002), Austen-Smith and Banks (2005), Barberà et al. (2017)). A preference relation over a policy space is said to be *single-peaked* if the policies can be ordered as points on a line; and if the preference relation has a maximum point; and if points farther away from this maximum point are less preferred. To make this definition precise, let us assume that all the policies are ordered by a binary relation denoted $\gg$, and that all agents perceive these policies as being arranged in this order. An agent i has a single-peaked preference $R_i(\theta)$ if there exists a policy or peak p_i^* such that: (1) for any other policy $p \neq p_i^*$, $p_i^* P_i(\theta) p$; and (2) for any policy $p, q \in A$, if $[p \gg q \gg p_i^*$ or $p_i^* \gg q \gg p]$, then $q R_i(\theta) p$. We will assume that the set of policies is the set of the ideal points of agents.

Our main result is that there exists at least one and at most two reciprocity equilibria when agents having single-peaked preferences. In addition, when the number of agents is odd, there is only one equilibrium policy, and the latter coincides with the ideal policy of the median voter.

Proposition 3. *Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta) \rangle$ be a political economy. Assume that preferences are single-peaked over A, and $\mathcal{C}$ is majority rule. Then, there exists at least one and at most two reciprocity equilibria. In addition, if n is odd, there exists only one reciprocity equilibrium.*

Proof of Proposition 3. Given that the dimension of A is 1, we can assume that the relation $\gg$ is the natural ordering $\geq$ (and $>$ the strict part of $\geq$). Let $p \in A$ be a policy and define $S(p)$ as the number of agents for whom p is the peak. Consider the real-valued functions f and g defined on A as follows: for any $q \in A$, $f(q) = \sum_{q \geq p} S(p) - \frac{n}{2}$, and $g(q) = \sum_{p \geq q} S(p) - \frac{n}{2}$. Let $A_f = \{p \in A : p$ is a peak and $f(p) \geq 0\}$ and $A_g = \{p \in A : p$ is a peak and $g(p) \geq 0\}$. Note that neither A_f, nor A_g is empty. Given the fact that A_f is finite and the fact that f is an (strictly) increasing function, there exists a unique peak q_1^* that minimizes f over A_f. Similarly, given the fact that A_g is finite and the fact that g is a (strictly) decreasing function, there exists a unique peak q_2^* that minimizes g over A_g. We note the following facts:

Fact 1: For any $q \in A$, $f(q) + g(q) = S(q)$. This is derived from the definitions of f and g.

Fact 2: There is no peak q such that either $q_1^* > q > q_2^*$ or $q_2^* > q > q_1^*$.

Fact 3: If $q_1^* \neq q_2^*$, then $q_2^* > q_1^*$.

We prove that q_1^* and q_2^* are the only equilibria. We consider three cases.

Case 1: n is even and $q_1^* = q_2^* = q^*$. Then, the analysis claims that q^* is the only equilibrium: $\mathfrak{E}(\mathcal{PE}(\theta)) = \{q^*\}$. Let $q' \in A$ be a peak. If $q^* > q'$, by definition of q^*, $\sum_{q' \geq p} S(p) < \frac{n}{2}$ and $\sum_{p \geq q^*} S(p) > \frac{n}{2}$, which implies that q^* is majority-preferred to q' in a pairwise majoritarian vote. Similarly, if $q' > q^*$, we show in the same way that q^* is majority-preferred to q' in a pairwise majoritarian vote. It follows that q^* is majority-preferred to any other peak in a pairwise majoritarian vote. Since there is no other option that is majority-preferred to q^*, then there is no justified plausible objection against q^*. Therefore, $\mathfrak{E}(\mathcal{PE}(\theta)) = \{q^*\}$.

Case 2: n is even and $q_1^* \neq q_2^*$. In this case, we show that, q_1^* and q_2^* are the only equilibria. First, note that neither q_1^*, nor q_2^* is majority-preferred to any other policy in a pairwise majoritarian vote. Furthermore, in a pairwise majoritarian vote opposing q_1^* and q_2^*, neither will win, since $q_2^* > q_1^*$, and $\sum_{q_1^* \geq p} S(p) + \sum_{p \geq q_2^*} S(p) = n$, each will receive exactly $\frac{n}{2}$ votes. Now, let $q' \in A$ be a peak that is distinct from q_1^* and q_2^*. It follows from Facts 2 and 3 that, either $q_1^* > q'$ or $q' > q_2^*$. If $q_1^* > q'$, then q_1^* is majority-preferred to q' in a pairwise majoritarian vote since the number of ballots obtained by q_1^* against q' is at least equal to $\sum_{p \geq q_1^*} S(p) > \sum_{p \geq q_2^*} S(p) \geq \frac{n}{2}$. Based on the fact that q_1^* is not majority-preferred to any other policy, then $q_1^* \gtrdot q'$. If $q_1^* > q'$, then q_1^* is majority-preferred to q' in a pairwise majoritarian vote since the number of ballots obtained by q_1^* against q' is at least equal to $\sum_{p \geq q_1^*} S(p) > \sum_{p \geq q_2^*} S(p) \geq \frac{n}{2}$. Based on the fact that q_1^* is not majority-preferred to any other policy, and $q_1^* \gtrdot q'$. We conclude that $\mathfrak{E}(\mathcal{PE}(\theta)) = \{q_1^*, q_2^*\}$.

Case 3: n is odd. We show that $q_1^* = q_2^*$. Consider the contrary, meaning that $q_1^* \neq q_2^*$. Fact 3 implies that $q_2^* > q_1^*$. Since n is odd and there is no agent whose peak is strictly comprised within the interval (q_1^*, q_2^*) (Fact 2), the number of agents whose peak is weakly to the left of q_1^* (i.e., $\sum_{q_1^* \geq p} S(p)$) is different from the number of agents whose peak is weakly to the right of q_2^* (i.e., $\sum_{p \geq q_2^*} S(p)$). Two situations can arise: 1) Suppose that $\sum_{q_1^* \geq p} S(p) > \sum_{p \geq q_2^*} S(p)$. Recall that by the definition of q_2^*, $\sum_{p \geq q_2^*} S(p) \geq \frac{n}{2}$, then, $\sum_{q_1^* \geq p} S(p) > \frac{n}{2}$. This implies that $\sum_{q_1^* \geq p} S(p) + \sum_{p \geq q_2^*} S(p) > n$, which is a contradiction, because $\sum_{q_1^* \geq p} S(p) + \sum_{p \geq q_2^*} S(p) = n$. 2) If $\sum_{q_1^* \geq p} S(p) < \sum_{p \geq q_2^*} S(p)$, then, using the definition of q_1^* ($\sum_{p \geq q_1^*} S(p) \geq \frac{n}{2}$), it follows that $\sum_{q_1^* \geq p} S(p) > \frac{n}{2}$. Consequently, $\sum_{q_1^* \geq p} S(p) + \sum_{p \geq q_2^*} S(p) > n$, which is a contradiction, given that $\sum_{q_1^* \geq p} S(p) + \sum_{p \geq q_2^*} S(p) = n$. Therefore, $q_1^* = q_2^*$. Referring to **Case 1**, q^* is the unique equilibrium (note that in Case 1, the proof does not necessarily use the assumption that n is even). This concludes that $\mathfrak{E}(\mathcal{PE}(\theta)) = \{q^*\}$. $\square$

Proposition 3 is a generalization of the median voter theorem to the reciprocity mechanism. Indeed, when the number of agents is odd, the unique equilibrium that exists is the ideal policy of the median voter. When the number of agents is even, a median agent may not exist, and in this case, there are two reciprocity equilibria that are very close to the center of the political spectrum.[22] Example 5 set out below on the admission of refugees illustrates these two situations.

Example 5. (Admitting refugees into a peaceful country)
How many refugees should the government of a peaceful country admit? We assume that such a decision is made by the legislators of this country. The country derives utility from the number of refugees that it admits. The utility can be in terms of the publicity (warm glow or the internal

[22]Remark that this result is not a consequence of the median voter theorem for one-shot games. In fact, the reciprocity set is in general larger than the equilibrium notion–the core–for which the original median voter theorem is proved. However, Proposition 3 shows that the reciprocity set coincides with the core when we consider the conditions that sustain the median voter theorem.

feelings of warmth and satisfaction) that it receives, or in terms of the skills, knowledge, and experience brought by the refugees. Each member of the country has a different perception of the utility that he or she or the country receives from admitting refugees. We assume that these considerations are reflected in the legislators' preferences that are represented by utility functions. The net utility received by each legislator i from p refugees being admitted is: $v_i(p) = u_i(p) - s_i p$, where u_i is an increasing and strictly concave function, and s_i is the fraction of the total cost of refugee admission incurred by the constituency of legislator i. Assume that $\mathcal{C}$ is majority rule, and the preference profile $R(\theta) = (R_i(\theta))_{i \in N}$ is defined as follows: for p and q numbers of refugees, and $i \in N$, $pR_i(\theta)q$ if $v_i(p) \geq v_i(q)$. Assume, for illustration, that there are $n = 5$ legislators who have to decide on the number of refugees to be admitted, and the utility function of legislator i is given by: $v_i(p_i) = b_i \ln(p_i) - \frac{1}{n} p_i$, with $b_i = 6 - i$, $i \in \{1, 2, 3, 4, 5\}$, and the policy space is the set of legislators' ideal number of refugees or peaks. Under the reciprocity mechanism, the unique reciprocity equilibrium is $p_3^* = 15$, the peak of the median legislator (the Condorcet winner). However, for $n = 6$ and $b_6 = 6$, there is no median legislator, and the peaks $p_2^* = 24$ and $p_3^* = 18$ are the only reciprocity equilibria. We relegate details of the calculations to the Supplement Materials.

5 Pareto-Efficiency and the Reciprocity Set

In this section, we show that every reciprocity equilibrium is **Pareto-efficiency**, which is a highly desirable property. Theorem 6 shows that the social choice correspondence $\mathfrak{E}$ satisfies Pareto-efficiency when the domain of preferences for each agent consists of strict or weak ordering over discrete or continuous policy spaces.

Theorem 6. *Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta) \rangle$ be a political economy. Assume A is either discrete or continuous, and $R_i(\theta)$ is weak or strict, for all $i \in N$. Then, every $x \in \mathfrak{E}(\mathcal{PE}(\theta))$ is Pareto-efficient.*

Proof of Theorem 6. Let $x \in A$ be a policy such that $x \in \mathfrak{E}(\mathcal{PE}(\theta))$. If x is not Pareto-efficient, then there exists $y \in A$ such that $yP_N(\theta)x$. It follows that for any $z \in A$ such that $zP_T(\theta)y$ with $T \in W(\mathcal{C})$, we have $yP_T(\theta)x$ because $T \subset N$. Then, $not(yR_T(\theta)x)$ is always false. Therefore the implication $[not(yR_T(\theta)x) \Rightarrow zR_N(\theta)x]$ is always true—that is, (y, N) is a justified plausible objection against x, which contradicts the assumption that $x \in \mathfrak{E}(\mathcal{PE}(\theta))$. Thus, x is Pareto-efficient. $\square$

Theorem 6 implies that the voting procedure V^{rm} resolves the conflict between agent rationality and optimality in public goods provision. As we prove in Section 3.1.1, by inducing selfish agents to adopt reciprocal and pro-social behavior, the procedure prevents free-riding, which is one reason why the reciprocity set cannot select a non-Pareto-efficient policy. This property also distinguishes the reciprocity set from several prominent solution concepts as we show in Section 7 and in the Supplement Materials (Section 3).

6 Inclusiveness: Strategic Minority Protection, Rawlsian Justice, and Fairness Under the Reciprocity Mechanism

In this section, we uncover other desirable properties of the reciprocity mechanism and the reciprocity set. We study inclusiveness under this mechanism. We do it in two different ways. The first approach consists of studying the "decisive minority" phenomenon—where an option preferred by a *minority group* can win a democratic election. We observe that policies that are preferred by *minority groups* (or losing coalitions) under a given constitution can emerge as reciprocity equilibria. In the political economy literature, two main reasons have been provided to explain the "decisive minority" phenomenon, namely *preference intensity* and *free-riding*. Campbell (1999) provides a detailed description of these arguments. In Section 6.1, we provide an additional argument to the decisive minority phenomenon, which is voters' farsighted (or strategic) behavior.

The second approach to studying inclusiveness is to show that the reciprocity set is compatible with classical notions of fairness and Rawlsian justice. Indeed, in Section 6.2, following a normative perspective, we show that the reciprocity set of a distributive economy always contains outcomes of two prominent concepts of distributive justice—the Nucleolus (Schmeidler, 1969) and the Shapley value (Shapley, 1953)—that are well-known in economic theory to have other desirable properties towards "minorities" or the worse-off.[23] More precisely, the nucleolus maximizes the welfare of the worse-off and is therefore considered as a formalization of Rawlsian justice (see, for example, Serrano (2020), and the references therein), and the Shapley value is a well-known notion of fairness (Aguiar et al. (2020), Aguiar et al. (2018)).

6.1 Decisive Minority Under the Reciprocity Mechanism

In our treatment of minority interests, a *minority group* is simply a set of agents who, in a particular context, favor a policy alternative that a majority of agents dislike. Minority groups can also include a group of agents with a "fixed" identity, such as an ethnic or a religious group (see Example 6 below). In our framework, protecting minority interests means allowing policies that may only be preferred by minority groups in the society to be part of the set of reciprocity equilibria. It is worth pointing out that our goal is not to propose specific policies that might protect minority interests. We are interested in uncovering whether a *minority option* can be an equilibrium under the reciprocity mechanism. In what follows, we formally define the notions of *minority* and *majority options*. We also define the concept of *"strategic" minority protection*, which we distinguish from the notion of *"mechanical" minority protection*. In Definitions 4 and 5, we assume that a political economy $\mathcal{PE}(\theta)$ is any array $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V, R(\theta) \rangle$, where V is a voting procedure which consists of finite or infinite stages.

Definition 4. Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V, R(\theta) \rangle$ be a political economy, and $x \in A$ be an alternative.

[23] We thank the Editor Tilman Börger for encouraging us to examine the relationship between fairness and the reciprocity set.

1. x is said to be a *minority option* if there exists an alternative $y \in A$ that is preferred to x by a winning coalition $S \in W(\mathcal{C})$: $yP_S(\theta)x$. If a policy is not a minority option, we say that it is a *majority option*.

2. A minority option x is said to be *non-strategic* if there exist a **majority option** $y \in A$ and a winning coalition $S \in W(\mathcal{C})$ such that $yP_S(\theta)x$.

Definition 5. Let $\overline{\mathfrak{E}}$ be a solution concept.[24]

1. $\overline{\mathfrak{E}}$ is said to *mechanically protect minority interests* if there exist $\mathcal{PE}(\theta) \in \mathbb{P}$, and a minority option $x \in A$, such that $x \in \overline{\mathfrak{E}}(\mathcal{PE}(\theta))$.

2. $\overline{\mathfrak{E}}$ is said to *strategically protect minority interests* if $\overline{\mathfrak{E}}$ mechanically protects minority interests, and for any $\mathcal{PE}(\theta) \in \mathbb{P}$, the set $\overline{\mathfrak{E}}(\mathcal{PE}(\theta))$ never contains a non-strategic minority option.

Intuitively, the definition of the strategic protection of minority interests takes into account the fact that decision-makers are farsighted. We have the following result.

Proposition 4. *The reciprocity set strategically protects minority interests.*[25]

Proof of Proposition 4. Consider a political economy $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$, where $N = \{1, 2, 3, 4\}$, $A = \{x, y, z\}$, $\mathcal{C}$ is the majority rule, and preference profile $R(\theta)$ is defined as follows: $zP_1(\theta)yP_1(\theta)x$; $zP_2(\theta)yP_2(\theta)x$; $yP_3(\theta)xP_3(\theta)z$; and $xP_4(\theta)zP_4(\theta)y$. The policy y is majority-preferred to x. Then x is a minority option. Since $x \in \mathfrak{E}(\mathcal{PE}(\theta))$, it follows that the reciprocity set mechanically protects minority interests. Now, fix $\mathcal{PE}(\theta) = \langle N, A, V, \mathcal{C}, R(\theta)\rangle$ and x a non-strategic minority option. Then, there exist a majority option $y \in A$ and a winning coalition $S \in W(\mathcal{C})$ such that $yP_S(\theta)x$. It follows that (y, S) is a justified plausible objection against x and $x \notin \mathfrak{E}(\mathcal{PE}(\theta))$. □

Proposition 4 shows that a minority option can be preserved because of voters' *farsighted* behavior. Farsightedness is indeed one of the key features of rational behavior under the reciprocity mechanism. We however show that, under this behavior, not all minority options can be protected; in particular, minority options that are non-strategic (that is, minority options that are dominated by a majority option) cannot be selected by the reciprocity set. Remark also that the strategic protection of minority interests under the reciprocity mechanism can be achieved under majority rule. Our analysis therefore provides an additional explanation of the "decisive minority" phenomenon (Campbell, 1999), especially in a context where a minority group does not necessarily have a "fixed" identity.

Campbell (1999) considers a costly majority rule voting environment with private information. A random set of voters decide on two alternatives, with the possibility of abstention. A "zealous voter"

[24]A solution concept $\overline{\mathfrak{E}}$ is a social choice correspondence which maps any political economy $\mathcal{PE}(\theta)$ into a subset $\overline{\mathfrak{E}}(\mathcal{PE}(\theta)) \subseteq A$; $\overline{\mathfrak{E}}(\mathcal{PE}(\theta))$ is interpreted as the set of equilibrium outcomes of $\mathcal{PE}(\theta)$ when agents adopt the rational behavior that defines $\overline{\mathfrak{E}}$.

[25]Interestingly, it directly follows from this result that if there exists a *Condorcet policy*, it will be the only option of the reciprocity set. This derives from the fact that a Condorcet policy is majority-preferred to any other policy, implying that any policy that is different from the Condorcet policy is a *non-strategic minority* option and therefore cannot belong to the reciprocity set according to Proposition 4.

is an agent who has either large stakes in the outcome of the election or small costs of participating in the election. Campbell shows that a minority group of zealous voters secures their preferred policy with high probability in any Bayesian Nash equilibrium if the expected population is large. Free-riding and preference intensity sustain the predictions of minority options as equilibrium policies under Campbell's framework. In addition to the fact that our model differs from Campbell (1999), we propose *farsighted behaviors* as another argument for the "decisive minority" phenomenon.

In what follows, we illustrate the result of Proposition 4 in a multi-ethnic society faced with the task of choosing a language other than the national language(s) to be taught in schools and to be used in official communication. The main purpose of the government is to protect linguistic diversity.

Example 6. Assume that a country that consists of four ethnic groups has to choose from a set of three languages, the ones to be used both in school and in official communication. Let $N = \{1, 2, 3, 4\}$ be the set of ethnic groups and $L = \{l_1, l_2, l_3\}$ be the set of languages. Each ethnic group has one representative. Language l_3 is the most commonly spoken language in ethnic groups 1 and 2; l_2 is the native language of ethnic group 3; and l_1 is a minority language that is mostly spoken in ethnic group 4. We define the preference profile $R(\theta) = (R_1(\theta), R_2(\theta), R_3(\theta), R_4(\theta))$ over languages as follows. Ethnic group 1 : $l_3 P_1(\theta) l_2 P_1(\theta) l_1$; Ethnic group 2 : $l_3 P_2(\theta) l_2 P_2(\theta) l_1$; Ethnic group 3 : $l_2 P_3(\theta) l_1 P_3(\theta) l_3$; and Ethnic group 4 : $l_1 P_4(\theta) l_3 P_4(\theta) l_2$. We suppose that the constitution $\mathcal{C}$ is majority rule. Figure 7 shows the popularity relationship among the various languages (for example, l_3 is more popular than l_2 as the former is preferred by three out of the four ethnic groups). The

$$l_1 \xrightarrow{\{1,2,3\}} l_2 \xrightarrow{\{1,2,4\}} l_3$$

Figure 7: Popularity relationship among the languages

analysis demonstrates that there are two reciprocity equilibrium languages. These languages are l_1 and l_3. In fact, if l_3 is the status quo, it will persist indefinitely, since it is not majority-preferred by any other language. If l_2 is the status quo, the majority coalition $\{1, 2, 4\}$ will object and propose l_3, given the fact that each of its members strictly prefers l_3 over l_2. Despite the fact that ethnic group 3 is opposed to that plausible objection, it cannot succeed at forming a majority coalition that will formulate a plausible counter-objection that will lead to the replacement of l_3. It follows that $l_2 \notin \mathfrak{E}(\mathcal{PE}(\theta))$. If the status quo is l_1, no majority coalition will formulate a plausible objection against it. In fact, although the members of $\{1, 2, 3\}$ strictly prefer l_2 over l_1, they know that if they formulate a plausible objection $(l_2, \{1, 2, 3\})$ against l_1, $\{1, 2, 4\}$ will formulate an unfriendly plausible counter-objection $(l_3, \{1, 2, 4\})$ against $(l_2, \{1, 2, 3\})$. Indeed, the move from l_1 to l_2 will harm the interests of ethnic group 4, which is the reason for the plausible counter-objection. The latter is unfriendly because not all members of $\{1, 2, 3\}$ prefer l_3 over the status quo l_1. Therefore $l_1 \in \mathfrak{E}(\mathcal{PE}(\theta))$. It follows that $\mathfrak{E}(\mathcal{PE}(\theta)) = \{l_1, l_3\}$.

Interestingly, remark that l_1 is a minority language because, unlike l_3, it is spoken by only one ethnic group, and l_2 is majority-preferred over l_1. Also, l_1 is a "strategic" minority option, whereas l_2 is a "non-strategic" minority option although it is majority-preferred over l_1. This illustrates an instance in which minority interests are *strategically* protected under the reciprocity mechanism. This example also shows that the reciprocity mechanism encourages diversity. Indeed, the political authority could decide that only equilibrium languages will be used for public communication. In that case, languages l_1 and l_3 will be used, leading to an officially bilingual country.

6.2 Rawlsian Justice and Fairness Considerations: The Nucleolus and The Shapley value

The Nucleolus (Schmeidler, 1969) and the Shapley value (Shapley, 1953) are known in economic theory to have desirable properties towards "minorities" and the worse-off. These properties include normative principles such as *welfare-maximization* (the Nucleolus), and *fairness* (the Shapley value). Consider a political economy $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$, where A, the set of feasible allocations, is the simplex: $A = \{x \in [0,1]^n : \sum_{i=1}^{n} x_i \leq 1\}$, and the preference profile $R(\theta)$ is defined as follows: for x, $y \in A$, $yR_i(\theta)x$ if $y_i \geq x_i$. We need additional definitions and notations before introducing the result of Section 6.2.

Given a political economy $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle \in \mathbb{P}$, the pair $G^{\mathcal{C}}_{\theta} = (N, \mathcal{C})$ defines a **simple game**, where the function $\mathcal{C}$ maps each subset $S \in \overline{\mathcal{N}} = \mathcal{N} \cup \{\emptyset\}$ to either 0 or 1, with $\mathcal{C}(S) = 1$ if $S \in W(\mathcal{C})$, and $\mathcal{C}(S) = 0$ if $S \notin W(\mathcal{C})$. Let $x \in A$, and a $S \in \overline{\mathcal{N}}$, we denote

$$x(S) = \sum_{i \in S} x_i, \text{ and}$$

$$e(S,x) = x(S) - \mathcal{C}(S) = \begin{cases} x(S) - 1 \leq 0 & \text{if } S \in W(\mathcal{C}) \\ x(S) \geq 0 & \text{if } S \notin W(\mathcal{C}) \end{cases}.$$

$e(S,x)$ is the welfare (or satisfaction) of subset S at x. For $x \in A$, we denote $e(x) = (e(S,x))_{S \in \overline{\mathcal{N}}}$ and $e^*(x)$ the permutation of entries of $e(x)$ arranged in increasing order.

We say that $e(x)$ is leximin superior to $e(y)$—denoted by $e(x)P_{lxm}e(y)$—, for $y \in A$, if $e^*(x)$ is lexicographically superior to $e^*(y)$, i.e., there exists $k+1 \in \{1, 2, ..., 2^n\}$ such that $e^*_t(x) = e^*_t(y)$ for $t = 1, 2, ..., k$ and $e^*_{k+1}(x) > e^*_{k+1}(y)$.

In the political economy $\mathcal{PE}(\theta)$, we examine the relationship between the reciprocity set on the one hand and the Nucleolus and the Shapley value of the simple game $G^{\mathcal{C}}_{\theta} = (N, \mathcal{C})$ on the other hand. We recall the definitions of these concepts below.

Definition 6. The Nucleolus of the game $G^{\mathcal{C}}_{\theta} = (N, \mathcal{C})$ is:

$$nc(G^{\mathcal{C}}_{\theta}) =: \{x \in A : x(N) = 1 \text{ and } \nexists y \in A,\ e(y)P_{lxm}e(x)\}.$$

It has been shown that the Nucleolus always exists and it is a single-valued allocation rule (Schmeidler, 1969). The Nucleolus lexicographically maximizes the welfare $e(S,x)$, $(S,x) \in \overline{\mathcal{N}} \times$

A, when these numbers are ordered in an increasing magnitude. For this reason, the Nucleolus maximizes the welfare of the worse-off and is therefore considered as a formalization of the "the Rawlsian social welfare function to a society where each coalition's welfare is evaluated independently" (Iñarra et al., 2020, p.4).

Definition 7. The Shapley value of the game $G_\theta^{\mathcal{C}} = (N, \mathcal{C})$ is the vector $\boldsymbol{Sh}(G_\theta^{\mathcal{C}}) \in \mathbb{R}^n$:

$$\boldsymbol{Sh}_i(G_\theta^{\mathcal{C}}) =: \sum_{S: i \notin S} \frac{|S|!(|N| - |S| - 1)!}{|N|!}[\mathcal{C}(S \cup \{i\}) - \mathcal{C}(S)], \text{ for all } i \in N.$$

The Shapley value is considered as a *fair* distribution because it is the unique allocation that satisfies three normative principles, namely, *efficiency* ($\sum_{i \in N} \boldsymbol{Sh}_i(N, \mathcal{C}) = \mathcal{C}(N) = 1$), *symmetry* (for all permutations π of N, $\boldsymbol{Sh}_{\pi(i)}(N, \pi\mathcal{C}) = \boldsymbol{Sh}_i(N, \mathcal{C})$, where $\pi\mathcal{C}(S) = \mathcal{C}(\pi(S))$ for all S), and *marginality* ($\boldsymbol{Sh}_i(N, \mathcal{C}) \geq \boldsymbol{Sh}_i(N, \mathcal{C}')$ as along as $\mathcal{C}(S \cup \{i\}) - \mathcal{C}(S) \geq \mathcal{C}'(S \cup i) - \mathcal{C}'(S)$ for all S and $\mathcal{C}' \in \mathfrak{C}$).[26] In general, the Nucleolus does not coincide with the Shapley value, see, for instance Example 7 below. In fact, the Nucleolus satisfies efficiency and symmetry, but does not satisfy marginality. Proposition 5 follows.

Proposition 5. *Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta) \rangle \in \mathbb{P}$ be a political economy. Then, $nc(G_\theta^{\mathcal{C}}) \in \mathfrak{E}(\mathcal{PE}(\theta))$ and $\boldsymbol{Sh}(G_\theta^{\mathcal{C}}) \in \mathfrak{E}(\mathcal{PE}(\theta))$.*

Proof of Proposition 5. 1. Assume that $nu(G_\theta^{\mathcal{C}}) = \{x\}$. Suppose there exists a winning coalition T such that $x_i = 0$ for each $i \in T$. Then, $e(T, x) = x(T) - 1 = -1$, such that T is the subset with the minimum satisfaction at x. One may assume that the vector x is such that $x_i = 0$ if $i \in T$, and $x_i > 0$ if $i \notin T$. Now, consider the vector $\overline{x}$ and ϵ very small such that $\overline{x}_i > 0$ for each i and assume that: $\overline{x}_i = \epsilon$ if $i \in T$, and $\overline{x}_i = x_i - \frac{t\epsilon}{n-t}$ if $i \notin T$, $t = |T| < n$. The vector $\overline{x} \in A$ because: $\sum_{i \in N} \overline{x}_i = t\epsilon + \sum_{i \notin T} x_i - (n-t)\frac{t\epsilon}{n-t} = \sum_{i \notin T} x_i = 1$. Moreover, $e(T, \overline{x}) = \overline{x}(T) - 1 = t\epsilon - 1$. The parameter ϵ is chosen so that T still induces the minimum excess at $\overline{x}$. Given that $t\epsilon - 1$ is greater than -1, it follows that $e^*(\overline{x})$ is lexicographically superior to $e^*(x)$, which is a contradiction. Hence, it is not possible to have every agent in a winning subset at x who receives a null payoff. Let $N_0(x) = \{i \in N : x_i = 0\}$, then $N_0(x) \notin W(\mathcal{C})$. If $N_0(x) = \emptyset$, obviously $x \in \mathfrak{E}(\mathcal{PE}(\theta))$. Assume that $N_0(x) \neq \emptyset$. Let L^x be a subset of $N - N_0(x)$ and $S = N_0(x) \cup L^x$ such that $S \in W(\mathcal{C})$ with $N_0(x) \cap L^x = \emptyset$. Consider the vector y defined as follows: $y_i = 0$ if $i \notin S$; $y_i = x_i + \frac{\sum_{i \notin S} x_i}{|S|}$ if $i \in L^x$; and $y_i = \frac{\sum_{i \notin S} x_i}{|S|}$ if $i \in N_0(x)$. Note that $\sum_{i \in N} y_i = \sum_{i \in L^x} x_i + |L^x \cup N_0(x)|\frac{\sum_{i \notin S} x_i}{|S|} = \sum_{i \in N} x_i = 1$. Then, $y \in A$. The pair (y, S) represents a possible plausible objection against x. Denote $S^y = \{i \in N : y_i = 0\}$, and consider $T = S^y \cup L^y$, with $L^y \subset N_0(x)$, $L^y \cap S^y = \emptyset$, and $(N - N_0(x)) \cap S^y \neq \emptyset$, so that $T \in W(\mathcal{C})$. Consider the vector z defined as follows: $z_i = \frac{\sum_{i \notin T} y_i}{|T|}$ if $i \in S^y$; $z_i = 0$ if $i \notin T$; and $z_i = y_i + \frac{\sum_{i \notin T} y_i}{|T|}$ if $i \in L^y$. Note that $\sum_{i \in N} z_i = \sum_{i \in L^y} y_i + |L^y \cup S^y|\frac{\sum_{i \notin T} y_i}{|T|} = \sum_{i \in N} y_i = 1$. Then $z \in A$. Moreover, for each $i \in T$, $z_i > y_i$. This implies that $zP_T y$. There also exists

[26] See, for example, Young (1985), Aguiar et al. (2018), Aguiar et al. (2020), among others.

$j \in (N - N_0(x)) \cap S^y$ such that $x_j > 0$ and $y_i = 0$, meaning that $not(yR_Tx)$. The pair (z, T) represents a possible plausible counter-objection against the plausible objection (y, S). Given that there exists $k \in S \cap (N - N_0(x))$ and $k \notin T$, then $x_k > 0$ and $z_k = 0$. This means that for any possible plausible objection (y, S) against x, we can construct a plausible counter-objection (z, T) so that $not(zR_Sx)$. Therefore x is a reciprocity equilibrium, and $nc(G_\theta^\mathcal{C}) \in \mathfrak{E}(\mathcal{P}(\theta))$.

2. For any $G_\theta^\mathcal{C} = (N, \mathcal{C})$, the Shapley value assigns a non-negative value to agent i, i.e., $\boldsymbol{Sh}_i(G_\theta^\mathcal{C}) \geq 0$ for each $i \in N$. Efficiency implies that $\sum\limits_{i \in N} \boldsymbol{Sh}_i(G_\theta^\mathcal{C}) = \mathcal{C}(N) = 1$. Let $S \in \overline{\mathcal{N}}$, and $i \in N$. Agent i is decisive in S if $i \in S$, $S \in W(\mathcal{C})$ and $S - \{i\} \notin W(\mathcal{C})$. In the simple game $G_\theta^\mathcal{C}$, the Shapley value assigns 0 to agent i if and only if she is not decisive in any subset of N. Let $\overline{S} = \{i \in N : \boldsymbol{Sh}_i(G_\theta^\mathcal{C}) = 0\}$. Assume that $\overline{S} \in W(\mathcal{C})$. Then $\overline{S} \neq \emptyset$. There exists $j \in \overline{S}$ and $S^j \in 2^{\overline{S}} \cap W(\mathcal{C})$, such that $S^j - \{j\} \notin W(\mathcal{C})$. This implies that j is decisive in S^j. Therefore, $\boldsymbol{Sh}_j(G_\theta^\mathcal{C}) > 0$, a contradiction. We conclude that $\overline{S} \notin W(\mathcal{C})$. Using the same reasoning in part 1, we show that $\boldsymbol{Sh}(G_\theta^\mathcal{C}) \in \mathfrak{E}(\mathcal{PE}(\theta))$. □

Proposition 5 implies that the reciprocity mechanism is compatible with the Rawlsian idea of justice and fairness. Example 7 illustrates the reciprocity set, the Shapley value, and the Nucleolus for three simple voting games.

Example 7. Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta) \rangle \in \mathbb{P}$ be a political economy.

Dictator. Assume that $W(\mathcal{C}) = \{S : d \in S\}$. Then, agent d is a dictator. Let $x^d \in [0, 1]^n$ such that: $x_i^d = 1$ if and only if $i = d$. Then, $nc(G_\theta^\mathcal{C}) = \boldsymbol{Sh}(G_\theta^\mathcal{C}) = x^d$ and $\mathfrak{E}(\mathcal{P}(\theta)) = \{x^d\}$. In this case, the reciprocity set coincides with both the Nucleolus and the Shapley value.

Compromise. Consider an environment in which $N = \{1, 2, 3\}$ and $\mathcal{C}$ is such that $W(\mathcal{C}) = \{S : 1 \in S \text{ and } |S| \geq 2\}$. In the game, agent 1 needs either 2 or 3 to win, whereas 2 and 3 together lose. Then, $nc(G_\theta^\mathcal{C}) = \boldsymbol{Sh}(G_\theta^\mathcal{C}) = (\frac{8}{12}, \frac{2}{12}, \frac{2}{12}) \in \mathfrak{E}(\mathcal{PE}(\theta))$.

A Weighted majority game. Now, assume $N = \{1, 2, 3, 4\}$ and the constitution $\mathcal{C}$ is defined as follows. Agents have different weights: $q_1 = 6$, $q_2 = 4$, $q_3 = 3$, $q_4 = 2$, and there is a quota $q = 8$ such that $S \in W(\mathcal{C})$ if and only if $\sum\limits_{i \in S} q_i \geq q$. In this environment, $nc(G_\theta^\mathcal{C}) = (\frac{2}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$, $\boldsymbol{Sh}(G_\theta^\mathcal{C}) = (\frac{1}{2}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$, and both belong to $\mathfrak{E}(\mathcal{PE}(\theta))$. We note here that the Nucleolus gives more payoff to the worse-off agents than the Shapley value; the former is therefore more inclusive than the latter.

6.3 Extension of Findings to the Sophisticated Reciprocity Set

In Section 6.3, we extend the properties of the reciprocity set to the sophisticated reciprocity set. In particular, we focus on equilibrium existence and Pareto-efficiency. We have Theorem 7.

Theorem 7. *Let $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta) \rangle$ and $\mathcal{PE}^s(\theta) = \langle N, A, \mathcal{C}, V^s, R(\theta) \rangle$ be two political economies.*

*1. (**Equilibrium existence**)*

(a) If A is discrete and finite and preferences are linear, then $SR(\mathcal{PE}^s(\theta))$ is not empty.

(b) If A is compact and convex, and preferences are continuous, linear, and endowed with the topology of closed convergence, then $SR(\mathcal{PE}^s(\theta))$ is not empty.

2. *(***Pareto-Efficiency***) Any policy in $SR(\mathcal{PE}^s(\theta))$ is Pareto-efficient.*

Proof of Theorem 7. Let $x, y \in A$ be two policies, and $S \in W(\mathcal{C})$ be a winning coalition. We note that $y \gtrdot_S^{sd} x$ if and only if a) $\forall (z, T) \in A \times W(\mathcal{C})$, $S \neq T$, $[yR_T(\theta)x$ and $not(zR_S(\theta)x)]$ or $not(zR_T(\theta)y)$ implies $yP_S(\theta)x$; and, b) $[\forall (z, T) \in A \times W(\mathcal{C}),\ S \neq T,\ zP_T(\theta)y$ and $not(yR_T(\theta)x)]$ implies $[zR_S(\theta)x]$.

1. Equilibrium existence. The proof follows the demonstration of results (Theorems 4 and 5) that ensure the existence of equilibria under the reciprocity mechanism in Section 4.

2. Pareto-efficiency. Consider $x \in SR(\mathcal{PE}^s(\theta))$. Then $x \in \mathfrak{E}(\mathcal{PE}(\theta))$. Hence, x is Pareto-efficient thanks to Theorem 6. $\square$

A few additional comments are needed. First, we note that item (2) in Theorem 7 showing that any sophisticated reciprocity equilibrium is Pareto-efficient is a direct consequence of the fact that any sophisticated reciprocity equilibrium is a reciprocity equilibrium, and any reciprocity equilibrium is Pareto-efficient (Theorem 6). Second, although we do not show this formally for expositional purposes, we remark that agents whose behavior follow the sophisticated reciprocity set behave in a reciprocal manner, which implies that Theorem 1 generalizes to this refined solution concept. Third, like the reciprocity set, the sophisticated reciprocity set strategically protects minorities, which extends Proposition 4. This can be seen in Example 1, where policy y is a sophisticated reciprocity equilibrium despite the fact that it is a minority option. We note that while y is a "strategic" minority option, z is "non-strategic" minority option, which is why z does not belong to the sophisticated reciprocity set. These desirable properties of the (sophisticated) reciprocity set distinguish these solution concepts from other well-known concepts, something that we shown in Section 7.

7 Comparison with Other Voting Procedures and Solution Concepts

We compare the reciprocity mechanism and the reciprocity set (resp. sophisticated reciprocity set) to some classical voting procedures and their associated solution concepts following the *blocking approach* when the policy space is discrete. It is important for us to strongly emphasize that the *only* purpose of this comparison is to show that the reciprocity set is not a redundant solution concept. Our goal is absolutely not to highlight the relative merits of each approach, which is beyond the scope of this paper. Indeed, each of the solution concepts to which we compare the reciprocity set has its own merits, and clearly captures rationality under a distinct set of structural and behavioral assumptions. The reader will realise that our solution concept has properties that distinguish it from the other solution concepts, which is understandable because these other concepts do not aim to address the same questions as the reciprocity set. The supplement Materials contain more details on the comparative analyses. In what follows, we summarize the findings.

First, we modify the voting procedure V^{rm} to remove the *reciprocity clauses*. This alternative procedure is formalized as a two-stage process in which no policy withdrawal (or retraction) by a sponsoring coalition is possible. We find that, under this new procedure, a non-Pareto-efficient status quo can persist indefinitely. This result occurs because, even though some agents might be willing to remove a non-Pareto-efficient status quo, they fear that the new policy might in turn be replaced by another policy that is worse for them relative to the status quo. This fear prevents these agents from removing a status quo, a situation that is avoided under the procedure V^{rm}. Indeed, the withdrawal clause embedded in the procedure V^{rm} prevents free-riding and promotes pro-social behavior, leading to only Pareto-efficient policy alternatives being selected.[27]

Second, we compare the reciprocity set with some solution concepts designed to capture rationality in one-shot procedures (or games). The most prominent are the von Neumann-Morgenstern (vNM) stable sets (von Neumann and Morgenstern, 1944), the core (Gillies, 1959), the top cycle set (Schwartz, 1976), and the uncovered set (Miller, 1980). All these concepts follow the blocking approach, which makes the comparison easier. We find that, unlike the reciprocity set, one of the latter solution concepts might select a non-Pareto-efficient policy, while others might be empty, meaning that they might fail to yield an equilibrium policy.

Third, we extend our comparison to dynamic games, specifically farsighted coalitional games. In such games, agents or coalitions may bargain indefinitely. The most prominent equilibrium concepts in this class of games that follow the tradition of blocking approach are based on the notion of indirect domination (Harsanyi, 1974; Chwe, 1994), which is a modification of direct domination, whereby first movers anticipate future moves. Comparing the reciprocity set to these two concepts, we find that they may lead to an inefficient policy persisting indefinitely. In a more recent study, Ray and Vohra (2015) modify the farsighted stable sets of Harsanyi (1974) by only imposing coalitional sovereignty (the modified solution concept uses exactly the same domination relation that defines Harsanyi's farsighted stable sets). Although the farsighted behavior in Ray and Vohra (2015) follows the *blocking approach*, we differ in defining a "constitution or voting rule" as a tool to distribute voting powers (or veto rights) among coalitions. For this reason, it is not intuitively reasonable to compare the reciprocity set to the *farsighted stable set* à la Ray and Vohra (2015).

[27] One can show that the stability set, a solution concept introduced by Rubinstein (1980), is a plausible solution concept under the modified voting procedure (without withdrawal), which explains why this solution concept generally selects non-Pareto-efficient outcomes, as mentioned by Rubinstein (1980). Also, in results not shown here but available upon request, we find that the reciprocity set (and hence the sophisticated reciprocity set) significantly refines the stability set.

Table 3: Comparison of Solution Concepts

Solution Concepts	Existence and non-emptiness	Efficiency	SPMI
Reciprocity set	Yes	Yes	Yes
Sophitiscated reciprocity set	Yes	Yes	Yes
Core	No	Yes	No
vNM stable set	No	Yes	No
Top cycle set	Yes	No	No
Uncovered set	Yes	Yes	No
Harsanyi stable set	No	Yes	Yes
Largest consistent set	Yes	No	Yes

In Table 3, we summarize the comparisons of the aforementioned solution concepts and the reciprocity set in terms of their properties. We also note that the sophisticated reciprocity set satisfies similar properties as the reciprocity set, although the former is a refinement of the latter. We say that a solution concept $\overline{\mathfrak{E}}$ satisfies the "Pareto-efficiency" property if for any political economy $\mathcal{PE}(\theta) \in \mathbb{P}$, the set $\overline{\mathfrak{E}}(\mathcal{PE}(\theta))$ contains only Pareto-efficient policies. A solution concept $\overline{\mathfrak{E}}$ satisfies the "existence and non-emptiness (or stability)" property if there does not exist a political economy $\mathcal{PE}(\theta) \in \mathbb{P}$ such that the set $\overline{\mathfrak{E}}(\mathcal{PE}(\theta))$ does not exist or $\overline{\mathfrak{E}}(\mathcal{PE}(\theta))$ is empty.[28] The definition of "strategic protection of minority interests" (SPMI) is the one given in Definition 5. As shown in Table 3, among all the solution concepts analyzed in this paper, only the reciprocity set and the sophisticated reciprocity satisfy existence and non-emptiness, Pareto-efficiency, and SPMI. It follows that the reciprocity set and the sophisticated reciprocity set differ from all the other solution concepts in their motivation, incentives structure, and properties. We have also shown that these differences lead to different predictions. The reciprocity set is also compatible with classical notions of distributive justice.[29]

Other approaches in modeling strategic interactions in social environments that embed voting procedures include multilateral bargaining in legislatures. In this literature, an agenda-setting defines a protocol that grants a veto right to agents or coalitions, and agents discount preferences with a discount factor and they consider expected utilities when making decisions. Notable examples include the studies pioneered by Baron and Ferejohn (1989) such as Banks and Duggan (2000), and Diermeier et al. (2017). Our framework differs from multilateral bargaining games such as Baron and Ferejohn (1989) and the large literature that this study has inspired for at least three reasons that we describe below.

(i) First, the voting procedure V^{rm} allows first movers to "retract" motions that face opposition

[28]As remarked by Chwe (1994), non-emptiness and existence are technically different. For instance, the core (or the Nash equilibrium) always exists, but it might be empty; the vNM stable sets might not exist, but when they exist, they are always non-empty. The reciprocity set and the sophisticated reciprocity set always exist and are always non-empty.

[29]While it has been shown that the core always contains the Nucleolus, it may not contain the Shapley value. The reciprocity set contains both concepts.

and second movers to "oppose" motions they dislike and to "retaliate" in case of non-retraction by first-movers.

(ii) Second, the models inspired by Baron and Ferejohn (1989) study a class of "distributive" problems where agents have to split a unit of a divisible good by voting. While our framework covers this class of problems, it addresses a larger class of problems, including problems related to the allocation of non-divisible goods. The larger domain of the (sophisticated) reciprocity set allows us to examine a variety of voting situations.

(iii) Third, within the class of distributive problems, our predictions sharply differ from those of models inspired by the pioneering work of Baron and Ferejohn (1989). For example, Baron and Ferejohn show that any distribution of the good to be split can be supported as a subgame perfect equilibrium in an infinite-session legislature with a closed rule when there is a sufficient number of members and they are not too impatient (Baron and Ferejohn, 1989, Proposition 2, p. 1189). The implication is that a non-Pareto-efficient outcome can be selected in this legislature, whereas this cannot happen in our framework. Also, whereas a non-strategic minority option can be selected in this legislature, this is impossible under the reciprocity mechanism. Moreover, the proof of Proposition 5 shows that only a subset of possible allocations forms the reciprocity set. Again, these differences in our predictions come from the fact the reciprocity voting procedure is very different from the procedures studied in Baron and Ferejohn (1989).

8 Contributions

This study contributes to several literatures, including the literature on institutional design, the literature on reciprocity, and the literature on equilibrium concepts that follow the blocking and bargaining approaches in coalitional games.

We introduce the reciprocity voting procedure. Although it is new, it simply modifies well-known procedures by introducing reciprocity clauses (such as the right to oppose, withdraw and retaliate). In doing so, it is close enough to voting procedures employed in most legislative bodies. It provides an incentive structure for reciprocal actions. We do not assume that agents are altruistic or have other-regarding preferences. Although preferences may be "selfish" (meaning that agents may only intend to care about themselves), agents find it rational to care about others and to take pro-social and reciprocal actions. It follows that, in our setting, reciprocity is not *preference-based*, it is *mechanism-based*. This feature distinguishes our research from the extant theoretical and experimental literature on reciprocity which assumes that agents have other-regarding preferences including conditional altruism (see, for example, Rabin (1993), Fehr and Gächter (1998), Dufwenberg and Kirchsteiger (2004), and Hahn (2009)). Interestingly, in our framework, the modeler is free to incorporate other-regarding preferences and fairness considerations (see Example 4). Another domain that has been found to induce reciprocity and sustain cooperation under certain conditions is the domain of infinitely repeated games (Aumann and Sorin, 1989; Bruni et al., 2008). The reciprocity mechanism induces a finite horizon game and therefore does not require that a game be repeated indefinitely in order to induce reciprocity and cooperation among decision-makers.

We obtain the *reciprocity set* and the *sophisticated reciprocity set* by formalizing the rational behavior of agents under the reciprocity voting procedure following both the *blocking* and the *non-cooperative* approaches.[30] Defining these approaches, Dutta and Vohra write:

> (a) the blocking approach ... follows traditional cooperative game theory in abstracting away from the details of the negotiation process and relying on a coalitional game to specify what each coalition is able to accomplish on its own, and (b) the bargaining approach ... is based on noncooperative coalition bargaining and relies on specifying details such as a protocol that describes the order of moves. (Dutta and Vohra, 2017, p. 1192)

To our knowledge, the reciprocity set (and the sophisticated reciprocity set) are the first solution concepts in the tradition of the *blocking approach* that captures "rational reciprocal" behavior. Additionally, we develop a non-cooperative approach to implement these solutions concepts. Other studies have used a similar approach in farsighted coalitional games (see, for example, Herings et al. (2004), and Granot and Hanany (2016)). Unlike Granot and Hanany (2016) who defines the evolution of moves as an infinite extensive form game, our approach is finite. Moreover, contrary to Herings et al. (2004), our non-cooperative framework can be viewed as an independent sequential legislative game with no primitive connection to the reciprocity set under the blocking approach. It follows that our mechanism embeds both elements of cooperative and non-cooperative domains and incorporates the notion of *maximality* in farsighted coalition formation (Ray and Vohra, 2014; Dutta and Vohra, 2017; Kimya, 2020). In that direction, this paper also contributes to the Nash Program, a research agenda initiated by Nash (1953) to bridge the gap between the cooperative and non-cooperative counterparts of game theory; see Serrano (2020) for a recent survey.

One feature of the one-shot mechanism in political games is its ability of ensuring that the chosen policy is Pareto-efficient. Nevertheless, it has been criticized for its lack of political inclusiveness and its general failure to select equilibrium policies.[31] On the other hand, farsighted coalitional games may lead to the selection and persistence of inefficient policies. The reciprocity set and and the *sophisticated reciprocity set* resolve all of these pitfalls under mild conditions.

[30]The blocking approach goes back to von Neumann and Morgenstern (1944) and Gillies (1959) and has been followed in a large number of studies (see, for example, Harsanyi (1974), Chwe (1994), Ray et al. (2007), Ray and Vohra (2014), Ray and Vohra (2015), and Dutta and Vohra (2017)). Even the Nash equilibrium and hybrid solution concepts such as pairwise stability (Jackson and Wolinsky, 1996; Pongou and Serrano, 2016, 2013) can be viewed as following the tradition of the blocking approach in the sense that these concepts abstract away from the process that leads to a particular equilibrium, and an equilibrium is simply defined as a state that cannot be blocked by rational agents. The non-cooperative bargaining approach can be traced back to Ståhl (1977), Rubinstein (1982), and Baron and Ferejohn (1989). As several scholars have argued (Roth (2020)), the two approaches are not mutually exclusive, and can in fact be viewed as complementary (Ray and Vohra, 2014; Roth and Wilson, 2019; Kimya, 2020).

[31]In other words, the Nash equilibrium may not exist just as the core might be empty. See Austen-Smith and Banks (1999, 2005) for a thorough survey and exposition of solution concepts in political games.

The reciprocity set has a very different motivation and incentives structure than the other solution concepts that we review in this paper. These solutions are complementary and each has its own merits and advantages. We compare the reciprocity set to the leading solution concepts that follow the blocking approach and find that they are very different indeed (see Section 7). For instance, there exist political games where the prediction of the reciprocity set strictly refines some of these concepts. In general, none of these solution concepts simultaneously guarantees policy stability (or non-emptiness), efficiency, and the *strategic* protection of minority interests (see Table 3 in Section 7) and are consistent with classical notions of *distributive justice*. These differences are mainly justified by the fact that we are addressing a new question, generating new insights into the design of mechanisms or voting procedures that foster reciprocal actions and lead to desirable social alternatives.

9 Conclusion

This paper considers the problem faced by a political authority that has to design a legislative mechanism that guarantees equilibrium existence, Pareto-efficiency, and inclusiveness. To address this problem, we propose a voting procedure that embeds *clauses of reciprocity*. These clauses grant voters the right to oppose actions that are not in their interest, retract actions that face opposition, and punish actions that harm them. We prove this voting procedure incentivizes "selfish" agents to display reciprocal and pro-social behaviors. We study voters' strategic behavior under this procedure following both the blocking approach and the non-cooperative approach. Under the blocking approach, we introduce two new solution concepts—the *reciprocity set* and the *sophisticated reciprocity set*—. We then show that these solution concepts satisfy non-emptiness (also known as stability), Pareto-efficiency, and inclusiveness in the sense of strategically protecting minority interests, and that they are compatible with classical notions of fairness and Rawlsian justice. By preventing political failure, these solution concepts resolve the long-established conflict between agent rationality and Pareto-efficiency in political decisions. Following the non-cooperative approach, we provide an implementation of each of these two solution concepts, which makes them applicable in a wide range of legislative settings. We also extend them to effectivity functions, a large class of games that includes the class of strategic form games. Moreover, we show that they have several merits, especially when compared to other well-known solution concepts.

References

Abreu, D. and Sen, A. (1990). Subgame perfect implementation: A necessary and almost sufficient condition. *Journal of Economic Theory*, 50(2), 285–299.

Acemoglu, D., Egorov, G., and Sonin, K. (2009). Equilibrium refinement in dynamic voting games.

Acemoglu, D., Ticchi, D., and Vindigni, A. (2011). Emergence and persistence of inefficient states. *Journal of the European Economic Association*, 9(2), 177–208.

Aguiar, V., Pongou, R., and Tondji, J.-B. (2018). A non-parametric approach to testing the axioms of the Shapley value with limited data. *Games and Economic Behavior*, 111, 41–63.

Aguiar, V. H., Pongou, R., Serrano, R., and Tondji, J.-B. (2020). An index of unfairness. *In Handbook of the Shapley Value, Algaba, E. (Ed.), Fragnelli, V. (Ed.), Sànchez-Soriano, J. (Ed.)*.

Arrow, K. J. and Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, (pp. 265–290).

Aumann, R. J. (1959). Acceptable points in cooperative general n-person games, in A. W. Tucker and R. D. Luce (eds.). *Contributions to the Theory of Games IV, Annals of Mathematics Study*, 40, 287–324.

Aumann, R. J. and Maschler, M. (1964). The bargaining set for cooperative games. *Advances in Game Theory*, 52, 443–476.

Aumann, R. J. and Sorin, S. (1989). Cooperation and bounded recall. *Games and Economic Behavior*, 1(1), 5–39.

Austen-Smith, D. and Banks, J. S. (1999). *Positive political theory*, volume 1. Ann Arbor: University of Michigan Press.

Austen-Smith, D. and Banks, J. S. (2005). *Positive political theory*, volume 2. Ann Arbor: University of Michigan Press.

Axelrod, R. M. (1984). *The evolution of cooperation*. Basic books.

Banks, J. S. and Duggan, J. (2000). A bargaining model of collective choice. *American Political Science Review*, 94(1), 73–88.

Barberà, S., Berga, D., and Moreno, B. (2017). On the different forms of individual and group strategic behavior, and their impact on efficiency. In *Co-utility* (pp. 33–48). Springer.

Baron, D. P. and Ferejohn, J. A. (1989). Bargaining in legislatures. *American Political Science Review*, 83(04), 1181–1206.

Bergin, J. and Duggan, J. (1999). An implementation-theoretic approach to non-cooperative foundations. *Journal of Economic Theory*, 86(1), 50–76.

Bernheim, B. D., Peleg, B., and Whinston, M. D. (1987). Coalition-proof Nash equilibria I. Concepts. *Journal of Economic Theory*, 42(1), 1–12.

Black, D. (1948). On the rationale of group decision-making. *The Journal of Political Economy*, (pp. 23–34).

Bobbio, N. (1996). *Left and right: The significance of a political distinction*. University of Chicago Press.

Börgers, T. (2000). Is internet voting a good thing? *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, (pp. 531–547).

Borgers, T. (2004). Costly voting. *American Economic Review*, 94(1), 57–66.

Bruni, L., Gilli, M., and Pelligra, V. (2008). Reciprocity: theory and facts. *International Review of Economics*, 55(1-2), 1–11.

Campbell, C. M. (1999). Large electorates and decisive minorities. *Journal of Political Economy*, 107(6), 1199–1217.

Castles, F. G. and Mair, P. (1984). Left–right political scales: Some 'expert' judgments. *European Journal of Political Research*, 12(1), 73–88.

Chwe, M. S.-Y. (1994). Farsighted coalitional stability. *Journal of Economic Theory*, 63(2), 299–325.

Dasgupta, P., Hammond, P., and Maskin, E. (1979). The implementation of social choice rules: Some general results on incentive compatibility. *The Review of Economic Studies*, 46(2), 185–216.

Demeze-Jouatsa, G. H., Pongou, R., and Tondji, J.-B. (2021). A free and fair economy: A game of justice and inclusion. *Available at SSRN 3893929*.

Demuynck, T., Herings, P. J.-J., Saulle, R. D., and Seel, C. (2019). The myopic stable set for social environments. *Econometrica*, 87(1), 111–138.

Diermeier, D., Egorov, G., and Sonin, K. (2017). Political economy of redistribution. *Econometrica*, 85(3), 851–870.

Dubey, P. and Shubik, M. (1977). A closed economy with exogenous uncertainty, different levels of information, money, futures and spot markets. *International Journal of Game Theory*, 6(4), 231–248.

Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47(2), 268–298.

Dutta, B., Sen, A., and Vohra, R. (1994). Nash implementation through elementary mechanisms in economic environments. *Economic Design*, 1(1), 173–203.

Dutta, B. and Vohra, R. (2017). Rational expectations and farsighted stability. *Theoretical Economics*, 12(3), 1191–1227.

Edgeworth, F. Y. (1881). *Mathematical psychics: An essay on the application of mathematics to the moral sciences*. London: Kegan Paul.

Eguia, J. X. and Shepsle, K. A. (2015). Legislative bargaining with endogenous rules. *The Journal of Politics*, 77(4), 1076–1088.

Ehlers, L., Peters, H., and Storcken, T. (2002). Strategy-proof probabilistic decision schemes for one-dimensional single-peaked preferences. *Journal of Economic Theory*, 105(2), 408–434.

Einy, E. and Wettstein, D. (1999). A non-cooperative interpretation of bargaining sets. *Review of Economic Design*, 4(3), 219–230.

Evans, G., Heath, A., and Lalljee, M. (1996). Measuring left–right and libertarian–authoritarian values in the British electorate. *British Journal of Sociology*, (pp. 93–112).

Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2), 293–315.

Falkinger, J., Fehr, E., Gächter, S., and Winter-Ebmer, R. (2000). A simple mechanism for the efficient provision of public goods: Experimental evidence. *American Economic Review*, 90(1), 247–264.

Farquharson, R. (1969). *Theory of voting*. Yale Univ. Press, New Haven.

Fehr, E. and Gächter, S. (1998). Reciprocity and economics: The economic implications of homo reciprocans. *European Economic Review*, 42(3-5), 845–859.

Fehr, E. and Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *The Journal of Economic Perspectives*, 14(3), 159–181.

Fotso, A. F., Pongou, R., and Tchantcho, B. (2017). Conservative vs optimistic rationality in games: A revisitation. *Economics Letters*, 156, 42–47.

Giddens, A. (1994). *Beyond left and right: The future of radical politics*. Stanford University Press.

Gillies, D. B. (1959). Solutions to general non-zero-sum games. *Contributions to the Theory of Games*, 4(40), 47–85.

Grandmont, J.-M. (1978). Intermediate preferences and the majority rule. *Econometrica*, (pp. 317–330).

Granot, D. and Hanany, E. (2016). Subgame perfect farsighted stability. *Unpublished paper, Faculty of Engineering, Tel Aviv University.[172]*.

Greenberg, J. (1990). *The theory of social situations: an alternative game-theoretic approach*. Cambridge University Press, Cambridge.

Gul, F. (1989). Bargaining foundations of shapley value. *Econometrica*, (pp. 81–95).

Hahn, V. (2009). Reciprocity and voting. *Games and Economic Behavior*, 67(2), 467–480.

Harsanyi, J. C. (1974). An equilibrium-point interpretation of stable sets and a proposed alternative definition. *Management Science*, 20(11), 1472–1495.

Hart, S. and Mas-Colell, A. (1996). Bargaining and value. *Econometrica*, (pp. 357–380).

Herings, P. J.-J., Mauleon, A., and Vannetelbosch, V. J. (2004). Rationalizability for social environments. *Games and Economic Behavior*, 49(1), 135–156.

Herpich, N. (2020). *The conundrum for international students*. Editorial, The Havard Gazette.

Hildenbrand, W. (1974). Core and equilibria of a large economy. *Princeton, NJ: Princeton University Press*.

Inada, K. (1964). A note on the simple majority decision rule. *Econometrica*, (pp. 525–531).

Iñarra, E., Serrano, R., and Shimomura, K.-I. (2020). The nucleolus, the kernel, and the bargaining set: An update. *Revue économique*, 71(2), 225–266.

Jackson, M. O. (1992). Implementation in undominated strategies: A look at bounded mechanisms. *The Review of Economic Studies*, 59(4), 757–775.

Jackson, M. O. (2001). A crash course in implementation theory. *Social choice and welfare*, 18(4), 655–708.

Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1), 44–74.

Kalai, E., Postlewaite, A., and Roberts, J. (1979). A group incentive compatible mechanism yielding core allocation. *Journal of Economic Theory*, 20(1), 13–22.

Kimya, M. (2020). Equilibrium coalitional behavior. *Theoretical Economics*, 15(2), 669–714.

Lagunoff, R. D. (1994). A simple noncooperative core story. *Games and Economic Behavior*, 7(1), 54–61.

Legros, P. (1987). Disadvantageous syndicates and stable cartels: The case of the nucleolus. *Journal of Economic Theory*, 42(1), 30–49.

Mas-Colell, A. (1989). An equivalence theorem for a bargaining set. *Journal of Mathematical Economics*, 18(2), 129–139.

Maskin, E. (1977). Nash equilibrium and welfare optimality. *Mimeo. MIT*.

Maskin, E. (1985). The theory of implementation in Nash equilibrium: A survey. In *Hurwicz L, Schmeidler D, Sonnenschein H (eds) Social Goods and Social Organization*. Cambridge University Press. Cambridge, MA.

Miller, N. R. (1980). A new solution set for tournaments and majority voting: Further graph-theoretical approaches to the theory of voting. *American Journal of Political Science*, (pp. 68–96).

Moldovanu, B. and Winter, E. (1994). Core implementation and increasing returns to scale for cooperation. *Journal of Mathematical Economics*, 23(6), 533–548.

Moore, J. (1992). Implementation, contracts, and renegotiation in environments with complete information. In *Laffont JJ (eds) Advances in Economic Theory. VI World Congresss of the Econometric Society*, volume I. Cambridge University Press. Cambridge.

Moore, J. and Repullo, R. (1988). Subgame perfect implementation. *Econometrica*, (pp. 1191–1220).

Moulin, H. (1980). On strategy-proofness and single peakedness. *Public Choice*, 35(4), 437–455.

Nash, J. (1951). Non-cooperative games. *Annals of Mathematics*, (pp. 286–295).

Nash, J. (1953). Two-person cooperative games. *Econometrica*, (pp. 128–140).

Nash, J. F., Nagel, R., Ockenfels, A., and Selten, R. (2012). The agencies method for coalition formation in experimental games. *Proceedings of the National Academy of Sciences*, 109(50), 20358–20363.

Okada, A. (2012). Non-cooperative bargaining and the incomplete informational core. *Journal of Economic Theory*, 147(3), 1165–1190.

Osborne, M. J. and Rubinstein, A. (1994). *A course in game theory*. MIT press.

Palfrey, T. R. and Rosenthal, H. (1985). Voter participation and strategic uncertainty. *American Political Science Review*, (pp. 62–78).

Perez-Castrillo, D. and Wettstein, D. (2000). Implementation of bargaining sets via simple mechanisms. *Games and Economic Behavior*, 31(1), 106–120.

Perry, M. and Reny, P. J. (1994). A noncooperative view of coalition formation and the core. *Econometrica*, (pp. 795–817).

Pongou, R. and Serrano, R. (2013). Fidelity networks and long-run trends in HIV/AIDS gender gaps. *American Economic Review*, 103(3), 298–302.

Pongou, R. and Serrano, R. (2016). Volume of trade and dynamic network formation in two-sided economies. *Journal of Mathematical Economics*, 63, 147–163.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, (pp. 1281–1302).

Ray, D. et al. (2007). *A game-theoretic perspective on coalition formation*. Oxford University Press.

Ray, D. and Vohra, R. (2014). Coalition formation. In *Handbook of Game Theory*, volume 4 (H. Peyton Young and Shmuel Zamir, eds.) (pp. 239–326). North Holland: Elsevier.

Ray, D. and Vohra, R. (2015). The farsighted stable set. *Econometrica*, 83(3), 977–1011.

Roth, A. (2020). The Shapley Value, a giant legacy, and ongoing research agenda. *In Handbook of the Shapley Value, Algaba, E. (Ed.), Fragnelli, V. (Ed.), Sànchez-Soriano, J. (Ed.)*.

Roth, A. E. and Wilson, R. B. (2019). How market design emerged from game theory: A mutual interview. *Journal of Economic Perspectives*, 33(3), 118–43.

Rubinstein, A. (1980). Stability of decision systems under majority rule. *Journal of Economic Theory*, 23(2), 150–159.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, (pp. 97–109).

Schmeidler, D. (1969). The nucleolus of a characteristic function game. *SIAM Journal on applied mathematics*, 17(6), 1163–1170.

Schwartz, T. (1976). The logic of collective choice. *Mimeograph, Dept. of Government, University of Texas at Austin*, Part III, Chap. 6.

Selten, R. (1981). A noncooperative model of characteristic-function bargaining. In *Aumann R et al. (eds) Essay in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*. Bibliographisches Institut Mannheim.

Serrano, R. (1993). Non-cooperative implementation of the nucleolus: the 3-player case. *International Journal of Game Theory*, 22(4), 345–357.

Serrano, R. (1995). A market to implement the core. *Journal of Economic Theory*, 67(1), 285–294.

Serrano, R. (2020). Sixty-seven years of the Nash program: time for retirement? *SERIEs*, (pp. 1–14).

Serrano, R. and Vohra, R. (1997). Non-cooperative implementation of the core. *Social Choice and Welfare*, 14(4), 513–525.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, 2(28), 307–317.

Sprumont, Y. (1991). The division problem with single-peaked preferences: a characterization of the uniform allocation rule. *Econometrica*, (pp. 509–519).

Ståhl, I. (1977). An n-person bargaining game in the extensive form. In *Mathematical economics and game theory* (pp. 156–172). Springer.

Thomson, W. (1997). The replacement principle in economies with single-peaked preferences. *Journal of Economic Theory*, 76(1), 145–168.

von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*, volume 60. Princeton, NJ: Princeton University Press.

Young, H. P. (1985). Monotonic solutions of cooperative games. *International Journal of Game Theory*, 14(2), 65–72.

Supplement to "A Political Reciprocity Mechanism"*

Roland Pongou[1, 2] and Jean-Baptiste Tondji[3]

[1]Department of Economics, University of Ottawa

[2]Harvard T.H. Chan School of Public Health, Harvard University

[3]Department of Economics and Finance, The University of Texas Rio Grande Valley

September 20, 2021

In this supplement, we provide (a) the proof of the non-cooperative implementation of the reciprocity set (Theorem 3, p. 22); (b) more details on two examples (Example 4, p. 29, and Example 5, p. 33) covered in the paper: "A Political Reciprocity Mechanism"; and (c) we compare the reciprocity mechanism and the equilibrium concepts defined in this domain to capture farsighted behavior to other well-known solution concepts that follow the blocking approach. A *political economy* or *game* is an environment which consists of a set of agents, a policy space, a voting procedure, a constitution, and a preference profile over the policy space. We denote $\mathbb{P}$ the set of all political economies. In what follows, the political economy derived from the voting procedure with reciprocity clauses V^{rm} is denoted $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$, where the preference profile $R(\theta) = (R_i(\theta))_{i\in N}$ is defined over A with θ being a state of the world (the strict part of $R_i(\theta)$ is denoted $P_i(\theta)$, and indifference is denoted $I_i(\theta)$). A policy $x \in A$ is (weakly) Pareto-efficient if there is no $y \in A$ such that $yP_i(\theta)x$ for all $i \in N$.

1 Theorem 3: Implementation of the Reciprocity Set - Proof

We need some additional notation before proving Theorem 3. Let $x \in A$ be the status quo and $G^{\textsf{Na}} \in \mathcal{F}^{\textsf{Na}}(x,\theta)$ be an extensive form game. Let $\Omega = A \times W(\mathcal{C}) \times N \times W(\mathcal{C}) \times A$, and (y, S, i, T, z) be an element of Ω, with $i \in T - S$, and $y \neq x$. Let $E^x(y, S, i, T, z)$ denote a possible sequence of moves in the game $G^{\textsf{Na}}$ (given the status quo x, an agent is recognized and proposes y, $S \in W(\mathcal{C})$ consists of agents who vote to support y against x, **Nature** continues the game, agent $i \notin S$

*Corresponding authors: Roland Pongou (rpongou@uottawa.ca or rop103@harvard.edu) and Jean-Baptiste Tondji (jeanbaptiste.tondji@utrgv.edu).

opposes y, the bill y is not withdrawn, and agents form $T \in W(\mathcal{C})$ to support z against y). A strategy profile m in the game $G^{\textsf{Na}}$ which corresponds to a sequence $E^x(y, S, i, T, z)$ leads to either y, or z as the final policy, i.e., $g(m, x) \in \{y, z\}$.

(1) Let $\mathcal{PE}(\theta)$ be a political economy. Consider $G^{\textsf{Na}} \in \mathcal{F}^{\textsf{Na}}(x, \theta)$ and assume that $x \in \mathfrak{E}(\mathcal{P}(\theta))$. We show that $x \in$ SPE($G^{\textsf{Na}}$). Suppose that $x \notin$ SPE($G^{\textsf{Na}}$). Then, any strategy profile $\hat{m} \in M$ which is a constrained subgame perfect equilibrium of the game $G^{\textsf{Na}}$ does not yield x as the final policy, i.e., $g(\hat{m}, x) \neq x$. Let $\hat{m}$ be such a strategy profile. Then, there exists an element $(y, S, i, T, z) \in \Omega$, such that the sequence of moves $E^x(y, S, i, T, z)$ generated from $\hat{m}$ leads to either y or z as the final policy, i.e., $g(\hat{m}, x) \in \{y, z\}$, and $g(\hat{m}, x) \neq x$. Consequently, $yP_S(\theta)x$ and the pair (y, S) is a plausible objection against x. Let $(z', T') \in A \times W(\mathcal{C})$ be a plausible counter-objection against (y, S). Then, there exists $j \in T' - S$ such that $xP_j(\theta)y$ and $z'P_{T'}(\theta)y$. Let m be a strategy profile in the game $G^{\textsf{Na}}$ that yields the sequence of moves $E^x(y, S, j, T', z')$. It follows that $g(m, x) \in \{y, z'\}$ because Nature (**Na**) acts after agents form S through voting (V), and (y, S) is a plausible objection against x. Given that $z'P_{T'}(\theta)y$, then we must have $g(m, x) = z'$. Therefore, $z'R_S(\theta)x$, and the plausible counter-objection (z', T') against (y, S) is friendly. Hence, (y, S) is a justified plausible objection, which contradicts the fact that $x \in \mathfrak{E}(\mathcal{PE}(\theta))$. It follows that $x \in$ SPE($G^{\textsf{Na}}$).

Assume there exists $y \in$ SPE($G^{\textsf{Na}}$) such that $y \neq x$. Then, there exists a strategy profile $\hat{m} \in M$ which is a constrained subgame perfect equilibrium of the game $G^{\textsf{Na}}$ with $g(\hat{m}, x) = y$. According to the game $G^{\textsf{Na}}$, there exists $S \in W(\mathcal{C})$ such that $yP_S(\theta)x$. Assume that there exists $(z, T) \in A - \{y\} \times W(\mathcal{C})$, $T \neq S$, $T \cap (N - S) \neq \emptyset$, and $zP_T(\theta)y$. Let $i \in T \cap (N - S)$ and consider the strategy profile $\overline{m} = (\hat{m}_{-i}, \overline{m}_i)$, where agent i opposes y in Step 3, and if recognized in Step 5 or Step 6, proposes z and votes for z against y, and keeps all other actions according to $\hat{m}_i$. Given that $T \in W(\mathcal{C})$, $g(\overline{m}) = z$ and $z_iP_i(\theta)y$, a contradiction, since $y \in$ SPE($G^{\textsf{Na}}$). If there is no $(z, T) \in A \times W(\mathcal{C})$ such that $zP_T(\theta)y$, then (y, S) is a plausible objection, which contradicts the fact that $x \in \mathfrak{E}(\mathcal{PE}(\theta))$. Therefore, $x \in \mathfrak{E}(\mathcal{PE}(\theta))$ implies SPE($G^{\textsf{Na}}$) $= \{x\}$, for every $G^{\textsf{Na}} \in \mathcal{F}^{\textsf{Na}}(x, \theta)$.

(2) Let $x \notin \mathfrak{E}(\mathcal{PE}(\theta))$ and $G^{\textsf{Na}} \in \mathcal{F}^{\textsf{Na}}(x, \theta)$. There exist $y \in A - \{x\}$ and $S \in W(\mathcal{C})$ such that (y, S) is a justified plausible objection against $x : yP_S(\theta)x$. Assume that x is the final policy supported by the constrained subgame perfect equilibrium profile $\hat{m} = (\hat{m}_1, ..., \hat{m}_n) \in M$: $g(\hat{m}, x) = x$.

a) Assume that there is no plausible counter-objection against y. Let a_1 be the first agent in S according to the protocol ϕ. By definition, $\hat{m}_{a_1}$ is a best response given any history $h^t_{a_1}$, $t \in \mathbb{T}$. Consider the deviation in which agent a_1 proposes y in Step 1. There are only two scenarios that could yield x as the final policy from this deviation.

Case (a-1). The vote in Step 2 favors x against y.

Case (a-2). The vote in Step 2 favors y against x; Nature continues the game; an agent $i \in N-S$ opposes policy y; and members in S fail to maintain y.

Consider Case (a-1). Given that $yP_S(\theta)x$ with $S \in W(\mathcal{C})$, x cannot arise as the final policy of a subgame perfect equilibrium. Consider Case (a-2). Here, as in Case (a-1), the best response for each agent in S is to vote "yes" and given that $S \in W(\mathcal{C})$, x cannot arise as the final policy of a constrained subgame perfect equilibrium of the game $G^{\mathbf{Na}}$, which is a contradiction.

b) Assume that there exists a plausible counter-objection (z, T) against the plausible objection (y, S). Thus, there exists $i \in T - S$ such that $xP_i(\theta)y$, and $zP_T(\theta)y$. Given that the plausible objection (y, S) is justified, then $zR_S(\theta)x$. Following the same reasoning in Case a), a constrained subgame perfect equilibrium in the game $G^{\mathbf{Na}}$ can yield either y (because Nature acts after the vote between y and x) or z as the final policy, which is a contradiction.

2 More Elaboration on Examples

2.1 Example 4: Ultimatum game with counter-offer

A population of n agents must share an amount of 100 dollars, with each agent receiving a non-negative portion. The set of feasible allocations is the simplex:

$$A = \{x = (x_1, ..., x_n) \in [0, 100]^n : \sum_{i=1}^{n} x_i \leq 100\}.$$

Assume that the constitution $\mathcal{C}$ is the majority rule, and for the state $\theta \in \Theta$, the preference profile $R(\theta) = (R_i(\theta))_{i \in N}$ is defined as follows: for $x, y \in A$, and $i \in N$, $yR_i(\theta)x$ if $u_i(y) \geq u_i(x)$, where $u_i : A \longrightarrow \mathbb{R}$ is agent i's payoff function. The decision-making process is as follows. An arbitrator proposes a status quo allocation $x_0 = (0, 0, ..., 0)$. A majority coalition S may propose a plausible objection $(y, S) \in A \times W(\mathcal{C})$ against x_0. If there is no opposition, y is implemented. If there is any opposition, S may withdraw y, leading to each agent receiving zero. But if S chooses to maintain y following any opposition, another majority coalition T might react by proposing a plausible counter-objection (z, T) against (y, S). In this case, z is implemented and each agent i obtains and consumes his or her payoff z_i.

To derive the reciprocity set and the predicted reciprocity set, we assume that there are three agents who have the following linear utility functions: $u_1(x_1, x_2, x_3) = x_1 - \frac{1}{3}(x_2 - x_1) - \frac{1}{3}(x_3 - x_1)$; $u_2(x_1, x_2, x_3) = x_2 - \frac{1}{3}(x_1 - x_2) - \frac{1}{3}(x_3 - x_2)$; and $u_3(x_1, x_2, x_3) = x_3 - \frac{1}{3}(x_1 - x_3) - \frac{1}{3}(x_2 - x_3)$.

Assume, for instance, that agents 1 and 2 propose a distribution $x = (x_1, x_2, x_3)$ to agent 3, with $x_1 + x_2 + x_3 = 100$. He or she is indifferent between rejecting x and accepting $x_0 = (0, 0, 0)$ if $u_3(x) = 0$, which means that $x_3 = \frac{1}{5}(x_1 + x_2)$. Since $x_1 + x_2 + x_3 = 100$, we have $x_1 + x_2 = \frac{500}{6}$ and therefore $x_3 = \frac{100}{6}$. Hence, the utility of agent 3, if he or she accepts a share $(x_1, x_2, \frac{100}{6})$, with $x_1 + x_2 = \frac{500}{6}$ is zero. Given the fact that utilities are symmetric, the same conclusion applies to each agent. This proves that proposing an allocation that exhausts the total amount of money and gives less than 100/6 to an agent is rejected. Therefore, the set of the predicted allocations from the status quo $x_0 = (0, 0, 0)$ is:

$$\mathfrak{E}^{x_0}(\mathcal{PE}(\theta))) = \left\{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = 100 \text{ and } x_i \geq \frac{100}{6};\ i = 1, 2, 3\right\}.$$

What about the reciprocity set?

1. Any allocation $x = (x_1, x_2, x_3)$ such that $x_1 + x_2 + x_3 < 100$ is not a reciprocity equilibrium. Consider the allocation $A = (x_1 + \frac{100-x_1-x_2-x_3}{3}, x_2 + \frac{100-x_1-x_2-x_3}{3}, x_3 + \frac{100-x_1-x_2-x_3}{3})$. A is majority-preferred to x since it is preferred to x by everybody, and no one has an incentive to oppose A, meaning that $(A, \{1, 2, 3\})$ is a justified plausible objection against x.

2. Any allocation that gives all of the money to one agent is not a reciprocity equilibrium. For instance, consider the allocation $x = (100, 0, 0)$. If x is the status quo, then the majority coalition that consists of agents 2 and 3 will formulate a plausible objection $(y, \{2, 3\})$ against x, where $y = (34, 33, 33)$. Then, agent 1 will oppose the plausible objection. But agents 2 and 3 will not withdraw it since they have no incentive to do so. It is obvious that agent 1 can convince agent 2 to formulate a plausible counter-objection $((76, 34, 0), \{1, 2\})$ against $(y, \{2, 3\})$. Moreover, any plausible counter-objection (z, T) against y is friendly to $\{2, 3\}$, which implies that $(y, \{2, 3\})$ is a justified plausible objection against x.

3. Any allocation that gives a strictly positive amount to at least two agents is an equilibrium. Consider, for instance, the allocation $x = (x_1, x_2, 0)$, with $x_1 > 0$, $x_2 > 0$, and $x_1 + x_2 = 100$. If x is a status quo, then agent 3 can form a coalition with either agent 1 or agent 2 to initiate a plausible objection against x. Let us assume that the majority coalition $\{2, 3\}$ objects against x by proposing the allocation $y = (0, x_2 + a, x_1 - a)$, with $a > 0$ and $x_1 > a$. Agent 1 will oppose this allocation, causing $\{2, 3\}$ to either withdraw it or maintain it. But $\{2, 3\}$ will withdraw this allocation because, otherwise, agent 1 will form a coalition with either agent 2 or agent 3 to formulate a plausible counter-objection. Assume, for instance, that agents 1 and 3 form a coalition and propose the allocation $(x_1, 0, x_2)$. Then, this will lower the payoff of agent 2. Given the fact that agent 2 is aware of this possibility, he or she will withdraw from the coalition $\{2, 3\}$, which will invalidate the plausible objection $(y, \{2, 3\})$ against the status quo $x = (x_1, x_2, 0)$. Therefore, there is no plausible

objection against $x = (x_1, x_2, 0)$. In essence, given the fact that $x_1 + x_2 = 100$, it is impossible to formulate a plausible objection that improves the payoff of every agent. It therefore follows that $x = (x_1, x_2, 0)$ is a reciprocity equilibrium. Given the fact that agents have symmetric preferences, any allocation that gives a positive amount to at least two agents is a reciprocity equilibrium. The reciprocity set is therefore:

$$\mathfrak{E}(\mathcal{PE}(\theta)) = \left\{ \begin{array}{l} (x_1, x_2, x_3): \ x_1 + x_2 + x_3 = 100, \text{ and there do not exist two agents} \\ i, j \in \{1, 2, 3\}, \ (i \neq j) \text{ such that } x_i = x_j = 0; \ i, j = 1, 2, 3 \end{array} \right\}.$$

2.2 Example 5: Admitting refugees into a peaceful country

We assume that the decision to admit refugees in a peaceful country is made by the legislators of this country. The latter derives utility from the number of refugees that it admits. Each legislator has a different perception of the utility that he or she or the country receives from admitting refugees. We assume that these considerations are reflected in the legislators' utility functions. Assume, for illustration, that there are five legislators who have to decide on the number of refugees to be admitted. The net utility function of each legislator i is given by: $v_i(p_i) = b_i \ln(p_i) - \frac{1}{5}p_i$, where p_i is legislator i's peak. Assume that $\mathcal{C}$ is the majority rule, and the preference profile $R(\theta) = (R_i(\theta))_{i \in N}$ is defined as follows: for p and q number of refugees, and $i \in N$, $pR_i(\theta)q$ if $v_i(p) \geq v_i(q)$. Observe that this function is single-peaked, with legislator i's ideal position being obtained by solving $v_i'(x_i) = 0$; this leads to the solution $x_i^* = 5b_i$. Assume that $b_i = 6 - i$, $i \in N = \{1, 2, 3, 4, 5\}$. We can show that the legislators' peaks are respectively: $p_1^* = 25$, $p_2^* = 20$, $p_3^* = 15$, $p_4^* = 10$ and $p_5^* = 5$. The peak p_3^*, which is the optimum of the median legislator, is the Condorcet winner under the reciprocity mechanism. It follows that p_3^* is the only reciprocity equilibrium: $\mathfrak{E}(\mathcal{PE}(\theta)) = \{15\}$.

Now, assume that there are six legislators, with legislator i's utility function being $v_i(p_i) = b_i \ln(p_i) - \frac{1}{6}p_i$, where $b_i = 6 - i$, $i \in \{1, 2, 3, 4, 5\}$, and $b_6 = 6$. The legislators' peaks are: $p_6^* = 36$, $p_1^* = 30$, $p_2^* = 24$, $p_3^* = 18$, $p_4^* = 12$ and $p_5^* = 6$. In this scenario, there is no median legislator. The reciprocity set consists of p_2^* and p_3^*. In fact, $(p_3^*, \{1, 2, 3, 4, 6\})$ is a justified plausible objection against p_5^*; $(p_3^*, \{1, 2, 3, 6\})$ is a justified plausible objection against p_4^*; $(p_2^*, \{2, 3, 4, 5\})$ is a justified plausible objection against p_1^*; and $(p_2^*, \{2, 3, 4, 5\})$ is a justified plausible objection against p_6^*. Moreover, there is no justified plausible objection against either p_3^* or p_2^*, if each of these alternatives is selected as the status quo. We can conclude that $\mathfrak{E}(\mathcal{PE}(\theta)) = \{p_2^*, p_3^*\} = \{24, 18\}$. Therefore, either 18 or 24 refugees will be admitted into the country.

3 Comparison with Other Voting Procedures

3.1 A Modification of the Reciprocity Mechanism

In this section, we modify the voting procedure, V^{rm}, in the reciprocity mechanism to the procedure V^1 described below (also see Figure 1). Assume that a society wishes to reform a status quo $x \in A$ by voting. The voting procedure V^1 is comprised of two principal stages.

Stage 1 (**Objection**). If no winning coalition S proposes an objection (y, S) against x, then x remains in place, which ends the process. If an objection (y, S) against x exists, y is recognized as a bill, and another coalition might formulate a counter-objection during the second stage.

Stage 2 (**Counter-objection**). Suppose that an objection (y, S) against x exists. Another winning coalition T might propose a counter-objection or an amendment (z, T) against (y, S). If no counter-objection to (y, S) exists, then y becomes the new policy, which ends the process. If a counter-objection (z, T) exists, z becomes the new policy, which ends the process.

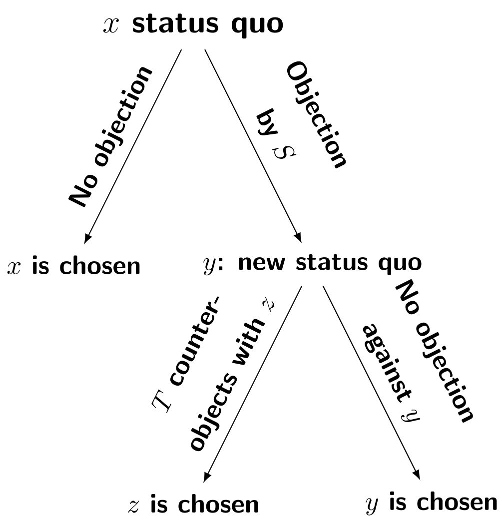

Figure 1: Description of Voting Procedure V^1

The procedure V^1 differs from the voting process V^{rm} in that it does not allow any sponsoring coalition S to withdraw a bill y once the latter has been introduced. This non-retraction clause is found in many legislatures. In the Canadian parliamentary system, for instance, there is no disposition that allows a sponsoring coalition to withdraw a bill before the final vote. In other systems, the withdrawal of a bill has to receive the unanimous consent of the members of the legislature to be approved. The procedure V^1 therefore gives any potential second mover T the possibility to free-ride on the action of S by deviating to a more preferred policy z from the bill y. This deviation still occurs even if the move from x to y initiated by S does not harm the interests of any member of T. If the second move is indeed likely to hurt S, S will anticipate this and will refrain from sponsoring the bill. The analysis reveals that this anticipation may result in a non-Pareto-efficient policy persisting indefinitely.

We define hereunder the rational behavior of agents under the game induced by the procedure V^1.

Definition 1. Let $\mathcal{PE}^1(\theta) = \langle N, A, \mathcal{C}, V^1, R(\theta)\rangle$ be a political economy, S be a winning coalition, and $x, y \in A$ be two policies.

1. **Plausible M-objection**: (y, S) is said to be a *plausible* M-objection against x if $yR_i(\theta)x$ for each $i \in S$ and $yP_j(\theta)x$ for some $j \in S$.

2. **Plausible M-counter-objection**: Let (y, S) be an plausible M-objection against x. A pair $(z, T) \in A \times W(C)$ is said to be an plausible M-counter-objection against (y, S) if $zR_i(\theta)y$ for each $i \in T$ and $zP_j(\theta)y$ for some $j \in T$.

3. **Unfriendly plausible M-counter-objection**: Let (y, S) be an plausible M-objection against x and $(z, T) \in A \times W(C)$ be a plausible M-counter-objection against (y, S). The plausible M-counter-objection (z, T) is said to be unfriendly if $not(zR_S(\theta)x)$.

4. **Justified plausible M-objection**: A plausible M-objection (y, S) against x is said to be justified if there is no unfriendly plausible M-counter-objection against (y, S).

5. **Non-reciprocity set**: The non-reciprocity set of the political economy $\mathcal{PE}^1(\theta)$, denoted by $NR(\mathcal{PE}^1(\theta))$, is the set of all of the policies in A against which no justified plausible M-objection exists.[1]

The rational behavior in games under the modified voting procedure V^1 is defined in accordance with the same logic as under the reciprocity mechanism. The main difference resides in the conditions under which a plausible counter-objection is formulated. Under the procedure V^1, any winning coalition that wishes to deviate from a sponsored bill to a new policy can formulate a plausible M-counter-objection. Under the reciprocity mechanism, there exists an additional condition, which is that, any winning coalition T that wishes to deviate from a sponsored bill to a more preferred policy can rationally do so only if the interests of some of its members would be harmed by the

[1] The plausible M-objection and M-counter-objection could be defined with strict preferences as for the reciprocity set. In this case, the set of equilibrium policies would be larger than the non-reciprocity set as currently defined. Since the non-reciprocity set generally contains non-Pareto-efficient policies as we show below, this larger set also contains non-Pareto-efficient policies.

enactment of the bill. In fact, if the enactment of the bill would not harm any member of T, any unfriendly plausible counter-objection would cause the first mover S to withdraw its plausible objection. Therefore the unfriendly plausible counter-objection would be nullified and the status quo would be maintained. This response would not benefit T. Under the reciprocity mechanism, it is therefore not rational for T to formulate an unfriendly plausible counter-objection if the plausible objection does not harm its interests, whereas under the game induced by the procedure V^1, this is possible. The result confirms that this absence of incentives for reciprocal actions in the voting procedure V^1 might cause a non-Pareto-efficient policy to persist indefinitely.

Proposition 1. *There exists a political economy $\mathcal{PE}^1(\theta) \in \mathbb{P}$ such that the non-reciprocity set $NR(\mathcal{PE}^1(\theta))$ contains a non-Pareto-efficient policy.*

Proof. Assume $N = \{1, 2, 3, 4, 5\}$ and $A = \{a, b, c, d, e\}$. Consider the constitution $\mathcal{C}$ defined by the winning coalitions $K = \{1, 3, 4\}$, $S = \{2, 4, 5\}$, $L = \{1, 2, 5\}$, $Q = \{2, 3, 4\}$, and any coalition that includes K, S, L, or Q, and the preference profile $R(\theta)$ is defined as follows:

$cP_1(\theta)bP_1(\theta)aP_1(\theta)eP_1(\theta)d$; $dP_2(\theta)cP_2(\theta)bP_{2(\theta)}aP_2(\theta)e$; $eP_3(\theta)bP_3(\theta)aP_3(\theta)dP_3(\theta)c$;

$eP_4(\theta)dP_4(\theta)bP_4(\theta)aP_4(\theta)c$; and $cP_5(\theta)dP_5(\theta)eP_5(\theta)bP_5(\theta)a$.

Let $\mathcal{PE}^1(\theta) = \langle N, A, \mathcal{C}, V^1, R(\theta)\rangle$. Figure 2 provides the graph of the popularity relationship among the policies (for example, the arrow from a to b means that the grand coalition N strictly prefers policy b to a.). The non-Pareto-efficient policy a is an equilibrium under the political economy $\mathcal{PE}^1(\theta)$. Even though all of the agents prefer b over a, no winning coalition will formulate a plausible M-objection against a since this will lead to the election of either d (if coalition L proposes a plausible M-objection) or e (if coalition S proposes a plausible M-objection). In either case, some member of the M-objecting coalition will regret, which is the reason why none of these coalitions will form in order to make a plausible M-objection. This implies that $a \in NR(\mathcal{PE}^1(\theta))$.[2]

Under the political economy $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$, coalition K, for instance, will formulate a plausible objection (b, K) against a. Once b is introduced, coalition L will not formulate a plausible counter-objection (L, c), knowing that such a plausible counter-objection is unfriendly as it harms the interest of agents 3 and 4 in K. In fact, if it does this, these agents will withdraw from the winning coalition K, therefore destroying the plausible objection (b, K) and causing a to remain in place. This response will harm the interest of agents 2 and 5 in the plausible counter-objecting

[2]Under the voting procedure V^1, in general, some agents who might be willing to participate in a coalition in order to remove a non-Pareto-efficient status quo fear that the interim policy will be replaced by a new policy which is worse for them than the status quo.

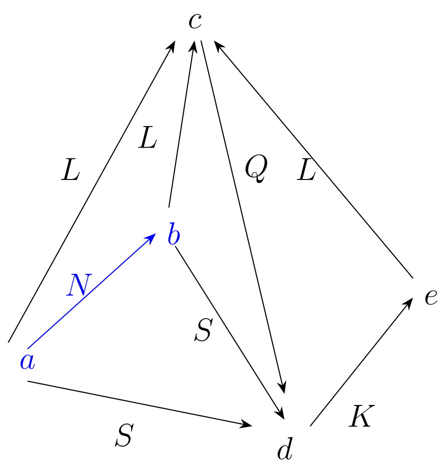

Figure 2: Popularity relationship among policies

coalition L. These latter agents will therefore not participate in coalition L and that coalition will never formulate a plausible counter-objection. Furthermore, coalition S will not formulate a plausible counter-objection (S, d) since agents 2 and 5 will not participate in that coalition. It follows that a is not a reciprocity equilibrium. We can show that the reciprocity set is $\mathfrak{E}(\mathcal{PE}(\theta)) = \{b, c, d, e\}$ and that $NR(\mathcal{PE}^1(\theta)) = \{a, b, c, d, e\}$. $\square$

3.2 One-shot Games

The one-shot game is derived from a voting procedure in which policy selection entails only one stage: a challenger is pitted against an incumbent policy, and if a winning coalition votes for change, the incumbent is replaced by the new policy and the selection process ends. Otherwise, the incumbent remains in place and the selection process ends. Let V^2 represents a one-stage voting procedure described above, and $\mathcal{PE}^2(\theta) = \langle N, A, \mathcal{C}, V^2, R(\theta) \rangle$ denotes a one-shot political economy. Many solution concepts have been defined to capture rationality in this class of games. The most prominent are the von Neumann-Morgenstern (vNM) stable set (von Neumann and Morgenstern, 1944), the core (Gillies, 1959), the top cycle set (Schwartz, 1976), and the uncovered set (Miller, 1980). A common feature of these solution concepts is that they are all based on the notion of direct domination. Their definitions are recalled below.

Definition 2. Let $\mathcal{PE}^2(\theta) = \langle N, A, \mathcal{C}, V^2, R(\theta) \rangle$ be a one-shot political economy, X be a subset of A, and x and y be two policies.

- y is said to directly dominate x if there exist a winning coalition $S \in W(\mathcal{C})$ that prefers y over x, that is, $yP_S(\theta)x$. If y directly dominates x, we say that y majority-defeats (or simply defeats) x.

- The core is the set of all of the policies that are not directly defeated.

- X is a vNM stable set if it satisfies the following stability conditions:

  1. (Internal stability): no policy in X is directly defeated by another policy in X; and
  2. (External stability): every policy not in X is directly defeated by some policy in X.

- The top cycle set, denoted as $TC(\mathcal{PE}^2(\theta))$, is the smallest subset of the policy space A, where every policy in $TC(\mathcal{PE}^2(\theta))$ defeats every other policy not in $TC(\mathcal{PE}^2(\theta))$.

- y is said to *cover* x if every policy z that is defeated by x is also defeated by y. An uncovered policy is such that there is no other policy that covers it. The uncovered set, denoted as $UC(\mathcal{PE}^2(\theta))$, is the set of uncovered policies in A.

It is easy to prove that a non-Pareto-efficient policy cannot belong to the core, or to a vNM stable set, or to the uncovered set. These equilibrium concepts share this feature with the reciprocity set. However, we prove hereunder that there exists a political economy for which the reciprocity set is not empty, the core is empty, a vNM stable set does not exist, and the top cycle set contains a non-Pareto-efficient policy.

Proposition 2. *There exists a one-shot political economy* $\mathcal{PE}^2(\theta) = \langle N, A, \mathcal{C}, V^2, R(\theta)\rangle \in \mathbb{P}$ *such that:*

1. *the reciprocity set* $\mathfrak{E}(\mathcal{PE}(\theta))$ *is not empty, where* $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$*;*
2. *the core of the one-shot game* $\mathcal{PE}^2(\theta)$ *is empty;*
3. *a vNM stable set of the one-shot game* $\mathcal{PE}^2(\theta)$ *does not exist; and*
4. *the top-cycle set of the one-shot game* $\mathcal{PE}^2(\theta)$ *contains a non-Pareto-efficient policy.*

Proof. Consider $N = \{1, 2, 3, 4, 5\}$, $A = \{a, b, c, d, e\}$, the constitution $\mathcal{C}$ defined by the winning coalitions $K = \{1, 3, 4\}$, $S = \{2, 4, 5\}$, $L = \{1, 2, 5\}$, $Q = \{2, 3, 4\}$, and any coalition that includes K, S, L, or Q, and the preference profile $R(\theta)$ defined as follows:

$cP_1(\theta)bP_1(\theta)aP_1(\theta)eP_1(\theta)d$; $dP_2(\theta)cP_2(\theta)bP_2(\theta)aP_2(\theta)e$; $eP_3(\theta)bP_3(\theta)aP_3(\theta)dP_3(\theta)c$;

$eP_4(\theta)dP_4(\theta)bP_4(\theta)aP_4(\theta)c$; and $cP_5(\theta)dP_5(\theta)eP_5(\theta)bP_5(\theta)a$.

Let denote $\mathcal{PE}^2(\theta) = \langle N, A, \mathcal{C}, V^2, R(\theta)\rangle$ and $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$.

1. As is shown in the proof of Proposition 1, the reciprocity mechanism is given by $\mathfrak{E}(\mathcal{PE}(\theta)) = \{b, c, d, e\}$. The reciprocity mechanism expels the non-Pareto-efficient policy a.

2. We can observe in Figure 2 that, for any policy $x \in A$, there exists another policy $y \in A$ and a winning coalition $S \in W(\mathcal{C})$ such that all agents in S prefer y over x. This shows that the core of the one-shot game $\mathcal{PE}^2(\theta)$ is empty.

3. In the one-shot game $\mathcal{PE}^2(\theta)$, there is no vNM stable set. In fact, c defeats a, c defeats b, c defeats e, b defeats a, d defeats a, d defeats b, d defeats c, e defeats d. Hence, no subset of the policy set A satisfies the vNM external and internal stability conditions.

4. The top cycle set of the one-shot game $\mathcal{PE}^2(\theta)$ contains all the policies in A. In fact, a policy does not belong to $TC(\mathcal{PE}^2(\theta))$ if and only if it is defeated by each policy in $TC(\mathcal{PE}^2(\theta))$; this condition is not satisfied for any of the policies in A. Then, $TC(\mathcal{PE}^2(\theta)) = A$, and it contains the non-Pareto-efficient policy a. □

We also have the following result comparing the uncovered set and the reciprocity set.

Proposition 3. *The following statements hold:*

1. *the uncovered set mechanically protects minority interests.*

2. *the uncovered set does not strategically protect minority interests.*

3. *There exists a one-shot political economy $\mathcal{PE}^2(\theta) = \langle N, A, \mathcal{C}, V^2, R'(\theta)\rangle$ such that the reciprocity set of the game $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R'(\theta)\rangle$ strictly refines the uncovered set: $\mathfrak{E}(\mathcal{PE}(\theta)) \subsetneqq UC(\mathcal{PE}^2(\theta))$.*

Proof. To prove statements 1. and 2., let consider $N = \{1, 2, 3, 4, 5, 6\}$, $A = \{a, b, c, d\}$, $\mathcal{C}$ the majority rule, and the preference profile $R(\theta)$ defined as follows:

$bP_1(\theta)dP_1(\theta)cP_1(\theta)a$; $dP_2(\theta)bP_2(\theta)cP_2(\theta)a$; $cP_3(\theta)bP_3(\theta)dP_3(\theta)a$;

$cP_4(\theta)bP_4(\theta)aP_4(\theta)d$; $aP_5(\theta)dP_5(\theta)cP_5(\theta)b$; and $aP_6(\theta)dP_6(\theta)cP_6(\theta)b$.

Let $S = \{1, 2, 3, 4\}$, $T = \{3, 4, 5, 6\}$, and $U = \{1, 2, 5, 6\}$. The domination or popularity graph among policies based on preferences is provided in Figure 3. Let $\mathcal{PE}^2(\theta) = \langle N, A, \mathcal{C}, V^2, R(\theta)\rangle$.

1. Let denote by $D(x)$ the set of all the alternatives that are defeated by the policy x. By the definition of the uncovered set, $x \in UC(\mathcal{PE}^2(\theta))$ if and only if there is no $y \in A$ such that $D(x) \subseteq D(y)$. We have $D(a) = \emptyset$, $D(b) = \{a\}$, $D(c) = \{a, b\}$, and $D(d) = \{c\}$. It follows that, either b or c covers a, and c covers b. Only c and d are uncovered, and the uncovered set

$$\begin{tikzcd} a \arrow[r, "S"] \arrow[rr, bend right, "S"'] & b \arrow[r, "T"] & c \arrow[r, "U"] & d \end{tikzcd}$$

Figure 3: Uncovered set versus Reciprocity set (the arrows indicate the direction of the popularity relationship; for instance b is a more popular policy than a because b is preferred over a by the majority coalition S)

is $UC(\mathcal{PE}^2(\theta)) = \{c, d\}$. Since the alternative c is a minority option and $c \in UC(\mathcal{PE}^2(\theta))$, the uncovered set mechanically protects minority interests.

2. The alternative c is not a non-strategic minority option, since d is a majority option that defeats c. Given that $c \in UC(\mathcal{PE}^2(\theta))$, it follows that the uncovered set does not strategically protect minority interests.

3. We modify the environment above only by changing the preference profile $R(\theta)$ to the profile $R'(\theta)$ described as follows:

$bP'_1(\theta)dI'_1(\theta)cI'_1(\theta)a$; $dP'_2(\theta)bP'_2(\theta)cI'_2(\theta)a$; $cP'_3(\theta)bP'_3(\theta)dP'_3(\theta)a$;

$cP'_4(\theta)bP'_4(\theta)aP'_4(\theta)d$; $aP'_5(\theta)dI'_5(\theta)cP'_5(\theta)b$; and $aP'_6(\theta)dI'_6(\theta)cP'_6(\theta)b$.

$$\begin{tikzcd} a \arrow[r, "S"] & b \arrow[r, "T"] & c \end{tikzcd}$$

Figure 4: Popularity Relationship Under $R'(\theta)$

The popularity relationship among policies based on the new preference profile $R'(\theta)$ is provided in Figure 4. Let $\mathcal{PE}^2(\theta) = \langle N, A, \mathcal{C}, V^2, R'(\theta) \rangle$ and $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R'(\theta) \rangle$. We have $D(a) = \emptyset$, $D(b) = \{a\}$, and $D(c) = \{b\}$. Thus, $\mathfrak{E}(\mathcal{PE}(\theta)) = \{c\}$ and $UC(\mathcal{PE}^2(\theta)) = \{b, c\}$. $\square$

Proposition 3 shows that the reciprocity set and the uncovered set have very different properties and predictions. For instance, consider a political economy $\mathcal{PE}^2(\theta) = \langle N, A, \mathcal{C}, V^2, R(\theta) \rangle$ that consists of a society that has four ethnic groups. This society has to choose from a set of three languages, the ones to be used both in school and in official communication. Let $N = \{1, 2, 3, 4\}$ be the set of ethnic groups and $L = \{l_1, l_2, l_3\}$ be the set of languages. Each ethnic group has one representative. Language l_3 is the most commonly spoken language in ethnic groups 1 and 2; l_2 is the native language of ethnic group 3; and l_1 is a minority language that is mostly spoken in ethnic group 4. The preference profile $R(\theta)$ is presented as follows. Ethnic group 1: $l_3P_1(\theta)l_2P_1(\theta)l_1$; ethnic group

2: $l_3 P_2(\theta) l_2 P_2(\theta) l_1$; ethnic group 3: $l_2 P_3(\theta) l_1 P_3(\theta) l_3$; and ethnic group 4: $l_1 P_4(\theta) l_3 P_4(\theta) l_2$. We suppose that the constitution $\mathcal{C}$ is the majority rule. Figure 5 shows the popularity relationship among the various languages (for example, l_3 is more popular than l_2 as the former is preferred by three out of the four ethnic groups). One can prove that $\mathfrak{E}(\mathcal{PE}(\theta)) = \{l_1, l_3\}$, whereas $UC(\mathcal{PE}^2(\theta)) = \{l_2, l_3\}$,

$$l_1 \xrightarrow{\{1,2,3\}} l_2 \xrightarrow{\{1,2,4\}} l_3$$

Figure 5: Popularity relationship among various languages

with $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta) \rangle$. Interestingly, both solution concepts select l_3 as an official language. However, while the reciprocity set selects the strategic minority option l_1, the uncovered set selects the non-strategic minority option l_2, which proves that the two solution concepts have a very different incentives structure.

3.3 Farsighted Coalitional Games

In farsighted coalitional games, agents or coalitions may bargain indefinitely. In a game, a status quo a_0 is randomly chosen from the set of policies. If no winning coalition replaces a_0, then it remains in place on an indefinite basis and the game ends. If a winning coalition S replaces a_0, say with a_1, then a_1 becomes the new status quo, and the process restarts, continuing until a policy has been reached to which no winning coalition is willing to object. Once that policy has been reached, each agent earns and consumes his or her payoff and the game ends. Figure 6 illustrates this voting procedure, that we denote by V^3. We also denote by $\mathcal{PE}^3(\theta) = \langle N, A, \mathcal{C}, V^3, R(\theta) \rangle$ a political economy or farsighted coalitional game. The most prominent equilibrium concepts in this class of games are based on the notion of indirect domination (Harsanyi (1974), Chwe (1994)), which is a modification of direct domination, whereby first movers anticipate future moves. We recall their definitions below.

Definition 3. Let $\mathcal{PE}^3(\theta) = \langle N, A, \mathcal{C}, V^3, R(\theta) \rangle$ be a farsighted coalitional game, K be a subset of A, and a and b be two policies.

A) Alternative b is said to farsightedly dominate a, denoted as $b \gg^H a$, if there exists a sequence of policies $a_0,\ a_1,\ \ldots,\ a_m \in A$ $(where\ a_0 = a\ and\ a_m = b)$ and a sequence of winning coalitions $S_0,\ S_1,\ \ldots, S_{m-1}$ such that a_{i+1} directly dominates a_i S_{i+1} and $bP_{S_i}(\theta)a_i$ for $i = 0, 1, \ldots, m-1$.

B) K is a farsighted stable set if it satisfies the following conditions:

1. (Internal stability): no policy in K is farsightedly defeated by any other policy in K; and

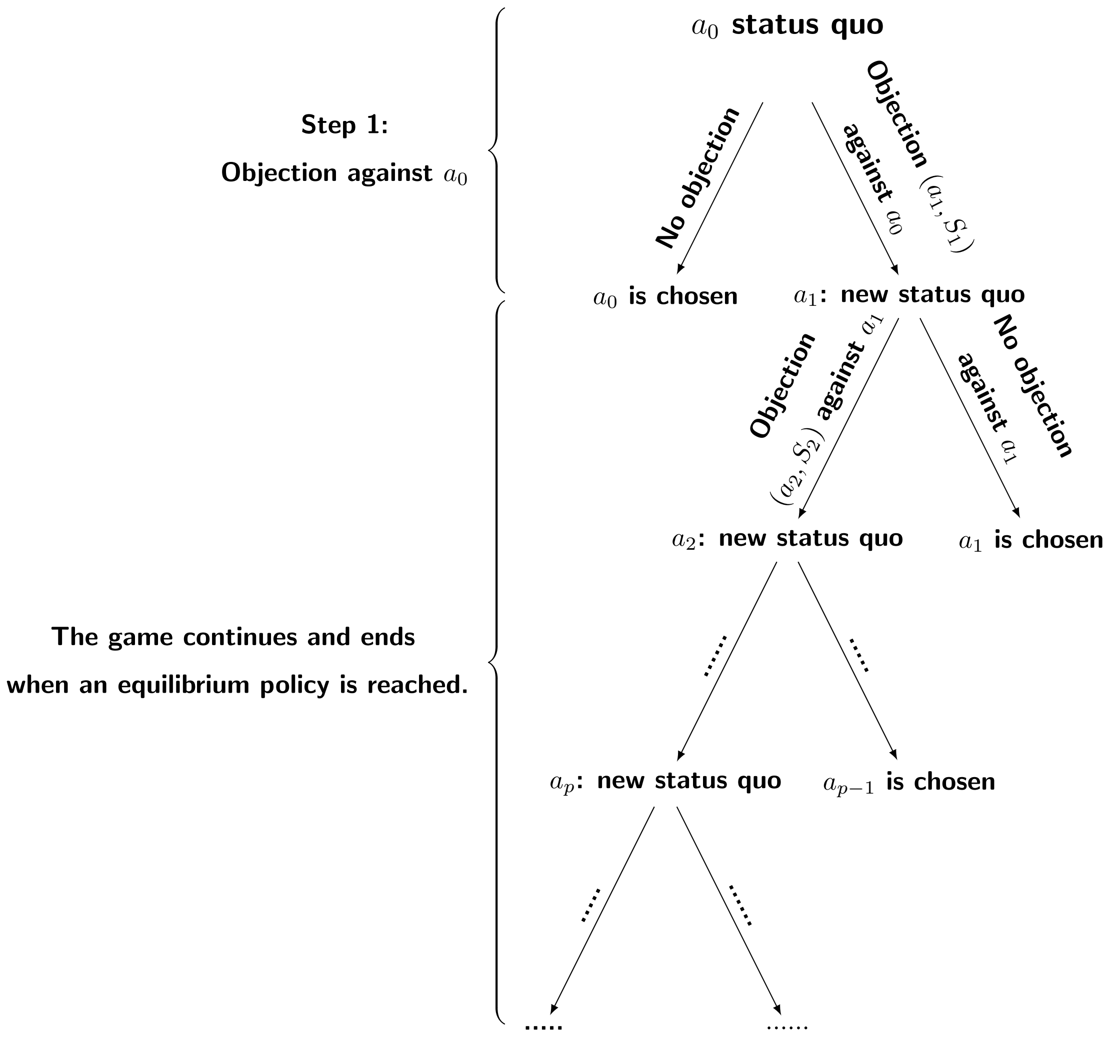

Figure 6: Description of Voting Procedure V^3

2. (External stability): every policy not in K is farsightedly defeated by some policies in K.

Definition 4. Let $\mathcal{PE}^3(\theta) = \langle N, A, \mathcal{C}, V^3, R(\theta)\rangle$ be a farsighted coalitional game, X be a subset of A, and a and b be two policies.

A) Alternative b is said to indirectly dominate a, denoted as $b \gg^C a$, if there exists a sequence of policies $a_0,\ a_1,\ ...,\ a_m \in A\ (where\ a_0 = a\ and\ a_m = b)$ and a sequence of winning coalitions $S_0,\ S_1,\ ..., S_{m-1}$ such that $a_i \longrightarrow_{S_i} a_{i+1}$ and $bP_{S_i}(\theta)a_i$ for $i = 0, 1, ..., m-1$. The relation $a \longrightarrow_S b$ means that, if a is a status quo, S has the power to make b the new status quo.

B) X is said to be consistent if:

$$f(X) = \begin{Bmatrix} a \in A : \forall d \in A, S \in W(\mathcal{C}),\ \exists e \in X,\ where \\ e = d\ or\ e \gg^C d\ and\ not(aP_S(\theta)e) \end{Bmatrix} = X.$$

C) The largest consistent set is the union of all the consistent sets of $\mathcal{PE}^3(\theta)$.

These solution concepts formalize the notion that a coalition that moves from a status quo to an alternative policy anticipates the possibility that another coalition might react. A third coalition might in turn react, and so on, without limit. It is therefore important to act in a way that does not lead a coalition to ultimately regret its action. Exactly what happens during the intermediate stages might not matter, as a coalition simply wants to be better off in terms of the final option relative to a status quo.

The procedure V^3 differs from V^{rm} in two major respects. First, V^{rm} is finite and involves only three stages. Second, under V^3, a coalition does not have the possibility to revise or withdraw its move. For these reasons, these procedures yield different equilibrium and welfare properties, as we show below.

Proposition 4. *There exists a political economy $\mathcal{PE}^3(\theta) = \langle N, A, \mathcal{C}, V^3, R(\theta)\rangle \in \mathbb{P}$ such that:*

1. *the reciprocity set $\mathfrak{E}(\mathcal{PE}(\theta))$ is not empty, where $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$;*
2. *there is no Harsanyi stable set in the farsighted coalitional game $\mathcal{PE}^3(\theta)$; and*
3. *the largest consistent set of the farsighted coalitional game $\mathcal{PE}^3(\theta)$ contains a non-Pareto-efficient policy.*

Proof. Consider $N = \{1, 2, 3, 4, 5\}$, $A = \{a, b, c, d, e\}$, the constitution $\mathcal{C}$ defined by the winning coalitions $K = \{1, 3, 4\}$, $S = \{2, 4, 5\}$, $L = \{1, 2, 5\}$, $Q = \{2, 3, 4\}$, and any coalition that includes K, S, L, or Q, and the preference profile $R(\theta)$ defined as follows:

$cP_1(\theta)bP_1(\theta)aP_1(\theta)eP_1(\theta)d$; $dP_2(\theta)cP_2(\theta)bP_2(\theta)aP_2(\theta)e$; $eP_3(\theta)bP_3(\theta)aP_3(\theta)dP_3(\theta)c$;

$eP_4(\theta)dP_4(\theta)bP_4(\theta)aP_4(\theta)c$; and $cP_5(\theta)dP_5(\theta)eP_5(\theta)bP_5(\theta)a$.

Let $\mathcal{PE}^3(\theta) = \langle N, A, \mathcal{C}, V^3, R(\theta)\rangle$ and $\mathcal{PE}(\theta) = \langle N, A, \mathcal{C}, V^{rm}, R(\theta)\rangle$. Figure 2 provides the graph of the popularity relationship among the policies. We can also view Figure 2 as the graph of "effectiveness relations ($\longrightarrow_S$)" (Chwe, 1994). It is straightforward to prove the following: $c \gg^C a$, $c \gg^C b$, $c \gg^C e$, $b \gg^C a$, $d \gg^C a$, $d \gg^C b$, $d \gg^C c$ and $e \gg^C d$. Note that in this environment, the binary relations $\gg^H$ and $\gg^C$ are equivalent.

1. It is already presented in the previous proof (Proposition 1) that the reciprocity set is non-empty and contains only Pareto-efficient policies: $\mathfrak{E}(\mathcal{PE}(\theta)) = \{b, c, d, e\}$.

2. Using Figure 2 and the farsighted domination $\gg^H$, we can conclude that there is no subset of policy set A that satisfies the internal and the external stability conditions in the game $\mathcal{PE}^3(\theta)$. Therefore, no Harsanyi stable set exists.

3. Now, we determine the largest consistent set. We start with $X = A$. It turns out that $f(X) = X$ and A is the largest consistent set. For instance, starting with policy a, there are three

possibilities: either coalition L will move from a to c, or coalition R will move from a to d, or all agents will move from a to b. For each initial move, there is a subsequent move that reaches either e, or d or c and in which some agent in the coalition that initiated the move from a is worse off. For these reasons, the inefficient policy a belongs to the largest consistent set. Following the same reasoning, we can show that the largest consistent set is equal to A. □

References

Chwe, M. S.-Y. (1994). Farsighted coalitional stability. *Journal of Economic Theory*, 63(2), 299–325.

Gillies, D. B. (1959). Solutions to general non-zero-sum games. *Contributions to the Theory of Games*, 4(40), 47–85.

Harsanyi, J. C. (1974). An equilibrium-point interpretation of stable sets and a proposed alternative definition. *Management Science*, 20(11), 1472–1495.

Miller, N. R. (1980). A new solution set for tournaments and majority voting: Further graph-theoretical approaches to the theory of voting. *American Journal of Political Science*, (pp. 68–96).

Schwartz, T. (1976). The logic of collective choice. *Mimeograph, Dept. of Government, University of Texas at Austin*, Part III, Chap. 6.

von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*, volume 60. Princeton, NJ: Princeton University Press.