2020

A Political Reciprocity Mechanism

Roland Pongou

Jean-Baptise Tondji
The University of Texas Rio Grande Valley

A Political Reciprocity Mechanism

ROLAND PONGOU* AND JEAN-BAPTISTE TONDJI†

February 8, 2020

This paper considers the problem faced by a political authority that has to design a legislative mechanism that guarantees the selection of policies that are stable, efficient, and inclusive in the sense of strategically protecting minority interests. Experimental studies suggest that some of these desirable properties can be achieved if decision-makers (e.g., legislators) are induced to display reciprocal and pro-social behavior. However, the question of how a voting mechanism can be designed to incentivize "selfish" individuals to display such behavior remains unresolved. We propose such a mechanism and find that it is a simplification of legislative procedures used in some democratic societies. This mechanism defines a new equilibrium concept–*the reciprocity set*–that satisfies all of the aforementioned properties under mild conditions, and it is easily implementable. In addition, it encourages positive reciprocity and generally protects minorities without having to make use of a supermajority rule as many real-world political institutions do. Finally, a comparative analysis shows that this mechanism has other desirable features and properties that distinguish it from other well-known political procedures.

KEYWORDS. Reciprocity, reciprocity set, political design, political bargaining, blocking approach, stability, efficiency, inclusiveness.
JEL CLASSIFICATION. P16, D72, C7, J15, H41.

*University of Ottawa and Harvard University: rpongou@uottawa.ca or rop103@harvard.edu
†The University of Texas Rio Grande Valley: jeanbaptiste.tondji@utrgv.edu

1 Introduction

In this paper, we consider the problem faced by a political authority that has to design a legislative mechanism that guarantees the selection of policies that are stable, efficient, and inclusive in that they generally protect minority interests. It is well known that the pursuit of self-interest often leads to the emergence and persistence of public policies that are inefficient and non-inclusive (Samuelson (1954), Falkinger et al. (2000), Acemoglu et al. (2011)). Experimental studies suggest that these issues can be resolved if individuals are induced to display reciprocal and pro-social behavior (Axelrod (1984), Fehr & Gächter (2000), Falk & Fischbacher (2006), Nash et al. (2012)). However, to our knowledge, in the context of voting, there does not currently exist any formal mechanism that is able to incentivize "selfish" individuals to display such behaviors. By providing an incentive structure for the adoption of policies that are pro-social and inclusive, such a mechanism would address two important problems—policy inefficiency and the tyranny of the majority—that have been identified by political scientists as characterizing many real-life voting systems. We propose such a mechanism, which we designate as a political *reciprocity mechanism*.

Our *reciprocity mechanism*, as described below, is a sequential procedure that embeds elements of both one-shot and infinite horizon games. A one-shot mechanism (e.g., von Neumann & Morgenstern (1944), Nash (1951), Gillies (1959), Aumann (1959), Schwartz (1976), Miller (1980)) entails only one stage of decision making in which a policy is pitted against the status quo or a reference point; in the context of voting, the status quo is replaced if a (qualified) majority of individuals are in favor of change. In infinite horizon games (see for e.g., Harsanyi (1974), Baron & Ferejohn (1989), Chwe (1994), Eguia & Shepsle (2015), among many others), individuals may object and counter-object until they reach a *stable* or an *equilibrium* outcome. It is well known that while the one-shot mechanism generally ensures that the chosen policy is Pareto-efficient, it has been criticized by political scientists and philosophers for its lack of political inclusiveness[1] and

[1]Discussion about the protection of minorities has received considerable attention since the

its general failure to select policies that are *stable* (e.g., Cox & Shepsle (2007), Brighouse & Fleurbaey (2010)). On the other hand, the infinite horizon mechanism (some solution concepts for this class of games are surveyed in Fotso et al. (2017)) may lead to the selection and persistence of inefficient policies. We will show that our *reciprocity mechanism* resolves all of these issues under mild conditions. Interestingly, by fostering positive reciprocity among decision-makers (voters, policymakers), it also reduces the expropriation of decision-making power by majority groups without necessarily having to resort to a supermajority rule to select policies.[2]

The *reciprocity mechanism* 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. During the first stage, a bill is introduced by a legislator or by a qualified majority of legislators.[3] During the second stage, the bill is debated, possibly leading to one or more modifications or amendments. In theory, this stage can last indefinitely. During the final stage, the possibly amended bill is put to a vote. If it passes, it becomes a new law, either immediately or after the completion of a final level of procedure in the form of an action by the head

works of Madison (1788) and Mill (1861).

[2]In real-life institutions, the tyranny of the majority is generally avoided by using supermajority rules to select policies (see, e.g., Buchanan & Tullock (1962), Guinier (1994), King (1996), McGann (2006)). To illustrate, suppose that a two-thirds majority rule is used to make decisions. Then a minority group that represents exactly one third of the population would be protected under such a rule if its members all oppose a particular policy. Our *reciprocity mechanism* protects minorities under supermajority rules. It also protects minorities under the majority rule, depending on voters' strategic considerations. In this case, minorities do not necessarily have a "fixed" identity, such as an ethnic group. Our analysis offers an additional explanation of the "decisive minority" phenomenon (Campbell, 1999), in which a social issue thought to be supported only by a minority group can win the support of a majority of voters.

[3]A proposal can be introduced by a single legislator in certain systems such as the United States Senate or the Canadian Parliament, whereas in many other systems including the European Parliament and the South Korean Parliament, a proposal can only be introduced if it is supported by a qualified majority of members.

of state or by his or her designated representative, specifically an approval that in many cases is provided as a non-discretionary formality (e.g., royal assent in the United Kingdom and Canada). If the bill fails to receive sufficient legislative support, the status quo remains in place. We show that this traditional decision-making procedure may lead to the selection of inefficient policies (see Section 7).

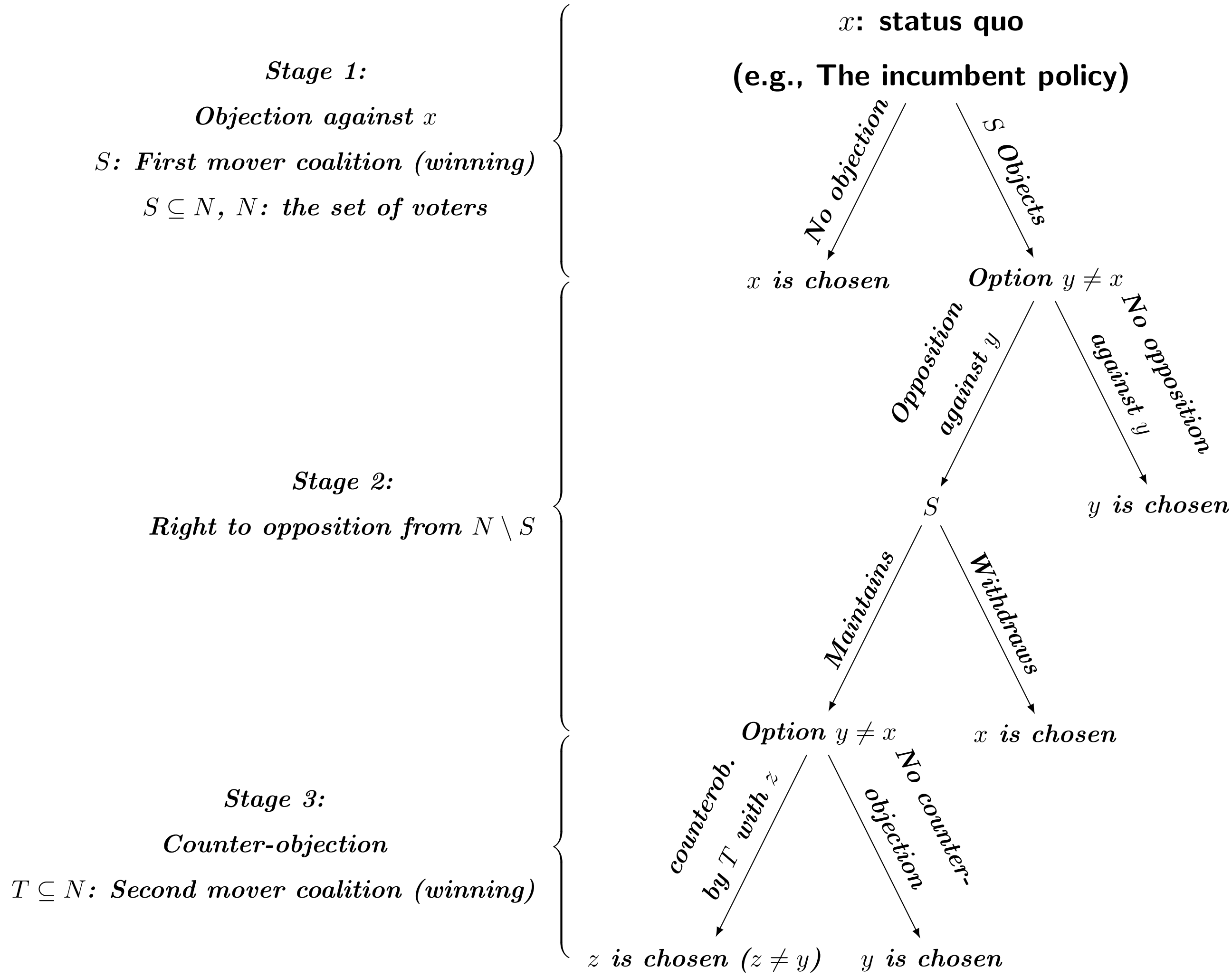


Figure 1: Reciprocity Mechanism

Unlike this traditional procedure which can carry on indefinitely, our mechanism comprises only three stages which we describe below and represent in Figure 1:

- During **stage 1**, a bill y is introduced to replace a status quo x. If this introduction is supported by a winning coalition S, y is submitted with the possibility of an amendment.[4]

- During **stage 2**, each individual who is not part of the sponsoring coalition S has the right to oppose the proposed policy y. If there is no opposition, y is adopted as the new policy and the process ends. If there is any opposition, the sponsoring coalition S is given the opportunity to withdraw the bill y. If S chooses to withdraw y, the status quo policy x remains in place and the process ends. But if S chooses to maintain y, the process moves to the final stage.

- During **stage 3**, the opposition has the right to propose an amendment z. If it does not do this, y becomes the new policy and the process ends. If it proposes an amendment z, a voting contest is organized between the proposed bill y and z; if another winning coalition T supports z against y, z becomes the new policy and the process ends; otherwise y is the new policy and the process ends.

We designate this political process a *reciprocity mechanism* because we show that it prevents free-riding by inducing rational (or farsighted) voters to act in a reciprocal manner towards one another. In fact, it provides an incentive for any second mover T to not oppose (or free-ride on) any action initiated by a first mover S that does not hurt any member of T. Similarly, it discourages any first mover S from initiating an action that will harm the interests of any individual who can successfully retaliate through an unfriendly amendment (Theorem 1).[5]

[4]A winning coalition S has the power to change the status quo to a new policy regardless of the preferences of the voters outside of S, whereas a losing coalition does not have this veto power. A formal definition of this concept is provided in Section 3.1.

[5]Importantly, this notion of reciprocity is *conceptually* unrelated to that in Rabin (1993), Dufwenberg & Kirchsteiger (2004), and Hahn (2009), among others. These studies model reciprocity through *preferences*, whereas reciprocity is induced by a *formal mechanism* or a voting protocol in our setting. We see this as a contribution to the literature. Section 2 provides more details on this distinction.

We formalize the rational behavior of agents under the reciprocity mechanism mainly following the tradition of the *blocking approach*. Ray & Vohra (2014) distinguish between two principal approaches to formalizing rational behavior in the game theory literature, namely the *blocking approach* and the *bargaining approach*. Defining these approaches, Dutta & 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 & Vohra (2017, p.1192).[6]

We remark that, notwithstanding its sequential nature, the reciprocity mechanism is not an extensive form game of perfect information. One reason for this is that there is no predetermined order in which players or coalitions move. Furthermore, the mechanism is silent with regard to the ways in which coalitions form. For this reason, the subgame perfect Nash equilibrium is not an appropriate solution concept for the game illustrated by Figure 1. In fact, the subgame perfect equilibrium cannot be used to solve several classes of sequential games, especially

[6]The blocking approach goes back to von Neumann & Morgenstern (1944) and Gillies (1959) and has been followed in a large number of studies (see, e.g., Harsanyi (1974), Chwe (1994), Ray et al. (2007), Ray & Vohra (2014), Ray & Vohra (2015), Dutta & Vohra (2017)). Even the Nash equilibrium and hybrid solution concepts such as pairwise stability (Jackson & Wolinsky (1996), Pongou & 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 bargaining approach can be traced back to Ståhl (1977), Rubinstein (1982), and Baron & Ferejohn (1989). The two approaches are not mutually exclusive, and can in fact be viewed as complementary (Ray & Vohra (2014), Kimya (2015)). Both approaches have been used to study individual behavior in political games and have sometimes been compared to each other (see, for instance, Banks & Duggan (2000) who compare the set of stationary equilibria in a bargaining model of collective choice to the core outcomes).

those that admit cycles (see, e.g., the political game depicted by Figure 10 in the appendix), unless additional assumptions are imposed. Indeed, if we impose some restrictive assumptions on our environment, we might be able to use the subgame perfect equilibrium as a solution concept.

We define a *stable* or *equilibrium policy* as a policy such that, if it is proposed as the status quo, no winning coalition will have an incentive to deviate from it.[7] The set of stable policies is called the *reciprocity set*. Any policy in the *reciprocity set* is called a *reciprocity equilibrium*.[8]

In Section 4, we examine the existence of stable policies under the reciprocity mechanism. The analysis distinguishes between discrete and continuous policy spaces. We find that a stable policy always exists if agents have strict preferences (Theorem 2 and Theorem 3). We also find that if preferences are single-peaked over a left-right political spectrum, and if the voting rule 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

[7]Our equilibrium concept is defined like the Nash equilibrium, with the difference that the latter concept is defined for one-shot games and it does not allow players to form coalitions. In the spirit of non-cooperative game theory, however, under the reciprocity mechanism, participation to a coalition is strictly an individual decision and there is no agreement that binds the members of a coalition. This implies that any individual participating to a coalition that forms to make a move is free to withdraw from that coalition or to join another coalition after the move is completed. In this paper, we introduce the *reciprocity equilibrium* concept in political games. However, all our results are robust to other environments, including non-cooperative games, transferable-utility games, or network games. Proofs are available upon request.

[8]In a similar spirit as the subgame perfect Nash equilibrium, we also define the *predicted reciprocity set*, which is a solution concept that predicts the outcome of a game taking place under the reciprocity mechanism, given a status quo. However, the *reciprocity set* is more robust than the *predicted reciprocity set* in the sense that a policy that belongs to the *reciprocity set* is the unique element of the *predicted reciprocity set* if it is proposed as the status quo, whereas a policy that belongs to the *predicted reciprocity set* (for a given status quo) does not necessarily belong to the *reciprocity set*. In other words, an element of the *reciprocity set* can be sustained as a steady state of an appropriately defined dynamic game, which is not the case for most elements of the *predicted reciprocity set*. Some examples given in Section 3.3 illustrate this point.

median voter's ideal point (Proposition 1). This finding generalizes the Median Voter Theorem to our sequential framework. Importantly, it does not follow from the Median Voter Theorem (Downs (1957)), as the reciprocity set is in general larger than the solution concept for which the original theorem is obtained.

Section 5 analyzes the welfare implications of stable policies. The key result is that all stable policies are efficient regardless of the nature of the policy space (Theorem 4). This finding implies that the reciprocity mechanism, in addition to strategically preserving minority interests (Proposition 2 in Section 6), resolves the well-known conflict between individual rationality (the pursuit of self-interest) and efficiency in political decision making. It therefore prevents political failure regardless of the degree to which political opinions are antagonistic.

In Section 7, we compare the reciprocity mechanism to other well-known mechanisms studied in the literature, namely the one-shot mechanism, a modified version of the reciprocity mechanism, and the infinite-horizon mechanism. To facilitate the comparison, we only consider the solution concepts that follow the tradition of the *blocking approach*. We find that the reciprocity mechanism has several comparative advantages. For example, as mentioned above, an equilibrium might not exist under the one-shot game, and the infinite-horizon mechanism might select an inefficient policy. These two pitfalls are avoided by our mechanism under mild conditions. Also, although our goal is not to provide a systematic comparison of the different mechanisms 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 mechanisms.[9]

In Section 8, we propose a refinement of the reciprocity set called the *sophisticated reciprocity set*. The reciprocity set assumes that agents are prudent or

[9] We would like 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 concept, which is beyond the scope of this paper. In fact, the other solution concepts we study were not developed to solve the problem the reciprocity set aims to address, which explains the main differences in their properties. Each solution concept indeed has its own merits, and captures rationality under a distinct set of assumptions.

ambiguity-averse in the face of uncertainty regarding the future of the game. When such uncertainty does not exist, agents can make moves that are more strategic. This consideration leads to our proposed refinement. Interestingly, this refinement preserves all the nice properties of the reciprocity set (Theorem 5). The next section situates our work in the literature.

2 Related Literature

This study contributes to four different literatures, namely, the literature on reciprocity, the literature on institutional design, the literature on game theory and bargaining, and the literature on equilibrium concepts for political games.

Preference-based *vs.* institution-induced reciprocity. Preference-based reciprocity has been extensively studied in the literature. Rabin (1993) argues that "fairness" is the main reason for returning favor for favor and harm for harm. He studies the existence of fairness equilibria in a two-person psychological game. In the latter, Rabin assumes that each individual's subjective expected utility depends on his or her strategy, his or her beliefs about the other individual's strategy choice, and his or her beliefs about the other individual's beliefs about his or her strategy. Several aspects of this paper have been generalized or tested in laboratory experiments (e.g., Falk & Fischbacher (2006), Bruni et al. (2008)). Experimental studies show that reciprocity promotes cooperation and increases social welfare (e.g., Komorita et al. (1991), Komorita et al. (1992), Fehr & Gächter (2000), Cox (2001)). Based on an experimental investment game, Berg et al. (1995) argue that reciprocity and trust are a primitive of human behavior.

Our paper differs from the aforementioned studies *conceptually* and in terms of its scope and primitive assumptions. Reciprocity in our setting is not *preference-based* as in this literature, it is *mechanism-based*. In fact, we do not assume that agents are altruistic or have other-regarding preferences. Instead, we propose a *formal mechanism* that induces "selfish" agents to display reciprocal and pro-social behavior (see Section 3.2.3 and Theorem 1.) In other words, although agents may only care about themselves, they find it rational to care about others and to take pro-social actions under the reciprocity mechanism. This feature indeed distinguishes our research from the extant theoretical literature on reciprocity which assumes that individuals have other-regarding preferences including conditional altruism (e.g., Rabin (1993), Dufwenberg & Kirchsteiger (2004), Hahn (2009)). In this literature, reciprocity emerges out of other-regarding preferences, whereas in our paper, it is the *reciprocity mechanism* that provides an incentive for taking reciprocal actions. This does not mean that preferences are only able to express selfishness in our model. In fact, preferences can be defined to incorporate other-regarding preferences and fairness considerations (see Example 2 in Section 3.3). However, this is not necessary for agents to display reciprocal behavior under the reciprocity mechanism. We view this as a contribution. Indeed, the reciprocity set is, to our knowledge, the first solution concept in the tradition of the blocking approach that captures "rational" reciprocal behavior.

Institutional design and reciprocity. Our study also contributes to the literature analyzing the various types of mechanisms that induce people to exhibit reciprocal behavior. In experimental studies, it has been shown that the sanction of non-cooperative behavior is a mechanism capable of inducing cooperation in public goods games (Fehr & Schmidt (1999), Fehr & Gächter (2000), Bochet et al. (2006)). Another popular mechanism that has been found to induce reciprocity and sustain cooperation is the repetition of a game (Aumann & Sorin (1989)). Bruni et al. (2008) argue that, in interactive situations where cooperation is the best long-run strategy and where defection and opportunism constitute the best one-shot strategy, reciprocity can emerge and sustain cooperation if the interactions are repeated indefinitely. The analysis described in this paper shares with this literature the assumption that individuals only care about themselves. Where this paper differs from the related literature, is in the fact that the mechanism presented in Figure 1 is quite static and therefore does not require that a voting game be repeated indefinitely in order to induce reciprocity and cooperation among decision-makers. The reciprocity mechanism is more appropriate in a legislative context. It has several merits, including the fact that it minimizes the time-cost of decision-making and guarantees the existence of stable, efficient, and inclusive outcomes.

Game theory and bargaining. The reciprocity mechanism shares some common elements with certain bargaining frameworks available in the literature (see, e.g., von Neumann & Morgenstern (1944), Aumann & Maschler (1964), Rubinstein (1980), Baron & Ferejohn (1989), Zhou (1994), Holzman et al. (2007), Pongou et al. (2008), and Eguia & Shepsle (2015)). Within these frameworks, negotiation often takes the form of a sequence of threats and counter-threats. All individuals can bargain together, with full knowledge of the process, and can settle upon a stable option that is based on the objections and counter-objections. We follow a similar approach, though the problem that we solve is completely different. We introduce a new solution concept, namely the *reciprocity set*, which identifies stable social policies that can emerge when people have possibly conflicting views on how society should be run. We compare this equilibrium concept with other well-known concepts. Like the core (Gillies (1959)) and the strong Nash equilibrium (Aumann (1959)), it only selects efficient policies. However, unlike these solution concepts, it cannot be empty when preferences are strict. The core and the (strong) Nash equilibrium are most suitable as solution concepts for one-shot games. If the game involves two or more stages, as is the case for several real-life games, they become inappropriate to capture rational behavior (see, e.g., Mas-Colell (1989), and the references therein). There exist other solution concepts defined for games which might go on indefinitely.[10] We provide a comparative analysis highlighting the strengths of the *reciprocity set* and those of the farsighted stable set (Harsanyi (1974))[11], and the largest consistent set (Chwe (1994)). The results indicate that an inefficient policy can persist indefinitely if individuals' rational behavior is captured by these concepts. This is explained by the fact that these concepts were not designed to solve the issues that we are addressing in this paper.

[10] See, e.g., Kimya (2015) and Fotso et al. (2017) for a review of some of these concepts.

[11] See Ray & Vohra (2015) for a modification of this solution concept.

Equilibrium Concepts for Political Games. Our paper contributes to the literature on solution concepts for political games. Most of these concepts were developed to address some limitations of the core. Farquharson (1969) introduces the concept of sophisticated voting, which uses backward induction in a multistage non-cooperative voting game in which voters are assumed to successively eliminate *majority-preferred* strategies. Schwartz (1976) defines the top cycle set and Miller (1980) introduces the uncovered set. These latter solution concepts are particularly useful for tournaments and majority voting in the absence of the core. These pioneering studies have generated a fascinating literature on the formalization of rational behavior in political games (see, e.g., McKelvey & Niemi (1978), Shepsle & Weingast (1984), McKelvey (1986), among others; see also Austen-Smith & Banks (1999, 2005) for a thorough survey and exposition of solution concepts in political games). This literature complements a large body of studies in sequential voting games (see, for instance, Banks (1985), Austen-Smith (1987) or Ordeshook & Schwartz (1987), Banks & Duggan (2000), and surveys like Miller (2013), and a more recent work by Barberà & Gerber (2017)).

The reciprocity set has a very different motivation and incentives structure. It is intended to solve a different problem than the aforementioned solution concepts. It is primarily meant to foster reciprocity among decision-makers facing a political decision. We show that the reciprocity mechanism encourages positive reciprocity and prevents actions that might trigger retaliation (see Section 3.2.3). In Section 7, we compare the reciprocity set to some of the aforementioned solution concepts and find that they are very different indeed. 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, efficiency, and the strategic protection of minority interests.[12] Obviously, as already noted, these differences are mainly justified by the fact that we are addressing a new question, generating new insights into the design

[12]Table 1 in Section 7 summarizes the analysis comparing the reciprocity set to the other solutions concepts reviewed in this paper.

of mechanisms or voting procedures that foster reciprocal actions.

3 A Reciprocity Mechanism in a Political Economy

A *political economy* is a society $N = \{1, 2, ..., n\}$, populated by a finite number of individuals, and endowed with a *constitution* $\mathcal{C}$ and a profile of individual preferences $(\succeq_i)_{i \in N}$ over a policy space A. It is assumed that each preference relation $\succeq_i$ is reflexive, transitive, and complete. The structure of the policy space depends on the nature of the decision to be made. Certain decisions are discrete in nature (e.g., the choice of a national language), whereas others are continuous (e.g., tax rates). Depending on the nature of the policy space, additional structures will be imposed on preferences, such as continuity in the case of a continuous policy space. The *constitution* and the reciprocity mechanism are formally described below.

3.1 The Constitution

A *constitution* is a distribution of decision-making power within the different subgroups or coalitions of the society. It is formalized as a function $\mathcal{C}$ that maps each subgroup (or coalition) S of the society into either 1 or 0; $\mathcal{C}(S) = 1$ means that S is a "winning coalition", and $\mathcal{C}(S) = 0$ means that S is a losing coalition. A winning coalition S has the power to change the status quo policy to a new alternative regardless of the preferences of agents outside of S, whereas a losing coalition does not have such a veto right. A simple example of a constitution is the majority rule, which is a rule under which a coalition of agents is winning if and only if it contains more than half of the population.

Throughout this paper, we denote by $\mathfrak{C}$ the set of winning coalitions. We assume that this set contains at least one winning coalition (i.e., $\mathfrak{C} \neq \{\emptyset\}$), and that the empty set cannot be a winning coalition (i.e $\emptyset \notin \mathfrak{C}$). In addition, we impose the following conditions on $\mathfrak{C}$: (1) for any non-empty coalitions S and T such that $S \subset T$, if $S \in \mathfrak{C}$, then $T \in \mathfrak{C}$; (2) for any coalition S, if $S \in \mathfrak{C}$, then $N \backslash S \notin \mathfrak{C}$. The first condition means that the addition of new voters to a winning coalition yields another winning coalition. The second condition means that the complementary set of a winning coalition is a losing coalition. This condition prevents trivial political deadlock by avoiding situations in which two non-overlapping winning coalitions have completely opposing positions on which policy should be adopted.

3.2 The Reciprocity Mechanism

This section describes the *reciprocity mechanism* and highlights some of its features. The description is presented in Section 3.2.1. Section 3.2.2 formalizes the rational behavior of agents under this mechanism, and introduces the *reciprocity set*, which is the set of *stable* policies. In Section 3.2.3, we show that the reciprocity mechanism provides incentives for rational agents to take *reciprocal actions*; this indeed justifies the designation of this mechanism as a *reciprocity mechanism*. Section 3.2.4 introduces the *predicted reciprocity set*, which is a solution concept aimed to predict the outcome of a political game from a given status quo, when agents adopt the rational behavior induced by the reciprocity mechanism. It is similar in spirit to the notion of subgame perfect equilibrium.

3.2.1 Description of the Reciprocity Mechanism

Consider reforming a given status quo $x \in A$ by means of a deliberative and voting mechanism comprised of the three stages described hereunder:

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

Stage 2 (**Right to opposition**): Each individual 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, 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 and the process ends. However, if S maintains y, then the opposing individuals have a right to formulate a counter-objection in the third and final stage.

Stage 3 (**Counter-objection**): Suppose that S maintains the bill y. Each opposing individual has the right to invite other individuals 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 counter-objection to the objection (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 is elected as a new policy and the process ends.

Interestingly, in order for the opposition to succeed in the final stage of the above-stated mechanism, it has to form a winning coalition T that supports the amendment z against the initial bill y, even by recruiting *opportunists* within the coalition S that initiated the move from x to y. This means that there is no agreement that binds the members of the sponsoring coalition S. Furthermore, if an opposition arises at stage 2, the individuals 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. An individual joins a coalition in order to achieve his or her personal interests. In addition, the mechanism 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. Legislators have full information about existing alternatives and individual 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.

The reciprocity mechanism shares some features of other models proposed in the literature (see, e.g., Aumann & Maschler (1964)). In these models, negotiation may take the form of a finite or infinite sequence of objections and counter-objections. Our mechanism differs in that it offers sponsoring coalitions the possibility to withdraw a bill. Arguably, its design also minimizes the time cost of decision-making and therefore is appropriate when decisions are subject to a deadline. The mechanism also has three other interesting properties, as is shown later in this study. First, it encourages reciprocal actions among selfish individuals. Second, it always leads to an equilibrium policy under natural assumptions on individuals' preferences. Third, it selects only efficient policies at equilibrium.

3.2.2 The Reciprocity Set

The formalization of the rational behavior of agents under the reciprocity mechanism answers the question of when a winning coalition will choose to formulate an objection against a status quo. This formalization takes into account the fact that agents are necessarily farsighted, given the sequential nature of our mechanism.

Let $\mathcal{P} = (N, A, (\succeq_i)_{i \in N}, \mathcal{C})$ be a political economy, $x \in A$ be a status quo, and $S \in \mathfrak{C}$ be a winning coalition that wishes to replace x with another policy y. The two following conditions should be satisfied:

1. Each member of S should prefer y over x (i.e., $y \succ_S x$; 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., $z \succ_T y$), if some member of T prefers a status quo x over the proposed policy y (i.e., $not(y \succeq_T x)$), then each member of S should weakly prefer z over x (i.e., $z \succeq_S x$). More formally, this condition is expressed as follows:

   $[\forall (z, T) \in A \times \mathfrak{C},\ S \neq T,\ z \succ_T y \text{ and } not(y \succeq_T x)] \text{ implies } [z \succeq_S x].$

The first condition is a natural requirement that appears in all two-stage bargaining models of rationality in the literature (see, e.g., Aumann & Maschler (1964) 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 cannot fully predict the future of the game after introducing the bill y. In fact, the game might end at y for several practical reasons. One reason is that 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 that policy z is not an inferior option for S than x, it is not clear that T will be formed. It is generally argued that many exogenous factors that are outside the control of individuals interfere with decision-making in real-life politics, and that such factors can prevent a coalition from forming. 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.[13]

The second condition says that, if, following an objection (y, S) against the status quo x, a winning coalition T formulates a counter-objection or an amendment (z, T), then S should weakly prefer z over x provided that certain members of T strictly prefer x over y. In fact, it is easy to see that if all of the members of T prefer y over x (i.e., $y \succeq_T x$) 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 counter-objection (z, T) against (y, S). Indeed, if T opposes y, then S will withdraw its 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 counter-objection (z, T) emerges if and only if T prefers z over y (i.e., $z \succ_T y$) and there is at least one member of T who prefers x over y

[13] 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. (2019) 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." (p. 2). See also Jackson & Wolinsky (1996) for similar arguments in the context of network formation. In Section 8, we propose a refinement of the reciprocity set that captures strategic behavior in a context where the kind of uncertainty that motivates the current definition does not exist. We call it the *sophisticated reciprocity set*, and we show that it preserves all the properties of the reciprocity set.

(i.e., *not*$(y \succeq_T x)$).

The definition of the *reciprocity set* is provided below. It is the set of all the *stable* or *equilibrium* policies, with a *stable* policy being a policy that cannot be replaced if it is a status quo under the reciprocity mechanism.

Definition 1. Let $\mathcal{P} = (N, A, (\succeq_i)_{i \in N}, \mathcal{C})$ be a political economy, S be a winning coalition, and $x, y \in A$ be two policies.

1. **Objection**: (y, S) is said to be an objection against x if $y \succ_S x$.

2. **Counter-objection**: let (y, S) be an objection against x. A pair $(z, T) \in A \times \mathfrak{C}$ is said to be a counter-objection against (y, S) if $z \succ_T y$ and *not*$(y \succeq_T x)$.

3. **Unfriendly counter-objection**: let (y, S) be an objection against x and $(z, T) \in A \times \mathfrak{C}$ be a counter-objection against (y, S). The counter-objection (z, T) is said to be unfriendly if *not*$(z \succeq_S x)$.[14]

4. **Justified objection**: an 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 objection against x, this is denoted by $y \gtrdot_S x$.

5. **Reciprocity set**: the reciprocity set of $\mathcal{P}$, denoted $R(\mathcal{P})$, is the set of all the policies in A against which no justified objection exists. Formally, $R(\mathcal{P}) = \{x \in A : \text{ there does not exist } (y, S) \in A \times \mathfrak{C} \text{ such that } y \gtrdot_S x\}$. Any policy in the reciprocity set is called a *reciprocity equilibrium* or a *stable policy* under the reciprocity mechanism.[15]

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

[15]Note that, in our framework, objections and counter-objections are formulated in terms of strict domination ($y \succ_S x$ if $y \succ_i x$ for each $i \in S$). All of the results of this paper continue to hold if we use weak domination ($y \succ_S x$ if $y \succeq_i x$ for each $i \in S$ and $y \succ_j x$ for some $j \in S$) instead, though the reciprocity set could be refined.

Under the reciprocity mechanism, all actors bargain and settle on a reciprocity equilibrium that emerges from the sequence of threats and counter-threats. The stability of an outcome, in general, reflects the power or influence of different population coalitions as well as the preferences of individuals. In our context, this power is provided by the constitution $\mathcal{C}$. The reciprocity set derives its name from the fact that *reciprocal actions*—returning favor for favor and harm for harm–are rational under our mechanism, as we show in the following section.

3.2.3 Incentives for Reciprocal Actions under the Reciprocity Mechanism

In this Section, we show that the reciprocity mechanism provides incentives for agents to take *reciprocal actions*. It 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 first-mover coalition S (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 (item (3) of Theorem 1).

Theorem 1. ***(Reciprocal actions are rational)*** *Let $\mathcal{P} = (N, A, (\succeq_i)_{i\in N}, \mathcal{C})$ be a political economy. For all winning coalitions $S, T \in \mathfrak{C}$ and distinct policies*

$x, y, z \in A$ such that $y \succ_S x$ and $z \succ_T y$, the assertions described hereunder are true:

1. *If not($y \succeq_T x$), then S will not formulate any objection (y, S) against x, unless $z \succeq_S x$. If not($y \succeq_T x$) and not($z \succeq_S x$), then S will not formulate an objection (y, S) against option x.*

2. *Assume that S formulates an objection (y, S) against x. If $y \succeq_T x$, then T will not formulate an unfriendly counter-objection (z, T) against (y, S) following an opposition.*

3. *As a corollary of 1. and 2., if $y \succeq_T x$, then S will formulate an objection (y, S) against x, even if not($z \succeq_S x$).*

It follows that the reciprocity mechanism 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 under our mechanism are induced by the possibility for a first-mover coalition to withdraw its move if an opposition arises (stage 2 of the reciprocity mechanism) and by the possibility for any second-mover coalition to retaliate against any first move that is harmful. This indeed justifies the proposed name for the voting process—*the reciprocity mechanism*—.

3.2.4 The Predicted Reciprocity Set

If a policy, say x, is the status quo, which policies will possibly be enforced at the end of the voting process? We introduce a solution concept—the *predicted reciprocity set*—that answers this question. This is the set of the policies that can be selected from a given status quo. This concept is intuitively similar to that of the subgame perfect equilibrium.

Clearly, if x is a reciprocity equilibrium (i.e., if $x \in R(\mathcal{P})$), then x will be enforced. But if x is not a reciprocity equilibrium, then it follows that there exist a justified objection (y, S) against x (i.e., $y \gtrdot_S x$). If there is no opposition or counter-objection against y, y will be chosen and thus is a possible predicted outcome at x. If there is any opposition and, in particular, if there is a counter-objection against y that may lead 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 objection against x, the latter counter-objection is not unfriendly to the coalition supporting the objection (i.e., S). The precise description of the *predicted reciprocity set* is provided in the next definition.

Definition 2. Let $\mathcal{P} = (N, A, (\succeq_i)_{i \in N}, \mathcal{C})$ be a political economy and $x \in A$ be a status quo. The *predicted reciprocity set* at x is defined as:

$P(x) = \{x\}$ if $x \in R(\mathcal{P})$; otherwise

$P(x) = \{y \in A :$ there exist $S \in \mathfrak{C},\ y \gtrdot_S x$ and there does not exist $z \in A,\ z \succ y\}$,
or

$P(x) = \{z \in A :$ there exist $S \in \mathfrak{C}$ and $y \in A,\ y \gtrdot_S x$ and $z \succ y\}$.

From this definition, as is stated above, each stable outcome is its own prediction, but there might exist predicted outcomes that are not stable.[16] The analysis indicates that the predicted reciprocity set from any given status quo policy is non-empty because the voting process is a finite sequential game. The next section provides some examples that illustrate the reciprocity mechanism and the key notions that have been defined thus far.

3.3 Examples

This section provides some examples to illustrate the reciprocity mechanism. In Appendix A, we provide more details on how we solve them. The first example set out below elucidates the difference between the reciprocity set and the predicted reciprocity set.

[16]The reciprocity set can be viewed as selecting *robust* options, as is argued in the Introduction. The robustness here is defined by the fact that such alternatives will persist in the economy for a long period of time.

Example 1. (**Elected vs. stable outcomes**)

Let $\mathcal{P} = (N, A, (\succeq_i)_{i \in N}, \mathcal{C})$ be a political economy where $N = \{1, 2, 3, 4, 5, 6\}$, $A = \{a, b, c, d\}$, and $\mathcal{C}$ is the majority rule. Preferences over the set A are given as follows: $b \succ_1 d \succ_1 c \succ_1 a$; $d \succ_2 b \succ_2 c \succ_2 a$; $c \succ_3 b \succ_3 d \succ_3 a$; $c \succ_4 b \succ_4 a \succ_4 d$; $a \succ_5 d \succ_5 c \succ_5 b$; and $a \succ_6 d \succ_6 c \succ_6 b$. 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 2 (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 reciprocity set is $R(\mathcal{P}) = \{b, d\}$.

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

Figure 2: Elected vs. stable outcomes

The predicted reciprocity sets are: $P(a) = \{c\}$, $P(b) = \{b\}$, $P(c) = \{d\}$, and $P(d) = \{d\}$. If the policy a is the status quo at the beginning of the voting process, then the policy c will be elected, since (c, S) is a justifed 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. If b is the status quo, it will be elected. If d is the status quo, it will also be elected. The alternatives b and d are the only stable policies in this economy. Plainly, an elected outcome is not necessarily a reciprocity equilibrium.

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.

Example 2. (Ultimatum game with counter-offer and other-regarding preferences)

A population of n individuals must share an amount of 100 dollars, with each individual receiving a non-negative portion. A feasible allocation is a vector:

$$x = (x_1, x_2, ..., x_n) \text{ such that } \ x_i \in [0, 100] \text{ and } \sum_{i=1}^{n} x_i \leq 100.$$

The set of feasible allocations is the simplex:

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

Only groups comprising more than half of the population have the right to formulate an objection or a counter-objection.

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 an objection $(y, S) \in A \times \mathfrak{C}$ against x_0. If there is no opposition, y is implemented. If there is any opposition, S may withdraw y, leading to each individual receiving zero. But if S chooses to maintain y following any opposition, another majority coalition T might react by proposing a counter-objection (z, T) against (y, S). In this case, z is implemented and each individual i obtains and consumes his or her payoff z_i.

The traditional ultimatum game is played between two individuals (a proposer and a responder) and, unlike the reciprocity mechanism, it does not allow for further negotiation after the first move. An interesting question is whether a reciprocity equilibrium exists. Assume that each individual i's utility function is strictly increasing in his or her payoff x_i, which is a natural assumption. Then, in this framework, a equilibrium always exists, though it might not be unique. Furthermore, if individuals have other-regarding preferences (with each individual i's utility depending on both his or her payoff x_i and other individuals' 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 efficient. To illustrate, suppose that there are three individuals who have the following linear utility functions:

$$u_1(x_1, x_2, x_3) = x_1 - \frac{1}{3}(x_2 - x_1) - \frac{1}{3}(x_3 - x_1),$$

$$u_2(x_1, x_2, x_3) = x_2 - \frac{1}{3}(x_1 - x_2) - \frac{1}{3}(x_3 - x_2),$$
$$u_3(x_1, x_2, x_3) = x_3 - \frac{1}{3}(x_1 - x_3) - \frac{1}{3}(x_2 - x_3).$$

Interpreting these utility functions, individuals have fairness considerations or other-regarding preferences in that fixing the payoff of an individual i and increasing the payoff of another individual have the effect of decreasing i's utility. The analysis proves that (see more details in Appendix A) the set of predicted allocations from the status quo x_0 is:

$$P(0,0,0) = \left\{ (x_1, x_2, x_3) : x_1 + x_2 + x_3 = 100 \text{ and } x_i \geq \frac{100}{6};\ i = 1, 2, 3 \right\},$$

and the reciprocity set is:

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

Note that in this case the reciprocity set strictly includes the predicted reciprocity set at the status quo $x_0 = (0, 0, 0)$.

The next example applies to the selection of a fiscal policy.

Example 3. (**Fiscal policies**)
Consider the problem of designing a taxation policy. Assume that the society consists of n individuals. There are also k categories. These categories might be defined by gender, age group, occupation, income group, marital status, ethnicity, or any combination of these factors. A tax policy is a vector $t = (t_1, ..., t_k) \in [0, 1]^k$, which means that a policy might affect each category differently (e.g., female hygiene products might be taxed less or more than male hygiene products).

The goal is to select a tax policy under the reciprocity mechanism. Each member of society typically belongs to several of these categories and might also be concerned about social justice. This implies that the individual utility derived from a tax policy $t = (t_1, ..., t_k)$ would depend on some or all of the components of t. We show that a reciprocity equilibrium tax policy exists for a wide class of preferences, and that all reciprocity equilibria are efficient.

The determination of all of the stable outcomes under the reciprocity mechanism is possible in certain contexts (e.g., see Examples 1 and 2). However, in certain economies, such as the fiscal policy game in Example 3, the process can be very complicated. The question then arises as to whether it is feasible to prove the existence and analyze the properties of stable policies under very general conditions. The next section provides a positive answer to this question. It establishes more general results on the existence of equilibrium policies.

4 Existence of Stable Policies under the Reciprocity Mechanism

The results of this section reveal that stable policies always exist under very natural assumptions regarding on individual preferences. Interestingly, when preferences have particular known structures, there might exist only one or two stable policies. 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 stable policies under this assumption. Generally, the existence of a stable policy depends on three exogenous factors: (1) the structure of preferences; (2) the nature of the political space; and (3) the size of the voting population.

4.1 Discrete Policies

This section deals with the existence of stable policies when individuals have strict or linear preferences over a finite set of policies. These are reflexive, antisymmetric, transitive, and complete binary relations over the policy space. The set of such preferences is denoted by $\mathcal{L}$. Preference linearity simply means that individuals have a strict ordering of all of the available policies. In this specific domain of preferences, a reciprocity equilibrium always exists.

Theorem 2. *Let $\mathcal{P} = (N, A, (\succeq_i), \mathcal{C})$ be a political economy where A is a finite policy space ($|A| < \infty$) and where preferences are strict. Then, a reciprocity equilibrium exists.*

Theorem 2 is illustrated in Example 5 on the preservation of minority interests in a multi-ethnic society. A multiplicity of reciprocity equilibria can emerge in this environment, as is also illustrated in this example. The empirical implications of this finding are that different policies might prevail in structurally identical economies. Indeed, there might not exist any reasonable explanation for why a minority language prevails as the official language in certain countries, but not in others.

4.2 Continuous Policies

We assume that the policy space A of a political economy $\mathcal{P} = (N, A, (\succeq_i), \mathcal{C})$ 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 assume that the policy space A is endowed with the topology of closed convergence (e.g., Hildenbrand (1974)). Similar assumptions have proven useful in the analysis of the existence of equilibria in continuous economies (e.g., Arrow & Debreu (1954)). We also denote by m the Lebesgue measure on the affine manifold spanned by policy space A. Given the nature of the policy space, there is a need to place additional structure on preferences. In particular, we assume that preferences are continuous. Preference continuity means that an individual 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.

We analyze the existence of a stable policy under strict individual preferences over a space of continuous policies. The theorem set out below reveals that, under the reciprocity mechanism, a stable policy always exists.

Theorem 3. *Let $\mathcal{P} = (N, A, (\succeq_i), \mathcal{C})$ be a political economy such that individual preferences are continuous, strict, and endowed with the topology of closed conver-*

gence. Then, a reciprocity equilibrium exists.

In the next section, we present an application to a classical and widely studied framework.

4.3 Left-Right Political Spectrum: Median Voter Theorem Under the Reciprocity Mechanism

We examine the predictions of the reciprocity mechanism when the policy space can be represented by a left-right political spectrum (e.g., Castles & Mair (1984), Giddens (1994), Bobbio (1996), Evans et al. (1996)), individual preferences are single-peaked, and the constitution is the majority rule.[17] 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 e.g., Black (1948), Inada (1964), Grandmont (1978), Moulin (1980), Sprumont (1991), Thomson (1997), Ehlers et al. (2002), 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 of the policies are ordered by a binary relation denoted $\gg$, and that all individuals perceive these policies as being arranged in this order. An individual i has a single-peaked preference $\succeq_i$ if there exists a policy a_i such that: (1) for any other policy $a \neq a_i$, $a_i \succ_i a$; and (2) for any policy $a, b \in A$, if $[a \gg b \gg a_i$ or $a_i \gg b \gg a]$, then $b \succ_i a$. We will assume that the set of policies is the set of the ideal points of individuals.

Our main result is that there exists at least one and at most two stable policies when individuals having single-peaked preferences. In addition, when the number of individuals is odd, there is only one stable policy, and this policy coincides with

[17] The majority rule has been shown to have very appealing properties (see, Dasgupta & Maskin (2008) and the references therein).

the optimal policy of the median voter.

Proposition 1. *Let $\mathcal{P} = (N, A, (\succeq_i), \mathcal{C})$ be a political economy in which preferences are single-peaked over A, and $\mathcal{C}$ is the 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.*

Proposition 1 is a generalization of the median voter theorem to the reciprocity mechanism. Indeed, when the number of individuals is odd, the unique equilibrium that exists is the ideal policy of the median voter. When the number of individuals is even, a median individual may not exist, and in this case, there are two reciprocity equilibria that are very close to the center of the political spectrum.[18] The example set out below on the admission of refugees illustrates these two situations.

Example 4. (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 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' utility functions. The net utility received by each legislator i from x refugees being admitted is:

$$v_i(x) = u_i(x) - s_i x,$$

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. Under

[18] 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 1 shows that the reciprocity set coincides with the core when the conditions that sustain the median voter theorem are assumed.

the reciprocity mechanism, there exists at least one reciprocity equilibrium and at most two reciprocity equilibria (details of the calculations are relegated to the appendix A).

The results in this section illustrate some of the conditions under which a stable policy exists under the reciprocity mechanism when the policy space is continuous. Example 2 on the ultimatum game with counter-offer and other regarding preferences and Example 3 on the selection of a fiscal policy most effectively illustrate the usefulness of the various findings. The next section examines the efficiency of stable policies.

5 Efficiency under the Reciprocity Mechanism

We analyze the welfare implications of stable policies under the reciprocity mechanism. The result set out below shows that any reciprocity equilibrium is Pareto-efficient (or simply *efficient*) regardless of whether individual preferences are strict or weak (reflexive, transitive, and complete).

Theorem 4. *Let $\mathcal{P} = (N, A, (\succeq_i), \mathcal{C})$ be a political economy where the policy space A is either discrete or continuous, and where preferences $(\succeq_i)$ over the policy space A are weak or strict. Then, any reciprocity equilibrium is Pareto-efficient.*

This finding implies that the reciprocity mechanism proposed in this paper resolves the conflict between individual rationality and optimality in public goods provision. By inducing selfish individuals to adopt reciprocal and pro-social behavior, the mechanism prevents free-riding, which is one reason why it cannot select an inefficient policy.

6 Decisive Minority: Strategic Protection of Minority Interests Under the Reciprocity Mechanism

Discussion about the protection of minority interests has received considerable attention in the literature since the works of Madison (1788) and Mill (1861).

Madison (1788) considered the involvement of minorities in electoral debates as having a credible value for political systems that foster external checks and balance. Mill (1861) valued minority participation primarily for political education. Nowadays, minority participation is considered an immediate criterion for fairness in democratic decision making, where each citizen's vote is equally worthwhile. However, Chwe (1999) shows that under certain circumstances, majorities can use the participation of minorities as a strategic device in the enforcement of their favored outcomes in elections. At the same time, it has been observed that some social issues thought to be supported only by a minority group can receive the support of a majority of voters (see Campbell (1999)). This is known as the "decisive minority" phenomenon. Two main reasons have been provided to explain why a minority option can win a democratic election, namely preference intensity and free-riding (see Campbell (1999) for a detailed description of these arguments).

In this section, we provide an additional explanation of the "decisive minority" phenomenon. We argue 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* (see Definition 3 below) cannot be selected by the reciprocity set.

It is important to note that in our treatment of minority interests, a *minority group* is not necessarily a group with a "fixed" identity such as an ethnic or a religious group. It is simply a set of individuals who, in a particular context, favor a policy alternative that is not favored by a majority of voters. Although such a set may have a fixed identity (see Example 5 below), this is not generally the case. In this regard, our analysis does not necessarily follow Madison (1788) and Mill (1861). Under the reciprocity mechanism, protecting a minority group with a fixed identity will generally require the use of a supermajority rule, although in some cases a minority group with a fixed identity will also be protected under the majority rule (see Example 5). We are more interested in uncovering conditions under which an option favored by a non-decisive group of individuals can win the support of a majority of voters under our mechanism.[19]

Below, we formally define the notions of minority and majority alternatives or options. We also define the concept of *strategic minority protection*, which we distinguish from the notion of *mechanical minority protection*.

Definition 3. Let $\mathcal{P} = (N, A, (\succeq_i), \mathcal{C})$ be a political economy, and x be an alternative in A.

1. The alternative 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 $S \in \mathfrak{C}$: $y \succ_S x$. 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 such that $y \succ_S x$.

Definition 4. Let $\mathcal{E}$ be a solution concept.[20]

1. $\mathcal{E}$ is said to *mechanically protect minority interests* if there exist a political economy $\mathcal{P} = (N, A, (\succeq_i), \mathcal{C})$, and a minority option $x \in A$, such that $x \in \mathcal{E}(\mathcal{P})$.

2. $\mathcal{E}$ is said to *strategically protect minority interests* if $\mathcal{E}$ mechanically protects minority interests, and for any political economy $\mathcal{P}$, the set $\mathcal{E}(\mathcal{P})$ 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 2. *The reciprocity set strategically protects minority interests.*[21]

[19]The determination of the minimal and the maximal size of a non-decisive group that can be protected under the majority rule is also of interest, but this is beyond the scope of this paper.

[20]A solution concept $\mathcal{E}$ is a correspondence which maps any political $\mathcal{P} = (N, A, (\succeq_i), \mathcal{C})$ into a subset $\mathcal{E}(\mathcal{P}) \subseteq A$; $\mathcal{E}(\mathcal{P})$ is interpreted as the set of stable outcomes of $\mathcal{P}$ when agents adopt the rational behavior that defines $\mathcal{E}$.

[21]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

Remark that the strategic protection of minority interests under the reciprocity mechanism can be achieved under the majority rule. That is, an outcome preferred by a minority group and disliked by a majority group may not be replaced, if proposed as the status quo under the reciprocity mechanism with the majority rule as the constitution. 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.

In what follows, we illustrate Proposition 2 in a multiethnic 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 5. Assume that a political economy $\mathcal{P}$ consists of a country that has four ethnic groups. This country 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. Preferences over languages are presented below:

$$\textit{ethnic group } 1 : l_3 \succ_1 l_2 \succ_1 l_1;$$

$$\textit{ethnic group } 2 : l_3 \succ_2 l_2 \succ_2 l_1;$$

$$\textit{ethnic group } 3 : l_2 \succ_3 l_1 \succ_3 l_3;$$

$$\textit{ethnic group } 4 : l_1 \succ_4 l_3 \succ_4 l_2.$$

We suppose that the constitution is the majority rule. Figure 3 shows the popularity relationship among the various languages (e.g., l_3 is more popular than l_2 as the former is preferred by three out of the four ethnic groups). The analysis

the Condorcet policy is a *non-strategic minority* option and therefore cannot belong to the reciprocity set according to Proposition 2.

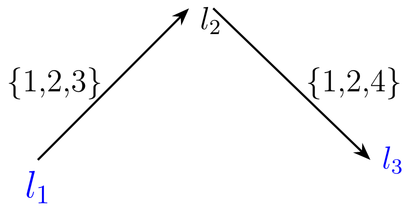


Figure 3: Popularity relationship among various languages

demonstrates that there are two stable languages under the reciprocity mechanism. 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 objection, it cannot succeed at forming a majority coalition that will formulate a counter-objection that will lead to the replacement of l_3. It follows that $l_2 \notin R(\mathcal{P})$. If the status quo is l_1, no majority coalition will formulate an 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 an objection $(l_2, \{1, 2, 3\})$ against l_1, $\{1, 2, 4\}$ will formulate an unfriendly 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 counter-objection. The latter is unfriendly because not all members of $\{1, 2, 3\}$ prefer l_3 over the status quo l_1. Therefore $l_1 \in R(\mathcal{P})$. It follows that $R(\mathcal{P}) = \{l_1, l_3\}$.

Interestingly, note 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. 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 stable 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.

7 Comparison with Other Mechanisms

We compare our mechanism to some classical mechanisms. It is important 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 concept, which is beyond the scope of this paper. Indeed, each of the solution concept 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 notice 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.

We begin this section by modifying our mechanism to remove the incentive for reciprocal actions. This alternative mechanism is formalized as a two-stage game in which no policy retraction by a sponsoring coalition is possible. We find that, under this alternative mechanism, an inefficient status quo can persist indefinitely (Proposition 3). This result occurs because, even though some individuals might be willing to remove an inefficient 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 individuals from removing a status quo, a situation that is avoided under the reciprocity mechanism. In Section 7.2, we compare the reciprocity set with some prominent solution concepts designed to capture rationality in one-shot mechanism games. 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 an inefficient outcome, while some others might fail to select a stable policy (Proposition 4). In Section 7.3, we consider infinite-horizon political mechanisms. We find that these mechanisms may lead to an inefficient policy persisting indefinitely (Proposition 6).

7.1 A Modification of the Reciprocity Mechanism

We modify the reciprocity mechanism to the mechanism M' described hereunder (also see Figure 4). Assume that a society wishes to reform a status quo $x \in A$. The decision-making mechanism 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.

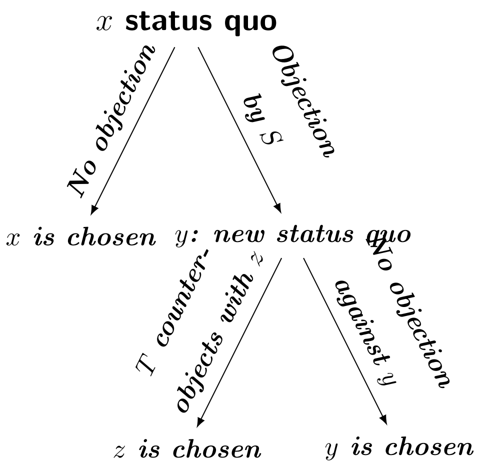


Figure 4: Non-reciprocity mechanism (M')

This mechanism differs from the reciprocity mechanism 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 mechanism 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 an inefficient policy persisting indefinitely.

We define hereunder the rational behavior of individuals under this mechanism.

Definition 5. Let $\mathcal{P} = (N, A, (\succeq_i)_{i \in N}, \mathcal{C})$ be a political economy, S be a winning coalition, and $x, y \in A$ be two policies.

1. **Objection**: (y, S) is said to be an objection against x if $y \succeq_i x$ for each $i \in S$ and $y \succ_j x$ for some $j \in S$.

2. **Counter-objection**: Let (y, S) be an objection against x. A pair $(z, T) \in A \times \mathfrak{C}$ is said to be a counter-objection against (y, S) if $z \succeq_i y$ for each $i \in T$ and $z \succ_j y$ for some $j \in T$.

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

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

5. **Non-reciprocity set**: The non-reciprocity set of the political economy $\mathcal{P}$, denoted by $NR(\mathcal{P})$, is the set of all of the policies in A against which no justified objection exists.[22]

[22] The objection and the counter-objection could be defined with strict preferences as for the reciprocity set. In this case, the set of stable outcomes would be larger than the non-reciprocity set as currently defined. Since the non-reciprocity set generally contains inefficient outcomes as we show below, this larger set also contains inefficient outcomes.

The rational behavior under the modified mechanism is defined in accordance with the same logic as under the reciprocity mechanism. The main difference resides in the conditions under which a counter-objection is formulated. Under the new mechanism, any winning coalition that wishes to deviate from a sponsored bill to a new policy can formulate a 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 enactment of the bill. In fact, if the enactment of the bill would not harm any member of T, any unfriendly counter-objection would cause the first mover S to withdraw its objection. Therefore the unfriendly 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 counter-objection if the objection does not harm its interests, whereas under the modified mechanism, this is possible. The result confirms that this absence of incentives for reciprocal actions in the modified mechanism might cause an inefficient policy to persist indefinitely.

Proposition 3. *Let $\mathcal{P} = (N, A, (\succeq_i), \mathcal{C})$ be a political economy under the non-reciprocity mechanism M'. Then, an inefficient policy might be stable.*

7.2 One-shot Games

The one-shot mechanism is a mechanism in which policy selection entails only one stage. Under this mechanism, a challenger is pitted against an incumbent policy. 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. 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 & 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 6. Let $\mathcal{P} = (N, A, (\succeq_i), \mathcal{C})$ be a political economy, V be a subset of A, and x and y be two policies.

- y is said to directly dominate x (labeled as $y >^d x$) if there exists a winning coalition S that strictly prefers y over x, that is, $y \succ_S x$. 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.
- V is a vNM stable set if it satisfies the following stability conditions:
  1. (Internal stability): no policy in V is directly defeated by another policy in V; and
  2. (External stability): every policy not in V is directly defeated by some policy in V.
- The top cycle set, $TC(\mathcal{P})$, is the smallest subset of the policy space A, where every alternative in $TC(\mathcal{P})$ defeats every alternative not in $TC(\mathcal{P})$.
- y is said to *cover* x if every alternative z that is defeated by x is also defeated by y. An uncovered alternative is an alternative such that there is no alternative that covers it. The uncovered set, $UC(\mathcal{P})$, is the set of uncovered alternatives in A.

It is easy to prove that an inefficient 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 an inefficient alternative.

Proposition 4. *There exists a political economy $\mathcal{P}$ such that the statements set out below are satisfied:*

1. *the reciprocity set $R(\mathcal{P})$ is not empty;*
2. *the core is empty;*
3. *a vNM stable set does not exist; and*
4. *the top-cycle set contains an inefficient policy.*

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

Proposition 5. *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 political economy $\mathcal{P} = (N, A, (\succeq_i), \mathcal{C})$ such that the reciprocity set strictly refines the uncovered set: $R(\mathcal{P}) \subsetneq UC(\mathcal{P})$.*

Proposition 5 shows that the reciprocity set and the uncovered set have very different properties and predictions. For instance, in Example 5, one can prove that $R(\mathcal{P}) = \{l_1, l_3\}$, whereas $UC(\mathcal{P}) = \{l_2, l_3\}$. 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.

Summarizing this subsection, we have shown that the reciprocity set is very different from all the aforementioned solution concepts. Table 1 presented in Section 7.3 below shows how these notions differ in terms of their properties.

7.3 Infinite-horizon Sequential Games

In infinite horizon games, individuals 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 individual

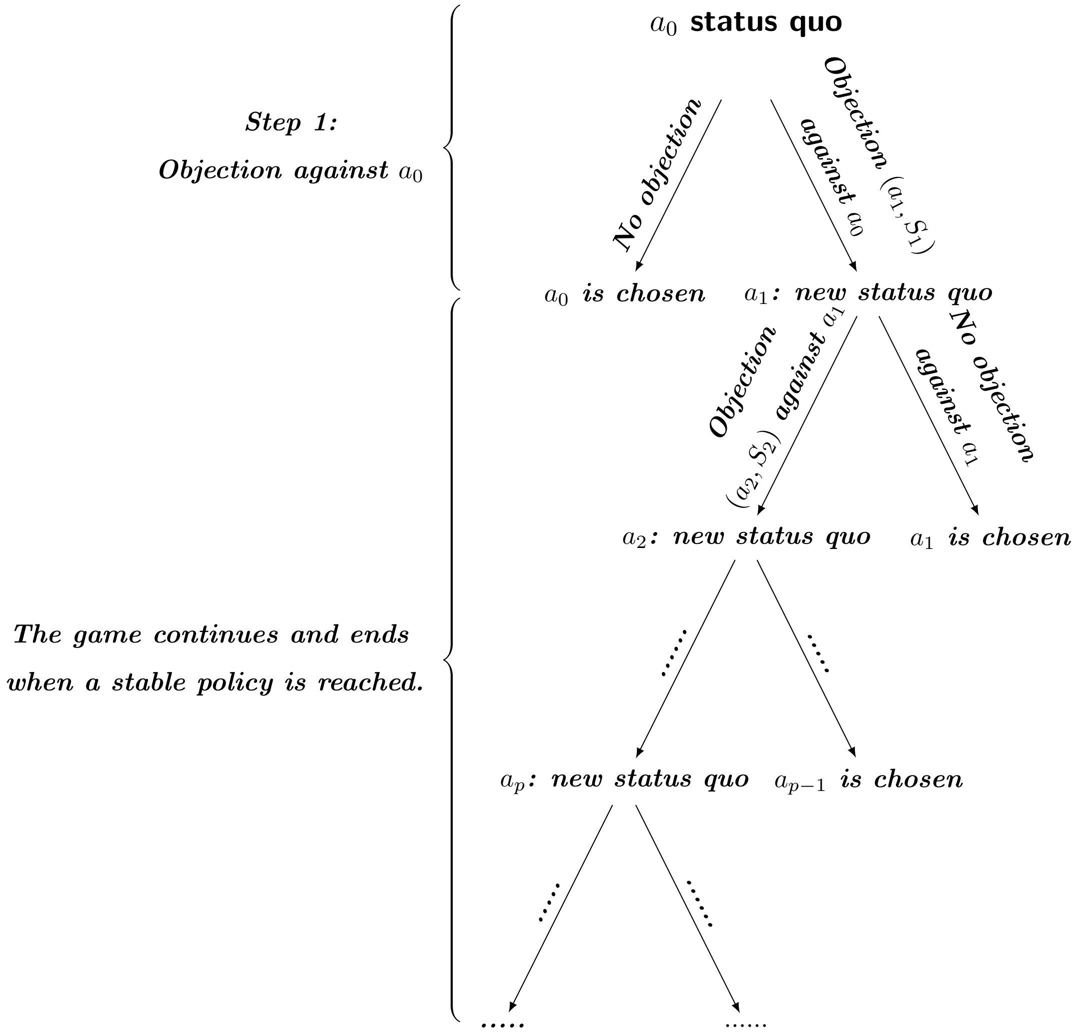


Figure 5: Illustration of an infinite-horizon sequential political game

earns and consumes his or her payoff and the game ends. Figure 5 illustrates this voting process. 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.[23] We recall their definitions below.

Definition 7. Let $\mathcal{P} = (N, A, (\succeq_i), \mathcal{C})$ be a political economy, 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 $a_0,\ a_1,\ ...,\ a_m \in A$ (*where* $a_0 = a$ *and* $a_m = b$) and a sequence of winning coalitions $S_0,\ S_1,\ ..., S_{m-1}$ such that $a_{i+1} >^d a_i$ via S_{i+1} and $b \succ_{S_i} a_i$ for $i = 0, 1, ..., m-1$.

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

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

Definition 8. Let $\mathcal{P} = (N, A, (\succeq_i), \mathcal{C})$ be a political economy, 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 $b \succ_{S_i} a_i$ for $i = 0, 1, ..., m-1$. The relation $a \longrightarrow_S b$ means that, if a is a status quo, S can make b be the new status quo.

B) X is said to be consistent if:

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

[23] Importantly, Baron & Ferejohn (1989) and Banks & Duggan (2000) also model sequential bargaining in a legislature, where there exists a rule that grants a veto right to individuals or coalitions. Their models follow the "bargaining approach" (Ray & Vohra (2014)), unlike Harsanyi (1974) andChwe (1994) who follow the *blocking approach* like the reciprocity set. Due to these conceptual differences, it is not easy to compare the predictions of the reciprocity mechanism to those of Baron & Ferejohn (1989) or Banks & Duggan (2000).

C) The largest consistent set is the union of all the consistent sets of $\mathcal{P}$.

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 infinite-horizon mechanism differs from the reciprocity mechanism in two major respects. First, the reciprocity mechanism is finite and involves only three stages. Second, under the infinite-horizon mechanism, a coalition does not have the possibility to revise or withdraw its move, unlike under the reciprocity mechanism. For these reasons, these mechanisms have different equilibrium and welfare properties, as is shown below.

Proposition 6. *There exists a political economy $\mathcal{P}$ such that:*

1. *the reciprocity set $R(\mathcal{P})$ is not empty;*
2. *there is no Harsanyi stable set; and*
3. *the largest consistent set contains an inefficient policy.*

In Table 1 below, 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* (treated in Section 8) satisfies similar properties as the reciprocity set, although the former is a refinement of the latter. We say that a solution concept $\mathcal{E}$ satisfies the "stability" property if there does not exist a political economy $\mathcal{P}$ such that $\mathcal{E}(\mathcal{P})$ is empty or $\mathcal{E}(\mathcal{P})$ does not exist. A solution concept $\mathcal{E}$ satisfies the "efficiency" property if for any political economy $\mathcal{P}$, $\mathcal{E}(\mathcal{P})$ contains only Pareto-efficient alternatives. The definition of "strategic protection of minority interests" (SPMI) is the one given in Definition 4. As shown in Table

1 below, among all the solution concepts analyzed in this paper, the reciprocity set is the only one that satisfies all of these three properties.

Table 1: Comparison of Solution Concepts

Solution Concepts	Stability	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
Larget consistent set	Yes	No	Yes

It follows that the reciprocity set differs from all the other solution concepts in its motivation, incentives structure, and properties. We have shown that these differences lead to different predictions.

8 A Refinement: The Sophisticated Reciprocity Set

The reciprocity set assumes that each member of a winning coalition S that formulates an objection (y, S) to a status quo policy x should prefer y to x. As acknowledged earlier, this assumption, which goes back to Aumann & Maschler (1964), can be viewed as formalizing a prudent behavior or an aversion to ambiguity when there is uncertainty regarding the continuation of the game after the first stage. In this section, we assume that such uncertainty does not exist, implying that players can act more strategically. Under this assumption, we modify the definition of an "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.

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 members of the sponsoring coalition know that the game 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 game will stop at y after this move. The game will stop at y under two different circumstances. The first is that there is no other winning coalition T that has an incentive to formulate a counterobjection (z, T) (that is, $not(z \succeq_T y)$ for any $T \in \mathfrak{C}$, $T \neq S$, $z \in A$). The second is that if such a coalition T exists, T must not be hurt by the move from x to y (i.e., $y \succeq_T x$) and S must be hurt by z relative x (i.e., $not(z \succeq_S x)$). 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 mechanism, which will hurt T. This reasoning is formalized in the definition of a "strategic objection" in Definition 9 below. The *sophisticated reciprocity set* is defined simply by replacing the notion of an objection in the definition of the reciprocity set (see Definition 1) by the notion of a strategic objection.

Definition 9. Let $\mathcal{P} = (N, A, (\succeq_i)_{i \in N}, \mathcal{C})$ 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 $(z, T) \in A \times \mathfrak{C}$, $S \neq T$, $[y \succeq_T x$ and $not(z \succeq_S x)]$ or $not(z \succeq_T y)$ implies $y \succ_S x$.

2. **Counter-objection**: let (y, S) be a strategic objection against x. A pair $(z, T) \in A \times \mathfrak{C}$, $S \neq T$, is said to be a counter-objection against (y, S) if $z \succ_T y$ and $not(y \succeq_T x)$.

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

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 $y \gtrdot_S^{sd} x$.

5. **Sophitiscated reciprocity set**: the sophitiscated reciprocity set of $\mathcal{P}$, denoted $SR(\mathcal{P})$, is the set of all the policies in A against which no justified strategic objection exists. Formally, $SR(\mathcal{P}) = \{x \in A :$ there does not exist $(y, S) \in A \times \mathfrak{C}$ such that $y \gtrdot_S^{sd} x\}$. Any policy in $SR(\mathcal{P})$ is called a sophisticated reciprocity equilibrium.

The following example illustrates the difference between the reciprocity set and the strategic reciprocity set.

Example 6. Let $\mathcal{P} = (N, A, (\succeq_i)_{i \in N}, \mathcal{C})$ be a political economy where $N = \{1, 2, 3, 4, 5, 6\}$, $A = \{x, y, z, t\}$, and $\mathcal{C}$ is the majority rule. Preferences over the set A are given as follows: $y \succ_1 t \succ_1 z \succ_1 x$; $y \succ_2 t \succ_2 z \succ_2 x$; $z \succ_3 y \succ_3 x \succ_3 t$; $t \sim_4 z \succ_4 x \succ_4 y$; $x \succ_5 t \succ_5 z \succ_5 y$; and $x \succ_6 t \succ_6 z \succ_6 y$. Let $S = \{1, 2, 3, 4\}$, $T = \{3, 4, 5, 6\}$, and $U = \{1, 2, 5, 6\}$. Figure 6 describes the domination or popularity graph among policies according to individual preferences. For the political

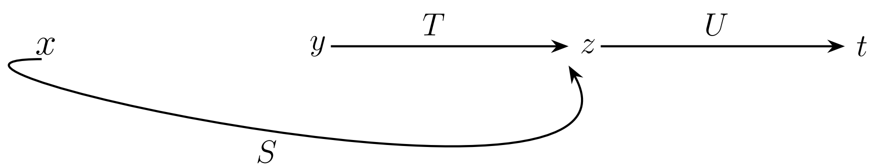


Figure 6: Reciprocity set vs. Sophitiscated reciprocity set

economy $\mathcal{P}$, the reciprocity $R(\mathcal{P})$ set is larger than the sophitiscated reciprocity set $SR(\mathcal{P})$. In fact, $R(\mathcal{P}) = \{x, y, t\}$ and $SR(\mathcal{P}) = \{y, t\}$. It is easy to see that if x is the status quo, it will not be replaced by any winning coalition if players behave according to the rationality defining the reciprocity set. However, the winning coalition S will replace x by y if players 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.[24]

Below, we show that the sophisticated reciprocity set refines the reciprocity set while at the same time preserving its nice properties.

Theorem 5. *Let $\mathcal{P} = (N, A, (\succeq_i)_{i \in N}, \mathcal{C})$ be a political economy.*

1. *(**Refinement**) $SR(\mathcal{P}) \subseteq R(\mathcal{P})$, and this inclusion may be strict.*

2. *(**Equilibrium existence**)*

   (a) If A is discrete and finite and preferences are strict, then $SR(\mathcal{P})$ is not empty.

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

3. *(**Efficiency**) Any policy in $SR(\mathcal{P})$ is Pareto-efficient.*

A few additional comments are needed. First, we note that item (3) in Theorem 5 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 4). Second, 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 2. This can be illustrated by Example 6 above, where policy y is a sophisticated reciprocity equilibrium despite the fact that it is a minority option. The sophisticated reciprocity set also protects minorities with a "fixed" identity using supermajority rules.

[24] Note that following the opposition by player 5, S will not withdraw y, giving the opportunity to 5 to form T and move from y to z, ending the game according to our mechanism.

9 Conclusion

This paper considers the problem faced by a political authority that has to design a legislative mechanism that guarantees the selection of policies that are stable, efficient, and inclusive in the sense of strategically protecting minority interests. We propose the *reciprocity mechanism*, and we prove that it satisfies all of these desirable properties under mild conditions. It does so by providing an incentive for "selfish" individuals to display reciprocal and pro-social behavior in voting. In particular, it fosters positive reciprocity and prevents negative reciprocity and free-riding.

Under certain conditions, the reciprocity equilibrium might be unique. Multiple reciprocity equilibria might also exist in other environments. By preventing political failure, the reciprocity mechanism resolves the long-established conflict between individual rationality and efficiency in political decisions. It is also very easy to implement, which makes it applicable in a wide range of legislative settings. We also show that it has several merits, especially when compared to other well-known mechanisms.

Acknowledgement

We are grateful to 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 for useful discussions and comments. We also thank 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.

References

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

Arrow, K. J. & 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. & Maschler, M. (1964). The bargaining set for cooperative games. *Advances in Game Theory*, 52, 443–476.

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

Austen-Smith, D. (1987). Interest groups, campaign contributions, and probabilistic voting. *Public choice*, 54(2), 123–139.

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

Austen-Smith, D. & 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. (1985). Sophisticated voting outcomes and agenda control. *Social Choice and Welfare*, 1(4), 295–306.

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

Barberà, S., Berga, D., & 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.

Barberà, S. & Gerber, A. (2017). Sequential voting and agenda manipulation. *Theoretical Economics*, 12(1), 211–247.

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

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1), 122–142.

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.

Bochet, O., Page, T., & Putterman, L. (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization*, 60(1), 11–26.

Brighouse, H. & Fleurbaey, M. (2010). Democracy and proportionality. *Journal of Political Philosophy*, 18(2), 137–155.

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

Buchanan, J. M. & Tullock, G. (1962). *The calculus of consent*, volume 3. Ann Arbor: University of Michigan Press.

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

Castles, F. G. & 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.

Chwe, M. S.-Y. (1999). Minority voting rights can maximize majority welfare. *American Political Science Review*, 93(1), 85–97.

Cox, G. W. & Shepsle, K. A. (2007). Majority cycling and agenda manipulation: Richard McKelvey's contributions and legacy. *Positive changes in Political Science: the legacy of Richard D. McKelvey's most influential writings*, (pp.~19).

Cox, J. C. (2001). *On the economics of reciprocity*. Technical report, Discussion paper, University of Arizona.

Dasgupta, P. & Maskin, E. (2008). On the robustness of majority rule. *Journal of the European Economic Association*, 6(5), 949–973.

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

Downs, A. (1957). An economic theory of political action in a democracy. *The Journal of Political Economy*, (pp. 135–150).

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

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

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

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

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

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

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

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

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

Fehr, E. & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, (pp. 817–868).

Fotso, A. F., Pongou, R., & 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).

Guinier, L. (1994). *The tyranny of the majority: fundamental fairness in representative democracy*. Free Pr.

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.

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

Holzman, R., Peleg, B., & Sudhölter, P. (2007). Bargaining sets of majority voting games. *Mathematics of Operations Research*, 32(4), 857–872.

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

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

Kimya, M. (2015). Equilibrium coalitional behavior. *Unpublished, Brown University.*

King, B. W. (1996). Use of supermajority voting rules in corporate america: Majority rule, corporate legitimacy, and minority shareholder protection. *Del. J. Corp. L.*, 21, 895.

Komorita, S. S., Hilty, J., & Parks, C. D. (1991). Reciprocity and cooperation in social dilemmas. *Journal of Conflict Resolution*, 35(3), 494–518.

Komorita, S. S., Parks, C., & Hulbert, L. (1992). Reciprocity and the induction of cooperation in social dilemmas. *Journal of Personality and Social Psychology*, 62(4), 607.

Madison, J. (1788). *The Federalist Papers.* Nos. 10, 51. New York.

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

McGann, A. J. (2006). *The logic of democracy: reconciling equality, deliberation, and minority protection.* University of Michigan Press.

McKelvey, R. D. (1986). Covering, dominance, and institution-free properties of social choice. *American Journal of Political Science*, (pp. 283–314).

McKelvey, R. D. & Niemi, R. G. (1978). A multistage game representation of sophisticated voting for binary procedures. *Journal of Economic Theory*, 18(1), 1–22.

Mill, J. S. (1861). *Considerations on representative government*. Parker, Son, and Bourn.

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).

Miller, N. R. (2013). A priori voting power and the us electoral college. In *Power, Voting, and Voting Power: 30 Years After* (pp. 411–442). Springer.

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

Ordeshook, P. C. & Schwartz, T. (1987). Agendas and the control of political outcomes. *American Political Science Review*, 81(1), 179–199.

Pongou, R., Lambo, L. D., & Tchantcho, B. (2008). Cooperation, stability and social welfare under majority rule. *Economic Theory*, 35(3), 555–574.

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

Pongou, R. & 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. *The American Economic Review*, (pp. 1281–1302).

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

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

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).

Samuelson, P. A. (1954). The pure theory of public expenditure. *The Review of Economics and Statistics*, 36(4), 387–389.

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

Shepsle, K. A. & Weingast, B. R. (1984). Uncovered sets and sophisticated voting outcomes with implications for agenda institutions. *American Journal of Political Science*, (pp. 49–74).

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

Zhou, L. (1994). A new bargaining set of an n-person game and endogenous coalition formation. *Games and Economic Behavior*, 6(3), 512–526.

10 Appendix

Appendix A: More Elaboration on the Examples

In this section, we provide more details on two examples (Examples 2 and 4) covered in the paper.

Example 2. We derive the reciprocity set of the ultimatum game with counter-offer and other-regarding preferences. We also derive the predicted reciprocity set at the status quo $x_0 = (0, 0, 0)$. Recall that the utility functions are:

$$u_1(x_1, x_2, x_3) = x_1 - \frac{1}{3}(x_2 - x_1) - \frac{1}{3}(x_3 - x_1),$$
$$u_2(x_1, x_2, x_3) = x_2 - \frac{1}{3}(x_1 - x_2) - \frac{1}{3}(x_3 - x_2),$$
$$u_3(x_1, x_2, x_3) = x_3 - \frac{1}{3}(x_1 - x_3) - \frac{1}{3}(x_2 - x_3).$$

Assume, for instance, that individuals 1 and 2 propose a distribution $x = (x_1, x_2, x_3)$ to individual 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 individual 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 individual. This proves that proposing an allocation that exhausts the total amount of money and gives less than 100/6 to an individual is rejected. Therefore, the set of the predicted allocations from the status quo $x_0 = (0, 0, 0)$ is:

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

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 objection against x.

2. Any allocation that gives all of the money to one individual 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 individuals 2 and 3 will formulate an objection $(y, \{2, 3\})$ against x, where $y = (34, 33, 33)$. Then, individual 1 will oppose the objection. But individuals 2 and 3 will not withdraw it since they have no incentive to do so. It is obvious that player 1 can convince individual 2 to formulate a counter-objection $((76, 34, 0), \{1, 2\})$ against $(y, \{2, 3\})$. Moreover, any counter-objection (z, T) against y is friendly to $\{2, 3\}$, which implies that $(y, \{2, 3\})$ is a justified objection against x.

3. Any allocation that gives a strictly positive amount to at least two individuals is stable. 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 individual 3 can form a coalition with either individual 1 or individual 2 to initiate an 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 $x_1 > a$. Player 1 will oppose this allocation, causing $\{2, 3\}$ to either withdraw it or maintain it. But $\{2, 3\}$ will withdraw this allocation because, otherwise, individual 1 will form a coalition with either individual 2 or individual 3 to formulate a counter-objection. Assume, for instance, that individuals 1 and 3 form a coalition and propose the allocation $(x_1, 0, x_2)$. Then, this will lower the payoff of individual 2. Given the fact that individual 2 is aware of this possibility, he or she will withdraw from the coalition $\{2, 3\}$, which will invalidate the objection $(y, \{2, 3\})$ against the status quo $x = (x_1, x_2, 0)$. Therefore, there is no objection against $x = (x_1, x_2, 0)$. In essence, given the fact that $x_1 + x_2 = 100$, it is impossible to formulate an objection that

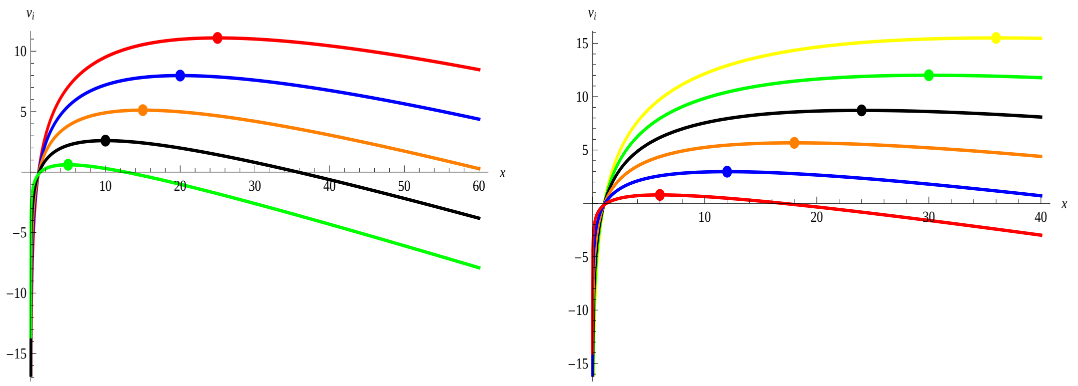
improves the payoff of every individuals. It therefore follows that $x = (x_1, x_2, 0)$ is a reciprocity equilibrium. Given the fact that individuals have symmetric preferences, any allocation that gives a positive amount to at least two individuals is a reciprocity equilibrium. The reciprocity set is therefore:

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

Example 4. Assume, for illustration, that there are five legislators who have to decide on the number of refugees to be admitted. The utility function of each legislator i is given by (see Figure 7):

$$v_i(x_i) = b_i \ln(x_i) - 1/5x_i.$$

Observe that this function is single-peaked, with legislator i's optimum being obtained by solving $v_i'(x_i) = 0$; this leads to the solution $x_i^* = 5b_i$. Figure 7(a) represents the legislator's utility functions, assuming $b_1 = 5$, $b_2 = 4$, $b_3 = 3$, $b_4 = 2$ and $b_5 = 1$. We can show that the legislators' optima are respectively: $x_1^* = 25$, $x_2^* = 20$, $x_3^* = 15$, $x_4^* = 10$ and $x_5^* = 5$.



(a) Five legislators (b) Six legislators

Figure 7: Legislators' utility functions, with the bundle on each curve representing legislator i's optimum point

We recall that the decision rule is the majority rule. The point x_3^*, which is the optimum of the median legislator, is the Condorcet winner under the reciprocity mechanism. It follows that x_3^* is the only reciprocity equilibrium: $R(\mathcal{P}) = \{15\}$.

Now, assume that there are six legislators, with legislator i's utility function being $v_i(x) = b_i \ln(x) - 1/6x_i$, where $b_1 = 5$, $b_2 = 4$, $b_3 = 3$, $b_4 = 2$, $b_5 = 1$ and $b_6 = 6$ (see Figure 7(b)). The legislators' optima are: $x_6^* = 36$, $x_1^* = 30$, $x_2^* = 24$, $x_3^* = 18$, $x_4^* = 12$ and $x_5^* = 6$. In this scenario, there is no median legislator. The reciprocity equilibria are x_2^* and x_3^*: $R(\mathcal{P}) = \{24, 18\}$. In fact, $(x_3, \{1,2,3,4,6\})$ is a justified objection against x_5; $(x_3, \{1,2,3,6\})$ is a justified objection against x_4; $(x_2, \{2,3,4,5\})$ is a justified objection against x_1; and $(x_2, \{2,3,4,5\})$ is a justified objection against x_6. Moreover, there is no justified objection against either x_3 or x_2, if each of these alternatives is selected as the status quo. We can conclude that $R(\mathcal{P}) = \{x_2^*, x_3^*\} = \{24, 18\}$. Therefore, either 18 or 24 refugees will be admitted into the country.

Appendix B: Proofs of Results

In order to facilitate the exposition of the proofs, we prove the lemma below, which gives an equivalent definition of the relation $\gtrdot$ used to formalize the notion of a justified objection (Definition 1).

Lemma 1. *Let $\mathcal{P} = (N, A, (\succeq_i)_{i\in N}, \mathcal{C})$ be a political economy, and $x, y \in A$ be two policies. Statements 1. and 2. below are equivalent:*

1. *$y \gtrdot x$.*
2. *There exists a winning coalition $S \in \mathfrak{C}$ such that:*

   a) $y \succ_S x$; and,

   b) $[\forall (z,T) \in A \times \mathfrak{C},\ S \neq T,\ z \succ_T y$ and not$(y \succeq_T x)]$ implies $[z \succeq_S x]$.

Proof. Let $M(\mathcal{P}) = \{x \in A :$ there is no $(y, S) \in A \times \mathfrak{C}$ such that $a)$ and $b)$ hold$\}$. We want to show that $R(\mathcal{P}) = M(\mathcal{P})$.

1. First, we show that $R(\mathcal{P}) \subseteq M(\mathcal{P})$. Let $x \in A$ be an option such that $x \notin M(\mathcal{P})$. Then, there exists a pair $(y, S) \in A \times \mathfrak{C}$ such that $a)$ and $b)$ hold. It follows that (y, S) is an objection against policy x. We demonstrate

that this objection against x is justified. If there is no counter-objection against (y, S), the proof is established. Assuming that there exists a counter-objection (z, T) against (y, S); the following assertions hold:

i) $z \succ_T y$; and

ii) $not(y \succeq_T x)$.

Given the fact that *b)* holds, then *i)* and *ii)* imply $z \succeq_S x$, and the counter-objection (z, T) against (y, S) is friendly. Therefore, a pair (y, S) is a justified objection against x by definition, meaning that $x \notin R(\mathcal{P})$. We conclude that $R(\mathcal{P}) \subseteq M(\mathcal{P})$.

2. Second, prove that $M(\mathcal{P}) \subseteq R(\mathcal{P})$. Let $x \in A$ be a policy such that $x \notin R(\mathcal{P})$. Then, there exists a justified objection $(y, S) \in A \times \mathfrak{C}$ against policy x : $y \gtrdot_S x$. We want to prove that conditions *a)* and *b)* hold, so that $x \notin M(\mathcal{P})$. Since $y \gtrdot_S x$, then it is obvious that $y \succ_S x$. Thus *a)* is satisfied.

   iii) If there is no winning coalition T and no policy z such that $z \succ_T y$, then $y \succ_S x$ and implication *b)* is satisfied since the right-hand side of this implication is false. It follows that $x \notin M(\mathcal{P})$.

   iv) Assume that there exists $(z, T) \in A \times \mathfrak{C}$ such that $z \succ_T y$. We have two possibilities described hereunder:

   - If $not(y \succeq_T x)$, then (z, T) is a counter-objection against (y, S). Since (y, S) is a justified objection, it follows that any counter-objection against (y, S) is friendly, leading to $z \succeq_S x$. Then, *b)* is satisfied and $x \notin M(\mathcal{P})$.
   - If $y \succeq_T x$, then, the assertion $not(y \succeq_T x)$ is false, which implies that the assertion $[not(y \succeq_T x) \Rightarrow z \succeq_S x]$ is true. Therefore, $x \notin M(\mathcal{P})$. We conclude that $M(\mathcal{P}) \subseteq R(\mathcal{P})$.

It follows from 1. and 2. that $R(\mathcal{P}) = M(\mathcal{P})$. $\square$

Proof of Theorem 1. Let $\mathcal{P} = (N, A, (\succeq_i)_{i \in N}, \mathcal{C})$ be a political economy, and S, T be two winning coalitions, and $x, y, z \in A$ be three distinct policies such that $y \succ_S x$ and $z \succ_T y$.

1. Suppose that *not*$(y \succeq_T x)$ and *not*$(z \succeq_S x)$. Assume by contradiction that S formulates an objection (y, S) against x. Then, since $z \succ_T y$, a member of T who is not a member of S will oppose the objection. Hence, S will either withdraw or maintain its objection following the prescription in stage 2. Clearly, S will withdraw its objection since, if it does not this, a 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*$(z \succeq_S x)$. It follows that x will remain in place, which implies that formulating an objection (y, S) against x will not profit S. The coalition S will therefore not formulate such an objection, which is a contradiction.

2. Assume that S formulates an objection (y, S) against x, and suppose that $y \succeq_T x$. 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 objection. Given the fact that the counter-objection (z, T) is a possibility if S does not withdraw y and is an unfriendly counter-objection by definition (since *not*$(z \succeq_S x)$), S will withdraw its objection, thus causing x to remain in place. But this does not benefit coalition T, since $y \succeq_T x$. Consequently, T will not formulate any unfriendly counter-objection (z, T) against (y, S), which is a contradiction.

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

benefit, thus justifying its objection (y, S) against option x.

Proof of Theorem 2. For all $x \in A$, define $f(x) = |\{y \in A; y \succ x\}|$ (where $|X|$ denotes the size or cardinality of the set X) and let $x_0 \in A$ such that $f(x_0) = \min_{x\in A}\{f(x)\}$. Prove that $x_0 \in R(\mathcal{P})$. If this assertion is not true, then there exists $y \in A$ and a winning coalition S, such that $y \gtrdot_S x_0$. It follows that:

$$\begin{cases} (\alpha)\ y \succ_S x,\ \ S \in \mathfrak{C} \\ (\beta)\ \forall (z, T):\ T \neq S,\ \ z \succ_T y, T \in \mathfrak{C} \text{ and not}(y \succeq_T x) \Rightarrow z \succeq_S x \end{cases} . \quad (1)$$

By definition of x_0 and given the fact that the preference $\succ$ is asymmetric, there exists $c \in A$ such that $c \succ y$ and *not* $(c \succ x_0)$. Thus, there exists a coalition T, with $T \in \mathfrak{C}$ such that $c \succ_T y$.

- If $y \succ_T x_0$, then we have $c \succ_T y$ and $y \succ_T x_0$, so that $c \succ_T x_0$ by transitivity, which is a contradiction because *not* $(c \succ x_0)$.
- Suppose that *not* $(y \succ_T x_0)$. Then according to the assertion (β) in equation (1), we have $c \succ_S x_0$, which is a contradiction.

We conclude that $x_0 \in R(\mathcal{P})$.

Proof of Theorem 3. Let f be the function defined over the policy A by $f(x) = m(\{y \in A; y \succ x\})$, where m is the Lebesgue measure on the manifold spanned by A. Given the fact that preferences are continuous, and A is compact and convex, there exists $x_1 \in A$ such that $f(x_1) = \min_{x\in A}\{f(x)\}$. Following the same reasoning as in the proof of Theorem 2, we can show that $x_1 \in R(\mathcal{P})$.

Proof of Proposition 1. Let $\mathcal{P}$ be a political economy in which each preference relation $\succeq_i$ is single-peaked over the policy set A. The proof is performed by construction. We recall that the constitution is defined here by the majority rule. Let $x \in A$ be a policy and define $S(x)$ as the number of individuals for whom x is the peak. Consider the functions f and g defined on A as follows: for any $q \in A$,

$$f(q) = \sum_{x \leq q} S(x) - \tfrac{n}{2},$$

and

$$g(q) = \sum_{x \geq q} S(x) - \tfrac{n}{2}.$$

Let $A_f = \{p \in A : p \text{ is a peak and } f(p) \geq 0\}$ and $A_g = \{p \in A : p \text{ is a peak and } g(p) \geq 0\}$. Notice that neither A_f, nor A_g is empty. In fact, if $\overline{p}$ and $\underline{p}$ are respectively the greatest and the smallest peaks in A, then $f(\overline{p}) = n - \frac{n}{2} = \frac{n}{2} > 0$ and $g(\underline{p}) = n - \frac{n}{2} = \frac{n}{2} > 0$, which implies that $\overline{p} \in A_f$ and $\underline{p} \in A_g$, in turn implying that $A_f \neq \emptyset$ and $A_g \neq \emptyset$.

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. In addition, for any peak $p < q_1^*$, $f(p) < 0$, which implies that $\sum_{x \leq p} S(x) < \frac{n}{2}$.

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. In addition, for any peak $p > q_2^*$, $g(p) < 0$, which implies that $\sum_{x \geq p} S(x) < \frac{n}{2}$.

In this proof, it is established that there exists at least one equilibrium and that q_1^* and q_2^* are the only (possible) equilibria. First, 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 strictly comprised within the interval (q_1^*, q_2^*). In fact, if $q_1^* = q_2^*$, the claim is proved. Consider that $q_1^* \neq q_2^*$. Two cases will be differentiated: $q_1^* < q_2^*$ and $q_1^* > q_2^*$.

- Suppose $q_1^* < q_2^*$ and by contradiction, that there exists a peak q^* between q_1^* and q_2^*. Using the definition of q_1^*, we have $f(q^*) > 0$, which implies that $\sum_{x \leq q^*} S(x) - \frac{n}{2} > 0$, and then $\sum_{x \leq q^*} S(x) > \frac{n}{2}$. This means the number of individuals whose peak is weakly on the left of q^* exceeds half of the population. We also know that $g(q_2^*) \geq 0$, which is equivalent to $\sum_{x \geq q_2^*} S(x) - \frac{n}{2} \geq 0$, and $\sum_{x \geq q_2^*} S(x) \geq \frac{n}{2}$. The latter inequality means that the number of individuals whose peak is weakly on the right of q_2^* is at least half of the population. Since $q^* < q_2^*$, there is no individual whose peak is weakly on the left of q^* and weakly on the right of q_2^* at the same time. Yet, both $\sum_{x \leq q^*} S(x) > \frac{n}{2}$ and $\sum_{x \geq q_2^*} S(x) \geq \frac{n}{2}$ imply that the number of individuals whose peak is weakly on the left of q^* or is weakly on the right of q_2^* strictly exceeds the total population n, which is a contradiction.

- Suppose that $q_1^* > q_2^*$ and by contradiction, that there exists a peak q^* between q_1^* and q_2^*. Using the definition of q_1^* and q_2^*, it follows that $f(q^*) < 0$ (since $q^* < q_1^*$) and that $g(q^*) < 0$, since $q^* > q_2^*$. This implies that $f(q^*) + g(q^*) < 0$, and that $S(q^*) < 0$, because $f(q^*) + g(q^*) = S(q^*)$, which is a contradiction.

Fact 3: If $q_1^* \neq q_2^*$, then $q_1^* < q_2^*$. By contradiction, consider that $q_1^* > q_2^*$. It follows from the definition of q_1^* and q_2^* that $f(q_2^*) < 0$ and that $g(q_1^*) < 0$. This implies that $\sum_{x \leq q_2^*} S(x) < \frac{n}{2}$ and $\sum_{x \geq q_1^*} S(x) < \frac{n}{2}$, and that $\sum_{x \leq q_2^*} S(x) + \sum_{x \geq q_1^*} S(x) < n$. In other words, the number of individuals whose peak is weakly on the left of q_2^* or is weakly on the right of q_1^* is strictly smaller than the size of the entire population n. Thus, there exists an individual whose peak is strictly comprised within the interval (q_1^*, q_2^*). This is a contradiction, because it is proven that there is no peak strictly comprised within the interval (q_1^*, q_2^*) (see Fact 2).

We prove that q_1^* and q_2^* are the only equilibria. Three cases are covered as described hereunder:

Case 1: n is even and $q_1^* = q_2^* = q^*$ (see Figure 8). Then, the analysis claims that q^* is the only equilibrium: $R(\mathcal{P}) = \{q^*\}$.

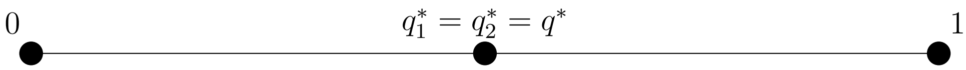


Figure 8: The voters' size is even and individuals have the same peak

Let $p \in A$ be a peak. If $p < q^*$, by definition of q^*, $\sum_{x \leq p} S(x) < \frac{n}{2}$ and $\sum_{x \geq q^*} S(x) > \frac{n}{2}$, which implies that q^* is majority-preferred to p in a pairwise majoritarian election. Similarly, if $p > q^*$, we show in the same way that

q^* is majority-preferred to p in a pairwise majoritarian election. It follows that q^* is majority-preferred to any other peak in a pairwise majoritarian election. Since there is no other option that is majority-preferred to q^*, then the implication b) in Lemma 1 is always true, thus implying that there is no justified objection against q^*. Therefore, $R(\mathcal{P}) = \{q^*\}$.

Case 2: n is even and $q_1^* \neq q_2^*$ (see Figure 9). In this case, it is shown that, q_1^* and q_2^* are the only equilibria.

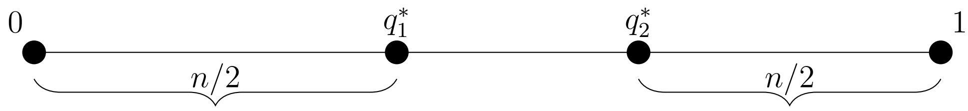


Figure 9: The voters' size is even and the minima peaks q_1^* and q_2^* are different

First, note that neither q_1^*, nor q_2^* is majority-preferred to any other policy in a pairwise majoritarian election. Furthermore, in a pairwise majoritarian election opposing q_1^* and q_2^*, neither will win, since $q_1^* < q_2^*$, and $\sum_{x \leq q_1^*} S(x) + \sum_{x \geq q_2^*} S(x) = n$, each will receive exactly $\frac{n}{2}$ votes. Now, let $p \in A$ be a peak that is distinct from q_1^* and q_2^*. It follows from Facts 2 and 3 that, either $p < q_1^*$ or $p > q_2^*$.

1. If $p < q_1^*$, then q_1^* is majority-preferred to p in a pairwise majoritarian election since the number of votes obtained by q_1^* against p is at least equal to $\sum_{x \geq q_1^*} S(x) > \sum_{x \geq q_2^*} S(x) \geq \frac{n}{2}$. Based on the fact that q_1^* is not majority-preferred to any other policy, and $q_1^* \gtrdot p$.
2. If $p > q_2^*$, then q_2^* is majority-preferred to p in a pairwise majoritarian election since the number of votes obtained by q_2^* against p is at least equal to $\sum_{x \leq q_2^*} S(x) > \sum_{x \leq q_1^*} S(x) \geq \frac{n}{2}$. In addition, given the fact that q_2^* is not majority-preferred to any other policy, it follows that $q_2^* \gtrdot p$.

   We conclude that $R(\mathcal{P}) = \{q_1^*, q_2^*\}$.

Case 3: n is odd. It is shown that $q_1^* = q_2^*$. Consider the contrary, meaning that $q_1^* \neq q_2^*$. Fact 3 implies that $q_1^* < q_2^*$. Since n is odd and since there is no individual whose peak is strictly comprised within the interval (q_1^*, q_2^*) (Fact 2), the number of individuals whose peak is weakly to the left of q_1^* (i.e., $\sum_{x \leq q_1^*} S(x)$) is different from the number of individuals whose peak is weakly to the right of q_2^* (i.e., $\sum_{x \geq q_2^*} S(x)$). Two cases are displayed:

1. Suppose that $\sum_{x \leq q_1^*} S(x) > \sum_{x \geq q_2^*} S(x)$. Recall that by the definition of q_2^*, $\sum_{x \geq q_2^*} S(x) \geq \frac{n}{2}$, then, $\sum_{x \leq q_1^*} S(x) > \frac{n}{2}$. This implies that $\sum_{x \leq q_1^*} S(x) + \sum_{x \geq q_2^*} S(x) > n$, which is a contradiction, because $\sum_{x \leq q_1^*} S(x) + \sum_{x \geq q_2^*} S(x) = n$.
2. If $\sum_{x \leq q_1^*} S(x) < \sum_{x \geq q_2^*} S(x)$, then, using the definition of q_1^* ($\sum_{x \geq q_1^*} S(x) \geq \frac{n}{2}$), therefore it follows that $\sum_{x \leq q_1^*} S(x) > \frac{n}{2}$. Consequently, $\sum_{x \leq q_1^*} S(x) + \sum_{x \geq q_2^*} S(x) > n$, which is a contradiction, given that $\sum_{x \leq q_1^*} S(x) + \sum_{x \geq q_2^*} S(x) = n$.

This being the case, $q_1^* = q_2^*$. Hence, referring to Case 1, as $q_1^* = q_2^* = q^*$, 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 $R(\mathcal{P}) = \{q^*\}$.

Proof of Theorem 4. Let $x \in A$ be a policy such that $x \in R(\mathcal{P})$. If x is Inefficient, then there exists $y \in A$ such that $y \succ_N x$. It follows that for any $z \in A$ such that $z \succ_T y$ with $T \in \mathfrak{C}$, we have $y \succ_T x$ because $T \subset N$. Then, not $(y \succeq_T x)$ is always false. Therefore the implication [*not* $(y \succeq_T x) \Rightarrow z \succeq_N x$] is always true–that is, $y \gtrdot_N x$ or (y, N) is a justified objection against x, which is a contradiction. Thus, x is Pareto-efficient.

Proof of Proposition 2 . It suffices to consider the political economy $\mathcal{P}$ defined in Example 5. The language l_2 is majority-preferred to l_1. The latter language is therefore a minority option. Since $l_1 \in R(\mathcal{P})$, it follows that the reciprocity set mechanically protects minority interests. Now, consider any political economy $\mathcal{P}' = (N, A, (\succeq_i), \mathcal{C})$, and x a non-strategic minority option. Then, there exist a majority option $y \in A$ and a winning coalition $S \in \mathfrak{C}$ such that $y \succ_S x$. It follows that (y, S) is a justified objection against x (or $y \gtrdot_S x$), i.e $x \notin R(\mathcal{P}')$.

Proof of Proposition 3. Consider the society $\mathcal{P}_0$ which consists of five individuals in the set $N = \{1, 2, 3, 4, 5\}$, and the policy space $A = \{a, b, c, d, e\}$. The individuals' preferences over these policies are defined as follows:

$$c \succ_1 b \succ_1 a \succ_1 e \succ_1 d;$$

$$d \succ_2 c \succ_2 b \succ_2 a \succ_2 e;$$

$$e \succ_3 b \succ_3 a \succ_3 d \succ_3 c;$$

$$e \succ_4 d \succ_4 b \succ_4 a \succ_4 c;$$

$$c \succ_5 d \succ_5 e \succ_5 b \succ_5 a.$$

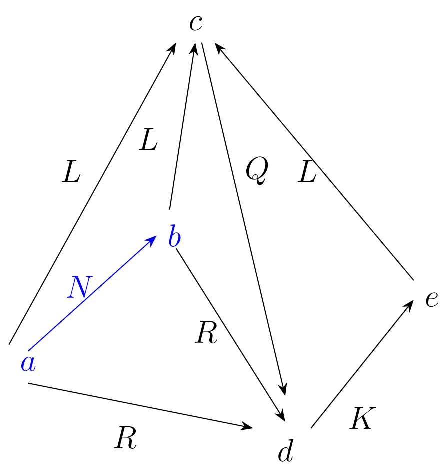


Figure 10: Popularity relationship among policies

The constitution is given by the winning coalitions $K = \{1, 3, 4\}$, $R = \{2, 4, 5\}$, $L = \{1, 2, 5\}$, $Q = \{2, 3, 4\}$, and any coalition that includes K, R, L, or Q. Figure 10 provides the graph of the popularity relationship among the policies (e.g., the arrow from a to b means that the grand coalition N strictly prefers policy b to a.).

The inefficient policy a is stable under the modified mechanism M'. Even though all of the individuals prefer b over a, no winning coalition will object to a since this will lead to the election of either d (if coalition L objects) or e (if coalition R objects). In either case, some member of the objecting coalition will regret, which is the reason why none of these coalitions will form in order to make an objection. This implies that $a \in NR(\mathcal{P}_0)$.[25]

Under the reciprocity mechanism, coalition K, for instance, will formulate an objection (b, K) against a. Once b is introduced, coalition L will not formulate a counter-objection (L, c), knowing that such a counter-objection is unfriendly as it harms the interest of individuals 3 and 4 in K. In fact, if it does this, these individuals will withdraw from the winning coalition K, therefore destroying the objection (b, K) and causing a to remain in place. This response will harm the interest of individuals 2 and 5 in the counter-objecting coalition L. These latter individuals will therefore not participate in coalition L and that coalition will never formulate a counter-objection. Furthermore, coalition R will not formulate a counter-objection (R, d) since individuals 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 $R(\mathcal{P}_0) = \{b, c, d, e\}$ and that $NR(\mathcal{P}_0) = \{a, b, c, d, e\}$.

Proof of Proposition 4. The political economy $\mathcal{P}_0$ considered in this proof is the same that we used in the proof of Proposition 3. We consider $\mathcal{P} = \mathcal{P}_0$.

1. As is shown in the proof of Proposition 3, the set of stable outcomes under the reciprocity mechanism is given by $R(\mathcal{P}) = \{b, c, d, e\}$. The reciprocity mechanism expels the inefficient policy a.

2. We can observe in figure 10 that, for any policy $x \in A$, there exists another policy $y \in A$ and a winning coalition S such that all individuals in S strictly prefer y over x. This shows that the core is empty.

3. In this political economy, 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.

[25] Under the modified mechanism, in general, some individuals who might be willing to participate in a coalition in order to remove an inefficient status quo fear that the interim policy will be replaced by a new policy which is worse for them than the status quo.

4. It can be shown that the top cycle set contains all the alternatives. In fact, an alternative does not belong to $TC(\mathcal{P})$ if and only if it is defeated by each alternative in $TC(\mathcal{P})$; this condition is not satisfied for any of the alternatives. Then, $TC(\mathcal{P}) = A$, and $TC(\mathcal{P})$ contains the inefficient policy a.

Proof of Proposition 5. To prove statements 1. and 2., we consider the

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

political economy described in Example 1 in which the popularity relationship among policies is reproduced below.

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, $x \in UC(\mathcal{P})$ 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 is $UC(\mathcal{P}) = \{c, d\}$. Since the alternative c is a minority option and $c \in UC(\mathcal{P})$, 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 $c \in UC(\mathcal{P})$, it follows that the uncovered set does not strategically protect minority interests. It can easily be shown that $R(\mathcal{P}) = \{b, d\}$. Consequently, the uncovered set contains the alternative c which is not a reciprocity equilibrium.

3. We modify the political economy described in Example 1 only by changing preferences. We assume that agents display the following preferences:

$$b \succ_1 d \sim_1 c \sim_1 a;$$

$$d \succ_2 b \succ_2 c \sim_2 a;$$

$$c \succ_3 b \succ_3 d \succ_3 a;$$

$$c \succ_4 b \succ_4 a \succ_4 d;$$

$$a \succ_5 d \sim_5 c \succ_5 b;$$

$$a \succ_6 d \sim_6 c \succ_6 b.$$

Denote $S = \{1, 2, 3, 4\}$, and $T = \{3, 4, 5, 6\}$. The popularity relationship among policies based on the new preferences is provided in the figure below. We

$$a \xrightarrow{\quad S \quad} b \xrightarrow{\quad T \quad} c$$

have $D(a) = \emptyset$, $D(b) = \{a\}$, and $D(c) = \{b\}$. It follows that $R(\mathcal{P}) = \{c\}$ and $UC(\mathcal{P}) = \{b, c\}$.

Proof of Proposition 6. Consider the political economy $\mathcal{P}_0$ defined in the proof of proposition 4. Figure 10 set out above can also be viewed as the graph of "effectiveness relations" (Chwe, 1994) which is equivalent to the graph of the popularity relationship defined in Section 7.3. 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 specific political economy, the binary relations $\gg^H$ and $\gg^C$ are equivalent.

1. It is already presented in the previous proof that the reciprocity set is non-empty and contains only efficient policies: $R(\mathcal{P}_0) = \{b, c, d, e\}$.

2. Using figure 10 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, meaning that no Harsanyi stable set exists.

3. Let us determine the largest consistent set. We start with set $X = A$. 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 individuals 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 individual 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.

Proof of Theorem 5. The lemma below gives an equivalent definition of the relation $\gtrdot^{sd}$ that we use to formalize the notion of a strategic justified objection (Definition 9).

Lemma 2. *Let $\mathcal{P} = (N, A, (\succeq_i)_{i\in N}, \mathcal{C})$ be a political economy, and $x, y \in A$ be two policies. Statements 1. and 2. below are equivalent:*

1. $y \gtrdot^{sd} x$.

2. There exists a winning coalition $S \in \mathfrak{C}$ such that:

a) $\forall (z, T) \in A \times \mathfrak{C}$, $S \neq T$, $[y \succeq_T x$ and not$(z \succeq_S x)]$ or not$(z \succeq_T y)$ implies $y \succ_S x$; and,

b) $[\forall (z, T) \in A \times \mathfrak{C},\ S \neq T,\ z \succ_T y$ and not$(y \succeq_T x)]$ implies $[z \succeq_S x]$.

The proof of Lemma 2 is similar to the proof of Lemma 1.

1. Let $x \in SR(\mathcal{P})$. Assume that $x \notin R(\mathcal{P})$. Then, there exists $(y, S) \in A \times \mathfrak{C}$ such that $y \gtrdot_S x$. It is straightforward to prove that $y \gtrdot_S^{sd} x$, which is a contradiction. Then, $SR(\mathcal{P}) \subseteq R(\mathcal{P})$. In Example 6, we provide a political economy $\mathcal{P}'$, such that $SR(\mathcal{P}') \neq R(\mathcal{P}')$.

2. The proof of item (2) follows the proofs of results (Theorems 2 and 3) that insure the existence of stable policies under the reciprocity mechanism in Section 4.

3. Consider $x \in SR(\mathcal{P})$. Then $x \in R(\mathcal{P})$. Hence, x is Pareto-efficient thanks to Theorem 4.