Universitat Pompeu Fabra Barcelona
Department of Economics and Business

Economics Working Paper Series

Working Paper No. 1540

Collective commitment

Christian Roessler
Sandro Shelegia
Bruno Strulovici

July 2016

Collective Commitment*

Christian Roessler

California State University, East Bay

Sandro Shelegia

Universitat Pompeu Fabra and Barcelona GSE

Bruno Strulovici

Northwestern University

*This paper subsumes earlier versions entitled "The Roman Metro Problem" and "Can Commitment Resolve Political Inertia?" We are grateful to Davide Debortoli, Wiola Dziuda, Georgy Egorov, Jeff Ely, Daniel Garcia, Michael Greinecker, Karl Schlag, Stephen Schmidt, Joel Watson, and numerous seminar participants for their comments. Strulovici gratefully acknowledges financial support from an NSF CAREER Award (Grant No. 1151410) and a fellowship from the Alfred P. Sloan Foundation.

Abstract

We consider collective decisions made by agents whose preferences and power depend on past events and decisions. Faced with an inefficient equilibrium and an opportunity to commit to a policy, can the agents reach an agreement on such a policy? We provide a consistency condition linking power structures in the dynamic setting and at the commitment stage. When the condition holds, commitment has no value: any agreement that may be reached at the outset coincides with the equilibrium without commitment. When the condition fails, as in the case of time-inconsistent preferences, commitment can improve outcomes. We discuss several applications.

JEL: D70, H41, C70

1 Introduction

In dynamic settings where information, preferences, and political influence evolve over time, successive decision-making by electorates, committees, or individuals often leads to suboptimal outcomes, such as the inability to implement needed reforms (Fernandez and Rodrik (1991)), the use of short-sighted monetary or fiscal policies (Kydland and Prescott (1977) and Battaglini and Coate (2008)), the stability of unpopular regimes (Acemoglu and Robinson (2005)), and inefficient adherence to the status quo (Volokh (2003)). Voters' behavior reflects in part their desire to protect themselves against such developments: for example, proponents of a moderate reform may fear that it will set the stage for further reforms they would no longer endorse, and thus refuse to support any change in the first place. In these situations, it would seem that the equilibrium outcome could be improved upon by giving the actors involved the chance to commit to a policy at the outset. In fact, that is the implicit premise of constitutions, laws, and other contracts that facilitate commitment. This paper studies formally when commitment can address dynamic inefficiency.

To make the issue concrete, consider a legislature having to decide whether to pass a moderate reform, whose adoption may be followed by a more radical expansion. As noted, some voters in favor of the initial reform may oppose it nonetheless, worried that it may create a "slippery slope" leading to the radical reform. The resulting deadlock could seemingly be resolved by a commitment to implement only the initial reform and rule out any further one. However, such a commitment is majority-preferred to the status quo if and only if it is itself majority-dominated by the policy consisting of implementing the initial reform and then expanding it if the expansion turns out to be desired by a majority of voters. This policy is, in turn, dominated by the status quo, thus creating a Condorcet cycle among policies.

The situation is depicted in Figure 1. Voters are equally divided into three groups (A, B, C) with terminal payoffs as indicated. A majority decision to implement the initial reform (Y) reveals with probability q that an expansion is feasible. In this case, a vote takes place on whether to stop at the moderate reform (M) or implement the radical one (R). For $q \in (2/3, 1]$, implementing the radical reform (YR) at the second stage is the unique equilibrium, while for $q \in [0, 1/3)$ keeping the moderate reform (YM) is the unique equilibrium. Moreover, YM beats the status quo (N) in majority voting and yields higher utilitarian welfare.

Figure 1 goes here

For $q \in (1/3, 2/3)$, however, the equilibrium outcome is the status quo, N, because voters from group A deem the risk of ending up with the radical reform too high, while voters from group C find the probability of getting the moderate reform (M), which they do not like, too high.[1] To remedy this situation, suppose that A tries to persuade B to use their joint majority to commit to policy YM (implement and keep the initial reform, regardless of what is learned later). Both favor this proposal over the status quo. However, C may approach B with a counteroffer to instead commit to policy YR (implement the initial reform, but expand it if feasible). Both prefer this proposal to YM. A could then remind C that both of them are better off with the status quo (opposing the initial reform), since YR corresponds exactly to the off-equilibrium path that A and C reject in the dynamic game without commitment. These arguments describe a Condorcet cycle among policies: $YM \prec YR \prec N \prec YM$. Thus, allowing the legislature to commit to a state-contingent plan at the outset is unlikely to resolve the problem. It only leads to disagreements over the plan to follow and, in particular, will not rule out the status quo as a viable option.[2]

We begin our analysis by showing that the conclusion of this motivating example extends to any setting in which all decisions (in the game and at the commitment stage) are made by simple majority rule. We show that given any payoffs, either the dynamic equilibrium is undominated or there is a Condorcet cycle with commitment, which involves both the policy corresponding to the dynamic equilibrium and the policy dominating it.

Allowing for commitment under the simple majority rule thus replaces any problem of *inefficiency* with one of *indeterminacy*, even when commitment carries no administrative or other contractual costs and is perfectly credible. Either commitment is unnecessary, as when the equilibrium is majority preferred to all other state-contingent policies, or it is impossible to agree on which

[1] The game is solved by backward induction, using the elimination of weakly dominated strategies as a refinement. If a modest reform is launched, and the expansion turns out to be feasible, then a majority consisting of B and C votes to institute a radical reform. Anticipating this, A initially votes for the reform if: $E(u_A) = -2q+1-q \geq 0 \iff q \leq \frac{1}{3}$. B votes for the initial reform regardless of q, because B benefits whether or not an expansion is feasible. C supports the reform if: $E(u_C) = q - 2(1-q) \geq 0 \iff q \geq \frac{2}{3}$. Overall, there is a majority in favor of the initial reform at the outset if $q \leq 1/3$ (supported by A and B) or $q \geq 2/3$ (supported by B and C). But in case $1/3 < q < 2/3$, A and C join forces, so that a majority opposes the reform at the outset.

[2] The problem persists even if the status quo is Pareto inefficient, as can be seen from a slight modification of the game: Suppose that, after implementing the reform Y there is a new action, P, giving 1/2 to all voters regardless of the resolution of uncertainty. The policy YP Pareto dominates the status quo, yet P is always majority dominated in the second stage. The status quo equilibrium and Condorcet cycle persist despite the status quo being Pareto dominated.

policy to commit to.

How can this result be reconciled with the apparent value of commitment indicated, for example, by the prevalence of contracts? To study this issue at a fundamental level, we proceed to consider general structures of political power. These may include the use of supermajority rules for some decisions and heterogenous allocations of power across agents. These details, it turns out, do not matter *per se* for the value of commitment.

Instead, what matters is a *power consistency* condition relating political power at the dynamic and commitment stages. When power consistency holds, the introduction of commitment suffers from the same problem as under majority voting: whenever it is potentially valuable, there is a cycle among all commitment policies. Furthermore, the power consistency condition is necessary for this negative result: when power consistency is violated, one may find a preference profile and a policy that dominates not just the equilibrium but also all other policies that are available with commitment.

We consider environments in which decisions in each period are binary and made according to arbitrary—possibly time-varying and state-dependent—voting rules.[3] Power consistency is defined by the following requirement: Consider two policies which are identical except for the decision made in a given period and for a given state (or subset of states) in this period. Then, the social ranking between these two policies must be determined by the same set of winning coalitions as the one arising in the dynamic game when that decision is reached. The condition thus rules out situations in which a subset of persons could impose one policy over another at the commitment stage, but would not be able impose that choice in the dynamic game.

We explore in detail when one should expect power consistency to hold and when it is likely to be violated. Power consistency reflects the notion that the importance of the decision is the same whether it is considered in the dynamic game or at the commitment stage. For example, if an important decision requires unanimity in the dynamic setting, power consistency rules out using simple majority at the commitment stage to compare policies differing only with respect to this decision. In other settings, power consistency captures a notion of fairness toward future

[3]The focus on binary decisions eliminates "local" Condorcet cycles in each period and thus also an important potential source of indeterminacy which may confuse the main points of the paper. Thanks to this assumption, the cycles which may arise among state-contingent policies have nothing to do with possible cycles in any given period. The focus on binary decisions also results in a unique equilibrium, which simplifies the statements of the paper.

generations. The condition prohibits current society members from committing to future actions which are contrary to the interest of future society members, who would normally be the ones deciding on these actions. Power consistency may also capture a notion of liberalism similar to the one described by Sen (1970): the social ranking of policies should respect the preferences of individuals who would naturally be making decisions in the dynamic setting.

Violations of power consistency are reasonable in some contexts. In particular, it is well-known that commitment is valuable for a time-inconsistent agent. Time inconsistency creates a particular form of power inconsistency which favors the first-period self (or preference) of the agent. Similarly, current actors or generations may be able to lock in future decisions in various macroeconomic and political economy contexts we discuss in Section 8. In these cases, commitment has value.

Even when power consistency holds, it may be possible to circumvent the indeterminacy result by imposing some restrictions on the set of available commitments. For example, if some players are ex ante symmetric, it seems reasonable to focus on "anonymous" (i.e., non-discriminatory) policies, which treat these players identically at the commitment stage, by giving each of them the same outcome distribution. To explore this idea, we develop a concept of "anonymous" policy which requires that similar agents be treated similarly—in a sense which we formalize—as well as another anonymity criterion based on a veil-of-ignorance argument (Section 7).[4] Both approaches can restore the value of commitment, either by removing policies which appeared in the cycle or by modifying the individual criteria used to assess policies.

We also show how some forms of commitment, such as a unilateral commitment or a commitment to vote in a particular way on a future collective decision, may be built into the underlying game—one advantage of the generality of the model considered here is precisely to allow for this. The equilibrium of the augmented game may then become efficient due to these commitments, removing indeterminacy in accordance with our results.

Arguments in favor of commitment, which appear in various literatures, have either presumed a violation of power consistency, such as the time inconsistency emphasized by Kydland and Prescott (1977) in macroeconomic settings, or they have not considered all possible policies to which one might commit, as discussed in Sections 5 and 8. In some cases, commitment restrictions such as

[4]Tabellini and Alesina (1990) use a similar argument to show that commitment to balanced budgets is valuable when agents do not know who will be in the position of power.

anonymity may be appealing. In general, however, our result calls for caution because, whenever power consistency holds and the set of policies under consideration is not restricted by some external criterion, there must be a cycle. Unless there is a reason why the switch from dynamic decisions to comparing policies puts a different group in charge, or certain commitments can be a priori ruled out, commitment will not resolve inefficiency.

Our result lends itself to two interpretations: a negative one pointing to the failure of commitment and a positive one emphasizing the value of the equilibrium policy, both of which find support in earlier literatures. The interpretation of commitment cycles as a failure to resolve equilibrium inefficiency is consistent with Boylan and McKelvey (1995), Boylan, Ledyard, and McKelvey (1996), and Jackson and Yariv (2015) who show that, when agents have heterogeneous discount factors, no agreement can be reached over consumption streams because no Condorcet winner exists. The absence of a Condorcet winner weakens the applicability and value of commitment, as in our paper. When cycles do *not* occur, our result casts a new light on positive results such as those of Acemoglu, Egorov, and Sonin (2012) and Acemoglu, Egorov, and Sonin (2015) who provide single-crossing conditions on agents' preferences guaranteeing that the equilibrium is undominated and a dynamic version of the median voter theorem holds. Indeed, our result shows that these single-crossing conditions—or any condition, for that matter—break all cycles only if the equilibrium itself is undominated.

While the paper's main conceptual contribution concerns the value of commitment in various applied settings, its analytical content contributes to voting theory in two ways. The literature on agenda setting has pointed out (e.g., Miller (1977)) that if the winner of a sequence of binary majority votes over alternatives depends on the order in which alternatives are compared, then there is no Condorcet winner among these alternatives.[5] Our framework extends the agenda setting literature by i) allowing uncertainty: the state of the world (physical reality, information, individual preferences) can evolve stochastically over time, and ii) considering general decision protocols in which decision rules and individual power may change and can depend on past decisions and events. These extensions are relevant in numerous applications—risky reforms, search by committees, theory

[5]In a static choice problem, Zeckhauser (1969) and Shepsle (1970) study the existence of Condorcet winners in voting over certain alternatives and lotteries over them. Zeckhauser shows that, if all lotteries over certain alternatives are in the choice set, no Condorcet winner can be found, even if there is such a winner among certain alternatives. In a comment on Zeckhauser, Shepsle demonstrates that a lottery can be a Condorcet winner against certain alternatives that cycle.

of clubs, to cite a few—and require a more sophisticated analysis.[6,7]

Sections 2 and 3 describe our main result for the simple majority rule and the general case, respectively. Interpretations of the power consistency condition, and violations thereof, are discussed in Section 4. Section 5 presents two applications. Generalizations of our model are considered in Section 6. Section 7 investigates two anonymity criteria which can be used to restore the value of commitment, even when the power consistency condition holds. Section 8 discusses the role of commitment in various literatures, demonstrating applications and violations of the power consistency condition introduced in this paper. The appendix contains an omitted proof. The online appendix reviews conditions for the existence of a Condorcet winner, shows how our ideas can be adapted to an infinite horizon, and gives an example illustrating the difference between our result and agenda setting approaches.

2 Simple Majority Rule

There are T periods and N (odd) voters. Each period starts with a publicly observed state $\theta_t \in \Theta_t$, which contains all the relevant information about past decisions and events. At each t, a collective decision must be made from some binary set $A(\theta_t) = \{\underline{a}(\theta_t), \bar{a}(\theta_t)\}$. This choice, along with the current state, determines the distribution of the state at the next period. Formally, each Θ_t is associated with a sigma algebra Σ_t to form a measurable space, and θ_{t+1} has a distribution $F_{t+1}(\cdot \mid a_t, \theta_t) \in \Delta(\Theta_{t+1})$. If, for instance, the state θ_t represents a belief about some unknown state of the world, θ_{t+1} includes any new information accrued between periods t and $t+1$ about the state, which may depend on the action taken in period t. The state θ_t may also include a physical component, such as the current stage of construction in an infrastructure-investment problem.

[6]Without uncertainty, any policy reduces to a single path in the dynamic game, and can be identified with its unique terminal node. Each policy then corresponds to an "alternative" in the agenda setting literature. With uncertainty, this relation breaks down because policies are state-contingent plans which can no longer be identified with terminal nodes. Choosing among policies at the commitment stage is then no longer equivalent to making a sequence of binary choices in the dynamic game. Notably, one may construct examples—as in the online appendix—in which reversing the order of moves in a dynamic game does not change the outcome, yet there is a Condorcet cycle among all state-contingent plans.

[7]In the theory of clubs (Roberts (2015)), an early decision to admit new members dilutes the power of preceding members and, hence, affects subsequent choices made by the club. Similarly, political power may be redistributed as agents learn their preferences through experimentation and form new political alliances, as in Strulovici (2010). These potential changes affect incentives early on and, without commitment, distort the equilibrium away from efficiency. Unlike the tournament literature (Laslier (1997)), the analytical framework considered here allows the set of winning coalitions for any given decision to depend on past decisions and events.

Let $\Theta = \bigcup_{t=1}^{T} \Theta_t$ and $A = \bigcup_{\theta \in \Theta} A(\theta)$ denote the sets of all possible states and actions. Each voter i has a terminal payoff $u_i(\theta_{T+1})$, which depends on all past actions and shocks, as captured by the terminal state θ_{T+1}. A *policy* $C : \Theta \to A$ maps at each period t each state θ_t into an action in $A(\theta_t)$.

If a policy C is followed by the group, then given state θ_t, i's expected payoff seen from period t is[8]

$$V_t^i(C \mid \theta_t) = E[u_i(\theta_{T+1}) \mid \theta_t, C].$$

From here onwards, as is standard in the tournaments literature, for simplicity we shall require that no voter is indifferent between the two actions in $A(\theta_t)$ at any state θ_t.[9]

Given a policy C and state θ, let C_θ^a denote the policy equal to C everywhere except possibly at state θ, where it prescribes action $a \in A(\theta)$.

Definition 1 (Voting Equilibrium)**.** *A profile* $\{C^i\}_{i=1}^N$ *of voting strategies forms a* Voting Equilibrium *if and only if*

$$C^i(\theta_t) = \arg \max_{a \in A(\theta_t)} V_t^i(Z_{\theta_t}^a \mid \theta_t)$$

for all $\theta_t \in \Theta$*, where* Z *is the policy generated by the voting profile:*

$$Z(\theta_t) = a \in A(\theta_t) \text{ if and only if } |C^i(\theta_t) = a| \geq \frac{N}{2}.$$

Z is defined by simple majority voting: at each time, society picks the action that garners the most votes. The definition captures the elimination of weakly dominated strategies: at each t, voter i, taking as given the continuation of the collective decision process from period $t+1$ onwards that will result from state θ_{t+1}, votes for the action that maximizes his expected payoff as if he were pivotal.

Because, by assumption, indifference is ruled out and the horizon is finite, this defines a unique voting equilibrium, by backward induction. The proof of this fact is straightforward and omitted.

[8]The utility function may depend arbitrarily on past states and decisions. For example, the formulation allows decision complementarities across periods and all forms of path-dependence, such as habit formation, addiction, taste for diversity, utility from memories, learning by experimentation, learning by doing, etc. Because the terminal state θ_{T+1} includes past states, this formulation includes the time-separable case where $u_i(\theta_{T+1}) = \sum_{t=1}^{T+1} u_{i,t}(\theta_t)$ for some period-utility functions $u_{i,t}$, as well as non-time-separable utility functions.

[9]The literature on tournaments assumes that preference relations across alternatives are asymmetric. See Laslier (1997). Without this strictness assumption, most of Theorem 1 still applies to "weak" Condorcet winners and cycles. See also Remark 1.

Proposition 1. *There exists a unique voting equilibrium.*

Commitment and Indeterminacy

Given a pair (Y, Y') of policies, we say that Y dominates Y', written $Y \succ Y'$, if there is a majority of voters for whom $V_1^i(Y \mid \theta_1) > V_1^i(Y' \mid \theta_1)$. A Condorcet cycle is a finite list of policies $Y_0, \ldots, Y_K$ such that $Y_k \prec Y_{k+1}$ for all $k < K$, and $Y_K \prec Y_0$. Finally, X is a Condorcet winner if, for any Y, either $X \succ Y$ or X and Y induce the same distribution over Θ_{T+1}.

Theorem 1. *Let Z denote the equilibrium policy.*

i) If there exists Y such that $Y \succ Z$, then there is a Condorcet cycle including Y and Z.

ii) If there exists a policy X that is a Condorcet winner among all policies, then X and Z induce the same distribution over Θ_{T+1}.

Remark 1. *If voters' preferences allow ties, Part i) still holds with a weak Condorcet cycle: there is a finite list of policies $Y_0, \ldots, Y_K$ such that $Y_k \preceq Y_{k+1}$ for all $k < K$, and $Y_K \prec Y_0$. Furthermore, Z continues to be a Condorcet winner in the sense that there does not exist another policy Y such that $Z \prec Y$.*

The proof is subsumed by the proof of Theorem 2 and is therefore omitted. The intuition of the proof is as follows: If a policy Y differs from the equilibrium policy Z, then Y must necessarily prescribe, for some states reached with positive probability, actions which the majority opposes. Using this observation, we iteratively construct a sequence of policies by gradually changing Y in these states, in the direction of the majority's will, so that each subsequent policy is majority preferred to the previous one. Because the game is finite, this process eventually ends with the policy Z where all actions follow the majority's preference. More explicitly, we start with the last period, t, for which Y differs from Z on some subset of states. We then create a new policy, Y_1, identical to Y except in some time-t state for which Y differs from Z. On these states (that are reached with positive probability), Y takes an action that is not supported by a majority, since Y and Z have the same continuation by definition of t, and Z was the equilibrium policy. Moreover, Y_1 is now closer to Z as it takes the same actions as Z on the state in which the change took place. We then apply the procedure to another time-t state for which Y_1 (and thus Y) prescribes

a different action from Z, creating a new policy Y_2, which is identical to Y_1 except for taking the majority preferred action in this state. By construction $Y_1 \prec Y_2$. Once all time-t states for which Y differs from Z have been exhausted by the procedure, we move to time $t-1$ and repeat the sequence of changes, constructing a chain of policies which are increasing in the majority ranking and getting gradually more similar to the equilibrium policy, Z. The process ends with a policy Y_K that coincides with Z. Because we know that Y is different from Z, $K \geq 2$, which creates a Condorcet cycle if and only if the initial policy Y dominated Z.

The cycles predicted by Theorem 1, whenever they occur, may be interpreted as follows: If the population were allowed, before the dynamic game, to commit to a policy, it would be unable to reach a clear agreement, as any candidate would be upset by some other proposal. If one were to explicitly model such a commitment stage, the outcome of this stage would be subject to well-known agenda setting and manipulation problems, and the agenda could in fact be chosen so that the last commitment standing at that stage be majority defeated by the equilibrium of the dynamic game.

Theorem 1 distinguishes two cases: when the equilibrium is undominated and when there is no Condorcet winner. These cases can often coexist in the same model, for different parameter values. This was the case in the slippery slope example, where the equilibrium is undominated for $q \in [0, 1/3] \bigcup [2/3, 1]$ and no Condorcet winner existed for $q \in (1/3, 2/3)$.

A more positive interpretation of Theorem 1 is that, even when the equilibrium policy is majority dominated by another policy, it must belong to the top cycle of the social preferences based on the majority ranking.[10] In the agenda-setting literature, it is well-known that the equilibrium must belong to the Banks set (Laslier (1997)). This need not be the case here, however, due to the presence of uncertainty, because the dynamic game does not give voters enough choice to compare all policies: the decision set is just not rich enough. In particular, with T periods agents make only T comparisons throughout the dynamic game, but policies, being state-contingent plans, are much more numerous when the state is uncertain. As a result, the equilibrium does not *per se* inherit the Banks-set property.

Another way of understanding the difference between the alternatives compared in the agenda-setting literature and the policies compared in our framework is that a state-contingent policy now corresponds to a *probability distribution* over terminal nodes, and in the dynamic voting game agents

[10] Even then, however, the equilibrium policy may be Pareto dominated by another policy, as shown in Footnote 2.

do not have rich enough choices to express preferences amongst all these distributions. Put in the more formal language of tournaments, the choice process along the dynamic game may not be summarized by a complete algebraic expression for comparing all policies (Laslier (1997)), leading to substantive differences between our model and the earlier literature.

3 General Voting Rules and Power Consistency

Collective decisions often deviate in essential ways from majority voting. In the slippery slope problem, for example, some decisions may be taken by a referendum and others by lawmakers. Another natural example concerns constitutional amendments in the United States, which require a supermajority rule. This section shows that our main result still holds for arbitrary decision rules, under a *power consistency* condition whose meaning and relevance are discussed in detail below.

The formal environment is the same as before except for the structure of political power.[11] Given a period t and state θ_t, the "high" action $\bar{a}(\theta_t)$ might, for instance, require a particular quorum or the approval of specific voters (veto power) to win against $\underline{a}(\theta_t)$. The decision rule may also depend on the current state and, through it, on past decisions. In many realistic applications, some voters may be more influential than others because they are regarded as experts on the current issue, or because they have a greater stake in it, or simply because they have acquired more political power over time.

To each state θ_t corresponds a set $\bar{\mathcal{S}}(\theta_t)$ of coalitions which can impose $\bar{a}(\theta_t)$ in the sense that, if all individuals in $S \in \bar{\mathcal{S}}(\theta_t)$ support $\bar{a}(\theta_t)$, then $\bar{a}(\theta_t)$ wins against $\underline{a}(\theta_t)$ and is implemented in that period. Likewise, there is a set $\underline{\mathcal{S}}(\theta_t)$ of coalitions which may impose $\underline{a}(\theta_t)$. These sets are related as follows: $\underline{\mathcal{S}}(\theta_t)$ contains all coalitions whose complement does not belong to $\bar{\mathcal{S}}(\theta_t)$, and vice versa. We impose the following condition: for any coalitions $S \subset S'$ and state θ, $S \in \bar{\mathcal{S}}(\theta) \Rightarrow S' \in \bar{\mathcal{S}}(\theta)$. This monotonicity condition implies that it is a dominant strategy for each individual to support their preferred action, for any given state: they can never weaken the power of their preferred coalition by joining it.

A *coalitional strategy* C^i for individual i is, as before, a map from each state θ_t to an action in $A(\theta_t)$. It specifies which action i supports in each state. Given any profile $\mathbf{C} = (C^1, \ldots, C^N)$

[11] The number of voters need not be odd any more.

of coalitional strategies and any state θ, there are two coalitions: those who prefer $\bar{a}(\theta)$ and those who prefer $\underline{a}(\theta)$, and one of them is a winning coalition that can impose its preferred action.[12] Let $a(\mathbf{C}, \theta)$ denote this action.

Definition 2 (Coalitional Equilibrium)**.** *A profile* $\{C^i\}_{i=1}^N$ *of coalitional strategies forms a* Coalitional Equilibrium *if and only if*

$$C^i(\theta_t) = \arg \max_{a \in A(\theta_t)} V_t^i(Z_{\theta_t}^a \mid \theta_t)$$

for all $\theta_t \in \Theta$*, where* Z *is is the policy generated by the profile:* $Z(\theta_t) = a(\mathbf{C}, \theta_t)$*).*

The definition is the same as for majority voting, except that now the action that wins in each period is the one supported by the strongest coalition. We maintain the assumption of the previous section that each voter has, for any policy and state θ_t, a strict preference for one of the two actions in $A(\theta_t)$. Because indifference is ruled out and the horizon is finite, this defines a unique coalitional equilibrium, by backward induction (the proof is omitted).

Proposition 2. *There exists a unique coalitional equilibrium.*

Commitment and Indeterminacy

Now suppose that voters are given a chance to collectively commit to a policy instead of going through the sequence of choices in the dynamic game. When can they agree on a policy that dominates the equilibrium? We need to specify the structure of power at the commitment stage. Given a pair (Y, Y') of distinct policies, say that S is a winning coalition for Y over Y' if $Y \succ Y'$ whenever all members of S support Y over Y' when the two policies are pitted against each other. A power structure specifies the set of winning coalitions for every pair of alternatives. Given a power structure and a profile of individual preferences over all distinct policies, one can then construct the social preference relation, which describes the pairwise ranking of every two alternatives: $Y \succ Y'$ if and only if there is a winning coalition S for Y over Y' all of whose members prefer Y to Y'.

Given the social preference relation $\succ$, say that a policy Y is a *Condorcet winner* if there is no other policy Y' strictly preferred over Y by a winning coalition. A *Condorcet cycle* is defined

[12] That is, the coalition of individuals preferring $\bar{a}(\theta_t)$ belongs to $\bar{\mathcal{S}}(\theta)$ if and only its complement does not belong to $\underline{\mathcal{S}}(\theta)$.

as in the previous section with the only difference that $\succ$ is used instead of the simple majority preference relation.[13]

Our main result relies on a consistency condition relating the power structures in the dynamic game and at the commitment stage.

Definition 3 (Power Consistency)**.** *Suppose that Y and Y' differ only on a set $\bar{\Theta}_t$ of states corresponding to some given period t and that S is a winning coalition imposing the action prescribed by Y over the one prescribed by Y' for all states in $\bar{\Theta}_t$. Then, S is also a winning coalition at the commitment stage, imposing Y over Y'.*

Although the power structure at the commitment stage must specify the set of winning coalitions for every pair of policies, the power consistency condition is only concerned with a much smaller subset of those pairs, namely the pairs for which the two policies are identical except on a subset of states in a single period.

Theorem 2. *Assume power consistency, and let Z denote the equilibrium of the coalitional game.*

i) If there exists Y such that $Y \succ Z$, then there is a Condorcet cycle including Y and Z.

ii) If some policy X is a Condorcet winner among all policies, then X and Z must induce the same distribution over Θ_{T+1}.

Proof. We fix any policy Y and let $\bar{\Theta}_T$ denote the set of states, in the last period, for which Y and the coalitional equilibrium Z prescribe different actions: $\bar{\Theta}_T = \{\theta_T \in \Theta_T : \ Z_T(\theta_T) \neq Y_T(\theta_T)\}$. For each $\theta_T \in \bar{\Theta}_T$, we denote by $\mathcal{S}_T(\theta_T)$ the coalition of individuals who prefer $Z_T(\theta_T)$ to $Y_T(\theta_T)$. $\mathcal{S}_T(\theta_T)$ must be a winning coalition when state θ_T is reached, since it imposes action $Z_T(\theta_T)$ in equilibrium. Finally, let $\mathcal{S}_T = \{\mathcal{S}_T(\theta_T) : \theta_T \in \bar{\Theta}_T\}$ denote the set of all such coalitions and n_T denote the cardinality of $\mathcal{S}_T$. Since there are finitely many possible coalitions, n_T is finite. We index the elements of $\mathcal{S}_T$ arbitrarily from S_1 to S_{n_T} and, for each $n \leq n_T$, let Θ_T^n denote the set of states $\theta_T \in \bar{\Theta}_T$ for which the set of individuals preferring $Z_T(\theta_T)$ to $Y_T(\theta_T)$ equals S_n *and* forms a winning coalition. In this final period, power consistency intuitively means that decisions picked by Z should be ranked higher than those by Y, because these decisions are preferred by winning

[13]These generalizations of majority-voting concepts to general tournaments are standard. See, e.g., Laslier (1997).

coalitions, and the ranking of policies should reflect this. To capture this intuition, we inductively construct the following sequence $\{Y_T^n\}_{n=1}^{n_T}$ of policies:

- Y_T^1 is identical to Y for all periods $t < T$, as well as for all states of period T except those in Θ_T^1 where Y_T^1 takes the same action as Z;
- for each $n \in \{2, \ldots, n_T\}$, Y_T^n is identical to Y_T^{n-1} for all periods $t < T$ and states of period T, except for those states of Θ_T^n where it takes the same action as Z.

By construction, Y_T^1 and Y differ only in period T, and do so over a set of states for which the winning coalition, S_1, prefers Y_T^1's action to Y's. By power consistency, this implies that $Y_T^1 \succeq Y$. Moreover, this preference is strict if and only if these states are reached with positive probability by policy Y. By induction, $Y_T^n \succeq Y_T^{n-1}$ for all $n \leq n_T$, with a strict inequality if and only if the set Θ_T^n of states for which Y_T^n and Y_T^{n-1} prescribe different actions is reached with positive probability under Y. Finally, $Y_T^{n_T}$ is identical to Z in the final period, because the constructed sequence has sequentially flipped each action of Y which differed from Z in that period.

The same transformation is applied, by backward induction, to each period from $T-1$ to 1. For period t, we let $\bar{\Theta}_t = \{\theta_t \in \Theta_t : \ Z_t(\theta_t) \neq Y_t(\theta_t)\}$ and, for $\theta_t \in \bar{\Theta}_t$, $\mathcal{S}_t(\theta_t)$ denote the coalition of individuals who prefer $Z_t(\theta_t)$ to $Y_t(\theta_t)$.[14] $\mathcal{S}_t(\theta_t)$ is a winning coalition at state θ_t because it imposes its preferred action in equilibrium. Moreover, $\mathcal{S}_t(\theta_t)$ prefers $Z_t(\theta_t)$ to $Y_t(\theta_t)$ *given that for both Z is the continuation policy from $t+1$ onwards* . This latter observation motivates our use of backward induction: after applying the transformation to periods T, $T-1$ down to $t+1$, all continuation policies of interest are identical to the coalitional equilibrium Z from time $t+1$ onwards.

Let $\mathcal{S}_t = \{\mathcal{S}_t(\theta_t) : \theta_t \in \bar{\Theta}_t\}$ and let n_t denote the cardinality of $\mathcal{S}_t$. We index the coalitions of $\mathcal{S}_t$ arbitrarily from S_1 to S_{n_t}, and let Θ_t^n denote the set of states in $\bar{\Theta}_t$ for which the set of individuals preferring $Z_t(\theta_t)$ to $Y_t(\theta_t)$ is equal to S_n *and* forms a winning coalition. As with period T, we iteratively construct a sequence $\{Y_t^n\}_{n=1}^{n_t}$ of policies increasing n within each period t, and then moving backward by one period: for each t,

- Y_t^1 is identical to $Y_{t+1}^{n_{t+1}}$ except in period t, over Θ_t^1, where it takes the same action as Z;
- for each $n \in \{2, \ldots, n_t\}$, Y_t^n is identical to Y_t^{n-1} except in period t, over Θ_t^p, where it takes the same action as Z.

[14] By construction, the last modified policy and Z coincide from period $t+1$ onwards, except, possibly, on states that cannot be reached with positive probability under either policy.

These policies have the same continuation, Z, from period $t+1$ onwards. By construction, moreover, $Y_t^{n+1} \succeq Y_t^n$ for all t, and $n < n_t$ and $Y_t^1 \succeq Y_{t+1}^{n_{t+1}}$ for all t, with a strict inequality whenever the set of states over which the policies being compared differ is reached with positive probability by policy Y.

The terminal policy, $Y_1^{n_1}$, which this algorithm generates, is by construction identical to Z. Let $\{Y_k\}_{k=1}^K$, $K \geq 1$ denote the sequence of *distinct* policies generated, starting from Y, by this construction, i.e., policies which induce distinct distributions over Θ_{T+1} (policies differing only at states which are not reached under either policy are not distinct).

If Y and Z are distinct, then necessarily $K \geq 2$, since the construction starts at Y and ends with Z. Moreover, we have

$$Y = Y_1 \prec Y_2 \cdots \prec Y_K = Z.$$

Therefore, a preference cycle must arise if $Z \prec Y$, which proves part i) of the theorem.

Moreover, the previous argument also shows that any Condorcet winner Y must be identical to Z except on a set of states which is reached with zero probability: otherwise the sequence $\{Y_k\}$ would include at least two elements and imply that Y is directly dominated by at least one policy, a contradiction. This proves part ii) of the theorem. □

Theorem 2 implies that when pairwise comparisons of policies are based on the same power structure as the one used in the binary decisions of the dynamic game, allowing commitment does not lead to an unambiguous improvement of the equilibrium. While some agenda setter may propose a commitment to resolve political inertia, such a commitment can be defeated by another commitment proposal, and so on, bringing us back to political inertia. While one may find some solace in the fact that the equilibrium policy is part of the top cycle among policies, it may be Pareto dominated by another policy and payoffs may be chosen so as to make the domination arbitrarily strong.

Remark 2. *As with Theorem 1, a modification of Theorem 2 based on weak Condorcet cycles and weak Condorcet winners holds when agents are allowed to have weak, instead of strict, preferences.*

The model of this section, by allowing history-dependent power structures, extends the agenda-setting and tournament literatures, which have assumed that the pairwise ranking of "alternatives" was prescribed by a single binary, complete, asymmetric relation (tournament), regardless of how

or when these alternatives were compared. In dynamic settings such as ours, where each decision affects the balance of power for future decisions, this invariance assumption is typically violated.

The Necessity of Power Consistency

When power consistency fails, one may find some policies which are unambiguously preferred to the equilibrium. More precisely, we will say that the power structures used in the dynamic and commitment stages are *inconsistent* if there exist policies Y and Y' and a coalition S such that i) Y and Y' are identical, except for a subset $\bar{\Theta}_t$ of states of some given period t, reached with positive probability under policy Y (and hence Y'), ii) whenever a state $\theta_t \in \bar{\Theta}_t$ is reached in the dynamic game, S is a winning coalition imposing the action prescribed by Y' over the one prescribed by Y, iii) at the commitment stage, S does not belong to the set of winning coalitions imposing Y' over Y.

Theorem 3. *Suppose that the power structures are inconsistent across stages. Then, there exist utility functions* $\{u_i(\theta_{T+1})\}_{i\in\{1,\dots,N\},\theta_{t+1}\in\Theta_{T+1}}$ *and a policy* X *such that the equilibrium* Z *is strictly dominated by* X *and* X *is a Condorcet winner.*

Combined with Theorem 2, Theorem 3 shows that power consistency captures the essence of the phenomenon studies in this paper: it characterizes the power structures in the dynamic game and at the commitment stage for which commitment is ineffective regardless of the players' payoffs, preferences, and information.

4 Interpreting Power Consistency

When does power consistency hold?

The simplest instance of our setting is when the same set of agents is making decisions at the dynamic and commitment stages, and these agents are time consistent. In this case, power consistency may be interpreted and justified in the following ways.

Expertise: Some decisions (choosing an energy policy, addressing international conflicts, setting monetary policy, etc.) require specific expertise. For these decisions, the power should lie with

experts, both when these decisions are made in the dynamic game and when comparing policies which differ only with respect to these decisions.

Liberalism: Some decisions primarily concern specific subgroups of the population (e.g., city or statewide decisions, rules governing some associations, etc.). It seems natural to let these groups have a larger say over these decisions both at the dynamic and the commitment stages. This consideration is related to Sen's notion of "liberalism" (Sen (1970)), a link explored further in this section. It may also be applied to minority rights.

Supermajority: Many constituencies require a supermajority rule to make radical changes to their governing statutes. For example, amendments to the United States constitution require two-thirds of votes in Congress, and substantive resolutions by the United Nations Security Council require unanimity. Power consistency says that rules should treat these radical changes consistently, whether they are part of a commitment or arise in the dynamic game.

Intergenerational altruism/liberalism: When a decision primarily concerns unborn generations, the relevant social preference relation may, normatively, take into account the preferences of these generations—which may depend on the future state—even though they are absent at the time of commitment. Today's generation is then guided by intergenerational altruism when considering commitments.

Departing generations: Conversely, some agents may die or leave the dynamic game following some actions or exogenous shocks. It is then reasonable to ignore them when comparing policies that differ only with respect to decisions arising after they left the game, which is captured by power consistency.

When is power consistency violated?

At the opposite extreme, another view of future generations is to simply ignore them in the social ranking of policies. This approach violates power consistency, and the current generation will typically find commitment valuable in this case.

Myopic/selfish generation: The current generation ignores the welfare and preferences of future generations. Power consistency is then violated, and this is exposed when the preferences of future generations are in conflict with those of the commitment-making generation.

Time inconsistency: Time-inconsistent agents violate power consistency because their initial ranking of social alternatives is not representative of their preferences when they make future decisions. One may think of a time-inconsistent agent as a succession of different selves, or agents, each endowed with preferences. At time t, the t-self of the agent is in power; he is the dictator and the unique winning coalition. When considering commitment at time 0, however, only the initial preferences of the agent are used to rank policies, which violates power consistency.

Commitment is deemed valuable in this case, but only because it is assessed from the perspective of the first-period agent. If one were to take the agent's preferences at various points in time into account, the value of commitment would be subject to the indeterminacy pointed out in Theorem 2.[15]

These observations extend to multiple agents. For example, a set of perfectly identical but time inconsistent agents would obviously face the same issues as a single time-inconsistent agent, regardless of the voting rule adopted in each period. Again, power consistency is violated if future selves have different preferences and their choices are not respected at time zero.

A similar source of time inconsistency concerns institutions whose government changes over time, bringing along different preferences. If an incumbent government can commit to a long-term policy which ties future governments' hands, it will typically find such a commitment valuable, and this commitment may increase overall efficiency. In Tabellini and Alesina (1990) and Alesina and Tabellini (1990), for instance, governments alternate because political power shifts over time (e.g., voting rights are gained by some minorities), changing the identity of the median voter, even though each voter taken individually has a time-consistent preference. The incumbent government borrows too much relative to the social optimum because it disagrees with how future governments will spend the remaining budget. When future governments cannot affect the choice of a commitment policy, the power consistency condition fails. When they can, our theorem applies and a cycle arises among

[15]The agent's preferences in the first period may incorporate his future preferences, and this very fact may be the source of the agent's time inconsistency, as in Galperti and Strulovici (2014). However, agent's future preferences do not *directly* affect his ranking of policies at time 1.

commitment policies. One way out of this cycle is to put all agents behind a veil of ignorance, as suggested by Tabellini and Alesina (1990). We explore this possibility in detail in Section 7.

Law of the current strongest: Another form of power inconsistency arises when some agents become more politically powerful over time. Their influence on future decisions in the dynamic game extends above and beyond their power at the commitment stage. These power changes may be foreseeable or random, depending on the economic or political fortunes of individuals at time zero. Regardless of the cause, commitment may be valuable as a way to insulate future decisions from the excessive power gained by a small minority. Power consistency is violated because the evolution of individual power is not included in the commitment decision.

Choosing future voting rules

In some applications (Barbera, Maschler, and Shalev (2001), Barbera and Jackson (2004)), earlier decisions determine the voting rule used for subsequent decisions. More generally, early decisions can affect each agent's weight in voting on future decisions. This possibility is allowed by our framework because the state θ_t includes any past decision and determines the set of winning coalitions at time t. Settings where the future allocation of political power is determined by current agents appear in the theory of clubs (Roberts (2015)) or in mayoral elections (Glaeser and Shleifer (2005)). Barbera, Maschler, and Shalev (2001) consider voters deciding on immigration policies that would expand their ranks, while Barbera and Jackson (2004) study the general problem of voters deciding today on voting rules that will be used in the future.

We now discuss in the context of an example whether power consistency should be expected to hold and what Theorem 2 means when power is endogenous. We start with a two-period model. In period 1, a first generation of voters, assumed for now to be homogeneous, chooses the voting rule for period 2, between simple majority and two-thirds majority. In period 2, the next generation votes on whether to implement a reform. It is assumed that a fraction $x \in [1/2, 2/3)$ of period-2 voters favors the reform. In this case, the period-1 generation can obtain whichever outcome it prefers for period 2, by choosing the voting rule appropriately. Whether power consistency holds is irrelevant, because period 2 voters really have no control over the outcome as they are split in their preferences and bound by the voting rule chosen by their elders. In particular, one may assume that the condition holds so that the conclusions of Theorem 2 apply. Here, the equilibrium is efficient for

the first generation and dominates any other policy from their perspective, so we are in the scenario where a Condorcet winner exists and coincides with the equilibrium, as predicted by the theorem.

Suppose next that there is a third period, and that the voting rule chosen by the first generation must also be used for the period-3 decision, with x taking the same value as in period 2. To make the problem interesting, we assume that the first generation wishes to implement the reform in period 2 but not in period 3. In this case, choosing a voting rule in period 1 cannot provide an efficient outcome from the first generation's perspective, and committing to a long-term policy clearly increases that generation's utility. Power consistency is violated because the third generation's power to choose the reform in the third period is not reflected in the social comparisons of policies, which are exclusively based on the first generation's preferences.

Finally, suppose that the three generations are in fact made up of the same individuals at different times. There is a fraction x of people who prefer the reform in the second period and the same fraction x of (partially different) people who support it in the third period. Also suppose that the first-period choice, deciding on which voting rule to use in later periods, is made according to the simple majority rule. If in equilibrium the first-period decision is to use the simple majority rule for future periods, the reform is adopted in both periods. If instead the two-thirds majority rule is chosen in the first period, no reform is adopted in later periods. Suppose that the two-thirds majority rule is chosen in equilibrium. This means that there is a majority of individuals who dislike the reform in at least one period, so much so that they prefer the status quo to having the reform in both periods, even though there is also a majority (x) of people who, in each period, prefer the reform to be implemented in that period.[16] If we use the simple majority rule when comparing any pair of policies other than the pairs differing only in one period, there is a cycle across policies: a majority of people prefer no reform at all (Z) to both reforms (Y), but a majority prefers reform in period 1 only (X) to Z, and a majority prefers reform in both periods (Y) to X, so that $Y \succ X \succ Z \succ Y$. Power consistency seems reasonable in this setting: whatever decision is made in the dynamic game reflects the preferences of the population at the beginning of the game. The theorem applies and, since the equilibrium is dominated by reform in either period, we get a

[16] For example, suppose that $x = 3/5$ and the $2/5$ who oppose the reform in any given period dislike it much more than they value the reform in the other period. By taking the sets of reform opponents to be completely disjoint across periods, we get $4/5$ of agents against the simple majority rule in period 1, which would cause the reform to be implemented in both periods.

Condorcet cycle.

Power consistency and liberalism

Sen (1970) has demonstrated that a social ranking rule cannot be both Pareto efficient and satisfy what Sen calls "Minimal Liberalism": for at least two individuals there exist two pairs of alternatives, one for each individual, such that the individual dictates the social ranking between the alternatives in his pair. By linking social preferences to individual decisions in a dynamic game, power consistency can capture Sen's notion of liberalism as a particular case.

Sen's setting concerns a static social choice problem, in which an "alternative" entails a complete description of all decisions in society. When these decisions (collective or individual) can be represented as a dynamic game, Sen's alternatives correspond to the policies studied here, and there are natural settings in which power consistency corresponds to liberalism.

To illustrate, consider Sen's main example which concerns two individuals, 1 (a 'pervert') and 2 (a 'prude'), and a book, *Lady Chatterley's Lover*. The prude does not want anyone to read the book but, should the book be read by someone (for simplicity, Sen does not allow both individuals to read the book), she prefers to be the one reading it. The pervert, by contrast, would like someone to read the book, and would also prefer the prude to read it rather than himself (the rationale being that he enjoys the idea of the prude having to read this subversive book). Let x, y, and z respectively denote the following alternatives: 2 reads the book; 1 reads the book; no one reads the book. The situation is captured in the game represented in Figure 2: 1 first decides whether to read the book, then 2 makes the same choice if 1 elected not to read the book.[17]

Figure 2 goes here

Power consistency implies that player 2 has the right to choose between reading the book or not. Player 1, too, is entitled to reading the book, regardless of what player 2 does. Thus, power consistency and Sen's version of liberalism are equivalent in this setting. In the coalitional equilibrium of this game, the pervert reads the book and the prude does not (y). Moreover, the Pareto condition of Sen's analysis may be translated into our setting by requiring unanimity for x to win against y. Because x Pareto dominates y given the players' individual preferences, the

[17] The reverse sequence of moves yields outcome x (the prude player reads the book) and thus does not capture the tension at the heart of Sen's theorem.

equilibrium y is defeated by the commitment to a policy in which the pervert does not read the book and the prude does (x). Theorem 2 implies the existence of a Condorcet cycle, which recovers Sen's result on the impossibility of a Paretian liberal.

5 Applications

Reforms

A common source of political inertia is the political risk associated with socially valuable reforms (Fernandez and Rodrik (1991)). Theorem 1 suggests that commitment may fail to resolve political inertia.

We build on the two-stage setting of Fernandez and Rodrik (1991). In the first stage, citizens of a country decide whether to institute a trade reform. If the reform is undertaken, each individual learns whether he is a winner or loser of the reform. In the second stage, citizens vote on whether to continue the reform, or to implement it if they hadn't done so in the first period. The game is represented on Figure 3. The reform imposes a (sunk) cost c on each individual that must be borne once, regardless of the duration of the reform. Voters are divided into three groups (*I*, *II*, and *III*) of equal size, and one of the groups is randomly (with uniform probability) chosen as the sole winner from the reform. Individuals in the winning group get a payoff of g per period for the duration of the reform, while the remaining individuals lose l per period. If the reform is implemented in the first period and continued in the second, we call it a long-term reform, whereas if it is revoked in the second period, it is a short-term reform.

Figure 3 goes here

Provided that g is sufficiently large relative to l, the long-term reform is socially valuable. It provides a higher expected payoff to everyone relative to the status quo of no reform. Commitment to the long-term reform is thus majority preferred to the status quo. However, any initial reform must be revoked in the second period, because two out of three groups find out that they are losers of the reform and have an incentive to end it in the second period. A status quo bias arises if the reform is not implemented at all in equilibrium even though committing to it would be socially beneficial.

Theorem 1 implies that, whenever the status quo bias arises, there must exist a Condorcet cycle over policies, and this cycle involves both the status quo and long-term experimentation. The status quo occurs in equilibrium when $g < 2l + 2c$, as in this case the expected payoff from the short-term reform (the reform is ended after the first period, because two groups of losers are identified) is negative for every type. The expected payoff from the long-term reform is positive for each type, provided that $g > 2l + 1.5c$. Other possible commitments, to short-term reform or to delayed reform in the second period also yield negative expected payoffs. Among these policies, the *long-term reform is a Condorcet winner*, which seems to contradict Theorem 1.

This paradox is explained by the consideration of other possible policies. For example, the commitment to a long-term reform is majority-dominated by the policy which consists of implementing the reform in the first period, and then revoking it only if group *I* is the winner. Groups *II* and *III* strictly prefer this policy to unconditional long-term reform. In turn, this policy is majority dominated by the commitment to continue the reform unless *I* or *II* is the winner. This differs from the previous policy only when *II* is the winner, and in that state of the world *I* and *III* benefit from ending the reform, so their expected payoffs increase. Now that *I* and *II* are at best temporary winners, both prefer the status quo (recall that the expected payoff from short-term reform is negative), which yields a cycle.

The policies in this Condorcet cycle may seem unfair because they single out certain groups. In Theorem 1, the Condorcet cycle over policies is constructed over the full set of feasible policies: any plan of action that conditions on states where voting takes place is under consideration. In this application, however, it may make sense to restrict commitment to anonymous policies, and there does exist a Condorcet winner among anonymous policies. This suggests a way of circumventing the negative results of Theorems 1 and 2 by restricting the policy space, an idea to which we return in Section 7.

Political Economy of Taxation

It is commonly observed in the political economy literature that commitment is valuable for the following reason: welfare-improving decisions are disregarded because the current government expects that such decisions would adversely affect its future political power. For instance, in Besley and Coate (1998), a public investment that increases citizens' ability is not undertaken because

it would lead to a change in preferences for redistribution. A commitment restricting future governments' decisions could a priori resolve this inefficiency. As illustrated below, however, the logic underlying our results suggests otherwise: allowing commitment need not yield a clear improvement over the equilibrium policy.

We consider a simplified version of Example 2 in Besley and Coate (1998), where for convenience we have eliminated ties. There are two periods in which citizens inelastically supply a unit of labor to the market and vote before each period on a linear redistribution scheme, (t, T), where t is the tax rate on labor income and T is the lump-sum redistribution obtained from the proceeds of the proportional tax (the budget is balanced in each period). A citizen with productivity a supplies one unit of labor, earns a, and receives utility $u = a(1-t) + T$. There are 3 types of citizens, assumed to be equal in numbers: "Low" types and "High" types have productivity a_L and a_H in both periods, respectively, while "Movers" have a productivity level a_L in the first period which increases to $a_L + \delta (< a_H)$ in the second period if a reform is implemented. In the first period, citizens decide on the tax scheme for this period and on whether to implement the reform. This decision is captured by the tuple (t_1, T_1, I), where $I = 1$ $(I = 0)$ means that the reform is (not) implemented. Implementing the reform is costless. In the second period, citizens vote on (t_2, T_2).

Besley and Coate model the political process as follows: Before each period, a citizen of each type decides—at virtually no cost—whether to run for office. Once entry decisions are made, citizens vote for one of the candidates, and the winning candidate implements his optimal policy. Candidates cannot make binding promises, and voters rationally anticipate the consequences of their vote. For example, if a High type wins in the second period, she will set taxes to zero because she gains nothing from any redistribution. By contrast, a Low type fully redistributes earnings by setting $t_2 = 1$.

Besley and Coate show that, even though the reform costs nothing and increases productivity, there are parameters for which the equilibrium entails no reform. Intuitively, Movers may prefer to side with the Low types to tax and fully redistribute the proceeds in both periods without the reform, rather than side with the High types who would support the reform but would prevent any redistribution. Besley and Coate suggest that, had commitment been available, the reform would have been undertaken. Our results suggest, instead, that there should be a cycle between commitments.

To see this clearly, we bypass Besley and Coate's running-decision stage and simply assume that exactly one candidate of each type runs for office. In each period, there is a closed primary between the Low and Mover candidates; the winner is then pitted against the High-type candidate in a general election. We set parameters to $a_L = 0$, $a_H = 30$, and $\delta = 20$ and assume that voters maximize their expected payoff, eliminating weakly dominated strategies, and that the majority rule is applied to the relevant electorate in each stage. Finally, we assume that, in case of tie in the primary, the Low-type candidate wins.[18] In equilibrium, the Low-type candidate wins the primary and the general election. The Low and Mover types then vote for the Low-type candidate in the general election, who taxes at 100% and does not implement the reform in the first period. This equilibrium policy, X_1, yields a payoff of 20 to everyone, assuming no discounting across periods: each group gets a third of the High type's output. Movers do not vote for the High type because he would not redistribute—the policy which rules out redistribution and implements the reform, X_2, yields payoffs 0,18,60 for the Low, Movers, and High types. In equilibrium, the Low-type candidate does not want to reform because she knows that it would prompt Movers to vote against her and, hence, against redistribution in the second period. This policy, in which redistribution takes place only in the first period and the reform is implemented, denoted X_3, yields payoffs 10,28,40. It achieves the optimal payoff for Movers. X_3 is majority preferred to the status quo (X_1). However, if redistribution is going to take place in a single period, and the reform is implemented, then clearly the High and Low types would prefer it to take place in the second period, in which Movers will contribute their earnings to the redistribution. The resulting policy, X_4, yields payoffs $16, 16, 46$. Notice however that X_4 is itself majority dominated by the status quo, X_1, resulting in the cycle predicted by our theorem.[19]

It seems intuitive, as suggested by Besley and Coate, that citizens would like to commit to a Pareto-improving policy in which the reform is undertaken: this reform increases global output at no cost. In particular, the policy X_5 that implements the same tax rate as the equilibrium (i.e., full redistribution) but implements the reform, clearly Pareto dominates the equilibrium by increasing

[18] Equivalently, one could assume that the Low types slightly outnumber Movers. Keeping an equal number of each type simplifies the computation.

[19] To fit the exact structure of our theorem, notice that each of the four policies above corresponds to an actual policy which would be implemented for some sequences of winning types across the two periods: X_1 is implemented if the Low type wins both elections, X_2 is implemented if the High type wins both, X_3 is implemented if Movers win both elections, and X_4 is implemented if a High type wins the first election and a Low type wins the second election.

the pie shared by all citizens. However, this policy, which yields 26 to each citizen, is majority dominated by policy X_3 and is thus part of a cycle.[20]

6 Extensions

Random proposers

In well-known agenda-setting protocols, voters may take turns making collective proposals, and may be chosen deterministically or stochastically to do so. These protocols are compatible with the setting of this paper. For example, for t odd the state θ_t would include, as well as past information, the identity of a proposer who chooses between two collective proposals. At the next, even, period, the new state θ_{t+1} includes the proposal just made and society decides whether to accept the proposal, given the possibility of future proposals.

Non-binary decisions

Our earlier focus on binary decisions in the dynamic game gets rid of Condorcet cycles at the stage-game level, which avoids confusion between these cycles and those at the heart of our result.[21] Many political problems do, in fact, have this binary structure. For example, choices such as referenda and initiatives take the form of binary decisions. Similarly, lawmakers introduce bills and amendments as "yes" or "no" choices.

With three or more alternatives to choose from at any time, one may attempt to resolve the potential Condorcet cycles by a "binarizing" procedure as in the agenda setting literature. The resulting game then becomes subject to the theorem of this paper: the binary choice sequence leads either to an undominated equilibrium or to an indeterminacy in the ranking of state-contingent policies.

[20] Policy X_5 is implemented if Movers win the first election and Low types win the second one.

[21] The approach is also used in the explicit protocol proposed by Acemoglu, Egorov, and Sonin (2012) (p. 1458). While collective decisions are all binary, that paper allows a player to make proposals among all possible states. This can be easily replicated here by a sequence of at most D periods, where D is the number of states, with each period corresponding to a new state being presented to the proposer, who makes a "no" decision until being presented with the state that he wants to propose to the group, at which point he votes "yes" and the proposal is made to the group.

Transfers

Transfers may be explicitly considered in the present setting. For instance, one may include periods at which the binary action corresponds to whether some player i makes a specific transfer to another player j. In this case, i is a 'dictator' over the decision, and the state θ_t keeps track of all past transfers entering players' payoffs at the end of the game.

7 Commitment and Anonymity

Many applications feature agents who play a symmetric role. This symmetry may be exploited to eliminate, on account of fairness, policies which discriminate among such agents. If these policies were involved in the Condorcet cycles predicted by our theorems, their removal may provide a normative resolution of indeterminacy at the commitment stage. This approach is formalized here, as well as a second approach, which compares policies behind a "veil of ignorance" (Harsanyi (1955); Atkinson (1970); Rawls (1971)) and is applicable even when all agents have asymmetric roles.

To illustrate these approaches, consider an asymmetric, three-agent variation of the Fernandez and Rodrik's resistance-to-reform example discussed in Section 5. The reform has a distinct effect on Agent *II*: for example, while each agent has an equal change of becoming the reform's winner, *II* may stand to gain more from the reform if he is the winner. This asymmetry is represented in Figure 4 by the use of different labels for II. For reasons that will soon be apparent, we do not yet specify agents' utility at terminal nodes.

Figure 4 goes here

Since agents *I* and *III* are symmetric, any policy that treats them differently intuitively violates anonymity. Because agent *II* is uniquely affected by the reform, however, a policy may treat him differently without necessarily violating any intuitive notion of anonymity.[22]

To capture this idea, anonymous policies are defined by grouping agents into homogeneous categories (in the previous game, there are two such categories: one for agent *II* and the other for agents *I* and *III*), and requiring an equal treatment of all agents with any given category.

[22]For example, if *II* was more skilled than the other two agents, stopping the reform when *II* is chosen by nature is not necessarily discriminatory.

This definition purposefully leaves aside agents' preferences: it does not require to provide the same utility to agents in the same category (such as I and III) unless these agents also have the same preferences. To formalize this separation, it is useful to distinguish *positions* in the game, and *preferences* over these positions. The separation can be implemented in any of the games considered in this paper, by adding to it a special structure on the state space, and is reminiscent of Harsanyi's "extended preferences" (Harsanyi, 1977, Chapter 4).[23]

Using Harsanyi's (1955) terminology, we consider a dynamic game with N voters, indexed by i, and N positions, indexed by j. Voters are characterized by their preferences, while positions correspond to possible physical realities agents might find themselves in. Because our notion of an anonymous policy is purely based on positions, we ignore for now voters' preferences and focus on how the game unfolds for each position. To each time t and position j corresponds a "positional state" θ_t^j, which summarizes the payoff-relevant history of the game up to time t for a voter in position j (e.g., j's employment history, the evolution of her savings, whether she is married to someone in position j' and the circumstance of j', as well as aggregate variables such as income inequality in the society where j lives at time t).[24] Given any position j, a policy C yields a probability distribution over terminal positional states θ_{T+1}^j.

Restricting commitment to anonymous policies

A set of positions belongs to the same *category* if switching labels for these positions does not alter the game for any position. For example, in the reform game all agents are unaffected if agent I is asked to take on III's role and vice versa: these two agents face exactly the same strategic situation and payoffs, and so does agent II.

Having partitioned the set of positions into categories, a policy is said to be anonymous if it generates the same distribution over terminal positional states for all positions within any given category.[25] Intuitively, this means that the policy does not use any preferential treatment to dis-

[23] Unlike Harsanyi's concept, ours does not require the spectator behind the veil of ignorance to consider other agents' preferences when evaluating positions, and is thus immune from the criticism which extended preferences have attracted (Adler (2014)) for forcing the spectator to imagine the "impossible state of affairs" of being someone she is not.

[24] Positional states contain redundancies among them (for example, the quantity of a public good at time t would be reflected in each θ_t^j). Moreover, the union of positional states at time t across all positions, $\{\theta_t^j : 1 \leq j \leq N\}$, contains the same information as did state θ_t in the original game of Section 3.

[25] For expositional simplicity, we exclude zero-probability events from this presentation.

criminate between otherwise (physically) identical agents. A policy that is ex post asymmetric for agents within the same category will still be anonymous as long as it is ex ante symmetric. For example, in the reform game a policy that randomizes between R and S, but does so with the same probabilities regardless of whether I or III is chosen, will be considered anonymous, even though I and III may be treated differently ex post.

Equipped with this definition, we can remove from consideration at the commitment stage all non-anonymous policies. The removal may be justified on normative grounds (it is unethical to treat otherwise identical people differently), as well as motivated by practical concerns (it may be difficult to dissociate individuals in any given category). The proofs of our theorems rely on the fact that all state-contingent policies are feasible. Removing discriminatory policies, we may destroy Condorcet cycles and thus identify a socially-preferred commitment.

To illustrate this approach, we return to the symmetric reform game of Section 5. Each position $j \in \{I, II, III\}$ gets a 0 payoff if no reform takes place, $g - c$ ($l - c$) if the reform is implemented for one period and j is a winner (loser), $2g - c$ ($2l - c$) if the reform lasts for two periods and j is a winner (loser). By symmetry, all three positions belong to the same category. Therefore, any policy that induces different payoff distributions across positions is discriminatory. In particular, position-specific reforms of the form "continue the reform only if the agent in position I is a winner" are discriminatory. The only non-discriminatory policies are the short-term reform in the first period, the short-term reform in the second period, the long-term reform, and the status quo. Among these, the long-term reform is the Condorcet winner.

Veil of ignorance

Our second approach has voters compare policies behind a veil of ignorance. First, each voter is assigned, with equal probability, one of the N positions in the game. Second, each voter evaluates the policy from a particular position against her preferences. For instance, agent 'II' in the asymmetric reform game (Figure 4) now evaluates policies by giving equal weight to the outcome distribution of position I (or III) as he does to the distribution associated with his actual position, and evaluates each of these distributions according to his own preferences.

To implement this approach, we specify a voter's preference for each position and policy in the

game: i's utility function, $\tilde{u}_i(\theta^j_{T+1})$, is defined over the set of terminal positional states.[26] Given a policy C, seen from time t, voter i's expected utility when she occupies a position j in the game is given by $\widetilde{V}^{ij}_t = E(\tilde{u}_i(\theta^j_{T+1}) \mid \theta_t, C)$.[27]

Assuming that voters maximize their expected utility, i's value from a policy C is, behind the veil of ignorance, given by

$$U_i(C) = \frac{1}{N}\sum_{j=1}^{N} E\tilde{u}_i(\theta^j_{T+1} \mid \theta_0, C). \tag{1}$$

Equipped with individual preference orderings, we can aggregate individual preferences, for example using majority voting. If all voters had identical preferences (i.e., identical $\tilde{u}_i$'s), the social ranking would select the utilitarian policy. In general, voters have heterogeneous preferences, and majority voting behind the veil of ignorance may lead to Condorcet cycles of a different nature from those predicted by Theorem 2.

Applying this second approach to the symmetric reform game produces the unconditional reform as the Condorcet winner, which is the utilitarian optimum (voters have identical preferences in this example). The veil-of-ignorance approach strips conditional reforms (such as continuing trade liberalization only if a specific industry sector benefits from it) from their appeal, by forcing voters to evaluate these policies without knowing their position in the game. Likewise, in Figure 4, if position *II* were discriminated against in the sense that all three voters would suffer were they to occupy this position, then all voters would take this into account behind the veil when evaluating policies. In effect, ranking policies behind the veil of ignorance amounts to a violation of power consistency, since voters behind the veil have different preferences from their realized counterparts outside the veil.

8 Discussion

Can political inertia and inefficient equilibria be resolved through the use of commitment? Introducing the option to commit to any state-contingent policy results in an unambiguous social gain only

[26] An agent's terminal state contains his state in all earlier periods of the game. His terminal utility can therefore be an arbitrary function of his per-period utility in each period, as explained in Section 6.

[27] For any given bijection $J(i)$ between the set of voters and positions we can then easily define the Coalitional Equilibrium as before by recovering our original utility function as $u_i(\theta_{T+1}) = \tilde{u}_i(\theta^{J(i)}_{T+1})$.

if the power structures used to compare policies under commitment and in the dynamic game are inconsistent. Under power consistency, attempts to improve on an inefficient equilibrium through commitment run into the problem of indeterminacy. This finding holds for general state processes and utility functions, allowing social learning, experimentation, and arbitrarily heterogeneous payoffs.

Examples abound in the literature of dynamic games where equilibria are inefficient in the absence of commitment. If we preserve the power structure of the game, then our theorem shows that commitment in general cannot solve the problem, unless there is a rationale for taking certain policies a priori off the table. Policies might be ruled out based on ethical considerations, such as anonymity, or because an agenda setter controls the options (see, e.g., Compte and Jehiel (2010) and Diermeier and Fong (2011)).

In some cases, it may be desirable to give power to decision-makers at the start of the game that they will not have later on. Commitment devices, such as contracts (backed by law enforcement), are created in order to facilitate power transfers. A contract requires mutual agreement at the outset from those with future decision rights. Renegotiation issues (e.g. Laffont and Tirole (1990)) arise precisely because, once the parties have locked in a series of choices, scenarios can arise where those with decision power would like to deviate from the agreed path.

Closely related are models of time inconsistency where inefficiencies arise because voters or governments anticipate that evolving states will create new policy biases in the future. (The general problem was highlighted by Kydland and Prescott (1977).) This includes the literature on redistribution (see, e.g., Campante and Ferreira (2007), Azzimonti, de Francisco, and Krusell (2008), Klein, Krusell, and Rios-Rull (2008)), where the ex-post allocation of productivity gains cannot be guaranteed, and therefore policies and actions are taken that fail to maximize overall welfare. In the representative democracy models of Persson and Svensson (1989), Besley and Coate (1998), Krusell (2002), or Saint-Paul, Ticchi, and Vindigni (2015), governments manipulate the state (debt, wealth), creating inefficiencies, in order to influence future political decisions. Our analysis shows that ex ante commitment to long-run policies can restore efficiency only when the set of feasible policies is restricted or when power consistency is violated. Even when power consistency fails, commitment may not be credible as commitment devices are unlikely to be available. For example, in overlapping or successive generations models allowing commitment to long-run macroeconomic

policy, future decision makers are not involved in selecting among policies simply because they are not around at the outset. While it is then possible to agree on a policy, power consistency is violated, and one should expect attempts by future generations to renege.

In extreme cases, the power shifts associated with commitment that is not power-consistent can lead to outcomes that are disastrous for the group that is deprived of its decision rights. Then, violating power consistency seems ethically hard to defend. Models with equilibria leading to immiseration have this character. In the original example given by Bhagwati (1958), immiserizing growth occurs when increases in the output of an export good reduce its cost so much that the country overall sustains a welfare loss from the rise in the relative cost of imports. Subsequent examples included ex ante optimal policies (such as tariffs in the trade literature, see Johnson (1967) and Bhagwati (1968)), rendered suboptimal by growth, or policies favored by the initial generation (Matsuyama (1991), Farhi and Werning (2007)) that harm future generations.

In all these cases, power consistency is implicitly violated by a long-run policy that ignores the interests of a group that should arguably have control at a future time. Immiseration results reflect that commitments are being allowed, or exclusively considered, that are only optimal with respect to initial preferences and seem infeasible to maintain because they become so objectionable over time.

Appendix

Proof of Theorem 3

We set the terminal payoffs equal to 0 for all policies, except when i) the action sequence until time t has followed policy Y (and hence Y') *and* ii) the state θ_t reached at time t belongs to $\bar{\Theta}_t$. In that case, members of coalition S (its complement $\bar{S}$) get 100 (10) if the time t action prescribed by Y' is played and followed by the continuation of policy Y (and hence Y'), with the reverse payoffs if instead the action prescribed by Y is played at time t and followed by the continuation of policy Y. If $\theta_t \notin \bar{\Theta}_t$, everyone's payoff is set to some small $\varepsilon > 0$ for the common continuation of Y and Y', and to zero for all other continuations.[28]

Let Z denote the equilibrium policy. Z must coincide with Y and Y' until period t since it is the only way, for any player, to get a nontrivial payoff. If the state $\theta_t \in \bar{\Theta}_t$, coalition S imposes at time t the action prescribed by Y', so as to achieve its highest possible payoff of 100, and it is in everyone's interest to implement the continuation corresponding to Y' from time $t+1$ onwards, so that even members of $\bar{S}$ get their second highest payoff of 10. Even if $\theta_t \notin \bar{\Theta}_t$, it is also in everyone's interest to follow the common continuation of Y and Y' so as to get ε. This shows that Z coincides with Y'.

We now consider the social comparisons of policies. Clearly, $\bar{S}$ imposes $Y \succ Y'$ since it has the power to do so and this achieves maximal expected payoff. Moreover, there cannot be any cycle among policies, since Y and Y' Pareto dominate all other policies. Thus, Y is the Condorcet winner[29] among all policies.

[28] Although this is largely irrelevant to the gist of the present argument (see Remark 2), one may add arbitrarily small, action-dependent payoffs to break ties at all points of the game for all histories as required by the non-indifference assumption of Theorem 2.

[29] Any other Condorcet winner X must be identical to Y except on a set of histories which has probability 0 when Y (or, equivalently, X) is followed.

References

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin. 2012. "Dynamics and Stability of Constitutions, Coalitions, and Clubs." *American Economic Review* 102 (4):1446–1476.

———. 2015. "Political Economy in a Changing World." *Journal of Political Economy* 123 (5):1038–1086.

Acemoglu, Daron and James A. Robinson. 2005. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press.

Adler, Matthew D. 2014. "Extended Preferences And Interpersonal Comparisons: A New Account." *Economics and Philosophy* 30 (02):123–162.

Alesina, Alberto and Guido Tabellini. 1990. "A Positive Theory of Fiscal Deficits and Government Debt." *The Review of Economic Studies* 57 (3):403–414.

Atkinson, Anthony B. 1970. "On the Measurement of Inequality." *Journal of Economic Theory* 2 (3):244–263.

Azzimonti, Marina, Eva de Francisco, and Per Krusell. 2008. "Production Subsidies and Redistribution." *Journal of Economic Theory* 142 (1):73–99.

Barbera, Salvador and Matthew O. Jackson. 2004. "Choosing How to Choose: Self-stable Majority Tules and Constitutions." *The Quarterly Journal of Economics* 119 (3):1011–1048.

Barbera, Salvador, Michael Maschler, and Jonathan Shalev. 2001. "Voting for Voters: A Model of Electoral Evolution." *Games and Economic Behavior* 37 (1):40–78.

Battaglini, Marco and Stephen Coate. 2008. "A Dynamic Theory of Public Spending, Taxation, and Debt." *American Economic Review* 98 (1):201–236.

Besley, Timothy and Stephen Coate. 1998. "Sources of Inefficiency in a Representative Democracy: a Dynamic Analysis." *American Economic Review* 88 (1):139–156.

Bhagwati, Jagdish. 1958. "Immiserizing Growth: A Geometrical Note." *The Review of Economic Studies* 25 (3):201–205.

Bhagwati, Jagdish N. 1968. "Distortions and Immiserizing Growth: a Generalization." *The Review of Economic Studies* 35 (4):481–485.

Boylan, Richard T., John Ledyard, and Richard D. McKelvey. 1996. "Political Competition in a Model of Economic Growth: Some Theoretical Results." *Economic Theory* 7 (2):191–205.

Boylan, Richard T. and Richard D. McKelvey. 1995. "Voting Over Economic Plans." *American Economic Review* 85 (4):860–871.

Campante, Filipe R. and Francisco H.G. Ferreira. 2007. "Inefficient Lobbying, Populism and Oligarchy." *Journal of Public Economics* 91 (5):993–1021.

Compte, Olivier and Philippe Jehiel. 2010. "Bargaining and Majority Rules: a Collective Search Perspective." *Journal of Political Economy* 118 (2):189–221.

Diermeier, Daniel and Pohan Fong. 2011. "Legislative Bargaining With Reconsideration." *The Quarterly Journal of Economics* 126 (2):947–985.

Farhi, Emmanuel and Ivan Werning. 2007. "Inequality and Social Discounting." *Journal of Political Economy* 115 (3):365–402.

Fernandez, Raquel and Dani Rodrik. 1991. "Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty." *American Economic Review* 81 (5):1146–1155.

Galperti, Simone and Bruno Strulovici. 2014. "Forward-Looking Behavior Revisited: A Foundation of Time Inconsistency." Working paper, Northwestern University.

Glaeser, Edward L. and Andrei Shleifer. 2005. "The Curley effect: The economics of shaping the electorate." *The Journal of Law, Economics, & Organization* 21 (1):1–19.

Harsanyi, John C. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63 (4):309–321.

———. 1977. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge University Press.

Jackson, Matthew O. and Leeat Yariv. 2015. "Collective Dynamic Choice: the Necessity of Time Inconsistency." *American Economic Journal: Microeconomics* 7 (4):150–178.

Johnson, Harry G. 1967. "The Possibility of Income Losses From Increased Efficiency or Factor Accumulation in the Presence of Tariffs." *Economic Journal* 77 (305):151–154.

Klein, Paul, Per Krusell, and Jose-Victor Rios-Rull. 2008. "Time-Consistent Public Policy." *Review of Economic Studies* 75 (3):789–808.

Krusell, Per. 2002. "Time-Consistent Redistribution." *European Economic Review* 46 (4):755–769.

Kydland, Finn E. and Edward C. Prescott. 1977. "Rules Rather Than Discretion: the Inconsistency of Optimal Plans." *The Journal of Political Economy* 85 (3):473–491.

Laslier, Jean-Francois. 1997. *Tournament Solutions and Majority Voting*. Springer, Berlin.

Matsuyama, Kiminori. 1991. "Immiserizing Growth in Diamond's Overlapping Generations Model: A Geometrical Exposition." *International Economic Review* 32 (1):251–262.

Miller, Nicholas R. 1977. "Graph-Theoretical Approaches to the Theory of Voting." *American Journal of Political Science* 21 (d):769–803.

Persson, Torsten and Lars E.O. Svensson. 1989. "Why a Stubborn Conservative Would Run a Deficit: Policy with Time-Inconsistent Preferences." *Quarterly Journal of Economics* 104 (2):325–345.

Rawls, John. 1971. *A Theory of Justice*. Harvard University Press.

Roberts, Kevin. 2015. "Dynamic Voting in Clubs." *Research in Economics* 69 (3):320–335.

Saint-Paul, Gilles, Davide Ticchi, and Andrea Vindigni. 2015. "A Theory of Political Entrenchment." *The Economic Journal* 126 (593):1238–1263.

Sen, Amartya. 1970. "The Impossibility of a Paretian Liberal." *Journal of Political Economy* 78 (1):152–157.

Shepsle, Kenneth A. 1970. "A Note on Zeckhauser's "Majority Rule with Lotteries on Alternatives": The Case of the Paradox of Voting." *Quarterly Journal of Economics* 84 (4):705–709.

Strulovici, Bruno. 2010. "Learning While Voting: Determinants of Collective Experimentation." *Econometrica* 78 (3):933–971.

Tabellini, Guido and Alberto Alesina. 1990. "Voting on the Budget Deficit." *American Economic Review* 80 (1):37–49.

Volokh, Eugene. 2003. "The Mechanisms of the Slippery Slope." *Harvard Law Review* 116:1026–1137.

Zeckhauser, Richard. 1969. "Majority Rule with Lotteries on Alternatives." *Quarterly Journal of Economics* 83 (4):696–703.

Figure legends

Figure 1: Slippery Slope. Solid circular nodes indicate majority decisions.

Figure 2: A representation of Sen's game. Individual preferences are indicated from the most preferred to the least preferred alternative.

Figure 3: Reform game from Fernandez and Rodrik (1991) for three voters and one winner from the reform.

Figure 4: Modified Fernandez-Rodrik reform game.