

# SOCIAL MOBILITY AND STABILITY OF DEMOCRACY: REEVALUATING DE TOCQUEVILLE\*

DARON ACEMOGLU  
GEORGY EGOROV  
KONSTANTIN SONIN

An influential thesis often associated with de Tocqueville views social mobility as a bulwark of democracy: when members of a social group expect to join the ranks of other social groups in the near future, they should have less reason to exclude these other groups from the political process. In this article, we investigate this hypothesis using a dynamic model of political economy. As well as formalizing this argument, our model demonstrates its limits, elucidating a robust theoretical force making democracy less stable in societies with high social mobility: when the median voter expects to move up (respectively down), she would prefer to give less voice to poorer (respectively richer) social groups. Our theoretical analysis shows that in the presence of social mobility, the political preferences of an individual depend on the potentially conflicting preferences of her “future selves,” and that the evolution of institutions is determined through the implicit interaction between occupants of the same social niche at different points in time. *JEL Codes: D71, D74.*

## I. INTRODUCTION

An idea going back at least to [Alexis de Tocqueville \(1835\)](#) relates the emergence of a stable democratic system to an economic structure with relatively high rates of social mobility. De Tocqueville, for example, argued:

In the midst of the continual movement which agitates a democratic community, the tie which unites one generation to another is relaxed or broken; every man readily loses the tract of the ideas of his

\*We thank four anonymous referees, the coeditor, Pol Antràs, and participants of the Stanford Institute for Theoretical Economics (SITE) conference on Dynamics of Collective Decision Making, CIFAR meeting, NBER Political Economy conference, Econometric Society World Congress in Montreal, Economic Theory conference at the University of Miami, Game Theory Congress in Maastricht, Warwick/Princeton Political Economy Conference in Venice, and seminars at UC Berkeley, Caltech, the University of Chicago, Higher School of Economics, MIT, New York University, University of Waterloo, University of Wisconsin–Madison, and the University of Zürich for helpful comments. Acemoglu gratefully acknowledges financial support from the Army Research Office, ARO MURI W911NF-12-1-0509, and the Carnegie Foundation.

© The Author(s) 2017. Published by Oxford University Press on behalf of the President and Fellows of Harvard College. All rights reserved. For Permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

*The Quarterly Journal of Economics* (2018), 1041–1105. doi:10.1093/qje/qjx038.  
Advance Access publication on November 6, 2017.

forefathers or takes no care about them. Nor can men living in this state of society derive their belief from the opinions of the class to which they belong, for, so to speak, there are no longer any classes, or those which still exist are composed of such mobile elements, that their body can never exercise a real control over its members. (De Tocqueville, 1835–40, Book 2, p. 120–121)

Lipset (1992) summarizes and further elaborates de Tocqueville's hypothesis as follows:

In describing "The Social Conditions of the Anglo-Americans" in *Democracy in America* Tocqueville concluded that the institutionalization of widespread individual social mobility, upward and downward, has "political consequences", the stabilization of the democratic order.

Many commentators have continued to view social mobility as a vital factor for the health of U.S. democracy. While Lipset and Bendix (1959) deem it to be "a critical, if not the most important, ingredient of the American democracy," Blau and Duncan's (1967) seminal study concluded "the stability of American democracy is undoubtedly related to the superior chances of upward mobility in this country" (similar ideas also appear in Sombart 1906; Pareto 1935; Moore 1966; Erikson and Goldthorpe 1992). This perspective suggests that greater social mobility—caused, for example, by improvements in the educational system, the dismemberment of barriers against occupational mobility, or technological changes—may improve the prospects of democracy's survival and flourishing. Indeed, some of the most stable democracies of the twentieth century are those that appear to have had relatively high rates of social mobility, such as the United States and Scandinavian countries.<sup>1</sup> At the same time, however, social mobility appears to have been high in Weimar Germany in the period preceding the

1. Though there is a debate about whether social mobility rates have been rising since the 1940s in the United States, intergenerational social mobility appears to be relatively high in the 1940s. For example, Chetty et al. (2014) estimate that the probability that the child of a father from the bottom quintile of the income distribution would reach the top quintile is close to 10% in the 1980s and thereafter. Aaronson and Mazumder (2008) and Hilger (2015) estimate similar rates of mobility in the 1940s. Estimates of the rates of intergenerational social mobility in Norway in the second half of the 1930s in Pekkarinen, Salvanes, and Sarvimaki (2016) are also similar to Chetty et al.'s numbers for the United States in the 1980s (compare Figure 2 in the former paper to Figure 1 in the latter).

rise of Hitler, perhaps the most spectacular collapse of a modern democratic system in history (e.g., Mann 2004).<sup>2</sup>

Despite its ubiquity in modern debates on democracy and in modern social theories, there has been little systematic formalization or critical investigation of the link between social mobility and the stability of democracy. The following example illustrates the basic intuition behind de Tocqueville's hypothesis.

**EXAMPLE 1.** Consider a society with  $n$  individuals, with  $\frac{2}{5}n$  or 40% of them rich,  $\frac{1}{5}n$  or 20% middle class, and  $\frac{2}{5}n$  or 40% poor. There are three possible political institutions: democracy, where decisions are made by the median voter who is a member of the middle class; left dictatorship, where all political decisions are made by the poor; and elite dictatorship, where all political decisions are made by the rich. Suppose that the economy lasts for two periods, and in each period, society adopts a single policy,  $p_t$ . There is no discounting between the two periods. All agents have stage payoffs given by  $-(p_t - b_i)^2$ , where political bliss points,  $b_i$ , for the poor, middle-class, and rich social groups are, respectively,  $-1$ ,  $0$ , and  $1$ . Society starts out with one of the three political institutions described above, and in the first period, a member of the politically decisive ("pivotal") social group decides both the current policy and the political institution for the second period. Then, in the second period, the group in power chooses policy.

Suppose we start with elite dictatorship. Without social mobility, the politically decisive rich prefer to keep their dictatorship so as to be able to set the policy in the second period as well.<sup>3</sup> Suppose, instead, that there is very high social mobility, involving complete reshuffling of all individuals across the three social groups. (At the time decisions are made, what will happen to a given individual is not known, so there is

2. Storer (2013), for example, points to significant progress in education, sciences, and arts and gains in women's labor force and political participation as evidence of greater social mobility in the Weimar Republic. Social mobility may have been high during the early Nazi years as well. Stachura (1993) writes: "In promoting the growth of industrial society, the [Nazi] regime destroyed the traditional class system bequeathed by the Weimar republic and encouraged social mobility on an unprecedented scale."

3. Throughout the article, when all current members of a social group have the same preferences, we interchangeably refer to a member of that social group or the entire social group.

no asymmetry of information or conflict of interest within a group.) If so, a rich individual expects to be part of the rich, the middle class, and the poor with probabilities  $\frac{2}{5}$ ,  $\frac{1}{5}$ , and  $\frac{2}{5}$ , respectively. His second-period expected utility is then  $-\frac{2}{5}(p_2 + 1)^2 - \frac{1}{5}p_2^2 - \frac{2}{5}(p_2 - 1)^2 = -p_2^2 - \frac{4}{5}$ . Thus, he prefers, in expectation,  $p_2 = 0$ . To achieve this, he would like next period's political institutions to be democratic.

However, the same theoretical example can also be used to highlight the opposite political forces in play.

**EXAMPLE 1 (CONTINUED).** Consider now a different pattern of social mobility:  $r$  middle-class agents become rich and  $r$  rich agents move down to the middle class between periods 1 and 2. Let  $\alpha = \frac{5r}{n}$  denote the share of the middle class that moves upward. Suppose that the society starts out as a democracy. Then, if sufficiently many members of the middle class move upward (i.e., if  $\alpha > \frac{1}{2}$ ), middle-class agents expect, on average, to have the preferences closer to those of the rich tomorrow, and hence prefer tomorrow's policy to be determined in elite dictatorship, making democracy unstable.

This example thus provides a simple (and as we will see, robust) reason that greater social mobility may undermine the stability of democracy:<sup>4</sup> if social mobility means that members of the politically pivotal middle class expect to change their preferences in a certain direction, they will have an incentive to change the political institutions in that direction as well.

Differently from this example, our model considers an infinite-horizon setting. This is for three reasons. First, in a two-period model, if the current decision makers could set policies for the next period (as in [Bénabou and Ok's 2001](#) analysis of the relationship between social mobility and redistribution), then there would be no need for institutional change. Second, such a model also precludes any effect of future social mobility on current

4. The fact that the social mobility in this example makes middle class agents more likely to move upward rather than downward is important—as we will see in our analysis. If they expected to move upward or downward symmetrically, then they would continue to prefer democracy to other political regimes because they would lose in expectation even more from elite (or left) dictatorship than they would gain.

preferences. Third and relatedly, we will see that beyond the two-period setting, what matters for the political equilibrium is not simply mobility next period but the interplay of the evolution of the preferences of an agent's "future selves" (because of evolving social mobility) and expectations about future institutions. This last feature is illustrated in the next example.

**EXAMPLE 2.** Consider the same setting as in Example 1, but now each agent maximizes her discounted utility over an infinite number of periods, and we take the discount rate to be  $\beta = \frac{4}{5}$ . In each period, the current decision maker determines next period's institution, and in between,  $r$  people move upward from the middle class, and  $r$  rich agents move downward. Let  $\alpha = \frac{5r}{n}$  again denote the share of the middle class moving upward.

In left dictatorship, the poor, who are not upwardly mobile, would maintain this political institution forever, and choose  $p_t = -1$  (their political bliss point) at all  $t$ . In elite dictatorship, the rich also have no incentive to change the political institutions. Middle-class preferences, on the other hand, depend on their expectations of future institutions and of how future middle-class agents will behave. Suppose that  $\frac{1}{4} < \alpha < \frac{1}{2}$ . Then a middle-class individual prefers her group to remain in power in the next period, but the rich to be in power after a few periods. (In the long run, the current middle-class expect to be rich  $\frac{2}{3}$  of the time and remain in the middle class  $\frac{1}{3}$  of the time.) Consequently, when today's middle class expects a transition to elite dictatorship tomorrow, it prefers to remain in democracy, and when it expects the survival of democracy, it prefers an immediate transition to dictatorship. This logic not only illustrates the interplay between the preferences and strategies of current and future "selves," but also shows that there is no pure-strategy Markovian equilibrium in this case because of this same interplay.

Our baseline framework corresponds to a straightforward generalization of the setup discussed in this example. Society consists of a finite number of social groups, each of which comprises a finite number of identical individuals. Individuals (and thus groups) are ordered with respect to their policy preferences. Social mobility results from well-defined stationary probabilities specifying how each individual transitions from one

social group to another. There is a finite set of alternative political institutions, which we refer to as states, and each state is represented by a set of weights assigned to individuals within each social group.<sup>5</sup> These weights determine the distribution of political power and the identity of the pivotal voter who chooses the current policy as well as next period's political state, which is equivalent to choosing next period's pivotal voter.

Our main results are of two sorts. First, we establish the existence and certain basic properties of Markov perfect equilibria in this economy. We focus on equilibria that are "monotone," which have the property that the equilibrium path starting from a state is always further to the right in the sense of first-order stochastic dominance relative to the equilibrium path starting from another state to the left. Though, as Example 2 suggests, pure-strategy equilibria may fail to exist, we demonstrate that mixed-strategy equilibria always exist, and that mixing takes a particularly simple form: (generically) there is mixing only between keeping the current institution and transiting to a uniquely defined alternative. This property implies, in particular, that the equilibrium direction of transition is always well defined. Similarly, the interplay between different selves of the current pivotal voter can lead to multiple equilibria. Nevertheless, we establish the uniqueness of equilibrium under a simple (even if somewhat demanding) within-person monotonicity condition, which imposes that the preferences of the future selves of an individual evolve monotonically. Specifically, this condition requires that as we consider selves further away from the present, preferences will either gradually shift to the left or to the right, and thus enable consistent aggregation of the preferences of future selves.

Second, we provide a comprehensive analysis of the relationship between social mobility and the stability of political institutions. Though our analysis applies to the stability of any political institution where a group or individual is pivotal, we focus on its implications for the stability of democracy, which was the

5. Focusing on a finite set of political institutions simplifies the analysis. More important than this is the feature that all political institutions we consider make one group (or one individual) pivotal—thus ruling out arrangements in which there are several groups or individuals with different preferences that share power and have to agree on policy or political decisions. The restriction to this type of political institutions implies that the current decision maker effectively chooses the identity of next period's decision maker, and this would continue to be true even if we had a continuum of possible political institutions.

motivating question for de Tocqueville and for the social science literature following him. We quantify the stability of democracy with the size of its basin of attraction along the equilibrium path.<sup>6</sup> Hence, we say that democracy is more stable under social mobility process  $M$  than  $M'$  if it is stable under  $M$  whenever it is stable under  $M'$ , and moreover, it is asymptotically stable under  $M$  whenever it is asymptotically stable under  $M'$ .<sup>7</sup> Example 1 provides an illustration of how social mobility may make democracy unstable—even starting in democracy, society will not stay there. Our main results, presented in Theorems 4 and 5, state that if the preferences of the median voter in democracy in the very distant future are close to her current preferences, then greater social mobility makes democracy more stable; otherwise, greater social mobility makes democracy less stable. When there is mobility between all social groups (so that the unique irreducible component of the social mobility process is the entire society), our main results become even simpler and more intuitive: social mobility increases the stability of democracy if the preferences of the median voter (i.e., median preferences) are close to the average of the preferences of all voters, and not otherwise.<sup>8</sup>

Our article is most closely related to the small literature on the interplay between social mobility and redistribution. The important article by [Bénabou and Ok \(2001\)](#), which has already been mentioned, shows how greater social mobility discourages redistributive taxation (see also [Wright 1986](#) for a similar argument in the context of unemployment benefits, and [Piketty 1995](#) for a related point in a model in which agents learn from their dynasties' experience about the extent of social mobility). The key economic mechanism in [Bénabou and Ok](#) is also linked to de Tocqueville's

6. Throughout the article, we treat social mobility as exogenous. Endogenous social mobility, and how the new forces identified here affect preferences over social mobility, is discussed in the working paper version ([Acemoglu, Egorov, and Sonin 2016](#)).

7. This notion of stability thus captures both the potential instability of democracy resulting from the median voter preferring, in the future, other political institutions to democracy, and other neighboring social groups wishing to keep society away from democracy (which would be relevant if society started in nondemocracy, or if political power randomly shifted to these groups or enabled them to mount actions against democracy).

8. More generally, our results highlight the following general criteria: (i) an institution is stable if and only if the distant future selves of pivotal decision makers under this institution prefer it to other institutions; (ii) under the same condition, greater social mobility increases the stability of this institution.

hypothesis—greater mobility makes the middle class less willing to tax the rich because they themselves expect to become rich in the future. They generate this effect in a two-period model by assuming that taxes are “sticky” (i.e., there is commitment to future taxes).<sup>9</sup> In [Bénabou and Tirole \(2006\)](#), beliefs about future social mobility support different equilibria—for example, “the American dream” equilibrium, in which a high level of effort stems from the belief in high social mobility (see also [Alesina and Glaeser 2004](#); [Alesina and Giuliano 2010](#)). Nevertheless, this literature does not consider the relationship between social mobility and support for and stability of political institutions, which is our main focus. More important, it neither incorporates the dynamic political trade-offs that are at the heart of our article nor does it feature the potentially destabilizing role of social mobility for democracy.

In an important precursor of our study, [Leventoğlu \(2005\)](#) augments the two-class model of [Acemoglu and Robinson \(2001\)](#) with social mobility; [Leventoğlu \(2014\)](#) introduces the middle class as an additional player in this framework. The main result of these two articles is consistent with de Tocqueville’s ideas—greater social mobility softens the distributional conflict in society and makes the transition to democracy and democratic consolidation more likely. These articles do not, however, develop a general framework similar to the one presented here and do not consider the possibility of the median voter in democracy choosing a different political regime, and as a consequence, they do not obtain our main result—that greater social mobility may destabilize democracy.

Our modeling approach overlaps with dynamic political economy models studying democratization, constitutional change, disenfranchisement (repression), and the efficiency of long-run institutional arrangements, including [Besley and Coate \(1998\)](#), [Bourguignon and Verdier \(2000\)](#), [Acemoglu and Robinson \(2000, 2001\)](#), [Lizzeri and Persico \(2004\)](#), [Gomes and Jehiel \(2005\)](#), [Lagunoff \(2006\)](#), [Acemoglu, Egorov, and Sonin \(2008, 2012, 2015\)](#), and [Roberts \(2015\)](#), though again none of this literature studies social mobility and the mechanisms that are at the heart of our article.

9. In fact, our model and results would be identical to theirs if we restrict ourselves to two periods, remove the choice over political institutions, and assume that taxes for the second period are decided in the first period.

Finally, the role of the implicit conflict between the current self and the future selves of the pivotal voter relates to a handful of papers considering time-inconsistency of collective or political decisions, most notably [Amador \(2003\)](#), [Gul and Pesendorfer \(2004\)](#), [Strulovici \(2010\)](#), [Bisin, Lizzeri, and Yariv \(2015\)](#), [Jackson and Yariv \(2015\)](#), and [Cao and Werning \(2016\)](#). Though some of these works also derive this time-inconsistency endogenously, none of them do so from social mobility or note the conflict between current and future selves resulting from social mobility.

The rest of the article is organized as follows. In [Section II](#) we introduce our setup. [Section III](#) solves the model and establishes existence of an equilibrium, provides conditions for uniqueness, and studies its main properties. [Section IV](#) contains our main results linking the speed of social mobility to the stability of democracy. [Section V](#) presents two sets of further results: first, we show how social mobility changes the nature of slippery slopes in dynamic political economy (whereby political changes that are beneficial in the short run are forsaken because of their medium-run or long-run consequences); second, we generalize our main results to alternative political decision-making rules that use weighted averages of the preferences of different players (rather than imposing weighted voting rules, which make one of the groups pivotal). [Section VI](#) concludes. [Appendix A](#) contains a more detailed presentation of the extensive-form game we use and the proofs of the main results presented in the text, while [Online Appendix B](#) includes the remaining proofs, several additional examples, and further results.

## II. MODEL

In this section, we introduce our basic model and our notion of equilibrium.

### *II.A. Society, Policies, and Preferences*

Time is discrete and infinite, indexed by  $t \geq 1$ . Society consists of  $n$  individuals split into  $g$  social groups,  $G = \{1, \dots, g\}$  with each group  $k$ ,  $1 \leq k \leq g$ , comprising  $n_k > 0$  agents (so  $\sum_{k=1}^g n_k = n$ ). The groups are ordered, and the order reflects their “economic” preferences (e.g., higher-indexed groups could be those that are richer). All individuals share a common discount factor  $\beta \in (0, 1)$ .

Preferences are defined over a policy space represented by the real line,  $\mathbb{R}$  (e.g., more left-wing policies could correspond to higher taxes or more public goods). We assume that individuals in each group have stage payoffs represented by the following quadratic function of the distance between current policy and their bliss point:

$$(1) \quad u_k(p_t) = A_k - (b_k - p_t)^2,$$

where  $p_t$  is the policy at time  $t$ ,  $b_k$  is the (political) bliss point of agents in group  $k$ , and  $A_k$  is an arbitrary constant, allowing for the possibility that some groups are better off than others (e.g., because they are richer).<sup>10</sup> In what follows,  $\mathbf{b} = \{b_k\}$  will denote the column vector of political bliss points. We assume that each  $b_k$  is different from the others, and order the groups so that  $\{b_k\}$  is (strictly) increasing.

Decision-making power depends on the current political state; in each period society makes decisions on the current policy  $p_t \in \mathbb{R}$  and on the next period's arrangement. We assume that there are  $m$  (political) states  $s \in S = \{1, \dots, m\}$ , which encapsulate the distribution of political power in society. In state  $s$ , individuals in group  $k$  are given weights  $w_k(s)$ , and political decisions are made by weighted majority voting as we specify below (this could be a reduced form for a political process involving legislative bargaining or explicit partial or full exclusion of some groups from voting via legislation or repression). The main restriction this formulation imposes is that, as noted in footnote 5, there are no political institutions that allow for several veto players.

We also assume that  $\sum_{k=1}^j w_k(s) \frac{n_k}{n} \neq \frac{1}{2}$  for all  $s \in S$  and all  $j \in G$ . This is a mild assumption adopted for technical convenience and holds generically within the class of weights. It ensures the pivotal group in each state  $s$ —namely, the group  $d_s$  such that  $\sum_{k=1}^{d_s} w_k(s) \frac{n_k}{n} \geq \frac{1}{2}$  and  $\sum_{k=d_s}^G w_k(s) \frac{n_k}{n} \geq \frac{1}{2}$ —is uniquely defined. Since, for our purposes, two states that have the same pivotal group are equivalent, without loss of any generality we can assume that each state has a different pivotal group, so  $\{d_s\}_{s \in S}$

10. For example, if all  $A_k = 0$ , then members of the middle class might not want to become rich when the political institution is democracy, because this would decrease their payoff from the policy choice without any compensating direct benefit from being richer.

are all different. We can then order states such that the sequence of pivotal groups,  $\{d_s\}$ , is increasing.

*II.B. Social Mobility*

We model social mobility by assuming that individuals can change their social group—corresponding to a change in their economic or social conditions and thus their preferences. This can be interpreted either as an individual becoming richer or poorer over time, or as her offspring moving to a different social group than herself (and the individual herself having dynastic preferences).

Throughout we assume that although there is social mobility, the aggregate distribution of population across different social groups is stationary. Since social mobility is treated as exogenous here, this assumption amounts to supposing that there exists a stationary aggregate distribution and that we start the analysis once society has reached this stationary distribution.<sup>11</sup>

Formally, we represent social mobility using a  $g \times g$  matrix  $M = \{\mu_{jk}\}$ , where  $\mu_{jk} \in [0, 1]$  denotes the probability that an individual from group  $j$  moves to group  $k$ , with the following natural restrictions:

$$(2) \quad \sum_{k=1}^g \mu_{jk} = 1 \text{ for all } j, \text{ and}$$

$$(3) \quad \sum_{j=1}^g n_j \mu_{jk} = n_k \text{ for all } k,$$

where the latter condition imposes the stationarity assumption requiring that the sizes of different groups remain constant. Since there is no within-group heterogeneity, the stochastic process for social mobility is the same for each individual within the same social group. Throughout the article, we impose the following assumption:

11. This assumption is both technical and substantive. Technically, it enables Markovian strategies to be “stationary”: if the aggregate distribution of population changed over time, it would have to be part of the payoff-relevant state variable, and the restriction to Markovian strategies would have little bite. Substantively, it enables us to focus on social mobility rather than the implications of changes in the social structure of society, which would be continuously ongoing if the aggregate distribution of population across social groups did not remain constant.

**ASSUMPTION 1. (Between-person monotonicity)** For two groups  $j_1$  and  $j_2$  with  $j_1 < j_2$ , the marginal probability distribution  $\{\mu_{j_1 \cdot}\}$  over  $G$  is first-order stochastically dominated by  $\{\mu_{j_2 \cdot}\}$ . Formally, for any  $l \in \{1, \dots, g - 1\}$ ,

$$(4) \quad \sum_{k=1}^l \mu_{j_1 k} > \sum_{k=1}^l \mu_{j_2 k}.$$

This assumption, which is quite weak, imposes that the distribution of a richer individual's future selves first-order stochastically dominates the distribution of a poorer individual's future selves. In essence, it rules out "deterministic reversals of fortune," where poorer people become (in expectation) richer than the currently richer individuals. We impose Assumption 1 in all of our analysis without explicitly stating it.<sup>12</sup> We next provide an example of a class of social mobility matrixes satisfying this assumption.

**EXAMPLE 3.** Let  $I$  be the identity matrix, so that  $M = I$  corresponds to a society with no social mobility. Let  $F$  be the matrix with elements  $\mu_{jk} = \frac{n_k}{n}$ ; it corresponds to full (and immediate) social mobility, as the probability of an individual becoming part of group  $k$  is proportional to the size of this group and does not depend on the identity of the original group  $j$ . Then for any  $\lambda \in (0, 1]$ ,  $\lambda I + (1 - \lambda)F$  is a matrix of social mobility satisfying Assumption 1.

Throughout the rest of the article, we use the standard notation  $M^\tau$  to denote the  $\tau$ th power of the social mobility matrix  $M$ . The element  $\mu_{jk}^\tau$  of this matrix stands for the probability that an individual currently in social group  $j$  will be in social group  $k$  in  $\tau$  periods time.

### II.C. Timing of Events

We present the extensive-form game describing how policy and political decisions are made in [Appendix A](#). Here we

12. This assumption can be further weakened to have a weak inequality in [equation \(4\)](#), but the version with strict inequality simplifies our exposition and proofs. In fact, [Example 1](#) only satisfies this assumption with weak inequality, but this is also just for simplicity, and having less than full reshuffling in that example would restore strict inequality without any substantive effect on any of its implications.

simplify the exposition by providing a “reduced-form” political decision-making rule.

In each period  $t$ , the society first makes a policy decision,  $p_t$ , and then a political decision over next period’s state,  $s_{t+1}$ . Both decisions are made by voting with the weighted majority rule, and the weights are given by the current state as  $\{w_k(s_t)\}$ . These weights determine which one of the  $g$  groups is pivotal; since preferences over policy in [equation \(1\)](#) are single-peaked and satisfy single-crossing, this pivotal group is well defined for policy choices, and because [Assumption 1](#) ensures that preferences over future states inherit these properties, this group is also pivotal for political decisions.

In the text, we simply assume that a member of this pivotal group is chosen at random and makes both the policy and political decisions. In the extensive-form game in [Appendix A](#), we explicitly model the agenda-setting stage where proposals are made and the voting stage where individuals vote in favor of or against each proposal. The equilibrium outcomes of this extensive-form game coincide with our reduced-form political decision-making rule here.

In addition, in [Section V.B](#), we consider an alternative political decision rule and show that our qualitative results continue to hold in this case.

## *II.D. Definition of Equilibrium*

We focus on symmetric monotone Markov perfect equilibrium (MPE). Symmetry requires that equilibria involve the same strategies for any individuals in the same social group. Monotonicity rules out equilibria in which the direction of political transitions is reversed.<sup>13</sup> As shown in [Example 2](#), pure-strategy equilibria may fail to exist, so we allow proposers or decision makers

13. Nonmonotone MPE exist for some parameter values (as we show in [Example B6](#) in [Online Appendix B](#)). But these are neither robust nor intuitive, and we believe they are not of much economic interest. If we did not rule them out by focusing on monotone equilibria, some of our results would be more complicated without changing the main insights. For example, the conditions for equilibrium uniqueness would become more cumbersome, but without major changes to the rest of our main results (see [Theorems B1](#) and [B2](#) in [Online Appendix B](#) for versions of our main results which apply without uniqueness, and [Theorem B4](#) for sufficient conditions for all equilibria to be monotone).

to mix between alternatives.<sup>14</sup> Thus, a strategy for player  $i$ , who is the decision maker after a certain history, is a mapping from the history (which codifies her current group affiliation and the current institution) into  $\Delta(S)$  (i.e., mixed strategies over the set  $S$ ). We next define our equilibrium concept more formally.

**DEFINITION 1. (Symmetric monotone MPE)** A subgame perfect equilibrium  $\hat{\sigma}$  is an MPE if the strategy of each player  $i$ ,  $\hat{\sigma}_i$ , is conditioned only on player  $i$ 's current social group and the current political institutions (in addition to the history of proposals and votes within the same stage).<sup>15</sup>

An MPE  $\sigma$  is symmetric if for any two individuals  $i$  and  $j$  in the same social group  $k$ ,  $\sigma_i = \sigma_j$ .

An MPE is monotone if for any two states  $x, y \in S$  such that  $x \leq y$ , the distribution of states in period  $\tau > t$  starting with  $s_t = x$  is first-order stochastically dominated by the distribution of states starting with  $s_t = y$ , that is, for any  $l \in [1, m]$ ,

$$(5) \quad \Pr(s_\tau \leq l \mid s_t = x) \geq \Pr(s_\tau \leq l \mid s_t = y).$$

In what follows, we refer to symmetric monotone MPE simply as equilibria. Moreover, although equilibria formally correspond to a complete list of strategies, it will also be more convenient to work with the policy choices and the equilibrium transitions (across different political states) induced by an equilibrium and not distinguish between equilibria that differ in terms of strategies but have the same equilibrium transitions.

Finally, we say that a state (or political institution)  $s$  is stable if  $s_t = s$  implies that  $s_{t+1} = s$ . We say that a state  $s$  is asymptotically stable if  $s_t \in \{s-1, s, s+1\} \cap S$  implies that  $\lim_{\tau \rightarrow \infty} \Pr(s_\tau = s) = 1$ , in other words if, starting from one of the neighboring states of  $s$ , the sequence of states induced in equilibrium converges to  $s$  with probability 1. This last definition is the analogue in discrete state space of the usual notion of asymptotic stability: starting with a small enough deviation from an asymptotically stable state, the equilibrium path will approach

14. With a slight abuse of notation, this definition applies both to the reduced-form game in the text, where strategies refer just to the actions of the decision maker from the pivotal group, and to the full extensive-form game in [Appendix A](#), where strategies specify proposals and votes over proposals.

15. Since ours is a complete information game, the definition of a subgame perfect equilibrium is standard.

the initial state arbitrarily closely with an arbitrarily high probability. For a monotone symmetric MPE, asymptotic stability of a state implies stability. We also quantify the notion of stability by saying that a state becomes more stable under a change in parameters: if (i) it remains stable whenever it was stable before the change of parameters, and (ii) it remains asymptotically stable whenever it was asymptotically stable before the change. The notion of less stable is defined analogously.

### III. ANALYSIS

In this section, we establish some basic properties of equilibria, like existence and conditions for uniqueness. We also introduce the notation that would be helpful for our main characterization results in [Section IV](#).

#### *III.A. Existence and Characterization*

The next theorem establishes the existence of an equilibrium (symmetric monotone MPE) and shows that an equilibrium can be represented by a sequence of policies and transitions that take a simple form, and the preferences of the current pivotal group play a critical role. This is a general result that applies to any pivotal-voter institution in the presence of social mobility. Using the general framework, we then study stability of a particular institution, democracy, which is defined as the political system in which the median voter (the social group containing the median voter) has political power. This analysis would have been impossible without having a general result first, as it allows us to evaluate what happens in the subgames after the democracy collapses (which, in turn, is a critical part of the agents' decision to abandon democracy in the first place).

**THEOREM 1. (Existence and characterization)** There exists an equilibrium. Moreover, in every equilibrium:

- i. The equilibrium policy coincides with the bliss policy of the current pivotal group at each  $t$ . That is, if the current state at time  $t$  is  $s$ , then the policy is  $p_t = b_{d_s}$ .
- ii. The next state maximizes the expected continuation utility of current members of the current pivotal group. That is, if we define the transition correspondence  $Q = Q(\sigma)$  by

$q_{sz} = \Pr(s_{t+1} = z \mid s_t = s)$ , then  $q_{sz} > 0$  implies

$$(6) \quad z \in \arg \max_{x \in S} \sum_{j \in G} \mu_{d_s j} V_j(x),$$

where  $\{V_j(x)\}_{j \in G}^{x \in S}$  satisfies

$$(7) \quad V_j(x) = u_j(b_{d_x}) + \beta \sum_{y \in S} q_{xy} \sum_{k \in G} \mu_{jk} V_k(y).$$

- iii. The transitions induced by the equilibrium are strongly monotone: if  $x < y$  and  $q_{xa} > 0, q_{yb} > 0$  (i.e., transitions from  $x$  to  $a$  and from  $y$  to  $b$  may happen along the equilibrium path), then  $a \leq b$ .
- iv. Generically, mixing is only possible between two states, one of which is the current one. Specifically, for almost all parameter values, if  $q_{sx} > 0$  and  $q_{sy} > 0$  for  $x \neq y$ , then  $s \in \{x, y\}$ .

The first two parts of this proposition imply that, starting in the current state  $s$ , the political process induces a path of policies and transitions that maximizes the discounted utility of the pivotal group,  $d_s$ .<sup>16</sup> Note that this maximization naturally takes into account that the current pivotal group may not be pivotal in the future. This feature of our (monotone) equilibria will greatly simplify the rest of the analysis, and we will often work with the preferences of the current pivotal group (or with a slight abuse of terminology, the current decision maker).

Part iii establishes that (stochastic) equilibrium transitions are strongly monotone, meaning that transitions that have positive probability starting from a higher state will never fall below transitions that have positive probability starting from a lower state. This property implies that if a transition from  $x$  to  $a$  is possible in equilibrium, then from  $y > x$ , only transitions to states  $a, a + 1, \dots$  are possible. Notice that as the qualifier “strongly” suggests, this result significantly strengthens the monotonicity requirement of our symmetric monotone MPE, which only required first-order stochastic dominance of the equilibrium path when

16. There is an analogous result in [Roberts \(2015\)](#) in a nonstrategic environment (and without social mobility), and in [Acemoglu, Egorov, and Sonin \(2015\)](#), also in a setting without social mobility.

starting from a higher state. The result here instead establishes that when we start in a higher state, the lowest state to which we can transition is higher than the highest state to which we can transition starting from a lower state.

Finally, part iv will greatly simplify our subsequent analysis. It establishes that equilibria in mixed strategies take a simple and intuitive form: they involve mixing only between the current state and some other state. Mixed strategies arise as a way of slowing down the transition from today's state to some unique target state. This is intuitive; as Example 2 illustrated, pure-strategy equilibria may fail to exist because the current decision maker would like to stay in the current state if he expects the next decision maker to move away, and would like to move if he expects the next decision maker to stay. This was a reflection of the fact that the current decision maker prefers the current state but would like to be in a different state because he expects his preferences to change in the near future as a result of social mobility. Mixed strategies resolve this problem by slowing down transitions: when she expects the next decision maker to slowly move away (i.e., move away with some probability), the current decision maker is indifferent between moving toward her target state and staying put. This intuition also clarifies why, generically, there is only mixing between two states: the current decision maker can be indifferent between three states only with nongeneric preferences/probabilities.<sup>17</sup> The notion of genericity here is essentially that the set of parameter values for which the statement is not true is of measure zero (because it requires the decision maker to be exactly indifferent between three states).<sup>18</sup> Another implication of this

17. Mixing can take place between two non-neighboring states because the continuation utility of the current decision makers may be maximized at two non-neighboring states. Though this might at first appear to contradict the concavity of utility functions, Example B4 in [Online Appendix B](#) demonstrates that it may take place as a result of the conflict between near and distant future selves (in particular, near selves prefer to stay in the current state, while distant ones prefer to move to states farther away and rapidly, and at the same time, moving to a neighboring state makes none of the selves happy).

18. More formally, the genericity notion requires the parameters,  $\beta$ , the  $\mu$ 's and the  $b$ 's, to be such that no subset of them are roots of a (nontrivial) polynomial with rational coefficients (since the value functions will be shown to be polynomial with rational coefficients in these parameters, see the proof of Theorem 1 in [Appendix A](#)). As there is a countable set of such polynomials, each of which defines a set of (Lebesgue) measure zero, the union of such points has measure zero as well. This substantiates the claim that the statements that are true generically in

characterization is that even though there may be mixed strategies, this does not change the direction of transitions but will just affect their speed.

It is worth noting that the equilibria of the game as characterized in Theorem 1 are shaped by two kinds of conflicts of interest in our framework. One is between agents from different social groups, which results in all decisions being made by the group that is currently pivotal. A different, more subtle conflict is between today's decision maker (an individual in the pivotal group) and future decision makers who will occupy in the future the same social group as the current decision maker. This latter conflict arises from the fact that today's decision maker anticipates being in a different social group in the future. This conflict is not only essential for understanding the political implications of social mobility, it also highlights a new trade-off in dynamic political economy models: without social mobility, changing institutions entails delegating future political power to agents with different preferences, whereas with social mobility, even with unchanged institutions, future political power will be effectively delegated to agents with different preferences. It is also related to the conflict between the different selves of an individual (or more appropriately, of individuals who belong to the same social group in future dates), and yet its origins are not in time-inconsistent preferences, but in social mobility. It is this conflict of interest between the occupants of the same social group at different points in time that leads to nonexistence of pure-strategy equilibria as well as to multiplicity of equilibria.

### III.B. Multiplicity and Uniqueness

Let us define the bliss point of an agent currently in group  $j$  in  $\tau$  periods from now as

$$b_j^{(\tau)} = \sum_{k=1}^g \mu_{jk}^{\tau} b_k = (M^{\tau} \mathbf{b})_j;$$

consistent with this definition, we let  $b_j^{(0)} = b_j$  and  $b_j^{(\infty)} = \lim_{\tau \rightarrow \infty} (M^{\tau} \mathbf{b})_j$  (this limit exists by standard properties of

---

this and subsequent propositions hold for all parameters except a subset that is of measure zero.

stochastic matrixes).<sup>19</sup> The role of this notation is evident from the following transformation of the expected stage utility of a member of group  $j$  in  $\tau$  periods if policy  $p$  were to be implemented at that point:

$$\begin{aligned} \sum_{k=1}^g \mu_{jk}^\tau (A_k - (b_k - p)^2) &= - \left( \sum_{k=1}^g \mu_{jk}^\tau b_k - p \right)^2 + \left( \sum_{k=1}^g \mu_{jk}^\tau b_k \right)^2 \\ &\quad + \sum_{k=1}^g \mu_{jk}^\tau (A_k - b_k^2), \end{aligned}$$

where  $\mu_{jk}^\tau$  denotes the  $jk$ th element of  $M^\tau$ , the  $\tau$ th power of the mobility matrix  $M$ . The last two terms in this expression are constants (reflecting, after rearranging, the expectation of  $A_k$  and the variance of  $b_k$ ), which implies that the effective preferences of the  $\tau$ th-future self of a member of group  $j$  are quadratic with the bliss point  $b_j^{(\tau)}$ . Thus, the values  $\{b_j^{(\tau)}\}_{j \in G, \tau \geq 0}$  encode all the bliss points of all (current and future) agents, whose preferences influence, directly or indirectly, political decisions.

The next assumption is imposed in some of our results to ensure uniqueness.

**ASSUMPTION 2. (Within-person monotonicity)** For any social group  $k$ , the sequence  $b_k^{(0)}, b_k^{(1)}, b_k^{(2)}, \dots$  is monotone, meaning that either  $b_k^{(\tau)} \geq b_k^{(\tau+1)}$  for  $\tau = 0, 1, \dots$  or  $b_k^{(\tau)} \leq b_k^{(\tau+1)}$  for  $\tau = 0, 1, \dots$ .

19. This limit will play an important role in the characterization results in Sections IV and V, and it is worth emphasizing that it is easy to compute. Introduce the following notation: for every group  $j \in G$ , let  $L_M(j)$  be the set of all groups  $k$  such that  $\mu_{jk}^\tau > 0$  for some  $\tau \geq 1$ . In the language of Markov chains,  $L_M(j)$  is a component (communication class) of matrix  $M$ , and the set of components,  $\{L_M(j)\}$ , is a partition of  $G$  (i.e.,  $L_M(j_1) \cap L_M(j_2) \neq \emptyset$  and  $L_M(j_1) \cup L_M(j_2) \cup \dots = G$ ). Intuitively,  $L_M(j)$  includes all groups which a current member of group  $j$  may eventually reach. Condition (3) guarantees that a member of group  $j$  may (eventually) move to group  $k$  if and only if members of group  $k$  can move to group  $j$ . Hence, these two groups need to be part of the same component. Moreover, from Assumption 1, each component is connected, that is, whenever  $k_1 < k_2 < k_3$  and  $k_1, k_3 \in L_M(j)$ , we have  $k_2 \in L_M(j)$ . This enables us to write the preferences of individuals from group  $j$  in the very distant future as the average preferences of all agents within the same component:  $b_j^{(\infty)} = \frac{\sum_{k \in L_M(j)} n_k b_k}{\sum_{k \in L_M(j)} n_k}$ .

This within-person monotonicity assumption imposes that if an individual's preferences next period are moving to the right (left), then they will (weakly) move further right (left) in future periods. We show below that this is a sufficient condition for uniqueness. Without this assumption, multiple equilibria are possible, as demonstrated in Example B2 in [Online Appendix B](#) (notice that the multiplicity illustrated in this example is not just a multiplicity of equilibrium strategies but of induced equilibrium paths). In that example, multiplicity is a consequence of fast social mobility between the middle class and the rich and slow social mobility between the middle class and the poor. This makes the preferences of the middle class, which is pivotal, similar to those of the rich in the near term and to those of the poor in the longer term, thus highlighting the origins of multiplicity in the conflict between near and distant future selves of the current pivotal group. Under Assumption 2, however, the interests of the near and distant future selves are essentially aligned (as they agree on the direction of change), and the following theorem shows this ensures the uniqueness of equilibrium paths.<sup>20</sup>

20. Another way to view Assumption 2 is through median voter intuition. The within-person monotonicity condition and its role in uniqueness can be understood as an instance of aggregation of heterogeneous preferences—in particular, the preferences of all future selves. Consider the problem of the current decision maker comparing two states,  $x$  and  $y$ . This decision maker will be implicitly aggregating the preferences of her future selves with weights given by the discount factor and the social mobility process. Within-person monotonicity means that if self- $t$  and self- $t'$  prefer  $x$  to  $y$ , then the same is true for self- $t''$ , provided that  $t < t'' < t'$ . This order implies that each current agent acts as if she were a weighted median of her future selves. This guarantees that the preferences of future selves can be aggregated in a simple way and can be represented as the weighted median future self of the current decision maker. Since current decisions are made by the current (weighted) median voter, this implies that they will maximize the preferences of the weighted median future self of the current weighted median voter. This aggregation in turn further implies uniqueness of equilibrium—once more because of the uniqueness of the weighted median voter in the presence of such well-defined preferences. This argument also provides a complementary intuition for why within-person monotonicity is not needed when  $\beta$  is sufficiently low: in this case, tomorrow's self receives almost all of the weight, and the problem of aggregation of preferences of different future selves becomes moot. But at the same time, this problem does not disappear as  $\beta$  approaches 1, and indeed Example B2 in [Online Appendix B](#) shows that there are multiple equilibria for  $\beta$  arbitrarily close to 1: even though the preferences of all future selves are closely aligned in this case, the coordination problem that the different selves need to solve does not vanish.

**THEOREM 2. (Uniqueness)** The equilibrium is generically unique (meaning that decisions on current policy and transitions in each state are determined uniquely within the class of symmetric monotone MPE, except for a set of parameters of measure zero) if either (i) the discount factor  $\beta$  is sufficiently low, or (ii) Assumption 2 (within-person monotonicity) is satisfied.

That the equilibrium is generically unique when the players are very myopic (have a very low discount factor) follows readily from the fact that such myopic players will simply maximize their next period utility, which generically has a unique solution. It is also of limited interest, since we are more concerned with situations in which the discount factor takes intermediate values so that the current decision maker takes into account the preferences of all of her future selves. For these cases, within-person monotonicity provides a sufficient condition for uniqueness as anticipated by our previous discussion.

### III.C. Farsighted Stability

In this subsection, we characterize the conditions under which democracy (or in fact any political institution) is stable when  $\beta$  is arbitrarily close to 1. This will give us the farsighted stability condition, which will play a critical role in how social mobility impacts the stability of democracy for any discount factor. Our result in this section necessitates an additional assumption, which we state next.

**ASSUMPTION 3. (Sufficiently rich set of states)** For each group  $j \in G$ , if state  $s_j \in \arg \min_{s \in S} |b_{d_s} - b_j|$ , then  $\mu_{j d_{s_j}}^\tau > 0$  for some  $\tau > 0$ .

This assumption states that individuals in every social group have a positive probability of moving to the social group that is pivotal in their current ideal state (i.e., the state with induced policy choice maximizing the stage payoff of this individual). This assumption is not particularly restrictive as it holds automatically either if for each group, there is a state in which it is pivotal (i.e.,  $S = G$ ), or if the social mobility matrix  $M$  is ergodic (meaning that there is a positive probability that an individual from any social group can eventually reach any other social group).

For the next theorem, we impose Assumption 2 as well, and also assume the uniqueness of equilibrium explicitly to simplify the exposition. Under Assumption 2, Theorem 2 already ensures generic uniqueness, but we impose it as an additional assumption for emphasis and to avoid further reference to generic parameter values.

**THEOREM 3. (Farsighted stability of institutions)** Suppose that Assumptions 2 and 3 hold and the equilibrium is unique. Then there exists  $\tilde{\beta} < 1$  such that for any  $\beta \in (\tilde{\beta}, 1)$ , the following is true:

- i. Starting from state  $s_1$ , the sequence of states along the equilibrium path,  $s_1, s_2, \dots$ , converges, with probability 1, to state  $z$  that minimizes  $|b_{d_z} - b_{d_{s_1}}^{(\infty)}|$ .
- ii. State  $s \in S$  is stable (that is,  $q_{ss} = 1$ ) if and only if

$$(8) \quad s \in \arg \min_{z \in S} |b_{d_z} - b_{d_s}^{(\infty)}|.$$

- iii. Denote democracy by  $x$ . Then democracy is stable if and only if the farsighted stability condition,

$$(9) \quad \frac{b_{d_{x-1}} + b_{d_x}}{2} \leq b_{d_x}^{(\infty)} \leq \frac{b_{d_x} + b_{d_{x+1}}}{2}$$

holds.<sup>21</sup>

The first part of this theorem states that when players are sufficiently patient (farsighted), the sequence of equilibrium states converges to a state  $z$  that minimizes  $|b_{d_z} - b_{d_{s_1}}^{(\infty)}|$ . Put differently, the equilibrium sequence will necessarily go to state  $z$ , the most preferred state of the very distant selves of the current decision maker (group  $d_{s_1}$ ).<sup>22</sup>

21. In this condition, to formally cover the cases in which the political institutions are the lowest and highest feasible ones, that is, 1 and  $m$ , respectively, we set  $b_{d_0} = -\infty$  and  $b_{d_{m+1}} = +\infty$ , which ensures that for these lowest and highest political institutions, condition (9) is only relevant on one side.

22. Intuitively, if the equilibrium sequence did not take society to state  $z$ , then the current decision maker would have an incentive to move there immediately, because with a sufficiently large discount factor, she would be willing to sacrifice the utilities of her selves in the near future for achieving the most preferred state for her very distant selves. However, this does not imply that in equilibrium the transition to state  $z$  will be immediate even when  $\beta$  is very close to 1, as she might

More important for our focus are the second and the third parts of this theorem. The second part, which is a direct corollary of the first part, establishes that a state is stable if and only if it guarantees a policy outcome closer to the (group-size weighted) average of the political bliss points of groups to which the current decision makers can move in the future (compared to the policy outcome that will follow from other institutions).

The third part of the theorem applies this result to democracy and derives the crucial farsighted stability condition (9). This condition imposes that the preferences of the current median voter in the very distant future are closer to his own current preferences than those of the decision makers in either neighboring state.<sup>23</sup> Single-peakedness and symmetry of preferences then imply that this condition is sufficient to guarantee that the median voter’s long-run future selves prefer democracy to any other political institution. Because  $\beta$  is arbitrarily close to 1, under this condition the median voter also prefers democracy to the alternatives today, ensuring the stability of democracy. Conversely, if this condition did not hold, the current median voter would prefer to change the prevailing political institution toward one that his long-run future self prefers.

A complementary interpretation of condition (8) and its particular case, condition (9), further clarifies the intuition. Note that  $b_{d_x}^{(\infty)}$  is the average bliss point within the component of the social mobility matrix  $M$  to which group  $x$  belongs. In the special case where this component corresponds to  $G$  (when there is possibly indirect social mobility from every group to every other group),  $b_{d_x}^{(\infty)}$  is simply the average bliss point in society. This implies that the condition that  $x \in \arg \min_{z \in S} |b_{d_x} - b_{d_x}^{(\infty)}|$  requires median preferences,  $b_{d_x}$ , which are those that will be implemented by democracy, to be sufficiently close to these average preferences,  $b_{d_x}^{(\infty)}$ .

#### IV. SOCIAL MOBILITY AND THE STABILITY OF DEMOCRACY

In this section, we present our main results on how social mobility affects the stability of democracy. Once again we

---

still prefer to spend the next several periods in the current state or other states that will still lead to state  $z$  in the long run.

23. This condition is equivalent to  $|b_{d_x} - b_{d_x}^{(\infty)}| \leq |b_{d_{x-1}} - b_{d_x}^{(\infty)}|$  and  $|b_{d_x} - b_{d_x}^{(\infty)}| \leq |b_{d_{x+1}} - b_{d_x}^{(\infty)}|$ .

simplify the exposition by assuming within-person monotonicity and uniqueness, relegating the results that relax these features to [Appendix A](#). Moreover, given our focus in this section, we fix all other parameters of the model and only vary the matrix of social mobility.

**DEFINITION 2. (Comparing the speed of social mobility)** Suppose we have two matrixes of social mobility  $M$  and  $M'$  with the same components (which implies that  $\mathbf{b}^{(\infty)} = \mathbf{b}'^{(\infty)}$ ). Then, we say that social mobility is faster under  $M'$  than under  $M$  if for each group  $j \in G$  and each  $t \geq 1$ , either  $b_j \leq b_j^{(t)} \leq b_j'^{(t)} \leq b_j^{(\infty)} = b_j'^{(\infty)}$  or  $b_j \geq b_j^{(t)} \geq b_j'^{(t)} \geq b_j^{(\infty)} = b_j'^{(\infty)}$ , with the inequality between  $b_j^{(t)}$  and  $b_j'^{(t)}$  being strict at least for some  $j$ .

Thus two matrixes  $M$  and  $M'$  are comparable in terms of the speed of social mobility only if the preferences of very distant future selves coincide, which is in turn guaranteed if they have the same components. Under this condition, mobility under  $M'$  is faster if the preferences of future selves at any time  $t$  are weakly closer to  $b_j^{(\infty)}$  (and weakly further from  $b_j$ ) than under  $M$ . This definition makes it clear that faster social mobility implies that the preferences of future selves will converge more rapidly to the preferences of the very distant self,  $b_j^{(\infty)}$ , which is the feature that will be responsible for the nature of the comparative statics we present in this section.

**EXAMPLE 4.** The simplest example of a collection of matrixes that can be ranked in terms of speed of mobility can be constructed as follows. Take some matrix  $M$  satisfying within-person monotonicity. Consider a family of matrixes of social mobility  $M(\gamma) = \gamma M + (1 - \gamma)I$ , where  $I$  is the identity matrix and  $\gamma \in (0, 1]$  is a parameter. Then social mobility for  $M(\gamma')$  is faster than that in  $M(\gamma)$  if and only if  $\gamma' > \gamma$ .

Another example is the following. Take some matrix  $Z$  that satisfies within-person monotonicity. Assume that individuals are reshuffled according to  $Z$  at random times determined according to a Poisson process with rate  $\lambda \in (0, \infty)$ . If so, the probabilities of transitions over an interval of time of unit length, corresponding to the interval between the two periods where political decisions are made, is given by  $M(\lambda) = e^{-\lambda} \left( \mathbf{I} + \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} Z^k \right)$ . In this case, social mobility for  $M(\lambda')$  is faster than  $M(\lambda)$  if and only if  $\lambda' > \lambda$ .

The next theorem shows that the relationship between social mobility and the stability of democracy depends on the preferences of the distant future selves of the current median voter.

**THEOREM 4. (When social mobility increases the stability of democracy)** Suppose that within-person monotonicity (Assumption 2) holds and the equilibrium is unique. Suppose also that social mobility under  $M'$  is faster than under  $M$ , and the farsighted stability condition (9) holds for either  $M$  or  $M'$  (these conditions are equivalent). Then democracy is more stable for  $M'$  than for  $M$ . More precisely, democracy is stable under both  $M$  and  $M'$ , and furthermore, if it is asymptotically stable under  $M$ , then it is also asymptotically stable under  $M'$ .<sup>24</sup>

The theorem thus supports de Tocqueville's hypothesis that social mobility contributes to the stability of democracy, provided that the farsighted stability condition (9) holds. The intuition for this result is that faster social mobility makes time run faster, making the preferences of all future selves closer to  $\mathbf{b}^{(\infty)}$ , and we know that under  $\mathbf{b}^{(\infty)}$ , the very distant selves of the median voter prefer democracy to other institutions. Put differently, with faster social mobility, individuals put less weight on events in the near future because the near future itself becomes more transient, and consequently, their preferences become more aligned with those of their distant selves, who favor democracy under condition (9). This implies that whenever democracy is stable under  $M$ , it will also be stable under  $M'$  (and the converse is not true).

Why does asymptotic stability under  $M$  guarantee asymptotic stability under  $M'$ ? To understand this result, recall that faster social mobility also implies that for any  $\beta$ , the preferences of all future selves of all social groups approach the preferences of their very distant selves, and because the preferences of the very distant selves are the same for all groups (within the component), the

24. The following stronger version of this result is also proved at the end of the proof of Theorem 4: let  $q_{sz}$  be the probability of transitioning from state  $s$  to state  $z$  under  $M$ , and  $q'_{sz}$  be the same probability under  $M'$ . Let us also denote democracy by  $x$ . Then  $q'_{x-1,x} \geq q_{x-1,x}$  (with strict inequality, unless  $q'_{x-1,x} = q_{x-1,x} = 1$ ) and  $q'_{x+1,x} \geq q_{x+1,x}$  (with strict inequality, unless  $q'_{x+1,x} = q_{x+1,x} = 1$ ), so that the speed of reaching democracy from neighboring states is greater under  $M'$  than under  $M$ . A similar strengthening of Theorem 5 can also be proved, but is omitted to save space.

preferences of all social groups approach each other as well. Recall from Theorem 3 that when condition (9) holds, the very distant selves of the current decision maker prefer democracy to any other political system; this is also true for any other group in the same component as the current decision maker, and consequently, faster social mobility makes neighboring groups (that are in the same component) also prefer democracy to any other political system. This ensures that asymptotic stability under  $M$  translates into asymptotic stability under  $M'$  (and once again, the converse not being true).

What if the farsighted stability condition (9) does not hold? In this case, the median voter would like to empower a group other than the one containing the median voter, which implies a deviation from democracy. This does not necessarily imply that she would want to go to this state immediately, and democracy may still be stable, because she may receive greater utility from staying in democracy than the entire transition path following a deviation. Nevertheless, it does imply that faster social mobility makes democracy less stable as we show in the next theorem.

**THEOREM 5. (When social mobility reduces the stability of democracy)** Suppose that Assumptions 2 and 3 hold and the equilibrium is unique. Suppose also that social mobility under  $M'$  is faster than under  $M$ , farsighted stability condition (9) does not hold (for  $M$  or, equivalently, for  $M'$ ), and

$$(10) \quad \frac{b_{d_{x-2}} + b_{d_{x-1}}}{2} \leq b_{d_{x-1}}^{(\infty)} \leq b_{d_{x+1}}^{(\infty)} \leq \frac{b_{d_{x+1}} + b_{d_{x+2}}}{2}.$$

Then democracy is less stable for  $M'$  than for  $M$ . More precisely, democracy is asymptotically stable at neither  $M$  nor  $M'$ , and if it is not stable at  $M$ , then it is not stable at  $M'$  either.

The substantive result of this theorem is that, when the farsighted stability condition (9) does not hold, faster social mobility has the opposite effect to that maintained by de Tocqueville's hypothesis: it makes democracy less stable. For social mobility matrixes that characterize either sufficiently slow or sufficiently fast mobilities (i.e., those that in matrix terms are sufficiently close to the unit matrix  $\mathbf{I}$  and the limit matrix  $M^\infty$  respectively), this results holds without additional conditions. If we impose the

additional condition (10), then stability of democracy decreases in the speed of social mobility monotonically (this would not hold more generally because of slippery slope effects, see Section V.A). The intuition for this result is closely related to that of Theorem 4. When faster social mobility aligns the preferences of the current median voter with her very distant selves who prefer an alternative institution, this may destabilize an otherwise stable democracy.<sup>25</sup>

Why does this theorem need condition (10)? The reason is the slippery slope considerations which will be discussed in greater detail in the next section: these considerations may make individuals unwilling to move to an institution that is more preferred in the short run because this transition might pave the way to yet more transitions, which may be less desirable for them. In this instance, as the speed of social mobility increases, institutions that lie between democracy and the institution most preferred by the very distant self may become unstable as well, and this might in turn make democracy stable because, due to slippery slope concerns, the current decision maker may not wish to move to these unstable institutions in the next period. Condition (10), on the other hand, ensures stability of the neighboring states, thus alleviating the slippery slope effect.

Even when condition (10) does not hold, similar conclusions to those of Theorem 5 hold, provided that social mobility under  $M'$  is sufficiently faster than social mobility under  $M$ , as shown in the next corollary.

**COROLLARY 1.** Suppose that within-person monotonicity (Assumption 2) holds and the equilibrium is unique. Suppose also that the farsighted stability condition (9) does not hold. Then there exist  $T_1$ ,  $T_2$ , and  $\varepsilon$  such that if  $|b_{d_x}^{(T_1)} - b_{d_x}| < \varepsilon$  and  $|b_{d_x}^{(T_2)} - b_{d_x}^{(\infty)}| < \varepsilon$ , that is, if social mobility under  $M$  is sufficiently slow and social mobility under  $M'$  is sufficiently fast, then democracy is less stable for  $M'$  than for  $M$ . Specifically, democracy is stable but not asymptotically stable under  $M$  and is neither stable nor asymptotically stable under  $M'$ .

25. Theorems B1 and B2 in Online Appendix B provide generalizations of the previous two theorems for the case in which the within-group monotonicity assumption, Assumption 2, is relaxed.

## V. FURTHER RESULTS AND EXTENSIONS

In this section we discuss slippery slope considerations and extend our main results to an environment without the within-person monotonicity assumption.

V.A. *Slippery Slopes*

We emphasized in the context of Theorem 5 how slippery slope considerations, which discourage a transition to a preferred state because of subsequent transitions that this would unleash, play a role in shaping when democracy may remain stable even when the preferences of future selves favor another state. More precisely, slippery slope considerations refer to the situation where in some state  $s$ , a winning coalition (e.g., a weighted majority) would obtain greater stage payoffs in some state  $x \neq s$  than in  $s$ , but in equilibrium stays in  $s$  because it anticipates further, less preferred transitions after the move to  $x$  (see Acemoglu, Egorov, and Sonin 2012). In models without social mobility, slippery slope considerations are more powerful when the discount factor is closer to 1 because in this case agents care little about the outcomes in the next period and a lot about future outcomes. Slippery slope considerations continue to be important in models of social mobility, but they arise not when the discount factor is high but when it is intermediate. The next theorem characterizes the extent of slippery slope considerations. Like all remaining results in the article, the proof of this theorem is in [Online Appendix B](#).

**THEOREM 6. (Slippery slopes)** Suppose that Assumptions 2 and 3 hold. There exist  $0 \leq \beta_0 < \beta_1 < 1$  such that for any  $\beta \in (0, 1) \setminus (\beta_0, \beta_1)$ , if some state  $s \in S$  is stable, then for any  $x \in S$ , the expected continuation utility of pivotal group  $d_s$  from staying in  $x$  forever cannot exceed their equilibrium continuation utility:

$$(11) \quad \sum_{t=1}^{\infty} \sum_{k \in G} \beta^t \mu_{d_s, k}^t u_k(b_{d_s}) \geq \sum_{t=1}^{\infty} \sum_{k \in G} \beta^t \mu_{d_s, k}^t u_k(b_{d_x}).$$

Furthermore, if for any states  $s \neq x$ ,  $b_{d_s}^{(1)} \neq \frac{b_{d_s} + b_{d_x}}{2}$ , then one can take  $\beta_0 > 0$ .

If, on the other hand,  $\beta \in (\beta_0, \beta_1)$ , condition (11) need not hold, and slippery slope considerations can prevent certain transitions.

In other words, this result suggests that for both high and low  $\beta$ , all stable states give higher expected utility to the current decision maker than any other state (with the expectation taken with respect to the social mobility process).<sup>26</sup> When slippery slope considerations are important, this need not be the case: there may be a state providing a higher expected utility to the current decision maker than the current state, but moving to this state would unleash another set of transitions that reduce the discounted continuation payoff of the current decision maker. Theorem 6 shows that such slippery slope considerations arise only for intermediate values of  $\beta$ . (See Example B1 in [Online Appendix B](#) for the second part of the theorem.)

The intuition for why slippery slope considerations do not play a role for myopic players (with low  $\beta$ ) is straightforward: myopic players care only about the next period's state, so the subsequent moves do not modify their rankings over states. That these considerations do not arise for very farsighted players (with high  $\beta$ ) is more interesting and perhaps surprising. Suppose a situation in which the current decision maker, who is pivotal in the current state  $s$ , prefers a different state,  $x$ , where by definition he will not belong to the pivotal group unless his preferences change due to social mobility. Such preferences are possible only when members of the current pivotal group have a positive probability of joining the group that is pivotal in state  $x$  (and conversely, those in the group pivotal in state  $x$  could move to the group that is pivotal in state  $s$ ). An implication is that even though the distribution of political power in states  $s$  and  $x$  have a conflict of interest today, because of social mobility their preferences in the distant future will be aligned. Therefore, with a sufficiently high discount factor, the current decision maker will not be worried about decision rights shifting to the group that is pivotal in state  $x$ , averting slippery slope considerations. In contrast, with intermediate discount factors, the loss of control in the near future can trigger concerns about slippery slopes, encouraging the current decision maker not to move in the direction of states that increase their immediate payoffs. Notably, this result is very different from that in [Acemoglu, Egorov, and Sonin \(2012\)](#), where slippery slope considerations became more important as the discount factor became larger. The difference is due to the fact that social mobility

26. The condition  $b_{d_s}^{(1)} \neq \frac{b_{d_s} + b_{d_x}}{2}$  in this theorem rules out situations where tomorrow's self is exactly indifferent between these two states.

changes the nature of the slippery slope concerns (and as social mobility limits to zero, we recover the result in [Acemoglu, Egorov, and Sonin 2012](#)).

### *V.B. Alternative Decision-Making Rules*

So far, we have focused on an extensive-form political game (or its reduced-form version in the text) that generalizes the standard median voter theorem to our environment. As a result, in each state a particular group emerges as the pivotal one that makes the policy and political decisions. In many political settings, including those with probabilistic voting and some lobbying models, political decisions are made as if a well-defined weighted social welfare function is being maximized (e.g., [Lindbeck and Weibull 1987](#); [Coughlin 1992](#); [Grossman and Helpman 1994](#)). In this subsection, we show that our main results generalize with minor modifications to this type of alternative decision-making rule.

We assume, as in our baseline model, that each state  $s$  is associated with a vector of weights  $\{w_k(s)\}$ , but now these are not voting weights, but weights in the “social welfare function” (itself presumably resulting from another political game, which we are not explicitly describing). More specifically, let us assume that the society, when in period  $t$  in state  $s = s_t$ , chooses policy  $p_t$  to maximize

$$(12) \quad p_t = \arg \max_{p \in \mathbb{R}} \sum_{k \in G} w_k(s) u_k(p),$$

and chooses state  $s_{t+1}$  as

$$(13) \quad s_{t+1} \in \arg \max_{x \in S} \sum_{k \in G} w_k(s) n_k \sum_{j \in G} \mu_{d_s j} V_j(x),$$

where  $\{V_j(x)\}_{j \in G}^{x \in S}$  are the value functions given in [equation \(7\)](#). We can interpret these decisions as being made by a fictitious decision maker, with policy preferences given by [equation \(12\)](#) and political preferences given by [equation \(13\)](#).<sup>27</sup> Different states lead

27. The use of the same weights in the two problems implies that different social groups (or their organizations) are equally influential in policy and political decisions.

to different weights  $\{w_k(s)\}$  and thus to different fictitious decision makers.

Equilibrium policy in state  $s$  is the bliss point of the fictitious decision maker, which is

$$\hat{b}_s = \frac{\sum_{k=1}^g w_k(s) n_k b_k}{\sum_{k=1}^g w_k(s) b_k}.$$

Generically,  $\hat{b}_s$  are different for different states, and then it is without loss of generality to assume that states are ordered such that  $\{\hat{b}_s\}$  is increasing in  $s$ .

Next, to determine political decisions, let us first define the “currently preferred” state of group  $k$  as  $\hat{s}^k$ , given as

$$\hat{s}^k = \arg \max_{s \in S} u_k(p(s)).$$

(Note that for this definition we are only considering stage payoffs and hence the reference to “currently preferred”.) This state is generically unique for each group  $k$ . Let us also define democracy as state  $\hat{s}^d$ , where group  $d$  contains the median voter.

We can also define the ideal policies of future selves of the fictitious decision makers. In particular, the ideal policy of the  $\tau$ -periods ahead future self of the fictitious decision maker in state  $s$  is

$$\hat{b}_s^{(\tau)} = \sum_{j=1}^g w_j(s) n_j \sum_{k=1}^g \mu_{jk}^\tau b_k = \sum_{j=1}^g w_j(s) n_j (M^\tau \mathbf{b})_j.$$

With this notation, we can now introduce the analogue of Assumption 2:

**ASSUMPTION 2.’ (Modified within-person monotonicity)** For each  $s$ ,  $\hat{b}_s^{(\tau)}$  is monotone in  $\tau$ .

The following two theorems are direct analogues of Theorems 4 and 5, and their proofs follow closely those of these two previous theorems and are thus omitted.

**THEOREM 7. (Social mobility and the stability of democracy under alternative decision-making rule)** Suppose that Assumption 2’ holds and the equilibrium is unique. Suppose also that social mobility under  $M'$  is faster than under  $M$ , and

inequality

$$(14) \quad \frac{\hat{b}_{x-1} + \hat{b}_{x+1}}{2} \leq \hat{b}_x^{(\infty)} \leq \frac{\hat{b}_x + \hat{b}_{x+1}}{2}$$

holds for either  $M$  or  $M'$ . Then democracy is more stable for  $M'$  than for  $M$ . More precisely, democracy is stable under both  $M$  and  $M'$  and furthermore, if it is asymptotically stable under  $M$ , then it is also asymptotically stable under  $M'$ .

**THEOREM 8. (Social mobility and the instability of democracy under alternative decision-making rule)** Suppose that Assumptions 2' and 3 hold and the equilibrium is unique. Suppose also that social mobility under  $M'$  is faster than under  $M$ , condition (14) does not hold (for  $M$  or, equivalently, for  $M'$ ), and

$$(15) \quad \frac{\hat{b}_{x-2} + \hat{b}_{x-1}}{2} \leq \hat{b}_{x-1}^{(\infty)} \leq \hat{b}_{x+1}^{(\infty)} \leq \frac{\hat{b}_{x+1} + \hat{b}_{x+2}}{2}.$$

Then democracy is less stable for  $M'$  than for  $M$ . More precisely, democracy is asymptotically stable at neither  $M$  nor  $M'$ , and if it is not stable at  $M$ , then it is not stable at  $M'$  either.

In other words, if the distant future selves of the fictitious decision maker in democracy prefer democracy to neighboring institutions, then greater social mobility makes democracy more stable. Otherwise (and provided that the fictitious decision makers in neighboring states prefer those states in the long run), greater social mobility makes democracy less stable.

## VI. CONCLUSION

An influential thesis often associated with Alexis de Tocqueville views social mobility as an important bulwark of democracy: when members of a social group expect to transition to some other social group in the near future, they should have less reason to exclude these other social groups from the political process. Despite the importance of this thesis for the evolution of the modern theories of democracy and its continued relevance in contemporary debates, it has received little attention in the modern political economy literature. This article has investigated the

link between social mobility and the dynamics of political institutions. Our framework provides a natural formalization of de Tocqueville's hypothesis, showing that greater social mobility can further enhance the stability of democracy for reasons anticipated by de Tocqueville.

However, more importantly, it also demonstrates the limits of this hypothesis. There is a robust reason greater social mobility can undermine the stability of democracy: when the median voter expects to move up (respectively, down), she would prefer to give less voice to poorer (respectively, richer) social groups, because she anticipates having different preferences than future agents who will occupy the same social station as herself.

We provided a tight characterization of these two competing forces and demonstrated that the impact of social mobility depends on whether the mean and the median of preferences over policy are close. When they are, not only is democracy stable (meaning that the median voter would not wish to undermine democracy), but it also becomes more stable as social mobility increases. Conversely, when the mean and median are not close, greater social mobility reduces the stability of democracy.

In addition to enabling a tight characterization of the relationship between social mobility and stability of democracy, our theoretical analysis also shows that in the presence of social mobility, the political preferences of an individual depend on the potentially conflicting preferences of her future selves, under certain conditions paving the way to multiple equilibria. When society is mobile, the current political institution may be disliked by the current decision makers not only because their future selves prefer another institution (which was at the root of the instability of democracy in the presence of high social mobility), but also because if the current institution were to continue, future decision makers might choose transitions that are not favored by the future selves of the current decision maker (which is a form of slippery slope consideration).

Motivated by this reasoning, we further characterized the conditions for general slippery slope considerations—which prevent certain institutional choices because of the additional series of changes that these choices would induce. But differently from other dynamic political economy settings, slippery slope concerns are more important when the discount factor takes intermediate values rather than when it is large. This is because in the presence of social mobility, high discount factors make current

decision makers not care about losing political power to another social group (since, in the long run, they will have preferences similar to the members of the group that will become pivotal in a different state). But with intermediate discount factors, they still care a lot about political developments in the next several periods, making slippery slope considerations relevant again.

There are many fruitful areas of research related to the political implications of social mobility. First, there is a clear need for systematic empirical analyses of the impact of social mobility (and perceptions thereof) on political attitudes and the resulting political behavior. Second, social mobility itself is not just determined by exogenous (technological or historical) factors but also determined by political choices of the same players deciding other dimensions of policy and future political institutions. The working paper version of our work, [Acemoglu, Egorov, and Sonin \(2016\)](#), showed how our results can be extended to an environment with endogenous social mobility under a number of simplifying assumptions. A more systematic analysis of endogenous social mobility is an obvious area for future theoretical study. Third, our framework can also be enriched to include individual decisions, such as on the quantity or quality of education, which affect the mobility of the members of a dynasty, while also shaping political attitudes. Fourth, the framework we presented here can be generalized to include political actions by different political coalitions (e.g., collective action, social unrest, or coups), which will be affected by social mobility as well. Finally, we also abstracted from structural change and social change which often accompany periods of rapid social mobility and impact the sizes of different social groups. An extension in this direction would be particularly interesting as it could improve our understanding of what types of structural changes contribute to the emergence and consolidation of democracy via both their direct effects and indirectly by changing the level of social mobility.

#### APPENDIX A: PROOFS OF MAIN RESULTS

We start by presenting an alternative extensive-form game, where agents from all groups can present proposals and vote. We then prove the results of the article for both this game and the one stated in [Section II](#).

A. *Extensive-Form Game for Policy and Political Decisions*

We describe in detail our extensive-form game which, as already noted, gives the same equilibrium outcomes as those implied by the more reduced-form game presented in the text. In the rest of this appendix, we use this full extensive-form game.

Take a fixed order of groups in each state,  $\pi_s : \{1, \dots, g\} \rightarrow G$ , which determines the sequence in which (representatives of) different groups make proposals, and in which group  $d_s$  is included among the proposers in state  $s$  (which is trivially satisfied if all groups have the opportunity to make proposals in each state).

The first period's state  $s_1$  is exogenously given, and so is some default policy,  $p_0$ , in the first period. Thereafter, denoting the group that individual  $i$  belongs to at time  $t$  by  $g_i^t$ , the timing in each period  $t \geq 1$  is as follows.

i. *Policy decision:*

- a. In each state  $s_t$ , we start with  $j = 1$  and the default option of preserving the previous period's policy,  $p_t^0 = p_{t-1}$ .
- b. A random agent  $i$  from group  $\pi_{s_t}(j)$  is chosen as the agenda setter and makes an amendment (policy proposal)  $\tilde{p}_t^j$ . (Since all members of social groups have the same preferences, which agent is chosen to do this is immaterial.)
- c. All individuals vote, sequentially, with each agent  $i$  casting vote  $v_i^p(j) \in \{Y, N\}$ .
- d. If  $\frac{\sum_{i=1}^n w_{s_t^i}(s_t) \mathbf{1}\{v_i^p(j)=Y\}}{\sum_{i=1}^n w_{s_t^i}(s_t)} > \frac{1}{2}$ , then the current proposal becomes the default policy ( $p_t^j = \tilde{p}_t^j$ ), otherwise the default policy stays the same ( $p_t^j = p_t^{j-1}$ ). The game returns back to stage i(b) with  $j$  increased by 1, unless  $j = g$ .
- e. The policy decided in the last stage is implemented:  $p_t = p_t^g$ .

ii. *Political decision:*

- a. In each state  $s_t$ , the default option to preserve the current institution,  $s_{t+1}^0 = s_t$ , is on the table, and we start with  $j = 1$ .
- b. A random agent  $i$  from group  $\pi_{s_t}(j)$  is chosen as the agenda setter and makes an amendment (proposal of political transition),  $\tilde{s}_{t+1}^j$ .

- c. All individuals vote, sequentially, with each individual  $i$  casting vote  $v_i^s(j) \in \{Y, N\}$ .
- d. If  $\frac{\sum_{i=1}^n w_{g_i^t}(s_t) \mathbf{1}\{v_i^s(j)=Y\}}{\sum_{i=1}^n w_{g_i^t}(s_t)} > \frac{1}{2}$ , then the current proposal becomes the default transition ( $s_{t+1}^j = \tilde{s}_{t+1}^j$ ), otherwise the default transition stays the same ( $s_{t+1}^j = s_{t+1}^{j-1}$ ). The game returns back to stage ii(b) with  $j$  increased by 1, unless  $j = g$ .
- e. The transition decided in the last stage is implemented:  

$$s_{t+1} = s_{t+1}^g.$$
- iii. *Payoffs*: Each individual  $i$  receives time- $t$  payoff of  $u_{g_i^t}(p_t)$ , given by (1).
- iv. *Social mobility*: At the end of the period, there is social mobility, so that individual  $i$  who belonged to group  $g_i^t$  in period  $t$  will start period  $t + 1$  in group  $k$  with probability  $\mu_{g_i^t, k}$ .

This specific game form, where proposals (for policies or political transitions) within a period are accepted temporarily and act as a status quo until the whole sequence of proposals is made, is similar to the “amendments” games discussed in [Austen-Smith and Banks \(2005\)](#).

In this game, we also focus on symmetric monotone MPE (Definition 1). The only difference is the definition of strategies; here, a strategy of player  $i$  is a mapping from history (which codifies her current group affiliation, the current institution, as well as the entire sequence of moves within the period). This mapping is into  $\mathbb{R}$  when player  $i$  is making a policy proposal; into  $\Delta(S)$  when she is proposing political transition; and into  $\{Y, N\}$  when she is at the voting stage.

The proofs presented next make it clear that the (subgame perfect) equilibrium of this extensive-form game will make players from a particular group pivotal in each state, and thus the results that follow from this extensive-form game are entirely analogous to those from our reduced-form modeling assumption in the text, where a group was directly specified as pivotal in each state, and policy and political decisions were made by a randomly chosen member of that group.

## B. Proofs of Main Results

We provide proofs of Theorems 1–5, for which we need a number of lemmas. Proofs of Lemmas A4–A8 are relegated to

**Online Appendix B.** To formulate intermediate results, which together establish that continuation utilities satisfy increasing differences, we will need the following notation. First, define two constants:

$$\bar{U} = \max_{j \in G} |A_j| + \max_{j, k \in G} (b_k - b_j)^2,$$

$$\bar{u} = \min_{j, k \in G: j \neq k} (b_k - b_j)^2.$$

In what follows, we say that a  $gm$ -dimensional vector  $v = \{v_j(x)\}_{j \in G}^{x \in S} \in \mathbb{R}^{gm}$  satisfies increasing differences if for  $j_1, j_2 \in G$  and  $x_1, x_2 \in S, j_1 < j_2$  and  $x_1 < x_2$  implies  $v_{j_2}(x_2) - v_{j_2}(x_1) > v_{j_1}(x_2) - v_{j_1}(x_1)$ . We call a subset  $X \subset S$  connected if  $X = [a, b] \cap S$  for some integers  $a, b$ . We also use the strong set order: that is, sets  $X, Y \subset S$  satisfy  $X \leq Y$  if  $\min X \leq \min Y$  and  $\max X \leq \max Y$ , and moreover, for  $X \subset S$  and  $y \in S, X \leq y$  if  $X \leq \{y\}$ . Other binary relations ( $<, \geq, >$ ) are defined similarly. We will use  $\Phi_s$  to denote the set of states to which the society can transition (in the next period) starting from state  $s$  in equilibrium, or more formally  $\Phi_s = \{x \in S : q_{sx} > 0\}$ .

**LEMMA A1.** Suppose that vector  $\{V_j(x)\}_{j \in G}^{x \in S} \in \mathbb{R}^{gm}$  satisfies increasing differences. Let

$$(16) \quad W_j(x) = \sum_{k \in G} \mu_{jk} V_k(x).$$

Then vector  $\{W_j(x)\}_{j \in G}^{x \in S} \in \mathbb{R}^{gm}$  also satisfies increasing differences.

*Proof of Lemma A1.* Take two states  $x, y \in S$  such that  $x < y$  and consider the difference

$$W_j(y) - W_j(x) = \sum_{k \in G} \mu_{jk} Z_k,$$

where  $Z_k = V_k(y) - V_k(x)$  is a sequence that is increasing in  $k$  by assumption. Let  $j, l \in G$  satisfy  $j < l$ . Since, by Assumption 1, the probability distribution  $\{\mu_j\}$  is first-order stochastically dominated by  $\{\mu_l\}$ , the expected values of a monotone sequence  $\{Z_k\}$

satisfy the inequality

$$\sum_{k \in G} \mu_{jk} Z_k < \sum_{k \in G} \mu_{lk} Z_k.$$

This implies

$$W_j(y) - W_j(x) < W_l(y) - W_l(x),$$

which proves that  $\{W_j(x)\}_{j \in G}^{x \in S}$  satisfies increasing differences. ■

LEMMA A2. Suppose that vector  $\{V_j(x)\}_{j \in G}^{x \in S} \in \mathbb{R}^{gm}$  satisfies increasing differences. Suppose that matrixes  $Q = \{q_{sz}\}_{s,z \in S}$  are such that for  $x < y$ , the distribution  $q_x$  is (weakly) first-order stochastically dominated by  $q_y$ . Then  $\{V'_j(x)\}_{j \in G}^{x \in S}$ , defined by

$$(17) \quad V'_j(x) = u_j(b_{d_x}) + \beta \sum_{y \in S} q_{xy} \sum_{k \in G} \mu_{jk} V_k(y),$$

satisfies increasing differences; moreover, if  $j, l \in G, x, y \in S$  and  $j < l, x < y$ , then

$$(18) \quad (V'_l(y) - V'_l(x)) - (V'_j(y) - V'_j(x)) \geq 2\bar{u}.$$

*Proof of Lemma A2.* Take two groups  $j, l \in G$  with  $j < l$ . For each  $s \in S$ , consider the following difference:

$$V'_l(s) - V'_j(s) = (u_l(b_{d_s}) - u_j(b_{d_s})) + \beta \sum_{z \in S} q_{sz} (W_l(z) - W_j(z)).$$

By Lemma A1, the term  $W_l(z) - W_j(z)$  is increasing in  $z$ . Take  $x, y \in S$  such that  $x < y$ ; then distribution  $q_x$  is (weakly) first-order stochastically dominated by  $q_y$ , and thus the expectation of  $W_l(z) - W_j(z)$  is weakly smaller when evaluated with the former distribution than with the latter, that is,

$$\sum_{z \in S} q_{xz} (W_l(z) - W_j(z)) \leq \sum_{z \in S} q_{yz} (W_l(z) - W_j(z)).$$

We thus have

$$\begin{aligned}
 & (V'_l(y) - V'_j(y)) - (V'_l(x) - V'_j(x)) \\
 &= (u_l(b_{d_y}) - u_j(b_{d_y})) - (u_l(b_{d_x}) - u_j(b_{d_x})) \\
 &+ \beta \left( \sum_{z \in S} q_{yz} (W_l(z) - W_j(z)) - \sum_{z \in S} q_{xz} (W_l(z) - W_j(z)) \right) \\
 &\geq 2(b_l - b_j)(b_{d_y} - b_{d_x}) \geq 2\bar{u}. \quad \blacksquare
 \end{aligned}$$

LEMMA A3. Suppose that vector  $W = \{W_j(x)\}_{j \in G}^{x \in S} \in \mathbb{R}^{gm}$  satisfies increasing differences. Suppose that  $X, Y$  are connected subsets of  $S$  and  $X \leq Y$ . Suppose  $j, k \in G$  and  $j < k$ , and suppose  $x \in \arg \max_{z \in X} W_j(z)$  and  $y \in \arg \max_{z \in Y} W_k(z)$ . Then  $x \leq y$ .

*Proof of Lemma A3.* Suppose, to obtain a contradiction, that  $x > y$ . Since  $X$  and  $Y$  are connected and  $X \leq Y$ , this implies that  $x, y \in X \cap Y$ . Now,  $x \in \arg \max_{z \in X} W_j(z)$  implies  $W_j(x) \geq W_j(y)$ , and since  $W$  satisfies increasing differences,  $x > y$  and  $k > j$ , it must be that  $W_k(x) > W_k(y)$ . However, this contradicts that  $y \in \arg \max_{z \in Y} W_k(z)$ .  $\blacksquare$

In the following proofs, we slightly abuse notation  $W_j(x, y, z, \dots)$  to denote the continuation value of group  $j$  when the sequence of states is  $x, y, z, \dots$

*Proof of Theorem 1.* We first establish the existence of a monotone symmetric MPE (existence of some MPE trivially follows from Kakutani’s theorem). We instead prove existence of a symmetric monotone MPE in a more general class of games, where some transitions are ruled out. This generality will be used in later proofs. Specifically, we require that all proposals  $x$  made in state  $s$  must satisfy  $x \in F_s$ , where  $F_s \subset S$ , and  $\{F_s\}_{s \in S}$  satisfies the following two conditions: (i) for each  $s, s \in F_s$  and (ii) if  $x < y < z$  or  $x > y > z, z \in F_x$  implies  $y \in F_x$  and  $z \in F_y$ . If we do so, then the statement of Theorem 1 follows immediately as a special case when all transitions are feasible (i.e.,  $F_s = S$  for all  $s \in S$ ).

We prove this claim in two steps. First, we construct a feasible monotone transition correspondence, that is, we construct a matrix  $\hat{Q}$  such that  $\hat{q}_{sx} > 0$  only if  $x \in F_s$ , and also  $\hat{q}_x$  weakly first-order stochastically dominates  $\hat{q}_y$  whenever  $x > y$ .

Second, we prove that there is an equilibrium  $\sigma$  such that  $Q(\sigma) = \hat{Q}$ .

Define  $\Pi \subset \mathbb{R}^{gm}$  by the following constraints:  $\{V_j(x)\}_{j \in G}^{x \in S} \in \Pi$  if and only if (i) for all  $j \in G, x \in S, |V_j(x)| \leq \frac{\bar{U}}{1-\beta}$  and (ii) for all  $j, k \in G$  such that  $j < k$  and for all  $x, y \in S$  such that  $x < y$ ,

$$(V_k(y) - V_k(x)) - (V_j(y) - V_j(x)) \geq 2\bar{u}.$$

This implies, in particular, that any  $\{V_j(x)\}_{j \in G}^{x \in S} \in \Pi$  satisfies strict increasing differences, and that  $\Pi$  is compact and convex.

Consider the following correspondence  $\Upsilon$  from  $\Pi$  into itself. Take a vector of values  $V = \{V_j(x)\}_{j \in G}^{x \in S} \in \Pi$ , and let  $W = \{W_j(x)\}_{j \in G}^{x \in S}$  be given by equation (16). For each state  $s \in S$ , let  $p_s$  be the ideal policy of pivotal group  $d_s$ , that is,  $p_s = b_{d_s}$ , and let  $\Psi_s$  be the expected utility of the members of pivotal group  $d_s$  from transitioning into state  $s$ , that is,  $\Psi_s = \arg \max_{x \in F_s} W_{d_s}(x)$ . Furthermore, let  $\lambda_s$  be any probability distribution over  $S$ , the support of which is a subset of  $\Psi_s$ , and let  $\Lambda_s$  be the set of such distributions. We also define  $\Upsilon(V) \subset \Pi$  to be such that  $V' \in \Upsilon(V)$  if and only if for each  $s \in S$  there is  $\lambda_s \in \Lambda_s$  such that for each  $j \in G$ ,

$$(19) \quad V'_j(s) = u_j(p_s) + \beta \sum_{x \in S} \lambda_s(x) W_j(x).$$

Let us prove that  $\Upsilon(V)$  is nonempty for any  $V \in \Pi$ . For each  $s$ , take any  $\lambda_s \in \Lambda_s$  (which exists, because  $\Lambda_s$  is nonempty), and define  $V'_j(s)$  as in equation (19). Then for all  $j \in G$  and  $s \in S$ ,

$$\begin{aligned} |V'_j(s)| &\leq |u_j(p_s)| + \beta |W_j(z_x)| \\ &\leq |u_j(p_s)| + \beta \sum_{k \in G} \mu_{jk} |V_k(z_x)| \\ &\leq \bar{U} + \beta \frac{\bar{U}}{1-\beta} = \frac{\bar{U}}{1-\beta}. \end{aligned}$$

Furthermore, notice that since  $W$  satisfies increasing differences, for any  $x, y \in S$  where  $x < y$ , any  $a \in \Psi_x$  and  $b \in \Psi_y$  must satisfy  $a \leq b$  (by Lemma A3), and thus there is  $c \in S$  such that  $\Psi_x \leq \{c\} \leq \Psi_y$ , which implies that any  $\lambda_x \in \Lambda_x$  is (weakly) first-order stochastically dominated by any  $\lambda_y \in \Lambda_y$ . Lemma A2 now implies

that  $V'$  satisfies equation (18). Therefore,  $V' \in \Pi$ , which means that  $\Upsilon(V)$  is nonempty for any  $V \in \Pi$ .

We now prove that  $\Upsilon(V)$  is convex for all  $V$ . Suppose  $V', V'' \in \Upsilon(V)$ . Let the corresponding probability distributions in  $\Lambda_s$  be  $\lambda'_s$  and  $\lambda''_s$ , respectively. For any  $\alpha \in (0, 1)$ ,  $\alpha\lambda'_s + (1 - \alpha)\lambda''_s$  is a probability distribution in  $\Lambda_s$ , and in particular its support is in  $F_s$ , and moreover,

$$u_j(p_s) + \beta \sum_{x \in S} (\alpha\lambda'_s + (1 - \alpha)\lambda''_s) W_j(x) = \alpha V'_j(s) + (1 - \alpha) V''_j(s).$$

Thus, for any  $\alpha \in (0, 1)$ ,  $\alpha V' + (1 - \alpha)V'' \in \Upsilon(V)$ , which implies convexity of  $\Upsilon(V)$ .

We next prove that  $\Upsilon(\cdot)$  is an upper-hemicontinuous correspondence. Notice that it is a composition of the following mappings: (i)  $\arg \max_{x \in F_s} W_{d_x}(x)$ , which is a mapping from  $\Pi$  to  $2^S \setminus \{\emptyset\}$ , the set of nonempty subsets of  $S$  (and has a closed graph when  $2^S \setminus \{\emptyset\}$  is endowed with discrete topology); (ii) a mapping from  $2^S \setminus \{\emptyset\}$  to  $\Delta(S)$ , where each subset  $X \in 2^S \setminus \{\emptyset\}$  is mapped to the set of probability distributions on  $S$  with support in  $X$ , which also has a closed graph; and (iii) a mapping from  $\Delta(S)$  to  $\Pi$ , which is linear and thus continuous. Because a composition of upper-hemicontinuous correspondences is upper-hemicontinuous,  $\Upsilon(V)$  also satisfies this property.

Since  $\Upsilon(\cdot)$  is upper-hemicontinuous and  $\Upsilon(V)$  is nonempty and convex-valued for all  $V \in \Pi$ , and  $\Pi$  is compact and convex, Kakutani's theorem implies that there is  $V \in \Pi$  such that  $V \in \Upsilon(V)$ . By definition of  $\Upsilon(V)$  there are  $\{\lambda_s\}_{s \in S}$  that satisfy

$$V_j(s) = u_j(p_s) + \beta \sum_{x \in S} \lambda_s(x) W_j(x).$$

Define the matrix  $\hat{Q}$  by setting  $\hat{q}_{sx} = \lambda_s(x)$ , then we have

$$(20) \quad V_j(s) = u_j(b_{d_s}) + \beta \sum_{x \in S} \hat{q}_{sx} \sum_{k \in G} \mu_{jk} V_k(x).$$

We now prove that this transition matrix  $\hat{Q}$  defines a feasible monotone transition correspondence. It is feasible by construction, since  $\hat{q}_{sx} > 0$  only if  $x \in \Psi_s$ , which is only possible if  $x \in F_s$ . It is monotone, because we proved above that for any choice of  $\{\lambda_s\}_{s \in S}$ ,  $x < y$  implies that  $\lambda_x$  is (weakly) first-order stochastically

dominated by  $\lambda_y$ , which means this is also true for  $\hat{q}_x$  and  $\hat{q}_y$ . This proves that both properties of  $\hat{Q}$  are satisfied.

We now construct an equilibrium  $\sigma$  that has transition matrix  $Q(\sigma)$  equal to  $\hat{Q}$ . Consider the game  $\Gamma_{s,p}$  that takes place in a period where the current state is  $s_t = s$  and the default policy is  $p_{t-1} = p$ . Define utilities of player  $i$  who is currently in group  $j \in G$  by

$$\begin{aligned} U_j(p_t, s_{t+1}) &= u_j(p_t) + \beta W_j(s_{t+1}) \\ &= u_j(p_t) + \beta \sum_{k \in G} \mu_{jk} V_k(s_{t+1}), \end{aligned}$$

where  $\{V_j(x)\}_{j \in G}^{x \in S}$  are defined as the unique solution to equation (20). We construct strategies of the players as follows. Denote the stage where a representative from group  $d_s$  makes a proposal by  $J$ , so  $\pi_s(J) = d_s$ .

In what follows, we proceed by backward induction, and in every stage we define strategies that are identical in isomorphic subgames (thus ensuring that the strategy profile is Markovian) and that are identical for different players that currently belong to the same group (thus ensuring symmetry). Following the logic of backward induction, we start with the political decision. In stages  $l > J$ , we allow proposers and voters to choose any pure strategy consistent with backward induction, with the only restriction being the following: if in stage  $l$ , the current status quo  $s_{t+1}^{l-1} \in \Lambda_s$ , then a weighted majority votes against the new proposal  $\tilde{s}_{t+1}^l$ . Specifically, if in the subgame that follows acceptance of alternative  $\tilde{s}_{t+1}^l$ , the ultimate decision is  $s_{t+1} = \tilde{s}$ , then individuals from all groups  $j \leq d_s$  vote  $N$  in case  $\tilde{s} \geq s_{t+1}^{l-1}$ , and individuals from all groups  $j \geq d_s$  vote  $N$  in case  $\tilde{s} < s_{t+1}^{l-1}$ ; these voting strategies ensure that any proposal made in such situation is rejected. In stage  $l = J$ , the representative from  $d_s$  chosen to make a proposal randomizes over proposals in  $\Psi_s$  and proposes  $x \in \Psi_s$  with probability  $\hat{q}_{sx} = \lambda_s(x)$  (and makes any other proposal with probability 0), and any proposal  $\tilde{s}_{t+1}^l \in \Psi_s$  is then accepted by voters. Specifically, if rejecting the current proposal would ultimately lead to decision  $\tilde{s}$ , then individuals from all groups  $j \leq d_s$  vote  $Y$  in case  $\tilde{s}_{t+1}^l \leq \tilde{s}$ , and individuals from all groups  $j \geq d_s$  vote  $Y$  in case  $\tilde{s}_{t+1}^l > \tilde{s}$ . If some proposal  $\tilde{s}_{t+1}^l \notin \Psi_s$  is made at this stage, then individuals make any voting choices consistent with backward induction. Finally, in

stages  $l < J$ , individuals make any proposals and any votes consistent with backward induction. It is easy to see that strategies constructed in this way form an SPE in the subgame where the political decision is made, and because they only depend on payoff-relevant histories, they are Markovian. Indeed, if these strategies are followed, then transition to state  $x \in S$  happens with probability  $\hat{q}_{sx}$ , and the decisions made in stages  $l < J$  are irrelevant for the outcome; then at stage  $J$ , proposals in  $\Psi_s$  are made with these respective probabilities and are accepted. Finally, in each of the subsequent stages, no alternative from  $\Psi_s$  is ever voted down, even by another alternative from  $\Psi_s$ .

To define strategies in the stage where the policy decision is made, we again solve the game by backward induction. For  $l > J$ , we choose any pure strategies (again, identical in isomorphic subgames and symmetric across players in the same group). This ensures that if the current status quo is  $p_t^{l-1} = b_{d_s}$ , then any alternative  $\tilde{p}_t^l \neq b_{d_s}$  will not be accepted. For  $l = J$ , we require that the representative from  $d_s$  chooses  $b_{d_s}$ , which is subsequently accepted; if another proposal is chosen, then any pure strategies consistent with backward induction are allowed. Finally, for  $l < J$ , we allow any proposals and votes to be made. We thus get a symmetric MPE in the within-period game, where policy  $p_t = b_{d_s}$  is chosen with probability 1, and a transition to alternative  $x$  takes place with probability  $\hat{q}_{sx}$ .

Denote the resulting profile of strategies  $\sigma_s$  (by construction, it does not depend on  $t$  explicitly, as we were choosing Markovian strategies). Taking these profiles for all values of  $s$ , we get strategy profile  $\sigma$ , which prescribes strategies for all players in the original game  $\Gamma$ . By construction, the corresponding transition mapping is  $Q(\sigma) = \hat{Q}$ , and if profile  $\sigma$  is played, continuation utilities of each player in each subgame are equal to the corresponding continuation utility in the corresponding game  $\Gamma_{s_t, p_{t-1}}$ . Furthermore,  $\sigma$  is an SPE: by the one-shot deviation principle, if there is a deviation, there must be a deviation in some period  $t$  where the current state is  $s_t = s$ ; but this contradicts that  $\sigma_s$  is an SPE in the game  $\Gamma_{s_t, p_{t-1}}$ . Thus,  $\sigma$  is an MPE in  $\Gamma$ . Since in the construction of  $\sigma_s$ , the strategies were defined identically for different players in the same group, the MPE is symmetric, and because  $\hat{Q}$  is feasible and monotone, these properties are also retained by  $\sigma$ . Thus,  $\sigma$  is an equilibrium with the desired properties. This completes the proof of existence of a symmetric monotone MPE for any combination

of feasible transitions  $\{F_s\}_{s \in S}$  that we allow, and in particular for  $F_s = S$  for all  $s \in S$ , as stated in the theorem.

We next prove the remaining claims in the theorem.

*Proof of Part i.* Take a symmetric MPE,  $\sigma$ . Consider period  $t$  where the current state is  $s_t = s$ , and the previous period's policy is  $p_{t-1} = p$ . Notice that the society's decision on  $p_t$  does not affect equilibrium actions when choosing the transition, nor does it affect any actions in subsequent periods, because strategies in  $\sigma$  are Markovian. Thus, without loss of any generality, we can suppress the policy decision and endow each group  $j$  with payoff  $u_j(p_t)$  at time  $t$ .

As before, let  $J$  denote the stage where group  $d_s$  makes a proposal. Let us prove the following statement by backward induction: if at some stage  $l \geq J$  the decision made (status quo for the next stage)  $p_t^l = b_{d_s}$ , then the ultimate policy decision  $p_t = b_{d_s}$ . The base is trivial: in the last stage, where  $l = g$ , the new status quo automatically becomes the policy decision, so  $p_t = p_t^l = b_{d_s}$ . Step: take  $l < g$ , and suppose this statement is true for  $k > l$ ; let us prove it for stage  $l$ . Suppose that  $p_t^l = b_{d_s}$  and consider stage  $l + 1$ . Suppose, to obtain a contradiction, that  $p_t \neq b_{d_s}$  with a positive probability. By induction, this is only possible if  $p_t^{l+1} \neq b_{d_s}$  with positive probability. For this to be true, it must be that at stage  $l + 1$ , the representative of group  $\pi_s(l + 1)$  with positive probability makes proposal  $x \neq b_{d_s}$ , which is subsequently accepted, and after that  $p_t \neq b_{d_s}$  with a positive probability. Let  $H$  be the distribution of  $p_t$  conditional on  $x$  becoming the new status quo  $p_t^{l+1}$  after stage  $l + 1$ ; notice that if  $p_t^{l+1} = b_{d_s}$ , then  $p_t = b_{d_s}$  by induction. Now, if  $\mathbb{E}H < b_{d_s}$  then all individuals in groups  $j \geq d_s$  prefer  $b_{d_s}$  to  $H$  (because of quadratic utility); similarly, if  $\mathbb{E}H > b_{d_s}$  then all individuals in groups  $j \leq d_s$  prefer  $b_{d_s}$  to  $H$ . Last, if  $\mathbb{E}H = b_{d_s}$  then all individuals in all groups  $j \geq d_s$  prefer  $b_{d_s}$  to  $H$ , because, by assumption, under  $H$ ,  $p_t \neq b_{d_s}$  with a positive probability, which implies that  $H$  has positive variance, which makes the expectation  $b_{d_s}$  preferable to  $H$  for all agents. In all cases, a weighted majority strictly prefers  $b_{d_s}$  to  $H$ , and hence in a sequential voting  $x$ , leading to  $H$ , cannot be the outcome. This contradiction proves the induction step.

Let us now prove that in the subgame starting with stage  $J$ ,  $p_t = b_{d_s}$ . To show this, it suffices to prove that  $p_t^J = b_{d_s}$  with probability 1. Notice that if  $\tilde{p}_t^J = b_{d_s}$  is proposed, then  $p_t = b_{d_s}$ ; indeed, if this were not the case, then a weighted majority would prefer to have  $p_t = b_{d_s}$  to any distribution  $H'$  of  $p_t$  conditional on

the proposal being rejected (the argument is similar to the one in the previous paragraph), and thus in the sequential voting, agents will ensure that the new status quo is  $p_t^J = b_{d_s}$ . Now suppose that  $p_t \neq b_{d_s}$  with a positive probability; this is only possible if group  $d_s$  proposes  $\tilde{p}_t^J \neq b_{d_s}$  with a positive probability. However, in this case it has a profitable deviation, which is proposing  $b_{d_s}$  and thus  $p_t = b_{d_s}$ . This contradiction proves that in the subgame starting with stage  $J$ ,  $p_t = b_{d_s}$ .

The last result holds regardless of the play in stages  $l < J$ . Consequently, in equilibrium  $\sigma$ ,  $p_t = b_{d_s}$  with probability 1, which completes the proof. ■

*Proof of Part ii.* Take an equilibrium  $\sigma$ , and consider period  $t$  where the current state is  $s_t = s$ . Notice that by the time the political decision is made, the policy is already decided (and in equilibrium, it is  $p_t = b_{d_s}$ ) and the continuation utility of a player from group  $j$  is given by  $W_j(s_{t+1})$ . In what follows, let  $\bar{W} = \max_{x \in S} W_{d_s}(x)$ ; it equals  $W_{d_s}(y)$  for  $y \in \Psi_s$ .

Let us first prove that in any equilibrium, the vector of continuation utilities  $V = \{V_j(s)\}_{j \in G}^{s \in S}$  satisfies increasing differences. Indeed, if  $Q$  is the transition correspondence in equilibrium  $\sigma$ , then  $V$  is the unique solution to equation (20), and it may be obtained through infinite iteration of mapping equation (17), because, given  $\beta < 1$ , this mapping is a contraction on  $\mathbb{R}^{gm}$  in the  $L_1$ -metric. Since for any  $V$  that satisfies increasing differences,  $V'$  also does by Lemma A2, the limit point  $V$  must satisfy increasing differences.

By Lemma A1, the vector  $W = \{W_j(s)\}_{j \in G}^{s \in S}$  also satisfies increasing differences. As before, let  $\Psi_s = \arg \max_{x \in S} W_{d_s}(x)$  (the maximum is taken over  $S$  because all transitions are feasible). Also, as before, let  $J$  be the stage where group  $d_s$  makes the proposal.

Suppose first that  $J = g$ , so group  $d_s$  is the last to propose. In that case,  $d_s$  can ensure that it gets the payoff  $\bar{W}$ . Indeed, if the current status quo is  $s_{t+1}^{J-1} \in \Psi_s$  then it can propose the same alternative  $\tilde{s}_{t+1}^J = s_{t+1}^{J-1}$ , in which case it will be implemented regardless of how people vote. On the other hand, if  $s_{t+1}^{J-1} \notin \Psi_s$ , then it can propose  $\tilde{s}_{t+1}^J \in \Psi_s$ ; then in the voting subgame, this alternative  $\tilde{s}_{t+1}^J$  must be accepted, because a weighted majority (all groups  $j \leq d_s$  if  $\tilde{s}_{t+1}^J < s_{t+1}^{J-1}$  and all groups  $j \geq d_s$  if  $\tilde{s}_{t+1}^J > s_{t+1}^{J-1}$ ) prefer it to  $s_{t+1}^{J-1}$ . Since group  $d_s$  can ensure its maximum payoff  $\bar{W}$  in this

subgame, it will do so; consequently, alternatives  $x \notin \Psi_s$  cannot be implemented as  $s_{t+1}$ .

Consider the other case, where  $J < g$ . Let  $\pi_s(g) = j$ ; suppose, without loss of generality, that  $j < d_s$  (the case  $j > d_s$  is considered similarly). In this case, we first prove the following: in any subgame that includes stage  $g$  (where  $j$  makes a proposal), the ultimate political decision  $s_{t+1}$  satisfies  $s_{t+1} \in \Xi_s$ , where  $\Xi_s = \Psi_s \cup \{x \in S: x < \min \Psi_s\}$ . Indeed, consider possible values of the current status quo  $s_{t+1}^{g-1}$ . If  $s_{t+1}^{g-1} \in \Psi_s$ , then no proposal  $\tilde{s}_{t+1}^g$  made by group  $j$  may be accepted in equilibrium, unless  $\tilde{s}_{t+1}^g \in \Psi_s$  as well. Therefore, in this case the statement is correct. If  $s_{t+1}^{g-1} \notin \Psi_s$ , consider two possibilities. Suppose that  $s_{t+1}^{g-1} > \min \Psi_s$ . Then if group  $j$  instead proposes  $\tilde{s}_{t+1}^g = \min \Psi_s$ , with a similar argument to above, it will be accepted. Moreover, since  $W_{d_s}(\min \Psi_s) \geq W_{d_s}(y)$  for any  $y \in S$ , including  $y > \min \Psi_s$ , then  $j < d_s$  implies  $W_j(\min \Psi_s) > W_j(y)$  for such  $y$ . Consequently, if  $s_{t+1}^{g-1} > \min \Psi_s$ , then group  $j$  prefers to propose  $\min \Psi_s$  as compared to proposing any alternative  $y > \min \Psi_s$ . If it proposes an alternative  $y < \min \Psi_s$  that is subsequently rejected, then  $s_{t+1}^g = s_{t+1}^{g-1}$  and again group  $j$  is better off proposing  $\min \Psi_s$ . Thus, the only alternative action that group  $j$  may (weakly) prefer to proposing  $\min \Psi_s$  is proposing  $y < \min \Psi_s$  that is subsequently accepted. Consequently, if  $s_{t+1}^{g-1} > \min \Psi_s$ , then in equilibrium either group  $j$  proposes  $\min \Psi_s$ , which is accepted, or some  $y < \min \Psi_s$  that is accepted; in either case  $s_{t+1}^g \in \Xi_s$ . Finally, consider the possibility  $s_{t+1}^{g-1} < \min \Psi_s$ . The statement may only fail if group  $j$  proposes, with a positive probability, some alternative  $y > \min \Psi_s$ ,  $y \notin \Psi_s$ , which is subsequently accepted. In that case, however, group  $j$  has a profitable deviaton: it would do better by proposing  $\min \Psi_s$ , since this proposal will be accepted, and  $W_{d_s}(\min \Psi_s) > W_{d_s}(y)$  implies, since  $j < d_s$ , that  $W_j(\min \Psi_s) > W_j(y)$  as well. This is impossible in equilibrium, which proves that in all cases,  $s_{t+1}^g \in \Xi_s$ .

Since  $p_t^g \in \Xi_s$  in all subgames, we can prove the following statement by backward induction: if at some stage  $l$ ,  $0 \leq l \leq g$ ,  $s_{t+1}^l = \max \Psi_s$  (which also equals  $\max \Xi_s$ ), then  $s_{t+1}^g \in \Psi_s$ . The base ( $l = g$ ) is trivial. To establish the inductive step, suppose that this is true for stage  $l$ , and consider stage  $l - 1$ . We have that the current status quo is  $s_{t+1}^{l-1} = \max \Psi_s$ . Suppose, to obtain a contradiction, that  $s_{t+1}^g \notin \Psi_s$  with a positive probability. By induction, this is only possible if  $s_{t+1}^l \neq \max \Psi_s$ , which, in turn, is only

possible if proposal  $\tilde{s}_{t+1}^l \neq \Psi_s$  is made and is accepted. However, we showed that a subsequent subgame will result in  $s_{t+1}^g$  having some distribution with support in  $\Xi_s$  and, moreover, with some  $y \notin \Psi_s$  having a positive probability. Notice, however, that all  $y \in \Xi_s$  satisfy  $W_k(y) \leq W_k(\max \Psi_s)$  for all  $k \geq d_s$ , and the inequality is strict if  $y \in \Xi_s \setminus \Psi_s$  for all such  $k$ . Thus, in this case a weighted majority strictly prefers to reject proposal  $y$ , which is a contradiction proving the induction step.

To complete the proof, notice that if group  $d_s$  proposes in stage  $J$ , then it can always guarantee utility  $\bar{W}$ : if preserving current status quo  $s_{t+1}^{J-1}$  (by proposing  $\tilde{s}_{t+1}^J = s_{t+1}^{J-1}$ , so  $s_{t+1}^J = s_{t+1}^{J-1}$ ) results in  $s_{t+1} \notin \Psi_s$  with a positive probability, group  $d_s$  can propose  $\max \Psi_s$ , which will be accepted, since all groups  $k \geq d_s$  strictly prefer  $s_{t+1}^J = \max \Psi_s$  to  $s_{t+1}^J = s_{t+1}^{J-1}$ . Consequently,  $s_{t+1} \in \Psi_s$  with probability 1, which completes the proof. ■

*Proof of Part iii.* Take equilibrium  $\sigma$ , and take  $x, y \in S$  such that  $x < y$ . Suppose that  $q_{x,a} > 0$  and  $q_{y,b} > 0$ ; by part ii, this implies  $a \in \Psi_x = \arg \max_{z \in S} W_{d_x}(z)$  and  $b \in \Psi_y = \arg \max_{z \in S} W_{d_y}(z)$ . We proved already that in equilibrium,  $\{W_j(x)\}_{j \in G}^{x \in S}$  satisfy increasing differences. Since  $x < y, d_x < d_y$ , and by Lemma A3 (where we set  $X = Y = S$ ), we have  $a \leq b$ . ■

*Proof of Part iv.* To prove this part of the theorem, we show that for every equilibrium in which there is a possible transition to more than two states (i.e.,  $s \in S, |\Phi_s \setminus \{s\}| \geq 2$ ), the model parameters  $(\{b_k\}_{k \in G}, \{A_k\}_{k \in G}, \{\gamma_k\}_{k \in G}, \{\mu_{jk}\}_{j,k \in G}, \beta)$  satisfy a nontrivial polynomial equation with rational coefficients (we achieve this by showing that if this were not the case, some equilibrium condition would be violated). Then because the set of nontrivial polynomials with rational coefficients is countable, the set of such parameters has measure 0. (In fact, we establish a stronger result, that the parameters must satisfy one of a finite subset of such polynomials). In what follows, let  $\mathbb{F}$  denote the smallest field that contains  $\mathbb{Q}$  and all the above-mentioned parameters (e.g., Hungerford 1974). Furthermore, let  $\bar{\mathbb{F}}$  be the set of all solutions to polynomial equations with coefficients in  $\mathbb{F}$ . Then standard arguments show that  $\bar{\mathbb{F}}$  is an algebraically closed field.

Suppose, to obtain a contradiction, that for some parameter values that do not satisfy a nontrivial polynomial equation with coefficients in  $\mathbb{Q}$ , the statement is nevertheless wrong. Without loss of generality, suppose that the set of states  $S$  contains the fewest elements among any such examples. Then the groups

$\{d_s\}_{s \in S}$  belong to the same irreducible component of matrix  $M$  (otherwise there are at least two groups of states without transition between them, and we can remove one such group), and we can without loss of generality assume that there is no other component (preferences of individuals in the other groups, if they exist, are irrelevant).

Let  $Z$  be the (nonempty) set of states  $s$  such that  $|\Phi_s \setminus \{s\}| \geq 2$ . Consider first the case where there is  $s$  such that  $\Phi_s \cap [s + 1, m] \geq 2$ . Then  $s = 1$  (otherwise all states to the left of  $x$  could be removed, thus violating the assumption that the number of states in  $S$  is minimal). Furthermore, for every state  $x < m$ , there is  $y > x$  with  $y \in \Phi_x$  (otherwise, monotonicity implies that for all  $z \leq x$ , transitions to states greater than  $x$  are impossible, and then those states may be removed). If so, for all  $x \in S$ ,  $\Phi_x \geq x$  (otherwise, if we take the smallest  $x$  for which this is violated, we would get a contradiction with monotonicity). Consequently, for all  $x \in (1, m)$ , there is a unique  $y \in \Phi_x$  such that  $y > x$  (for  $x = 1$  there are two such  $y$ , and for  $x = m$  there is none). Now let  $A \subset S$  be defined by  $A = \{x \in S: |\Phi_x| \geq 2\}$ . Now for each  $x \in A$ , let  $\rho_x = \max \Phi_x$  and let  $\lambda_x = \max(\Phi_x \setminus \{\rho_x\})$ ; notice that for  $x > 1$ ,  $\lambda_x = x$ , and for  $x = 1$ ,  $\lambda_x > 1$ . In what follows, for  $x \in A$ , let  $\alpha_x = q_{x\lambda_x}$ , then for  $x \in A \setminus \{1\}$ ,  $q_{x\lambda_x} = 1 - \alpha_x$ .

Let us prove by backward induction over the set of elements in  $A$  that the following is true for every element  $x$  in  $A$ . First, the equilibrium utility of group  $d_x$  in state  $x$ ,  $W_{d_x}(\Lambda_x)$ , is not equal to its utility if transitions correspondence was  $\tilde{Q}$  such that  $\tilde{q}_y = q_y$  for  $y \neq x$  and  $\tilde{q}_{x\lambda_x} = 1$ , while  $\tilde{q}_{xy} = 0$  for  $y \neq x$ . Second, if  $x \neq 1$ , then the transition probability to  $\Lambda_x$ ,  $\alpha_x$ , satisfies a nontrivial polynomial equation with coefficients being polynomials in the parameters of the model. Third, if  $x \neq 1$ , then for any group  $j \in G$  and any state  $y < \Lambda_x$ , let  $H_{jx}(b_{d_y}) = \frac{1}{b_{d_y}} \left( W_j(\Lambda_x) - \sum_{\tau=1}^{\infty} \beta^{\tau-1} \mu_{jd_y}^{(\tau)} b_{d_y}^2 \right)$  be a function of  $b_{d_y}$  for all other parameters of the model fixed; then it is a well-defined real analytic function in the neighborhood of the true parameter  $b_{d_y}$ , and any analytic continuation of this function is bounded at  $\infty$  (more precisely, there is  $C, K > 0$  such that  $|b_{d_y}| > K$  implies  $|H_{jx}(b_{d_y})| < C$ ). Indeed, if we prove this for all  $x$ , then the first property applied to  $x = 1$  would imply that the equilibrium utility of group  $d_1$  is not equal to its utility when the society immediately transits to  $\lambda_1 > 1$ , which would be a contradiction.

Base: If  $x = \max A$ , then equating  $W_{d_x}(\Lambda_x)$  to  $W_{d_x}(x, x, x, \dots)$  results in a nontrivial polynomial equation (it is nontrivial,

because  $W_{d_x}(\Lambda_x) - \sum_{\tau=1}^{\infty} \beta^{\tau-1} \mu_{d_x d_x}^{(\tau)} b_x^2$  is linear in  $b_x$ , as the society never gets to state  $x$  on a path starting from  $\Lambda_x$ , and the only terms that are quadratic in  $b_x$  come from individuals from group  $d_x$  being in this group in the future, whereas  $W_{d_x}(x, x, x, \dots) - \sum_{\tau=1}^{\infty} \beta^{\tau-1} \mu_{d_x d_x}^{(\tau)} b_x^2$  has quadratic terms in  $b_x$ , and the coefficient is nonzero because it is polynomial in other parameters, and it cannot be equal to 0 if the parameters are generic). Now, continuation values  $\{V_j(s)\}_{j \in G}^{s \geq x}$  solve equation (20) with  $q_{..}$  as linear functions of  $\alpha_x$ , which implies that  $\{V_j(s)\}_{j \in G}^{s \geq x}$  can be expressed as ratios of polynomials of  $\alpha_x$ ; then equating  $W_{d_x}(\Lambda_x)$  to  $W_{d_x}(x)$  results in a nontrivial polynomial of  $\alpha_x$  (it is nontrivial, because it holds for some  $\alpha_x$  as there is such an equilibrium, but not for some other, say  $\alpha_x = 0$ , as in that case  $W_{d_x}(x) = W_{d_x}(x, x, x, \dots) \neq W_{d_x}(\Lambda_x)$ , as we just proved). Finally, the function  $H_{jx}(b_{d_y})$  does not depend on  $\alpha_x$ , and the result follows immediately by evaluating  $W_j(\Lambda_x)$ .

Step: suppose that the result is proven for  $z \in A$  such that  $z > x$ , let us prove it for  $x$ . Notice that as before, equating  $W_{d_x}(\Lambda_x)$  to  $W_{d_x}(x, x, x, \dots)$  would give rise to a polynomial equation in all parameters of the model and  $\{\alpha_y\}_{y \in A, y > x}$ . As before, this equation is nontrivial, because  $H_{d_x x}(b_{d_x})$  is bounded for  $b_{d_x}$  high enough by induction (since  $x < \Lambda_x$ ), while  $W_{d_x}(x, x, x, \dots) - \sum_{\tau=1}^{\infty} \beta^{\tau-1} \mu_{d_x d_x}^{(\tau)} b_x^2$  has quadratic terms and thus is unbounded, even after dividing by  $b_{d_x}$ . Now suppose  $x > 1$ ; as before, we get that equating  $W_{d_x}(\Lambda_x)$  to  $W_{d_x}(x)$  gives rise to a polynomial equation in  $\alpha_x$  with coefficients in all parameters of the model and also  $\{\alpha_y\}_{y \in A, y > x}$  (which is nontrivial for the same reasons as before). Since  $\bar{\mathbb{F}}$  is algebraically closed,  $\alpha_x$  must satisfy a polynomial equation with coefficients in  $\mathbb{F}$ . Moreover, since  $\mathbb{F}$  consists of ratios of polynomials of the parameters of the model with coefficients in  $\mathbb{Q}$ , we can multiply by the common denominator to prove the second part of the statement.

Finally, we need to prove that if  $x > 1$ , then for any group  $j \in G$  and any state  $y < \Lambda_x$ ,  $H_{jx}(b_{d_y})$  is bounded for  $|b_{d_y}|$  large enough. Notice that  $H_{jx}(b_{d_y})$  depends on  $b_{d_y}$  explicitly (and it is a linear function), and also through  $\{\alpha_y\}_{y \in A, y > x}$ , which can appear in both the numerator and the denominator. It now suffices to prove that the denominator does not tend to 0 as  $b_{d_y}$  tends to  $\infty$ . Since each  $\alpha_y$  satisfies a polynomial equation with coefficients that are polynomials in the parameters of the model, either  $\alpha_y$  does not depend on  $b_{d_y}$  explicitly, or there is only a finite number of limit

points (including  $\infty$ ) of the solutions to this equation as  $b_{d_s}$  tends to  $\infty$ . If for at least one of these limit points, the denominator tends to 0, this yields a polynomial equation on  $\alpha_z$  for  $z$  being the smallest element in  $A$  greater than  $x$ . This means that there are two polynomial equations on  $\alpha_z$  that have a common root, which is only possible if their resultant equals 0, which again gives a polynomial condition on the parameters of the model. Since by assumption such a condition cannot be satisfied, we have proved the induction step.

This backward induction leads to a contradiction, as it means that the society may not be indifferent between transitioning from state 1 to  $\lambda_1$  and  $\Lambda_1$ . This proves that there is no state  $s \in Z$  such that  $\Phi_s \cap [s + 1, m] \geq 2$ . We can similarly prove that there is no state  $s \in Z$  such that  $\Phi_s \cap [1, s - 1] \geq 2$ . Consequently, if  $Z$  is nonempty, there must exist  $s \in Z$  and  $x, y \in \Phi_s$  such that  $x < s < y$ . In this case, we can follow a very similar logic and arrive at a similar contradiction. This implies that  $Z$  is empty, which completes the proof. ■

The following lemmas will be used in several proofs. (Proofs of these and all subsequent lemmas are relegated to [Online Appendix B](#).)

**LEMMA A4.** Suppose that  $\sigma$  is an equilibrium with transition correspondence  $Q$  in a game where the set of states is  $S$  and set of feasible transitions is  $F$ . Suppose that  $S' \subset S$  is such that for any  $x \in S'$  and  $y \in S \setminus S'$ ,  $q_{xy} = 0$  (i.e.,  $S'$  is such that  $Q$  does not include transitions out of it, which is true, for example, if  $S' = S$ ), and suppose that the set of feasible transitions  $F'$  on  $S'$  is such that for  $x, y \in S'$ ,  $q_{xy} > 0$  implies  $y \in F'_x$ , and  $y \in F'_x$  implies  $y \in F_x$  (i.e.,  $F'$  is more restrictive than  $F$ , but is nevertheless consistent with  $Q$ ). Then there is an equilibrium  $\sigma'$  in a game where the set of states is  $S'$  and the set of feasible transitions is  $F'$  (and other parameters are the same) such that its transition correspondence  $Q'$  satisfies  $q'_{xy} = q_{xy}$  for any  $x, y \in S'$ .

**LEMMA A5.** Let  $Q = \{q_{x,y}\}_{x,y \in S}$  be a monotone transition correspondence, and suppose that for any  $a \in S$  and  $b \in \Phi_a$ ,  $W_{d_a}(b) = \tilde{W}_{d_a}$ , which does not depend on  $b$ . Suppose that for some  $x', y' \in S$ , we have  $W_{d_{x'}}(y') > \tilde{W}_{d_{x'}}$ . Then there also exist  $x, y \in S$  such that  $W_{d_x}(y) > \tilde{W}_{d_x}$  and, in addition, the

correspondence  $Q' : S \rightarrow S$  given by

$$(21) \quad q'_{sa} = \begin{cases} q_{sa} & \text{if } s \neq x, \\ 1 & \text{if } s = x \text{ and } a = y, \\ 0 & \text{if } s = x \text{ and } a \neq y \end{cases}$$

is monotone.

LEMMA A6. Suppose that for some  $j$ , the sequence  $b_j^{(t)}$  is nondecreasing (respectively, nonincreasing). Then if in state  $s \in S$ ,  $j = d_s$ , then  $\Phi_s \geq s$  (respectively,  $\Phi_s \leq s$ ).

LEMMA A7. Suppose that for some  $j$ , the sequence  $b_j^{(t)}$  is nondecreasing (respectively, nonincreasing). Furthermore, suppose that some state  $s \in S$  satisfies  $j = d_s$ , and  $\arg \min_{z \in S} |b_j^{(\infty)} - b_{d_z}| = \{s\}$ . Then  $\Phi_s = \{s\}$ .

LEMMA A8. Suppose that for some  $j$ , the sequence  $b_j^{(t)}$  is nondecreasing and, moreover, there is some  $\tau \geq 1$  such that  $b_j^{(\tau)} \neq b_j^{(\tau+1)}$ . Fix a state  $s$  where  $j = d_s$  and consider any monotone set of mappings  $Q = \{q_{xy}\}$  for  $x \neq s$ . Suppose that for some  $x > s$ ,  $\Phi(x) \geq x$ . For any  $\alpha$ , denote the continuation utility of individuals from current group  $j$  from moving to state  $x$  by  $W_j^{(\tau)}(x)$ , and from staying in  $s$  and moving to  $x$  with probability  $\alpha$  in each period thereafter by  $W_j^{(\tau)}(s; \alpha)$ . Let

$$f(\alpha) = W_j(s; \alpha) - W_j(x).$$

Then  $f$  satisfies the following strict single-crossing property: if for some  $\alpha$ ,  $f(\alpha) = 0$ , then  $f(\alpha') > 0$  for  $\alpha' > \alpha$  and  $f(\alpha') < 0$  for  $\alpha' < \alpha$ .

*Proof of Theorem 2.* Uniqueness when  $\beta$  is sufficiently small follows from the following argument. Let us show that in equilibrium, from state  $s$  the society must transit to state  $z$  that minimizes  $|b_{d_z} - b_{d_s}^{(1)}|$ . Let  $\beta_0$  be defined by  $\beta_0 = \frac{\zeta}{\zeta + 2U}$ , where

$$\zeta = \min_{s, y, z \in S, |b_{d_y} - b_{d_s}^{(1)}| > |b_{d_z} - b_{d_s}^{(1)}|} \left( (b_{d_y} - b_{d_s}^{(1)})^2 - (b_{d_z} - b_{d_s}^{(1)})^2 \right).$$

Suppose, to obtain a contradiction, that for some  $s \in S$ , a transition to a state  $z$  that does not minimize  $|b_{d_z} - b_{d_s}^{(1)}|$  occurs. This means that for some  $y \in S$ ,  $|b_{d_y} - b_{d_s}^{(1)}| < |b_{d_z} - b_{d_s}^{(1)}|$ . Now consider the utility of individuals from group  $d_s$  if they transitioned to  $y$  instead. Their gain in utility (after factor  $\beta$ ) would be

$$\begin{aligned} W_{d_s}(y) - W_{d_s}(z) &= \sum_{k \in G} \mu_{d_s k} (V_k(y) - V_k(z)) \\ &= \sum_{k \in G} \mu_{d_s k} (A_k - (b_k - b_{d_y})^2 - A_k + (b_k - b_{d_z})^2) + \beta(\dots) \\ &\geq \sum_{k \in G} \mu_{d_s k} \left( (b_k - b_{d_z})^2 - (b_k - b_{d_y})^2 \right) + \frac{\beta}{1 - \beta} 2\bar{U} \\ &= (b_{d_y} - b_{d_z}) \sum_{k \in G} \mu_{d_s k} (2b_k - b_{d_y} - b_{d_z}) + \frac{\beta}{1 - \beta} 2\bar{U} \\ &= (b_{d_y} - b_{d_z}) (2b_{d_s}^{(1)} - b_{d_y} - b_{d_z}) + \frac{\beta}{1 - \beta} 2\bar{U} \\ &= (b_{d_s}^{(1)} - b_{d_z})^2 - (b_{d_s}^{(1)} - b_{d_y})^2 + \frac{\beta}{1 - \beta} 2\bar{U} > 0, \end{aligned}$$

provided that  $\beta \in (0, \beta_0)$ . Therefore, a transition to  $z$  does not maximize the continuation utility of the pivotal group  $d_s$  (they would be better off moving to  $y$ ), which contradicts part ii of Theorem 1. This shows that for any  $s$ , transition to state  $z$  that minimizes  $|b_{d_z} - b_{d_s}^{(1)}|$  must occur. Because this state is generically unique (in the space of social mobility matrixes), the equilibrium is generically unique for  $\beta < \beta_0$ .

We now turn to generic uniqueness under Assumption 2, which will be proved in several steps.

Step 1. Suppose that there are two equilibria  $\sigma_1$  and  $\sigma_2$ , and let  $Q^1$  and  $Q^2$  be the corresponding transition matrixes. Then, for generic parameter values, if  $Q^1 \neq Q^2$ , then there are at least two states  $x, y \in S, x \neq y$ , such that the distributions  $q_x^1 \neq q_x^2$  and  $q_y^1 \neq q_y^2$ . In other words, it is impossible that transition probabilities from only one state are different.

*Proof.* Suppose not, so there is a unique state  $s$  such that  $q_s^1 \neq q_s^2$ . Let us first prove that generically for set  $\Omega = (\Phi_s^1 \cup \Phi_s^2) \setminus \{s\}$ ,  $|\Omega| = 1$ . Indeed, if  $\Omega$  is empty,  $\Phi_s^1 = \Phi_s^2 = \{s\}$ , hence  $q_s^1 = q_s^2$ , which

contradicts the choice of  $s$ . On the other hand, suppose that there are  $x, y \in \Omega$  such that  $x \neq y$ ; without loss of generality,  $x < y$ . Without loss of generality, suppose  $x \in \Phi_s^1$ . Then by part iv of Theorem 1, for generic parameter values,  $y \notin \Phi_s^1$ , which means that  $y \in \Phi_s^2$ , which, again by part iv of Theorem 1, implies  $x \notin \Phi_s^2$  for generic parameter values. Now consider three possibilities. If  $x < s < y$ , then, from part iii of Theorem 1,  $x \in \Phi_s^1$  implies  $\Phi_z^1 \leq x$  for  $z < s$ ; moreover, for such  $z$ ,  $q_z^2 = q_z^1$ . Therefore, if society moves from state  $s$  to  $x$ , the continuation utilities of group  $d_s$  should be the same for both equilibria:  $W_{d_s}^1(x) = W_{d_s}^2(x)$ . Similarly,  $y \in \Phi_s^2$  implies  $\Phi_z^2 \geq y$  for all  $z > s$ ; moreover, for such  $z$ ,  $q_z^1 = q_z^2$ . Thus, if the society moves from state  $s$  to  $y$ , the continuation utilities again coincide:  $W_{d_s}^1(y) = W_{d_s}^2(y)$ . But by part ii of Theorem 1, we have  $W_{d_s}^1(x) \geq W_{d_s}^1(y) = W_{d_s}^2(y) \geq W_{d_s}^2(x) = W_{d_s}^1(x)$ , which implies that both inequalities hold with equality, in particular,  $W_{d_s}^1(x) = W_{d_s}^1(y)$ . This means  $x, y \in \Psi_s^1$ , which, as shown in the proof of part iv of Theorem 1, is impossible for generic parameter values. The remaining possibilities are  $x < y < s$  and  $s < x < y$ ; they are considered similarly.

We have therefore proved that there is a unique  $x \in \Omega$ . Suppose that  $x > s$  (the case of  $x < s$  is entirely analogous). Notice that  $q_{sx}^1 \neq q_{sx}^2$ ; otherwise, since  $\Phi_s^1 \subset \{s, x\}$  and  $\Phi_s^2 \subset \{s, x\}$  we would have  $q_{ss}^1 = q_{ss}^2$ , again meaning that  $q_s^1 = q_s^2$  and contradicting the choice of  $s$ . Without loss of generality, assume  $q_{sx}^1 < q_{sx}^2$ , so in equilibrium  $\sigma_1$  the society stays in  $s$  longer than in equilibrium  $\sigma_2$ , in expectation; this means, in particular,  $q_{sx}^1 < 1$  and  $q_{sx}^2 > 0$ . It must be that the sequence  $b_{d_s}^{(t)}$  is nondecreasing and, moreover, it is nonstationary, for otherwise  $q_{sx}^2 > 0$  would contradict Lemma A6.

Let  $j = d_s$ . The continuation utilities from moving to  $x$  are the same in both equilibria:  $W_j^1(x) = W_j^2(x)$ , because the transition probabilities are identical thereafter. Moreover, in equilibrium  $\sigma_2$ , transiting to  $x$  is a best response, so  $W_j^2(x) \geq W_j^2(s)$ , and in equilibrium  $\sigma_1$ , staying is a best response, so  $W_j^1(s) \geq W_j^1(x)$ . We thus have  $W_j^1(s) \geq W_j^1(x) = W_j^2(x) \geq W_j^2(s)$ , meaning that the utility of individuals from group  $j$  from staying is at least as high under  $\sigma_1$  as under  $\sigma_2$ . Denote  $W_j(s; \alpha)$  the utility of staying in  $s$  if the subsequent equilibrium play has probability  $\alpha$  of moving to  $x$ ; then  $W_j(s; q_{sx}^1) = W_j^1(s)$  and

$W_j(s; q_{sx}^2) = W_j^2(s)$ . By Lemma A8, the function  $f(\alpha) : [0, 1] \rightarrow \mathbb{R}$ , defined by  $f(\alpha) = W_j(s; \alpha) - W_j(x)$ , satisfies the strict single-crossing condition. Now, if  $f(q_{sx}^1) = 0$ , then  $f(q_{sx}^2) > 0$ , meaning that  $W_j(s; q_{sx}^2) > W_j(x)$ , which contradicts that moving to  $x$  is a best response in  $\sigma_2$ . Similarly, if  $f(q_{sx}^2) = 0$ , then  $f(q_{sx}^1) < 0$ , meaning that  $W_j(s; q_{sx}^1) < W_j(x)$ , which contradicts that staying at  $s$  is a best response in  $\sigma_1$ . If  $f(q_{sx}^1) \neq 0$  and  $f(q_{sx}^2) \neq 0$ , then since staying in  $s$  is a best response in  $\sigma_1$ , we must have  $f(q_{sx}^1) > 0$ ; similarly, we must have  $f(q_{sx}^2) < 0$ . But then by continuity there is  $\alpha \in (q_{sx}^1, q_{sx}^2)$  such that  $f(\alpha) = 0$ . In that case, it must be that  $f(q_{sx}^1) < 0 < f(q_{sx}^2)$ . But this would contradict Lemma A8. This contradiction completes the proof of Step 1.

Step 2. Let  $m$  be the minimal number of states for which there are two equilibria,  $\sigma_1$  and  $\sigma_2$ . Then  $m = 2$ .

*Proof.* Suppose not, then either  $m = 1$  or  $m \geq 3$ . If  $m = 1$ , there is only one possible transition mapping:  $Q$  with  $q_{11} = 1$ . Suppose  $m > 3$  and let  $Q^1$  and  $Q^2$  be the transition matrixes in equilibria  $\sigma_1$  and  $\sigma_2$ . Let  $Z \subset S$  be the set of  $z \in S$  such that  $q_z^1$  and  $q_z^2$  are different distributions; from Step 1 it follows that  $|Z| \geq 2$ . In what follows, let  $L = \{s \in S : \Phi_s^1 \leq s, \Phi_s^2 \leq s\}$  and  $R = \{s \in S : \Phi_s \geq s, \Phi_s^2 \geq s\}$ . By Lemma A6,  $L \cup R = S$ ; let us denote  $I = L \cap R$ .

First, we show that if  $s \in S$  and  $1 < s < m$ , then  $s \notin I$ . Indeed, otherwise, we would have  $\Phi_s^1 = \Phi_s^2 = \{s\}$ . Take  $x \in Z \setminus \{s\}$ . If  $x < s$ , then by Lemma A4 there exist two equilibria  $\sigma_1|_{[1, s]}$  and  $\sigma_2|_{[1, s]}$  in the game with the set of states  $S' = S \cap [1, s]$ . Similarly, if  $x > s$ , then there are two equilibria  $\sigma_1|_{[s, m]}$  and  $\sigma_2|_{[s, m]}$  in the game with the set of states  $S' = S \cap [s, m]$ . In either case, we get a contradiction with  $m$  the lowest number of states where multiple equilibria are possible.

Second, let  $x = \min(Z \setminus \{1\})$  and  $y = \max(Z \setminus \{m\})$  (both are well defined because  $|Z| \geq 2$ ). We must have  $x \in L$ . Indeed, suppose not, then  $x \in R$ . If  $x = m$ , then we have  $\Phi_x^1 = \Phi_x^2 = \{x\}$  by definition of  $R$ , and then  $x \notin Z$ , a contradiction. If, on the other hand,  $x \in R$  and  $x < m$ , then, again using Lemma A4, we get that there exist two different equilibria  $\sigma_1|_{[x, m]}$  and  $\sigma_2|_{[x, m]}$  in the game with the set of states  $S' = S \cap [x, m]$ , a contradiction. We can similarly prove that  $y \in R$ .

There are two possibilities. If  $Z \neq \{1, m\}$ , then  $x = \min(Z \setminus \{1\}) = \min(Z \cap [2, m - 1]) \leq \max(Z \cap [2,$

$m - 1$ ) =  $\max(Z \setminus \{m\}) = y$ . This means, again by Lemma A4 that  $\sigma_1|_{[x,y]}$  and  $\sigma_2|_{[x,y]}$  are two different equilibria on  $[x, y]$ , which again contradicts the choice of  $m$ . The remaining case to consider is  $Z = \{1, m\}$ . Since  $m \geq 3$ ,  $2 \notin \{1, m\}$ . Then if  $2 \in L$ , then we have two equilibria  $\sigma_1|_{[1,2]}$  and  $\sigma_2|_{[1,2]}$  on  $[1, 2]$  and if  $2 \in R$ , we have two different equilibria  $\sigma_1|_{[2,m]}$  and  $\sigma_2|_{[2,m]}$  on  $[2, m]$ . In either case, we get a contradiction; this contradiction proves that  $m = 2$ .

*Completing the proof.* We have shown that there is a game with two states,  $S = \{1, 2\}$ , and two equilibria. Moreover, the set of states  $Z$  where  $q_z^1$  and  $q_z^2$  are different is the whole set  $S$ . Without loss of generality, suppose  $q_{11}^1 > q_{11}^2$ . Since  $q_{11}^2 < 1, q_{12}^2 > 0$  and in a monotone equilibrium we must have  $q_{22}^2 = 1$ ; this means  $q_{22}^1 < 1$ , and thus  $q_{21}^1 > 0$  and again by monotonicity  $q_{11}^1 = 1$ . From Lemma A6, this implies that the sequence  $b_{d_1}^{(t)}$  is nondecreasing (because equilibrium  $\sigma_2$  exists) and  $b_{d_2}^{(t)}$  is nonincreasing (because equilibrium  $\sigma_1$  exists). Suppose  $b_{d_1}^{(\infty)} < \frac{b_{d_1} + b_{d_2}}{2}$ , then Lemma A7 implies that  $q_{11}^1 = q_{11}^2 = 1$ , which contradicts  $q_{11}^1 > q_{11}^2$ . If  $b_{d_2}^{(\infty)} > \frac{b_{d_1} + b_{d_2}}{2}$ , then we get a similar contradiction. Since  $b_{d_1}^{(\infty)} \leq b_{d_2}^{(\infty)}$  by Assumption 1, we must have  $b_{d_1}^{(\infty)} = b_{d_2}^{(\infty)} = \frac{b_{d_1} + b_{d_2}}{2}$ , which is nongeneric. This proves that under Assumption 2, for generic parameter values, we have a unique equilibrium. ■

*Proof of Theorem 3.* Part i. In this proof, let  $Z_s = \arg \min_{z \in S} |b_{d_z} - b_{d_s}^{(\infty)}|$ ; this set is either a singleton or consists of two adjacent states. The result follows from the following three steps.

Step 1. Denote

$$\xi = \min_{s,y,z \in S, |b_{d_y} - b_{d_s}^{(\infty)}| > |b_{d_z} - b_{d_s}^{(\infty)}|} \left( (b_{d_y} - b_{d_s}^{(\infty)})^2 - (b_{d_z} - b_{d_s}^{(\infty)})^2 \right),$$

$$\Xi = b_m - b_1,$$

and take  $\varepsilon = \frac{\xi}{4\Xi}$ . For such  $\varepsilon$  there exists  $T \geq 1$  such that for all  $s \in S$  and  $t > T$ ,  $|b_{d_s}^{(t)} - b_{d_s}^{(\infty)}| < \varepsilon$ . Let  $\tilde{\beta} = (1 - \frac{\xi}{4\Xi^2})^{\frac{1}{T}}$ . Then for  $\beta \in (\tilde{\beta}, 1)$ , if for  $s \in S$ , some state  $z \in Z_s$  is stable (satisfies  $\Phi_z = \{z\}$ ), and the equilibrium path starting from state  $x$  never reaches the set  $Z_s$ , then the decisive group in  $s, d_s$ , strictly prefers moving to  $z$  to moving to  $x$ :  $W_{d_s}(z) > W_{d_s}(x)$ .

*Proof.* Consider the following difference:

$$\begin{aligned}
 &W_{d_s}(z) - W_{d_s}(x) \\
 &= \sum_{k \in G} \mu_{d_s,k} (V_k(z) - V_k(x)) \\
 &= \sum_{t \geq 1} \sum_{k \in G} \sum_{y \in S} \beta^{t-1} \mu_{d_s,k}^{(t)} \Pr(s_t = y) \left( A_k - (b_k - b_{d_z})^2 - A_k + (b_k - b_{d_y})^2 \right) \\
 &= \sum_{t \geq 1} \sum_{k \in G} \sum_{y \in S \setminus Z_s} \beta^{t-1} \mu_{d_s,k}^{(t)} \Pr(s_t = y) \left( (b_k - b_{d_y})^2 - (b_k - b_{d_z})^2 \right) \\
 &= \sum_{t \geq 1} \sum_{k \in G} \sum_{y \in S \setminus Z_s} \beta^{t-1} \mu_{d_s,k}^{(t)} \Pr(s_t = y) (b_{d_z} - b_{d_y}) (2b_k - b_{d_y} - b_{d_z}) \\
 &= \sum_{t \geq 1} \sum_{y \in S \setminus Z_s} \beta^{t-1} \Pr(s_t = y) (b_{d_z} - b_{d_y}) \left( 2b_{d_s}^{(t)} - b_{d_y} - b_{d_z} \right) \\
 &= \sum_{t \geq 1} \sum_{y \in S \setminus Z_s} \beta^{t-1} \Pr(s_t = y) \left( (b_{d_z} - b_{d_y}) \left( 2b_{d_s}^{(\infty)} - b_{d_y} - b_{d_z} \right) \right. \\
 &\quad \left. + 2(b_{d_z} - b_{d_y}) \left( b_{d_s}^{(t)} - b_{d_s}^{(\infty)} \right) \right) \\
 &= \sum_{t \geq 1} \sum_{y \in S \setminus Z_s} \beta^{t-1} \Pr(s_t = y) \left( \left( b_{d_s}^{(\infty)} - b_{d_y} \right)^2 - \left( b_{d_s}^{(\infty)} - b_{d_z} \right)^2 \right. \\
 &\quad \left. + 2(b_{d_z} - b_{d_y}) \left( b_{d_s}^{(t)} - b_{d_s}^{(\infty)} \right) \right) \\
 &\geq \frac{\beta}{1 - \beta} \xi - 2 \frac{\beta(1 - \beta^T)}{1 - \beta} \Xi^2 - 2 \frac{\beta^{T+1}}{1 - \beta} \Xi \varepsilon \\
 &> \frac{\beta}{1 - \beta} \left( \xi - 2(1 - \tilde{\beta}^T) \Xi^2 - 2\Xi \varepsilon \right) \\
 &= \frac{\beta}{1 - \beta} \left( \xi - 2 \frac{\xi}{4\Xi^2} \Xi^2 - 2\Xi \frac{\xi}{4\Xi} \right) = 0.
 \end{aligned}$$

Thus,  $W_{d_s}(z) > W_{d_s}(x)$ .

**Step 2.** Suppose that  $\beta$  is sufficiently close to 1, and in some equilibrium, for state  $s \in S$ , at least one of the states  $z \in Z_s$  is stable. Then with probability 1 the society starting from  $s$  will end up in one of these states (in some  $z \in Z_s$  that is stable).

*Proof.* If  $s \in Z_s$  and is stable, then the statement is trivial.

Suppose  $s \in Z_s$  and is not stable. Without loss of generality,  $s < z$  (where  $z$  is the stable state from  $Z_s$ ). Then  $\Phi_s \leq z$  due to monotonicity. On the other hand, from Step 1 it follows that  $\Phi_s \geq s$ , as otherwise members of  $d_s$  would be strictly better off moving to  $z$ . Thus, starting from  $s$ , only  $s$  and  $z$  may be reached, and since  $s$  is unstable,  $z$  is reached with a positive probability every period. Thus, it is reached with probability 1.

Finally, suppose  $s \notin Z_s$ . Again, without loss of generality,  $s < Z_s$ . From Step 1 it follows that  $\Phi_s \geq s$ . If  $\Phi_s \neq \{s\}$ , then  $y \in \Phi_s$  for some  $y > s$ , and then by monotonicity  $y \leq z$  for  $z \in Z_s$  that is stable. Moreover, the last inequality holds for all states that may be reached from  $y$ . But such paths must reach  $Z_s$  with probability 1 (otherwise it would contradict the result in Step 1), and once they do, they must reach a stable state in  $Z_s$ . The only remaining possibility is  $\Phi_s = \{s\}$ , so  $s$  is stable. But this is impossible from Step 1. This proves that a stable state from  $Z_s$  is reached with probability 1.

Step 3. For sufficiently high  $\beta$ , there exists an equilibrium such that for each state  $s \in S$ , at least one of the states  $z \in Z_s$  is stable:  $\Phi_z = \{z\}$ .

*Proof.* First, notice that for all states  $z \in Z_s$ , the corresponding bliss point of the decision makers' distant future selves is the same,  $b_{d_z}^{(\infty)} = b_{d_s}^{(\infty)}$ , and thus  $Z_z = Z_s$ . This follows from Assumption 1, which implies that each component is a connected set (intersection of  $S$  with an interval), and, for each state  $x$  in this component,  $b_{d_x}^{(\infty)}$  lies in the convex hull of the current selves' bliss points.

We now define the following set of feasible transitions, so as to make use of the more general result established in the proof of Theorem 1. Suppose first that  $Z_s$  is a singleton  $\{z\}$ . Then define the set of feasible transitions  $\{F_x\}_{x \in S}$  in the following way:  $y \in F_x$  if either  $x < z$  and  $y \leq z$ , or  $x > z$  and  $y \geq z$ , or  $x = y = z$  (in other words, we postulate that state  $z$  is stable, and allow any transitions that do not lead from the left of  $z$  to the right of  $z$  or vice versa). We established that this game has an equilibrium, with a corresponding transition matrix  $\tilde{Q}$ ; let  $\tilde{\Phi}_s$  be the set  $\{x \in S : \tilde{q}_{sx} > 0\}$ . By construction,  $\tilde{q}_{zz} = 1$ , so  $\tilde{\Phi}_z = \{z\}$ . If there exists a symmetric monotone MPE in the original game (without restricted transitions) that also gives rise to transition matrix  $\tilde{Q}$ , the result is proven. If not, then by Lemma A5 there must exist a monotone deviation, namely, states  $x, y, a \in S$  such that  $\tilde{q}_{xa} > 0$ ,

$W_{d_x}(y) > W_{d_x}(a)$  and, in addition, the correspondence  $q' : S \rightarrow S$  defined by equation (21) (replacing  $q$  with  $\tilde{q}$ ) is monotone.

Notice that it must be that  $x = z$ . If not, then without loss of generality assume  $x > z$ , and monotonicity implies  $y \geq z$  and  $a \geq z$  (because  $z$  is stable under  $\tilde{Q}$ ), but then  $W_{d_x}(y) > W_{d_x}(a)$  would be equivalent to  $\tilde{W}_{d_x}(y) > \tilde{W}_{d_x}(a)$  as the paths would be identical in the two games, with or without restriction on transitions. But the last equation would contradict that  $\tilde{Q}$  is a transition matrix of a MPE. Thus,  $x = z$ , and then  $a = x = z$  ( $\tilde{q}_{za} > 0$  implies  $a = z$ ). Now,  $W_{d_x}(y) > W_{d_x}(a)$  implies  $W_{d_x}(y) > W_{d_x}(z)$ , so  $y \neq z$ . Without loss of generality, assume  $y > z$ . But by monotonicity of this deviation, we must have  $\tilde{\Phi}(y) \geq y$ , and therefore all paths that start from  $y$  never reach  $z$ . But then  $W_{d_x}(y) > W_{d_x}(z)$  contradicts Step 1, because, as argued above,  $Z_z = Z_s = \{z\}$ . This contradiction completes the proof in this case.

Now assume that  $Z_s$  consists of two points,  $z < z'$ . Here, we need an auxiliary step. Introduce the set of feasible transitions  $F'$  in the following way:  $(x, y) \in F'$  if either  $x < z$  and  $y \leq z'$ , or  $x > z'$  and  $y \geq z$ , or  $x, y \in Z_s$ . As before, there is an equilibrium  $\sigma'$  that gives rise to a transition mapping  $Q'$ . By feasibility, it is only possible to transition from  $z$  and  $z'$  onto this set, and monotonicity implies that at least one of the states  $z$  and  $z'$  is stable in this equilibrium. Without loss of generality, suppose that state  $z$  is stable; then from  $z'$  it may only be possible to stay in  $z'$  or transit to  $z$ . Now, let us lift the restriction on transitions. If matrix  $Q'$  corresponds to an equilibrium in the original game, the result is proven. Otherwise, as before, by Lemma A5 there must exist a monotone deviation. For the tuple  $(x, y, a)$  that constitutes a deviation, it is impossible that  $x < z$  or  $x > z'$  (there is no monotone deviation that would not be feasible under  $F'$ ). Suppose instead that  $x \in Z_s = \{z, z'\}$ . A deviation within  $Z_s$  (i.e.,  $y \in Z_s$ ) cannot yield a higher utility to  $d_s$ , because it was feasible under  $F'$ . Thus, the remaining case to consider is  $y \notin Z_s$ . If  $y < z$ , then this deviation leads to a path that never reaches  $Z_s$ , which contradicts Step 1. If  $y > z$ , then monotonicity of deviation implies that from state  $y$  it is impossible to move to any state  $b < y$ , and in particular to return to  $Z_s$ , which again contradicts Step 1. This contradiction proves Step 3 for the case where  $Z_s$  consists of two points. This completes the proof of part i of the theorem.

Part ii. By part i, there exists an equilibrium with the desired properties, and since the equilibrium is unique, the result follows.

Part iii. This follows immediately from part ii. ■

*Proof of Theorem 4.* Let us first prove that there is an equilibrium with stable democracy under both  $M$  and  $M'$ ; since we consider only the cases of unique equilibria, it would imply that democracy is stable under both  $M$  and  $M'$ . Let us do this in case of  $M$ . Impose the following restrictions on transitions: transition from  $y$  to  $z$  is infeasible if  $y \leq x < z$  or  $z < x \leq y$  and feasible otherwise. In the proof of Theorem 1, we established that there is an equilibrium in this case under some transition probability matrix  $Q$ ; since transitions from  $x$  were ruled out,  $q_{xx} = 1$ . Let us now lift the requirement on feasibility of transitions and assume that all transitions are feasible. Lemma A5 implies that if matrix  $Q$  does not correspond to an equilibrium, then there must be a deviation at state  $x$ . From the farsighted stability condition (9), we have  $\frac{b_{d_{x-1}} + b_{d_x}}{2} \leq b_{d_x}^\infty \leq \frac{b_{d_x} + b_{d_{x+1}}}{2}$ , and  $\frac{b_{d_{x-1}} + b_{d_x}}{2} \leq b_{d_x} \leq \frac{b_{d_x} + b_{d_{x+1}}}{2}$ , so under Assumption 2, we can conclude that  $\frac{b_{d_{x-1}} + b_{d_x}}{2} \leq b_{d_x}^{(t)} \leq \frac{b_{d_x} + b_{d_{x+1}}}{2}$  for any  $t$ . Therefore, there is no deviation that would make group  $d_x$  better off. This implies that there is an equilibrium with transition matrix  $Q$ , that is, an equilibrium where democracy  $x$  is stable. This proves that democracy is stable under  $M$  and, analogously, under  $M'$ .

Suppose that democracy is asymptotically stable under  $M$ . Consider equilibrium  $\sigma$ . Denote democracy by  $x$  and take  $y = x - 1$ ; if  $y \in G$ , then asymptotic stability implies that  $q_{yx} > 0$ . This means that  $b_{d_y}^{(t)}$  is nondecreasing: otherwise, Assumption 2 would imply that it is nonincreasing, and then by Lemma A6 applied to matrix  $M$ , would imply that  $q_{yx} > 0$  is impossible.

Let us prove that  $q'_{yx} > 0$ . Suppose not; then since  $x$  is stable, this is only possible if  $\Phi'_y \leq y$ . Now, applying Lemma A6 to matrix  $M'$ , we have  $\Phi'_y \geq y$ ; consequently, the only possibility is  $\Phi'_y = \{y\}$ , so  $y$  is stable under  $M'$ . If  $\Phi'_y = \{y\}$  in equilibrium, then  $W'_{d_y}(y) \geq W'_{d_y}(x)$  by part ii of Theorem 1. Now for matrix  $M$ , taking into account that  $b_{d_y}^{(t)} \leq b'^{(t)}_{d_y}$  for every  $t \geq 0$ , single-crossing implies that  $W_{d_y}(y, y, \dots) \geq W_{d_y}(x)$ , where the first term is the utility of members of  $d_y$  if the society stays in  $y$  forever. But this implies, by Lemma A4, that if the set of states is restricted to  $\{x, y\}$ , then under  $M$  there is an equilibrium where both  $x$  and  $y$  are stable. On the other hand, the same Lemma A4 implies that there is also an equilibrium  $\sigma|_{\{x, y\}}$ , where  $x$  is stable, but  $y$  is not. However, existence of two such equilibria would contradict Lemma A8 (which

is applicable because strict inequality  $b_j^{(t)} < b_j^{(t)}$  for some  $t$  implies that  $b_j^{(t)} < b_j^{(\infty)}$ , thus  $b_j^{(\tau)} < b_j^{(\tau+1)}$  for some  $\tau \geq t$ . This contradiction implies that the hypothesis that  $q'_{yx} = 0$  is wrong, and in fact  $q'_{yx} > 0$ . Now,  $b_{d_y}^{(t)}$  being nondecreasing implies that  $\Phi_y \geq y$ , and since  $\Phi_x = \{x\}$ , we must have  $\Phi_y \subset \{x, y\}$ . Consequently, with probability 1, starting from  $y$  there is convergence to  $x$ . The case of  $y = x + 1$  is considered similarly.

Finally, we prove that convergence to democracy is faster under  $M'$  than under  $M$  as claimed in note 24. Consider convergence from  $y = x - 1$  (the case of convergence from  $z = x + 1$  is considered similarly). We need to prove that  $q'_{yx} > q_{yx}$ . Since  $x$  is asymptotically stable,  $q'_{xa} > 0$  and  $q'_{ya} > 0$  are possible for  $a \in \{x, y\}$  only. Therefore, we have (using the same calculus as in the proof of Step 1 in Theorem 3):

$$\beta (W_j(x) - W_j(y)) = \sum_{t=1}^{\infty} \beta^t \left( (q_{yx}^{(t)} - q_{xx}^{(t)}) (b_j^{(t)} - b_{d_x})^2 + (q_{yy}^{(t)} - q_{xy}^{(t)}) (b_j^{(t)} - b_{d_y})^2 \right).$$

Notice that  $q_{yx}^{(t)} = 1 - (1 - q_{yx})^{t-1}$ ,  $q_{xx}^{(t)} = 1$ ,  $q_{yy}^{(t)} = (1 - q_{yx})^{t-1}$ , and  $q_{xy}^{(t)} = 0$ ; this implies

$$\begin{aligned} W_j(x) - W_j(y) &= \sum_{t=1}^{\infty} (\beta (1 - q_{yx}))^{t-1} \left( (b_j^{(t)} - b_{d_y})^2 - (b_j^{(t)} - b_{d_x})^2 \right) \\ &= \sum_{t=1}^{\infty} (\beta (1 - q_{yx}))^{t-1} (2b_j^{(t)} - b_{d_x} - b_{d_y}) (b_{d_x} - b_{d_y}) \\ &= (b_{d_x} - b_{d_y}) \sum_{t=1}^{\infty} (\beta (1 - q_{yx}))^{t-1} (2b_j^{(t)} - b_{d_x} - b_{d_y}). \end{aligned}$$

Let us denote  $\alpha = q_{yx}$ . In terms of notation of Lemma A8 we have

$$\begin{aligned} f(\alpha) = W_j(y) - W_j(x) &= -(b_{d_x} - b_{d_y}) \sum_{t=1}^{\infty} (\beta (1 - \alpha))^{t-1} \\ &\quad \times (2b_j^{(t)} - b_{d_x} - b_{d_y}). \end{aligned}$$

If, instead of transition matrix  $M$  we used matrix  $M'$ , but with the same probability  $\alpha$  of transition from  $y$  to  $x$  equal, we would obtain (similarly)

$$f'(\alpha) = -(b_{d_x} - b_{d_y}) \sum_{t=1}^{\infty} (\beta(1 - \alpha))^{t-1} (2b_j^{(t)} - b_{d_x} - b_{d_y}).$$

Since we have  $b_j^{(t)} \geq b_j^{(t)}$  for all  $t$  with at least one strict inequality, we have  $f'(\alpha) < f(\alpha)$ .

Notice that if  $q'_{yx} = 1$ , the result is proven (either  $q_{yx} = q'_{yx} = 1$  or  $q_{yx} < 1 = q'_{yx}$ ), so assume  $q'_{yx} < 1$  from now on. Consider two cases. If  $q_{yx} < 1$ , then since  $q_{yx} > 0$  (as  $x$  is asymptotically stable under  $M$ ),  $\alpha = q_{yx}$  must satisfy  $f(\alpha) = 0$ . This implies  $f(\alpha) < 0$ . Now, since  $q'_{yx} \in (0, 1)$ , it must satisfy  $f'(q'_{yx}) = 0$ , and then by Lemma A8 we must have  $q_{yx} = \alpha < q'_{yx}$ . Now consider the second case, where  $q_{yx} = 1$ . By part ii of Theorem 1, we must have  $f(\alpha) \leq 0$ , in which case  $f'(1) = f(\alpha) < 0$ . By Lemma A8 and continuity of  $f(\cdot)$ , we must have  $f(\xi) < 0$  for all  $\xi \in [0, 1]$ , and thus  $f'(q'_{yx}) < 0$ . Again by Theorem 1 this is only possible if  $q'_{yx} = 1$ . ■

*Proof of Theorem 5.* Suppose that  $b_{d_x}^{(\infty)} \leq b_{d_x}$  (the opposite case is analogous). Since condition (9) does not hold, we have that  $\frac{b_{d_x-2} + b_{d_x-1}}{2} \leq b_{d_x-1}^{(\infty)} \leq b_{d_x}^{(\infty)} < \frac{b_{d_x-1} + b_{d_x}}{2}$ . In this case,  $\frac{b_{d_x-2} + b_{d_x-1}}{2} \leq b_{d_x-1}^{(\infty)} < \frac{b_{d_x-1} + b_{d_x}}{2}$  implies, using the existence of equilibrium with restricted transitions (similarly to the proof of Theorem 4 and using Assumption 3) and then Lemma A5, that under both  $M$  and  $M'$  there are equilibria where state  $x - 1$  is stable. If so, democracy  $x$  is not asymptotically stable under either  $M$  or  $M'$ .

Suppose, to obtain a contradiction, that democracy is not stable under  $M$ , but is stable under  $M'$ . Denoting  $y = x - 1$ , Lemma A6 and the fact that  $y$  is stable implies that  $\Phi_x \in \{x, y\}$ . Then since  $x$  is not stable under  $M$ ,  $q_{xy} > 0$ , furthermore, since  $x$  is stable under  $M'$ ,  $q'_{xx} = 1$ . Since  $b_{d_x}^{(\infty)} < b_{d_x}$ , the fact that mobility under  $M'$  is faster than under  $M$  implies that  $b_{d_x}^{(t)} \geq b_{d_x}^{(t)}$  for all  $t \geq 1$ , with at least one strict inequality. If so, taking the equilibrium under  $M$  and changing transition probabilities so that  $x$  is stable would give another equilibrium under  $M$  (with the set of states restricted to  $\{x, y\}$ ), similarly to the proof of Theorem 4. However existence of two such equilibria contradicts Lemma A8; thus if democracy is not stable under  $M$ , then it is not stable under  $M'$  either. ■

*Proof of Corollary 1.* Without loss of generality, assume that the sequence  $b_{d_x}^{(t)}$  is nondecreasing. We first show that state  $x$  is not asymptotically stable. If it is not stable, this follows immediately, so suppose it is stable. Consider the state  $x + 1$ . If  $b_{d_{x+1}}^{(t)}$  is also nondecreasing, then  $\Phi_{x+1} \geq x + 1$  by Lemma A6, and thus starting from  $x + 1$  the society cannot reach  $x$ . If  $b_{d_{x+1}}^{(t)}$  is nonincreasing, then either  $\Phi_{x+1} = \{x + 1\}$  or  $x \in \Phi_{x+1}$ , because  $x$  is stable. In the former case, the result is proved, so suppose  $x \in \Phi_{x+1}$ . But then, since condition (9) is violated,  $b_{d_{x+1}}^{(\infty)} = b_{d_x}^{(\infty)} > \frac{b_{d_x} + b_{d_{x+1}}}{2}$ , therefore,  $b_{d_{x+1}}^{(t)} > \frac{b_{d_x} + b_{d_{x+1}}}{2}$  for all  $t$ . By Lemma A7 we then have  $\Phi_{x+1} = \{x + 1\}$ , a contradiction. This contradiction shows that  $x$  is not asymptotically stable.

Let us show that there exist  $\varepsilon_1$  and  $T_1$  such that  $|b_{d_x}^{(T_1)} - b_{d_x}| < \varepsilon_1$  implies that  $x$  is stable. Take  $\varepsilon_1 = \frac{b_{d_{x+1}} - b_{d_x}}{4}$  and  $T_1$  satisfying  $T_1 > -\frac{\log\left(1 + \frac{\Xi^2}{8\varepsilon_1^2}\right)}{\log \beta}$ , where again  $\Xi = b_m - b_1$ . Consider the set of transitions  $F$  such that  $b \in F_a$  if and only if either  $a > x$  or  $b \leq x$ . Then by Theorem 1 there exists equilibrium  $\sigma$  under this set of transitions. Furthermore, Lemma A6 implies that  $x$  is stable in this equilibrium. To verify that there is equilibrium  $\sigma'$  where  $x$  is stable, by Lemma A5 it suffices to verify that individuals from group  $d_x$  would not benefit from deviating to any  $z > x$ . Indeed, we have, similar to the proof of Theorem 3, that

$$\begin{aligned} W_{d_x}(x) - W_{d_x}(z) &= \sum_{k \in G} \mu_{d_x k} (V_k(x) - V_k(z)) \\ &= \sum_{t \geq 1} \sum_{y \geq z} \beta^{t-1} \Pr(s_t = y) \left( (b_{d_y} - b_{d_x}^{(t)})^2 - (b_{d_x} - b_{d_x}^{(t)})^2 \right) \\ &\geq \frac{1 - \beta^{T_1}}{1 - \beta} \left( (3\varepsilon_1)^2 - (\varepsilon_1)^2 \right) - \frac{\beta^{T_1}}{1 - \beta} \Xi^2 \\ &= \frac{1}{1 - \beta} \left( 8\varepsilon_1^2 - \beta^{T_1} (8\varepsilon_1^2 + \Xi^2) \right) > 0, \end{aligned}$$

where the probability is over the distribution of states following transition to  $z$ . This implies that a deviation to  $z$  is not profitable

from group  $d_x$ . Consequently, there is an equilibrium in the original game where  $x$  is stable, and given uniqueness, this must be true in this unique equilibrium.

Let us now show that there exist  $\varepsilon_2$  and  $T_2$  such that  $\left| b_{d_x}^{(T_2)} - b_{d_x}^{(\infty)} \right| < \varepsilon_2$ , which implies that  $x$  is unstable. Let state  $s \in \arg \min_{y \in S} \left| b_{d_y} - b_{d_x}^{(\infty)} \right|$  (or the leftmost one if the maximum is attained at two states); this implies, in particular, that  $b_{d_s}^{(\infty)} = b_{d_x}^{(\infty)}$ . This means that there is an equilibrium where  $s$  is stable, so  $q_{s,s} = 1$ . Now take  $\varepsilon_2 = \min \left( \frac{b_{d_s} - b_{d_s-1}}{2} \right)$  and  $T_2 = 1$ . Suppose, to obtain a contradiction, that  $x$  is stable. Since under matrix  $M'$  all future selves of individuals from group  $d_x$  prefer state  $s$  to  $x$  and  $s$  is stable, they would have a profitable deviation in the form of transiting to  $s$ . This cannot be the case in equilibrium, which proves that  $x$  is unstable. Setting  $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$ , we see that  $\varepsilon, T_1, T_2$  satisfy the required properties, which completes the proof.

MIT

NORTHWESTERN UNIVERSITY

UNIVERSITY OF CHICAGO AND HIGHER SCHOOL OF ECONOMICS,  
MOSCOW

#### SUPPLEMENTARY MATERIAL

An [Online Appendix](#) for this article can be found at *The Quarterly Journal of Economics* online.

#### REFERENCES

- Aaronson, Daniel, and Bhaskar Mazumder, "Intergenerational Economic Mobility in the United States, 1940 to 2000," *Journal of Human Resources*, 43 (2008), 139–172.
- Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin, "Political Selection and Persistence of Bad Governments," *Quarterly Journal of Economics*, 125 (2008), 1511–1575.
- , "Dynamics and Stability of Constitutions, Coalitions and Clubs," *American Economic Review*, 102 (2012), 1446–1476.
- , "Political Economy in a Changing World," *Journal of Political Economy*, 123 (2015), 1038–1086.
- , "Social Mobility and Stability of Democracy: Re-evaluating De Tocqueville," NBER Working Paper 22174, 2016.
- Acemoglu, Daron, and James Robinson, "A Theory of Political Transitions," *American Economic Review*, 91 (2001), 938–963.
- , "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective," *Quarterly Journal of Economics*, 115 (2000), 1167–1199.

- Alesina, Alberto, and Edward Glaeser, *Fighting Poverty in the US and Europe: A World of Difference* (Oxford: Oxford University Press, 2004).
- Alesina, Alberto, and Paola Giuliano, "Preferences for Redistribution" in *Handbook of Social Economics*, Vol. 1A, Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, eds. (Netherlands: North-Holland, 2010), 93–131.
- Amador, Manuel, "Essays in Macroeconomics and Political Economy," PhD thesis, MIT, 2003.
- Austen-Smith, David, and Jeffrey S. Banks, *Positive Political Theory II: Strategy and Structure* (Ann Arbor: University of Michigan Press, 2005).
- Besley, Timothy, and Stephen Coate, "Sources of Inefficiency in a Representative Democracy: A Dynamic Analysis," *American Economic Review*, 88 (1998), 139–156.
- Bénabou, Roland, and Efe Ok, "Social Mobility and the Demand for Redistribution: The POUM Hypothesis," *Quarterly Journal of Economics*, 116 (2001), 447–487.
- Bénabou, Roland, and Jean Tirole, "Belief in a Just World and Redistributive Politics," *Quarterly Journal of Economics*, 121 (2006), 699–746.
- Bisin, Alberto, Alessandro Lizzeri, and Leeat Yariv, "Government Policy with Time-Inconsistent Voters," *American Economic Review*, 105 (2015), 1711–1737.
- Blau, Peter, and Otis Duncan, *The American Occupational Structure* (New York: Wiley, 1967).
- Bourguignon, Francois, and Thierry Verdier, "Oligarchy, Democracy, Inequality and Growth," *Journal of Development Economics*, 62 (2000), 285–313.
- Cao, Dan, and Ivan Werning, "Dynamic Savings Choices with Disagreements," NBER Working Paper No. w22007, 2016.
- Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner, "In the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility," *American Economic Review*, 104 (2014), 141–147.
- Coughlin, Peter J., *Probabilistic Voting Theory* (Cambridge: Cambridge University Press, 1992).
- De Tocqueville, Alexis, *De la démocratie en Amérique* (Paris: Librairie de Charles Gosselin); English translation, *Democracy in America* (Chicago: University of Chicago Press, 1835).
- Erikson, Robert, and John Goldthorpe, *The Constant Flux: A Study of Class Mobility in Industrial Societies* (Oxford: Clarendon Press, 1992).
- Gomes, Armando, and Philippe Jehiel, "Dynamic Processes of Social and Economic Interactions: On the Persistence of Inefficiencies," *Journal of Political Economy*, 113 (2005), 626–667.
- Grossman, Gene M., and Elhanan Helpman, "Protection for Sale," *American Economic Review*, 84 (1994), 833–850.
- Gul, Faruk, and Wolfgang Pesendorfer, "Self Control, Revealed Preferences and Consumption Choice," *Review of Economic Dynamics*, 7 (2004), 243–264.
- Hilger, Nathaniel, "The Great Escape: Intergenerational Mobility since 1940," NBER Working Paper 21217, 2015.
- Hungerford, Thomas, *Algebra* (New York: Springer, 1974).
- Jackson, Matthew O., and Leeat Yariv, "Collective Dynamic Choice: The Necessity of Time Inconsistency," *American Economic Journal: Microeconomics*, 7 (2015), 150–178.
- Lagunoff, Roger, "Markov Equilibrium in Models of Dynamic Endogenous Political Institutions," Mimeo, Georgetown University, 2006.
- Leventoğlu, Bahar, "Social Mobility and Political Transitions," *Journal of Theoretical Politics*, 17 (2005), 465–496.
- , "Social Mobility, Middle Class, and Political Transitions," *Journal of Conflict Resolution*, 58 (2014), 825–864.
- Lindbeck, Assar, and Jörgen W. Weibull, "Balanced-Budget Distribution as the Outcome of Political Competition," *Public Choice*, 52 (1987), 273–297.

- Lipset, Seymour, "Political Consequences of Social Mobility," in *Social Mobility and Political Attitudes*, Frederik C. Turner ed. (New Brunswick: Transaction Publishers, 1992).
- Lipset, Seymour, and Reinhard Bendix, *Social Mobility in Industrial Society* (Berkeley: University of California Press, 1959).
- Lizzeri, Alessandro, and Nicola Persico, "Why Did the Elites Extend the Suffrage? Democracy and the Scope of Government, with an Application to Britain's 'Age of Reform'," *Quarterly Journal of Economics*, 119 (2004), 705–763.
- Mann, Michael, *Fascists* (Los Angeles: University of California Press, 2004).
- Moore, Barrington, *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World* (Boston: Beacon Press, 1966).
- Pareto, Vilfredo, *The Mind and Society* (New York: Harcourt, Brace, 1935).
- Pekkarinen, Tuomas, Kjell Salvanes, and Matti Sarvimaki, "The Evolution of Social Mobility: Norway over the 20th Century," Mimeo, 2016.
- Piketty, Thomas, "Social Mobility and Redistributive Politics," *Quarterly Journal of Economics*, 110 (1995), 551–583.
- Roberts, Kevin, "Dynamic Voting in Clubs," *Research in Economics*, 69 (2015), 320–335.
- Sombart, Werner, *Warum gibt es in den Vereinigten Staaten keinen Sozialismus? (Why Is There No Socialism in the United States?)* (New York: Sharpe, 1976, 1906).
- Stachura, Peter, "National Socialism and the German Proletariat, 1925–1935: Old Myths and New Perspectives," *Historical Journal*, 36 (1993), 701–718.
- Storer, Colin, *A Short History of the Weimar Republic* (London: I.B.Tauris, 2013).
- Strulovici, Bruno, "Learning while Voting: Determinants of Collective Experimentation," *Econometrica*, 78 (2010), 933–971.
- Wright, Randall, "The Redistributive Roles of Unemployment Insurance and the Dynamics of Voting," *Journal of Public Economics*, (1986), 377–399.