*Social Mobility and Stability of Democracy: Reevaluating De Tocqueville**

NBER WORKING PAPER SERIES

SOCIAL MOBILITY AND STABILITY OF DEMOCRACY: RE-EVALUATING DE TOCQUEVILLE

Daron Acemoglu
Georgy Egorov
Konstantin Sonin

Working Paper 22174
http://www.nber.org/papers/w22174

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2016

We thank participants of 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, Warwick/Princeton Political Economy Conference in Venice, and seminars at the University of Chicago, Higher School of Economics, MIT, University of Waterloo, University of Wisconsin-Madison, and the University of Zurich for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Social Mobility and Stability of Democracy: Re-evaluating De Tocqueville
Daron Acemoglu, Georgy Egorov, and Konstantin Sonin
NBER Working Paper No. 22174
April 2016
JEL No. D71,D74

ABSTRACT

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 paper, 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. When social mobility is endogenized, our model identifies new political economic forces limiting the amount of mobility in society – because the middle class will lose out from mobility at the bottom and because a peripheral coalition between the rich and the poor may oppose mobility at the top.

Daron Acemoglu
Department of Economics, E52-446
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and CIFAR
and also NBER
daron@mit.edu

Konstantin Sonin
Irving B. Harris School of
Public Policy Studies
University of Chicago
1155 E60th St
Chicago, IL 60637
ksonin@uchicago.edu

Georgy Egorov
Kellogg School of Management
Northwestern University
2001 Sheridan Road
Evanston, IL 60208
and NBER
g-egorov@kellogg.northwestern.edu

1 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 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 [1862], Book 2, pp. 120-121).

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

> "In describing 'The Social Conditions of the Anglo-Americans' in the *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 American 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 seminal (1967) study concluded "the stability of American democracy is undoubtedly related to the superior chances of the upward mobility in this country" (similar ideas also appear in Pareto, 1935, Barrington Moore, 1966, Sombart, 1906, and Erikson and Goldthorpe, 1992). This perspective suggests that a 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.

Despite its ubiquity in modern debates on democracy and in modern social theories, there has been little systematic formalization or critical investigation of this idea. The next example illustrates not only why this idea is intuitive, but also why greater social mobility may actually destabilize democracy.

Example 1 Consider a society with n individuals, with $\frac{2}{5}n$, or 40 percent of them, rich, $\frac{1}{5}n$ or 20 percent, middle class, and $\frac{2}{5}n$ or 40 percent 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 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.[1] 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 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 $2/5, 1/5$, and $2/5$, 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.

The same 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 upwards. Suppose that the society starts out as a democracy. Then, if sufficiently many members of the middle class move upwards (i.e., if $\alpha > 1/2$), the middle-class agents expect, on average, to have the preferences of the rich tomorrow, and hence prefer policy tomorrow to be determined in elite dictatorship, making democracy unstable.

This example thus provides a simple (and as we will see, robust) reason why greater social mobility may undermine the stability of democracy: 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.[2]

Differently from this example, our main model will consider 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 Benabou and Ok's, 2001, analysis of the relationship between social mobility and

[1] Throughout the paper, when all current members of a social group have the same preferences, we will interchangeably refer to a member of that social group or the entire social group.

[2] The fact that the social mobility in this example makes middle-class agents more likely to move upwards rather than downwards is important, as we will see in our analysis. If they expected to move upwards or downwards 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.

redistribution), then there would be no need for institutional changes. Second, such a model also precludes any effect of future social mobility on preferences, e.g., from the fact that middle-class agents may not only move up to the next social group in the next period, but move yet further up or even possibly down in subsequent periods. 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.[3]

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 $\beta = 4/5$. In each period, the current decision-maker determines next period's institution, and in-between, r people move upwards from the middle class, and r rich agents move downwards. Let $\alpha = \frac{5r}{n}$ again denote the share of the middle class moving upwards.

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 $1/4 < \alpha < 1/2$. 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 2/3 of the time and remain in the middle class 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 the 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

[3] This example and our analysis below highlight the importance of two kinds of conflicts of interest: between agents from different social groups; and between today's decision-maker and future decision-makers who will occupy in the future the same social group as the current decision-maker. The latter conflict arises from the fact that today's decision-maker anticipates being in different social class in the future. This conflict is not only essential for understanding the political implications of social mobility, but 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 class in future dates), and yet its origins are not in time-inconsistent preferences, but in social mobility.

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. 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 to the further 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 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 requires 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 democracy. We quantify the stability of democracy with the size of its basin of attraction along the equilibrium path. 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'.[4] 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), the condition on the preferences of the median voter takes an even more intuitive form: it requires the preferences of the median of the society to be close to the average of the preferences of all voters.

Our paper is most closely related to the small literature on the interplay between social mobility and redistribution. The important paper by Benabou and Ok (2001), which has already been

[4] This notion of stability thus captures both the potential instability of democracy resulting from the median voter preferring 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).

mentioned, shows how greater social mobility (or expectations thereof) 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 Benabou and Ok is linked to De Tocqueville's hypothesis — greater mobility makes the middle class less willing to tax the rich because they expect to become rich in the future. They generate this effect by assuming that taxes are 'sticky' (i.e., there is some commitment to future taxes). In Benabou and Tirole (2006), beliefs about future social mobility support different equilibria — e.g., 'the American dream' equilibrium, in which high level of efforts stems from the belief in high social mobility (see also Alesina and Glaeser, 2004, and Alesina and Giuliano, 2010). Nevertheless, this literature does not consider the relationship between social mobility and support for different types of political institutions. More importantly, it neither incorporates the dynamic political trade-offs that are at the heart of our paper nor does it feature the potentially destabilizing role of social mobility for democracy. Notably, Leventoglu (2005) investigates the link between social mobility and democracy in a world with three social groups, but only obtains the stabilizing role of social mobility due to various special assumptions.

Our modeling approach overlaps with dynamic political economy models studying democratization, constitutional change, repression and the efficiency of long-run institutional arrangements, including Besley and Coate (1998), Bourguignon and Verdier (2000), Acemoglu and Robinson (2000 and 2001), Lizzeri and Persico (2004), Gomes and Jehiel (2005), Lagunoff (2006), Acemoglu, Egorov, and Sonin (2010, 2015), and Roberts (2015), though again none of this literature studies social mobility and the mechanisms that are at the heart of our paper.

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 none of these works note the conflict between current and future selves resulting from social mobility or study the implications of this type of conflict for institutional change.

The rest of the paper is organized as follows. In Section 2 we introduce our setup. Section 3 solves the model and establishes existence of an equilibrium, provides conditions for uniqueness, and studies its main properties. Section 4 contains our main results linking the speed of social mobility to the stability of democracy. Section 5 contains 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); and second, we generalize our main results to environments with multiple equilibria. Section 6 endogenizes social mobility and studies how, in a simplified version of our baseline model, concerns about changes in future social mobility constrain equilibrium mobility decisions. Section 7 concludes. Ap-

pendix A contains the proofs of the main results presented in the text, while Appendix B, which is not for publication, includes the remaining proofs, several additional examples and further results.

2 Model

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

2.1 Society, policies and preferences

Time is discrete and infinite, indexed by $t \geq 1$. Society consists of n individuals split into g social groups, $G = \{1, ..., g\}$ 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., lower-indexed groups could be those that are richer and prefer lower taxes). All individuals share a common discount factor $\beta \in (0, 1)$.

Preferences are defined over a policy space represented by the real line, $\mathbb{R}$. 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:

$$u_k(p_t) = A_k - (b_k - p_t)^2, \tag{1}$$

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).[5] 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 both on the current policy $p_t \in \mathbb{R}$ and on the next period's arrangement. We assume that there are m (political) states $s \in S = \{1, \ldots, m\}$, 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).

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

[5] For example, if all $A_k = 0$, then members of the middle class might not want to become rich if the political institution is democracy, because this may hurt policy payoff. This is inconsequential when social mobility is exogenous, but would lead to unrealistic predictions once we endogenize social mobility.

pivotal group are equivalent, we can without loss of any generality assume that each state has a different pivotal group, so $\{d_s\}_{s\in S}$ are all different. We can then order states such that the sequence of pivotal groups, $\{d_s\}$, is increasing.

2.2 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 the her offspring moving to a different social group than herself (and the individual herself having dynastic preferences).

Throughout we assume that, though 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.[6]

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

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

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

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.[7] Throughout the paper, we also impose the following assumption:

Assumption 1 ***(Between-Person Monotonicity)*** *For two groups j_1 and j_2 with $j_1<j_2$, marginal probability distribution $\left\{\mu_{j_1\cdot}\right\}$ over G is first-order stochastically dominated by $\left\{\mu_{j_2\cdot}\right\}$. Formally, for any $l\in\{1,\ldots,g\}$,*

$$\sum_{k=1}^{l}\mu_{j_1k}>\sum_{k=1}^{l}\mu_{j_2k}. \tag{4}$$

[6] 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.

[7] Matrix M can be equivalently defined by using permutations $\pi\in S_N$ of individuals and assuming that in each period, Nature changes identities of individuals according to π with probability λ_π, such that $\sum_{\pi\in S_N}\lambda_\pi=1$. The symmetry requirement then becomes $\lambda_\pi=\lambda_{\sigma\circ\pi\circ\tau}$ for any $\sigma,\tau\in S_N$ that reshuffle individuals within groups only. In this case, denoting the set of individuals in group j by G_j, we have $\mu_{jk}=\frac{1}{n_j}\sum_{i\in G_j}\sum_{\pi\in S_N:\pi(i)\in G_k}\lambda_\pi$. The converse is also true: for any matrix $M=\left\{\mu_{jk}\right\}$ of nonnegative elements satisfying (2)–(3), there is a corresponding distribution λ over permutations π (this distribution may be not uniquely defined). This relatively minor generalization of the Birkhoff-von Neumann theorem for doubly stochastic matrices is proved in Lemma B2 in Appendix B.

This assumption, which is quite weak, imposes that the distribution of a richer individual's future selves first-order stochastically dominates the distribution of the 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.[8] We next provide an example of a class of social mobility matrices 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 paper, we use the standard notation M^τ to denote the τth power of the social mobility matrix M, and use μ_{jk}^τ to denote its generic element. The element μ_{jk}^τ of this matrix gives the probability that an individual currently in social group j will be in social group k in τ periods time.

2.3 Timing of events

To specify how political decisions are made, we assume that there is a fixed order of groups in each state, $\pi_s : \{1, ..., g\} \to G$, which determines the sequence in which (representatives of) different groups make proposals, and that 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.

1. *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).

[8] This assumption can be further weakened to have a weak inequality in (4), but the version with strict inequality simplifies our exposition and proofs. In fact, Example 1 in the Introduction only satisfies this assumption with weak inequality, but this is also just for simplicity, and having less than full reshuffling in that example would not affect the conclusions.

(c) All individuals vote, sequentially, with each agent i casting vote $v_i^p(j) \in \{Y, N\}$.

(d) If $\frac{\sum_{i=1}^n w_{g_i^t}(s_t)\mathbf{1}\{v_i^p(j)=Y\}}{\sum_{i=1}^n w_{g_i^t}(s_t)} > \frac{1}{2}$, 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 1(b) with j increased by 1, unless $j = g$.

(e) The policy decided in the last stage is implemented: $p_t = p_t^g$.

2. *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 2(b) with j increased by 1, unless $j = g$.

(e) The transition decided in the last stage is implemented: $s_{t+1} = s_{t+1}^g$.

3. *Payoffs:* Each individual i receives time-t payoff of $u_{g_i^t}(p_t)$, given by (1).

4. *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 are made, is similar to the "amendments" games discussed in Austen-Smith and Banks (2005).

2.4 Definition of equilibrium

We focus on symmetric monotone Markov Perfect Equilibrium (MPE for short). 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.[9] Since, as shown in Example 2 in the Introduction, pure-strategy equilibria may fail to exist, we allow proposers to mix between alternatives. Thus, 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

[9] Example B5 in Appendix B provides an example of a non-monotonic Markov Perfect Equilibrium, but Theorem B2 provides intuitive sufficient conditions for all equilibria to be monotone.

proposing political transition, and into $\{Y, N\}$ when she is at the voting stage. We next define our equilibrium concept more formally.

Definition 1 ***(Symmetric Monotone Markov Perfect Equilibrium)*** *A subgame perfect equilibrium $\hat{\sigma}$ is a Markov Perfect Equilibrium (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).*[10]

An MPE σ 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 $s_t = y$, i.e., for any $l \in [1, m]$,

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

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 (political) state 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 $\lim_{\tau \to \infty} \Pr(s_\tau = s) = 1$, 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 analog 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 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.

3 Analysis

In this section, we prove the existence of equilibrium, present some basic characterization results, and also provide conditions for uniqueness.

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

3.1 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.

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

1. *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 $p_t = b_{d_s}$;*

2. *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*

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

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

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

3. *The transitions induced by the equilibrium are strongly monotonic: 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 $a \leq b$;*

4. *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.[11] 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 3 establishes that (stochastic) equilibrium transitions are strongly monotonic, 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

[11] There is an analogous result in Roberts (2015) in a non-strategic environment (and without social mobility), and also in Acemoglu, Egorov and Sonin (2015) in a setting without social mobility.

from x to a is possible in equilibrium, then from $y > x$, only transitions to states $a, a+1, \ldots$ 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 starting from a higher state. The result here instead establishes that when we start in a higher state, the lowest state to which there can be a transition is higher than the highest state to which there can be a transition starting from a lower state.

Finally, Part 4 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 towards 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 non-generic preferences/probabilities.[12] 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).[13] One implication of this characterization is that even though there may be mixed strategies, this will not change the direction of transitions, but will just affect its speed.

Note also that the expected stage utility of an agent currently in group j in τ periods if policy p were to be implemented at that point is

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

[12] 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 B3 in 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).

[13] More formally, the genericity notion requires that the parameters, β, the μ'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 this and subsequent propositions hold for all parameters except a subset that is of measure zero.

where μ_{jk}^{τ} denotes the jkth element of M^{τ}, the τth 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). This implies that policy preferences can be equivalently represented by the square of the distance between the policy and the political bliss point of the self in τ periods given by

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

Let us also define $b_j^{(0)} = b_j$ and $b_j^{(\infty)} = \lim_{\tau\to\infty} \left(M^{\tau}\mathbf{b}\right)_j$ (this limit exists by standard properties of stochastic matrices).

Some of our results — specifically, the ones dealing with β close to 1 — are easiest to formulate under the following assumption. We will specifically note when we impose this assumption.

Assumption 2 ***(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 $\mu_{jd_{s_j}}^{(\tau)} > 0$ for some $\tau > 0$.*

This assumption states that every social group has a positive probability of becoming pivotal starting in its ideal state (i.e., the state with induced policy choice maximizing the stage payoff of individuals in this social group). 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).

Theorem 2 ***(Very myopic or very patient players)***

1. *There exists $\beta_0 > 0$ such that for any $\beta \in (0, \beta_0)$, the equilibrium is such that if in period t the state is s, then the state in period $t+1$ is $z \in S$ that minimizes $\left|b_{d_z} - b_{d_s}^{(1)}\right|$. In other words, if agents are sufficiently myopic, then society immediately moves to a state where the resulting policy is closest to the bliss point of tomorrow's self of the current pivotal group, $b_{d_s}^{(1)}$.*

2. *Suppose in addition that Assumption 2 holds. There exists $\tilde{\beta} < 1$ such that for any $\beta \in (\tilde{\beta}, 1)$ there is an equilibrium such that if in period t the state is s, then the sequence of states along the equilibrium path $s_{t+1}, s_{t+2}, \ldots$ will converge, with probability 1, to state z that minimizes $\left|b_{d_z} - b_{d_s}^{(\infty)}\right|$.*

The first result is straightforward: sufficiently myopic players in the pivotal group will choose the political institution that maximizes the welfare of their immediate future selves. In fact, in this case, it can also be shown that the equilibrium is generically in pure strategies (where genericity is to rule out the cases in which tomorrow's ideal point is exactly half way between the policies that will follow from two adjacent states).

The second result is more subtle and already starts illustrating some of the reasoning that will play an important role in the rest of our analysis: if β is high, agents are patient and are willing to act in a way that will eventually lead to a state where the utilities of their distant future selves are maximized. Thus, if the equilibrium evolution did not take society to such a state, then the current decision-maker would have an incentive to move there immediately. Intuitively, when the discount factor is sufficiently large, agents care about the preferences of their current and near-future selves only inasmuch as this does not conflict with the preferences of their distant future selves. To complete the argument, one needs to show that the state z that minimizes $\left|b_{d_z} - b_{d_s}^{(\infty)}\right|$ is stable, so once the society gets there, it stays there forever. Though the mathematical argument is more involved, the intuition for this result is straightforward: in the long run, the distributions of future selves of individuals from groups d_s and d_z are the same, and therefore their interests are aligned. So decision-makers from group d_z prefer to maintain state z, which is exactly what group d_s, from the vantage point of the beginning of the game, wishes to achieve in the long run. Notice also that Theorem 2 does not imply immediate transition to the long-run stable state even when β is very close to 1 because current decision-makers might still prefer to spend the next several periods in the current state.

3.2 Multiplicity and Uniqueness

The same economic forces that lead to equilibria in mixed strategies also open the way to multiplicity as the next example demonstrates. The key feature of the example, responsible for multiplicity, is the presence of different aspects of social mobility that take place at different speeds. In the next example, there is 'fast social mobility,' meaning a high likelihood of the members of the middle class to move up, which will make them have preferences similar to the rich in the near future, but also 'slow social mobility,' meaning a lower but still positive probability for them to move down, which implies that their preferences in the very far future will coincide with those of the current poor.

Example 4 Consider an environment as in Example 2, but with the following changes: first, the discount factor, β, can take any value; and second, in each period, r members of the middle class become rich and an equal number of rich become middle class, while also r' other members of the middle class become poor, while an equal number of poor become middle class. Assume that $r = \frac{1}{8}n$ and $r' = \frac{1}{50}n$ (where n is any number divisible by 200). Notably, r is much larger than r', which will imply that for the middle class, mobility upwards is considerably faster than mobility downwards, though because they are part mixing with the middle class and the poor have the same preferences in the very distant future.

Current members of the middle class prefer policy 0 today, but for tomorrow, their bliss point is given by $\frac{5}{8} \times 1 + \frac{1}{10} \times (-1) = \frac{21}{40}$. Consequently, these individuals would prefer the rich to rule in the next period, which is a consequence of the fast mobility upwards. But in subsequent periods, because of slow mobility downwards, they again prefer democracy: for example, in the period after

next, their bliss point is $\left(\frac{5}{8} \times \frac{11}{16} + \frac{11}{40} \times \frac{5}{8}\right) \times 1 + \left(\frac{1}{10} \times \frac{19}{20} + \frac{11}{40} \times \frac{1}{10}\right) \times (-1) = \frac{1533}{3200} < \frac{1}{2}$. In fact, thereafter their bliss points decline monotonically towards zero, which is the bliss point of the very distant future self of all agents (computed as the weighted average of the bliss points of different social groups in the stationary distribution).

It can be verified that for $\beta > 0.373$, there is an equilibrium in which democracy is stable. In this equilibrium, the middle class can resist the temptation to transfer power to the rich, because this would be beneficial for only one period, and when $\beta > 0.373$, this is not sufficient to compensate for the lower utility thereafter. At the same time, for all $\beta \in (0, 1)$, and thus a fortiori for $\beta > 0.373$, there is an equilibrium in which democracy is unstable, and where the society immediately transitions to elite dictatorship and stays there forever. Intuitively, when the transition to elite dictatorship tomorrow is expected, the current middle class know that their transition decision affects the utility of their tomorrow's selves, but not the utility of their more distant selves, who will find themselves in elite dictatorship even if the middle class stays in democracy for the next period. Since only the utility of tomorrow's self is at stake, and this self prefers elite dictatorship to democracy, moving away from democracy is indeed a best response by the middle class. (In addition, when $\beta > 0.373$, there is also a third equilibrium involving mixing).

Notice that the multiplicity illustrated in this example is not just a multiplicity of equilibrium strategies but of induced equilibrium paths. The economic intuition comes from the interplay between the current decision-maker's strategies and her expectation of future behavior by both those who in the future will be in the same social group as herself and those in other social groups. Though this type of multiplicity can occur whenever there is social mobility at different speeds, a straightforward (though not necessarily weak) assumption is sufficient to rule it out. We next present this assumption, which will be imposed in some of our results to ensure uniqueness.

Assumption 3 ***(Within-person monotonicity)*** *For any social group k, the sequence $b_k^{(0)}, b_k^{(1)}, b_k^{(2)}, \ldots$ is monotone, meaning that either $b_k^{(\tau)} \geq b_k^{(\tau+1)}$ for $\tau = 0, 1, ...$ or $b_k^{(\tau)} \leq b_k^{(\tau+1)}$ for $\tau = 0, 1, ...$*

To understand the implications of within-person monotonicity, let us revisit the reasoning of the current decision-maker. This agent, by choosing the state tomorrow, is indirectly deciding the sequence of states at all future dates. Imagine a situation in which she expects her preferences to first move to the right and then to the left (thus violating within-person monotonicity). In this case, she might be happy to stay in the original state in order to balance the interests of all future selves. However, if she expected future decision-makers to move right in the next period, she would prefer to do so immediately, because tomorrow's self is the only one that benefits from such a move. This paves the way for multiplicity. If, on the other hand, the within-person monotonicity condition is satisfied, this sort of multiplicity is not possible: her tomorrow's self wishes a move to the right more

than her current self, and if future decision-makers are more likely to move to the right, the current self, who tries to balance the interest of all future selves, becomes less likely to move the right.

The intuition that within-person monotonicity should ensure uniqueness (in the sense of uniqueness of equilibrium paths) is confirmed by the next theorem.

Theorem 3 ***(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 β is sufficiently low, or (ii) Assumption 3 (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.

It is also worth recalling from Example 4 that in the absence of within-person monotonicity, multiplicity of equilibria does not disappear even as β approaches 1. The reason is that even if the current and long-run selves have similar preferences, they still need to coordinate so that the pivotal voter at each point chooses policies in line with their long-run preferences, not their short-run incentives. As β approaches 1, these short-run incentives become less and less important, but the coordination problem does not vanish.

The within-person monotonicity condition and its role in uniqueness can be understood alternatively as an instance of aggregation of heterogeneous preferences — in particular, the preferences of all future selves. Consider the problem of a 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; moreover, the weights of all future selves are the same across individuals. 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 β 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.

3.3 Farsighted stability of institutions

If agents are sufficiently farsighted, Theorem 2 yields two corollaries, which are both interesting in their own right and will also be crucial for the rest of our analysis (even though this analysis will be for the case in which β takes an arbitrary value).

Theorem 2 implies that when β is very high, the preferences of very distant future selves $\mathbf{b}^{(\infty)}$ play a key role. These distant preferences are straightforward to compute. Let us 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., $L_M(j_1) \cap L_M(j_2) \neq \varnothing$ and $L_M(j_1) \cup L_M(j_2) \cup \ldots = G$. 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:

$$b_j^{(\infty)} = \frac{\sum_{k \in L_M(j)} n_k b_k}{\sum_{k \in L_M(j)} n_k}. \tag{8}$$

The next several results will be stated under the assumption that the equilibrium is unique. Since we will also adopt Assumption 3, Theorem 3 already ensures generic uniqueness. We impose equilibrium uniqueness as an additional assumption both for emphasis and to avoid further reference to generic parameter values.

Corollary 1 ***(Farsighted stability of institutions)*** *Suppose that Assumptions 2 and 3 hold and the equilibrium is unique. Then state $s \in S$ is stable for sufficiently high β (formally, there exists $\tilde{\beta} < 1$ such that for any $\beta \in (\tilde{\beta}, 1)$, $q_{ss} = 1$) if and only if*

$$s \in \arg\min_{z \in S} \left| b_{d_z} - b_{d_s}^{(\infty)} \right|. \tag{9}$$

This result states that when players are sufficiently farsighted, 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 which the current decisions can move to than the policy choice that will result in any other state. Applying this result to democracy, we can conclude that democracy is stable if and only if the

median voter's long-run future self would still prefer democracy over any other institution — i.e., if his political bliss point lies closer to the policy that the median voter will choose under democracy than to the policy that the decisive voter under any other institution would choose. Given single-peakedness (and symmetry) of preferences, it is sufficient to compare policies under democracy and under the two neighboring institutions. More precisely, we have the following corollary:[14]

Corollary 2 ***(Farsighted stability of democracy)*** *Suppose that Assumptions 2 and 3 hold and the equilibrium is unique. Denote democracy by x. Then democracy is stable for sufficiently high β if and only if*

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

This corollary, which follows directly from Corollary 1, provides a simple, and as it will turn out powerful, characterization of the stability of democracy when the discount factor, β, is sufficiently close to 1. Intuitively, it requires 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.[15] When this is the case, the current median voter prefers to delegate future decisions to future median voters (in democracy). When it is not, he would like to empower a group other than the one containing the median voter, which implies a deviation from democracy. We will see in the next section that this condition not only determines whether or not democracy is stable for high values of the discount factor β, but also shapes the comparative statics of democracy with respect to the speed of social mobility (for any value of β).

A complementary interpretation of conditions (9) and (10) further clarifies the intuition. Note from (8) 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 each group to every other group), $b_{d_x}^{(\infty)}$ is simply the average bliss point in society, so the condition that $x \in \arg\min_{z \in S} \left| b_{d_x} - b_{d_x}^{(\infty)} \right|$ requires *median preferences*, b_{d_x}, which are those which will be implemented by democracy, to be sufficiently close to these *average preferences*, $b_{d_x}^{(\infty)}$.

4 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 simplify the exposition by assuming within-person monotonicity, relegating the results that relax this assumption to the next section. Moreover, given our focus in this section, we fix all other parameters of the model, and only vary the matrix of social mobility.

[14] To formally cover the cases in which the political institutions are the lowest and highest feasible ones, i.e., 1 and m respectively, in what follows we set $b_{d_0} = -\infty$ and $b_{d_{m+1}} = +\infty$, which ensures that for these lowest and highest political institutions, condition (10) is only relevant on one side.

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

Definition 2 ***(Comparing the speed of social mobility)*** *Suppose we have two matrices 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 matrices 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 5 The simplest example of a collection of matrices 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 matrices 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 condition (10) introduced in Corollary 1.

Theorem 4 ***(When social mobility increases the stability of democracy)*** *Suppose that Assumption 3 holds and the equilibrium is unique. Suppose also that social mobility under M' is faster than under M, and inequality* (10) *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'.*[16]

[16] The following stronger version of this result is also proved at the end of the proof of Theorem 4: let q_{sz} be the

In the case where (10) holds, this theorem thus supports De Tocqueville's hypothesis that social mobility contributes to the stability of democracy. The intuition for this result is instructive about the workings of our model. We know from Corollary 1 that (10) holds, democracy is stable for β sufficiently close to 1, because the long-run preferences of the current median voter are close to the preferences of the median voter in the very far future. This does not guarantee stability for β significantly less than 1, however, because the current median voter may benefit sufficiently from shifting political power in the near future future to another social group. In this situation, faster social mobility makes 'time run faster', making the preferences of all future selves closer to $\mathbf{b}^{(\infty)}$. Put differently, with faster social mobility, individuals put less weight on events in the very near future because the very near future itself becomes more transient, and consequently, their preferences become more aligned with those of their distance selves. 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 β, the preferences of all future selves of all social groups approach the preferences of their very distant self, and because the preferences of the very distant self are the same for all groups (within the component), the preferences of all social groups approach each other as well. Since, from condition (10), the very distant self of the current decision-maker prefers 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, and thus implies that asymptotic stability under M translates into asymptotic stability under M' (and once again, the converse not being true).

What if (10) does not hold? In this case, the current median voter expects that her future selves in the very distant future will prefer another state. When the discount factor, β, is not too close to 1, this does not necessarily imply that she would want to go to this state immediately, and democracy may still be stable. 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 Assumption 3 holds, the equilibrium is unique and that social mobility under M' is faster than under M. Suppose also that for M, inequality* (10) *does not hold, but we have*

$$\frac{b_{d_{x-2}} + b_{d_{x-1}}}{2} \leq b^{(\infty)}_{d_{x-1}} \leq b^{(\infty)}_{d_{x+1}} \leq \frac{b_{d_{x+1}} + b_{d_{x+2}}}{2}. \tag{11}$$

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 limited to save space.

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 (10) does not hold, and under the additional condition given by (11), faster social mobility has the opposite effect to that maintained by De Tocqueville's hypothesis and to that characterized in Theorem 4: it makes democracy less stable.

The intuition for this result is closely related to that of Theorem 4. When (10) does not hold, democracy is not stable for sufficiently high β, but may still be stable for $\beta < 1$, because the current median voter benefits in the near term from preserving democracy. But then as in Theorem 4, faster social mobility aligns the preferences of the current median voter with her very distant selves, but this may not destabilize and otherwise-stable democracy.

Why does this theorem need condition (11)? 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 other 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 (11), on the other hand, ensures stability of the neighboring states, thus alleviating the slippery slope effect.

5 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.

5.1 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 paper, the proof of this theorem is in 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:*

$$\sum_{t=1}^{\infty} \sum_{k \in G} \beta^t \mu_{d_s k}^t u_k \left(b_{d_s} \right) \geq \sum_{t=1}^{\infty} \sum_{k \in G} \beta^t \mu_{d_s k}^t u_k \left(b_{d_x} \right). \tag{12}$$

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, $\beta \in (\beta_0, \beta_1)$*, (12) need not hold and slippery slope considerations can prevent certain transitions.*

In other words, this result suggests that for both high and low β, 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).[17] 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 β. (See Example B1 in 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 β) 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 β) 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.

[17] 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.

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 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).

5.2 Comparative statics without within-person monotonicity

We stated our main results, Theorems 4 and 5, under the assumption that the equilibrium was unique, which is ensured generically under Assumption 3 (within-person monotonicity). We next provide direct generalizations of these results when neither equilibrium uniqueness nor Assumption 3 is imposed. The substantive and intuitive economic content of these results are essentially identical, but the statements are a little more involved because the language has to be adjusted for possible multiplicity of equilibria.

Theorem 7 ***(Social mobility and stability of democracy without within-person monotonicity I)*** *Suppose that social mobility under M' is faster than under M, and* (10) *holds with strict inequalities. Then democracy is more stable for M' than for M. More precisely, democracy is stable in all equilibria under M and M', and, furthermore, if it is asymptotically stable in every equilibrium under M, then it is asymptotically stable in every equilibrium under M'.*

Theorem 8 ***(Social mobility and stability of democracy without within-person monotonicity II)*** *Suppose that social mobility under M' is faster than under M. Suppose also that for M, inequality* (10) *does not hold, but*

$$\frac{b_{d_{x-2}} + b_{d_{x-1}}}{2} < b^{(\infty)}_{d_{x-1}} \leq b^{(\infty)}_{d_{x+1}} < \frac{b_{d_{x+1}} + b_{d_{x+2}}}{2}.$$

Then democracy is less stable for M' than for M. More precisely, democracy is not asymptotically stable under any equilibrium under M or M', and if it is not stable in every equilibrium under M, then there is no equilibrium under M' where it is stable.

6 Endogenous social mobility

In this extension, we allow the society to choose the speed of social mobility (thus endogenizing the extent of social mobility). We show how political preferences over social mobility are formed, and how this introduces a new set of forces limiting equilibrium social mobility.

To simplify the analysis, we focus on a setting with only three social groups, the poor (P), the middle class (M), and the rich (R), with shares γ_P, γ_M and γ_R, respectively; $\gamma_P + \gamma_M + \gamma_R = 1$. We

also assume that $\gamma_P, \gamma_R < \frac{1}{2}$, so that the median voter belongs to the middle class. Let us denote the states where these groups are decisive by, respectively, l (left), d (democracy), and r (right).

For ease of exposition, we consider two alternative scenarios: social mobility at the bottom (i.e., between P and M while leaving R intact), and social mobility at the top (i.e., between M and R while leaving P intact). These two scenarios can be combined to obtain arbitrary patterns of social mobility in this three-class society, but we do not discuss this hybrid case so as to keep the choice over social mobility single-dimensional and to economize on space.

Let us normalize the preferences of the middle class to $b_M = 0$, and let $b_P < 0$ and $b_R > 0$ be the political bliss points of the poor and the rich, respectively. The constants $\{A_k\}$, which have so far played no major role, will be important because they will parameterize the direct benefits from social mobility (e.g., how important it is for the rich to remain rich rather than transition to the middle class). We normalize $A_M = 0$ and assume that $A_P < -b_P^2$ and $A_R > b_R^2$. These two natural assumptions impose that even when the poor rule, it is better to be in the middle class than the poor, and even when the middle class rule, it is still better to be rich than middle class.

The rest of the section proceeds as follows. In the next subsection we use our characterization results from Section 3 to derive the preferences of the three social groups over social mobility. In the following subsection we allow a one-time choice over social mobility and derive our main results on the interplay between the evolution of political institutions and endogenous social mobility. In the last subsection we discuss the case in which there are frequent choices over social mobility, corresponding to the decision over the speed of social mobility being made at each date together with the policy and institutional transition decisions.

6.1 Preferences for social mobility

Let us first suppose that the level of social mobility is chosen once at the beginning and remains constant thereafter. Under this assumption, the next two propositions characterize the preferences of the three social groups over the pace of social mobility.

We start with social mobility at the bottom — that is, between the poor and the middle class. Let θ^l be the share of middle class who become poor at the end of each period (accordingly, it is the probability that a given person moves down). Then the probability that a member of the poor moves up to the middle class is $\frac{\gamma_M}{\gamma_P}\theta^l$. The values of θ^l consistent with Assumption 1 are $\theta^l \in \left[0, \theta^l_{max}\right]$, where $\theta^l_{max} = \frac{1}{1+\frac{\gamma_M}{\gamma_P}}$.[18]

Proposition 1 ***(Preferences over mobility at the bottom)*** *If $\gamma_M > \gamma_P$, then a higher θ^l makes the poor better off and the middle class worse off, while the rich are indifferent over the speed of*

[18] When $\theta^l = \theta^l_{max}$ (and similarly, below when $\theta^h = \theta^h_{max}$), Assumption 1 holds only as a weak inequality. None of the results we use here depend on this assumption holding as a strict inequality, or we could specify that $\theta^l \in [0, \theta^l_{max} - \varepsilon]$ for some small ε (and similarly for θ^h), with no impact on our results.

social mobility.

If $\gamma_M < \gamma_P$, then a higher θ^l makes the poor better off and the middle class worse off. The rich become weakly worse off as θ^l increases; strictly worse off if θ^l increases within the interval $\left[\frac{1-\beta}{2-\beta\left(1+\frac{\gamma_M}{\gamma_P}\right)}, \frac{1}{2}\right]$ (because the probability of transitioning from democracy to left dictatorship increases on this interval); and their utility is constant outside of this interval.

The poor always value social mobility at the bottom, both for economic reasons (this enables them to transition to the middle class) and for political reasons (it can lead to institutional change from democracy to left dictatorship when they are more numerous than the middle class). In contrast, the middle class, which stands to transition to a lower social class, dislikes social mobility. The rich are not directly impacted by social mobility as long as democracy remains stable. This stability is guaranteed when $\gamma_M > \gamma_P$, and also holds when $\gamma_M < \gamma_P$ provided that social mobility is not very high. For higher θ^l, which corresponds to faster social mobility, democracy becomes unstable, making way to a left dictatorship. In this case, the rich lose out indirectly from greater social mobility — because it destabilizes democracy in favor of a left dictatorship.

We next turn to social mobility at the top. Let us now denote the share of middle class who become rich by θ^h, which then implies that the share of the rich that move to the middle class is $\frac{\gamma_M}{\gamma_R}\theta^h$. In this case, the values of θ^h consistent with Assumption 1 are $\theta^h \in \left[0, \theta^h_{max}\right]$, where $\theta^h_{max} = \frac{1}{1+\frac{\gamma_M}{\gamma_R}}$.

Proposition 2 ***(Preferences mobility at the top)*** *If $\gamma_M > \gamma_R$, then a higher θ^h makes the middle class better off and the rich worse off, while the poor are indifferent.*

If $\gamma_M < \gamma_R$, then a higher θ^h makes the middle class better off, and the utility of the poor is monotonically decreasing on the interval $\left[\frac{1-\beta}{2-\beta\left(1+\frac{\gamma_M}{\gamma_R}\right)}, \frac{1}{2}\right]$ (because the probability of transition from democracy to elite dictatorship on this interval) and is constant ouside of this interval. The utility of the rich is monotonically decreasing in θ^h if $\beta\frac{\gamma_M}{\gamma_R}\left(\frac{A_R}{b_R^2}+1\right) \geq 4$, and is nonmonotone in θ otherwise.[19]

Now conversely, the poor do not directly care about social mobility at the top, and they will not oppose it as long as it does not have institutional consequences. But they do so indirectly, because if it makes democracy less stable in favor of elites dictatorship, it makes the poor worse off. In

[19] More precisely, the utility of the rich is increasing on the interval $\left[\frac{1-\beta}{2-\beta\left(1+\frac{\gamma_M}{\gamma_R}\right)}, \min\left((1-\beta)\left(\sqrt{\beta\frac{\gamma_M}{\gamma_R}\left(\frac{A_R}{b_R^2}+1\right)}-\beta\left(1+\frac{\gamma_M}{\gamma_R}\right)\right)^{-1}, \frac{1}{2}\right)\right]$, which is in this case nonempty. At the lower end of the interval, the middle class are indifferent between staying in democracy forever and transiting to elite dictatorship. The interval may extend all the way up to $\frac{1}{2}$, where transition to the rich class becomes immediate, or to an interior point, in which case the rich do not benefit from a faster transition as it is achieved by excessively rapid social mobility.

contrast, social mobility at the top always benefits the middle class, but poses a trade-off for the rich: on the one hand, they may move to the middle class, which will make them worse off; on the other hand, the middle class may change the political institutions in their favor. Proposition 2 describes how this trade-off is resolved: a marginal increase in the pace of social mobility is favored only if it affects the probability of transition away from democracy, and within that range, it is more likely to have an impact for smaller θ^h. The rich are more likely to benefit from social mobility if inequality between M and R, as captured by A_R, is small. This is because, with limited inequality, they do not get much extra benefit from being rich in a world with middle class policies, but would benefit considerably from institutional change. If, in contrast, inequality is high, it is more important for the rich to stay rich than to secure a transition to elite dictatorship; hence a lower b_R, which corresponds to less conflict of interest between the middle class and the rich, decreases the chance that the rich will benefit from mobility. The rich are also more likely to benefit from mobility when γ_M/γ_R is small, because in this case rich agents are expected to remain rich longer even with more rapid social mobility.

6.2 Collective decisions over social mobility

We next turn to collective choices over social mobility. Suppose first that social mobility is decided once at the beginning of the game and society starts in democracy. Intuitively, this corresponds to the case in which social mobility choices are made much less frequently than decisions over political transitions, for example, because social mobility is primarily affected by the educational system, which can only be changed infrequently or has a slow-acting impact.

Formally, the game form from our main analysis is now augmented with a stage $t = 0$ where a (constant) level of social mobility is chosen for the rest of the game. We also specify, for completeness, the default level of social mobility θ_0, though as in our main analysis, this default does not impact equilibrium outcomes.

0. *Social mobility decision:*

   (a) The status-quo option is to preserve the default social mobility, $\theta^0 = \theta_0$, and we start with $j = 1$.

   (b) A random agent i from group $\pi_{s_t}(j)$ is chosen as the agenda setter, and proposes social mobility $\tilde{\theta}^j \in [0, \theta_{max}]$.

   (c) All individuals vote, sequentially, with each agent i casting vote $v_i^\theta(j) \in \{Y, N, A\}$.

   (d) If $\sum_{i=1}^n w_{g_i^t}(d)\,\mathbf{1}\left\{v_i^\theta(j) = Y\right\} > \sum_{i=1}^n w_{g_i^t}(d)\,\mathbf{1}\left\{v_i^\theta(j) = N\right\}$, then the current proposal becomes the new default ($\theta^j = \tilde{\theta}^j$); otherwise the default stays the same ($\theta^j = \theta^{j-1}$). The game returns back to stage 0(b) with j increased by 1, unless $j = g$.

(e) The social mobility decided in the last stage is implemented: $\theta = \theta^g$, and the game moves to stage 1 (described in Section 2)

Observe that social mobility stage is essentially identical in structure to that for policy and political decisions. Furthermore, to avoid uninteresting multiplicities, we assume that agents who are indifferent between supporting and opposing a proposal choose to abstain. This assumption ensures that pairwise plurality — meaning that one of the two groups that directly care about the type of social mobility under consideration is larger than the other — produces a unique winner. We start with endogenous social mobility at the bottom.

Proposition 3 ***(Endogenous social mobility at the bottom)*** *If $\gamma_M > \gamma_P$, the unique equilibrium choice of social mobility at the bottom $\hat{\theta}^l$ is 0. If $\gamma_M < \gamma_P$, then the unique equilibrium choice of social mobility at the bottom is $\hat{\theta}^l = \frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta}$, which is decreasing in β and increasing in the relative size of the middle class, $\frac{\gamma_M}{\gamma_P}$. In either case, democracy is stable.*

Here, the middle class and the poor are in direct conflict: the former want less social mobility, and the latter want more. As long as the rich are indifferent, the larger of the two groups will be able to implement their preferred policy. The rich, in turn, are indifferent between any social mobility as long as this level does not induce democracy to transition to left dictatorship. Thus, if the middle class is more numerous than the poor, they will be able to impose no social mobility, and when the poor are more numerous, they will be able to choose higher levels of social mobility, but only up to $\theta = \frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta}$ — the point where the middle class are indifferent between preserving democracy and abandoning it. The poor would not be able to go beyond this level because the rich would now start caring about the level of social mobility and vote against proposals increasing it beyond this level. Put differently, a coalition of middle class and rich would stop increases in social mobility beyond this level. This result thus highlights how concerns about the interplay between social mobility and the stability of democracy can act as a powerful force constraining the pace of social mobility.

The comparative statics in Proposition 3 follow from this observation. When $\gamma_M < \gamma_P$, social mobility is decreasing in β because the middle class is more likely to abandon democracy when they are more forward-looking, and this reduces the maximum threshold of social mobility that keeps democracy stable. The comparative statics with respect to $\frac{\gamma_M}{\gamma_P}$ are also intuitive: when $\gamma_M < \gamma_P$, a larger size of the middle class relative to the poor means that current middle class members would spend comparatively less time being poor for any given θ, making them less willing to abandon democracy. This then reduces the threshold of social mobility that keeps democracy stable. This result also implies that both a very large and a very small middle class is bad for social mobility at the bottom, which is greatest when the middle class is sufficiently large to value social mobility in the long run, but not too powerful to be able to stop it unilaterally.

We next turn to the mobility at the top. One new result in this case is that peripheral coalitions (between the rich and the poor) can form because both the rich and the poor may be opposed to high social mobility — for the rich, because of its direct costs, and for the poor because of its indirect cost in terms of its impact on stability of democracy.

Proposition 4 ***(Endogenous social mobility at the top)*** *If $\gamma_M > \gamma_R$, then the equilibrium choice of social mobility at the top is $\hat{\theta}^h = \theta^h_{max}$. If $\gamma_M < \gamma_R$ and*

$$\frac{A_R}{b_R^2} > (2-\beta)\left(\frac{\gamma_R}{\gamma_M} - 1\right) + 1 \tag{13}$$

holds, then $\hat{\theta}^h = 0$, and if (13) does not hold, then $\hat{\theta}^h = \frac{1}{2}$. In the first two cases, democracy is stable, whereas in the third case it is immediately abandoned in favor of elite dictatorship. Condition (13) is satisfied for a larger range of parameters if A_R is high, b_R is low, β is high, or $\frac{\gamma_M}{\gamma_R}$ is high.

When $\gamma_M > \gamma_R$, so that the middle class is more numerous than the rich, maximal social mobility (consistent with Assumption 1) will arise, and because the preferences of the middle class in this case are still sufficiently to the left of the rich, democracy remains stable. The situation changes dramatically, however, when $\gamma_M < \gamma_R$. In this case, so long as the poor are not opposed to their preferences, the rich can dictate its level. When (13) holds, the level of inequality between the rich and the middle class is relatively high,[20] and the rich prefer having no social mobility to inducing a transition to elite dictatorship (and because $\gamma_M < \gamma_R$, sufficiently high social mobility, in particular anything above $\theta = \frac{1}{2}$, induces the middle class to abandon democracy). This case can be viewed as a rich-poor coalition against the middle class (for had the poor supported the middle class, there would be positive social mobility). If, on the other hand, (13) does not hold, the rich prefer a transition to elite dictatorship to staying in democracy with zero social mobility. They then enter into a middle class-rich coalition in favor of high social mobility, but one that also makes the middle class abandon democracy.

Summarizing our results, we have seen that with social mobility at the bottom, democracy remains stable, because there will always be a middle class-rich coalition limiting social mobility below the level at which democracy might be endangered. In contrast, with social mobility at the top, democracy may collapse in favor of elite dictatorship when the middle class is sufficiently small, when inequality at the top is low, and when there is substantial conflict of interest between the rich and the middle class (b_R high). These results confirm the long-held hypothesis that a larger middle class is generally good for the stability of democracy.

[20] In addition to A_R, the level of inequality between the middle class and the rich, note that (13) depends on b_R^2, since this parameter captures the extent of conflict of interest between the middle class and the rich over policy.

6.3 Joint dynamics of institutions and social mobility

We have so far focused on the case in which social mobility is chosen only once. As already mentioned, this case corresponds to an environment in which decisions over the determinants of social mobility are made relatively infrequently (compared to other political decisions). The alternative, where both types of decisions are made at the same frequency, leads to a setup where social mobility will be chosen in each period at the same time as the decisions about next period's political state. A full analysis of this case is challenging, as it requires the treatment of political decisions on two dimensions, institutional transitions and social mobility, one of which does not even satisfy a natural single-crossing property. It is thus beyond the scope of the current paper. Nevertheless, some basic conclusions can be derived, and we now discuss them briefly.

As in the case with only a single decision over social mobility, in this environment elite dictatorship is stable, and left-wing dictatorship is stable whenever it is relevant (whenever the middle class would consider transitioning to it). Hence, the main question is when a democracy is stable and what level of social mobility will be chosen in this political regime. When we consider mobility at the bottom, the conclusions are similar to those of Proposition 3, in that democracy is always stable, and when it is more numerous than the poor, the middle class restrict social mobility because it is not in their interest. When the poor are more numerous than the middle class, as we have already seen, high levels of social mobility can trigger a transition to left dictatorship, and this motivates a coalition between the middle class and the rich to restrict social mobility. The only difference in this case is that this now permits a greater equilibrium level of social mobility. This is because the middle class are less keen on a transition to left dictatorship as they anticipate that such a transition will further increase social mobility once the poor become the decision-maker (since higher social mobility is always in the interest of the poor), and this makes them more willing to preserve democracy, and increases the level of social mobility that the rich are willing to tolerate (for the only reason why the rich oppose social mobility at the bottom is to prevent a transition to left dictatorship).

Likewise, when we consider mobility at the top, the conclusions are similar to those of Proposition 4, modified mainly because the middle class now expect a slower pace of social mobility following a transition to elite dictatorship. This translates into a greater level of social mobility at which the middle class would be willing to abandon democracy, and as in Proposition 3, this happens when the middle class is sufficiently small (γ_M/γ_R low); when inequality at the top is low (A_R low); and when there is substantial conflict of interest between the rich and the middle class (b_R high). Consequently, as in the previous subsection, there is a possible instability of democracy, though only at even higher levels of social mobility. Put differently, the instability of democracy no longer comes hand-in-hand with high levels of social mobility, but with a high level of social mobility at first, followed with a sharp decline in mobility once the rich come to power.

Summing up, with frequent decisions over social mobility, as with the case of infrequent decisions,

democracy remains stable when social mobility is at the bottom, but a middle class-rich coalition limits the extent of this type of social mobility. Instead, with social mobility at the top, democracy might make way to elite dictatorship, because the middle class may support a high level of social mobility but also wish to abandon democracy, particularly when it is small, when inequality at the top is low, and when there is substantial conflict of interest between the rich and the middle class. With both frequent and infrequent decisions, the interplay between social mobility and the stability of democracy is an additional force constraining the equilibrium level of social mobility.

7 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 paper 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 why 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 to have 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 slopes 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.

Finally, we also showed that when social mobility is endogenized (albeit in a simpler version of our model with only three social groups), our model identifies new political economic forces limiting the extent of mobility. First, the middle class, which tends to be pivotal, is generally opposed to social mobility at the bottom (between the middle class and the poor). Second, a peripheral coalition between the rich and the poor may emerge to limit social mobility at the top, because the rich dislike this type of mobility while the poor are wary that very high levels of this type of mobility may destabilize democracy in favor of elite dictatorship.

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, though we provided a first attempt at endogenizing social mobility, there is much more that can be done to study the interplay of endogenous social mobility and the impact of social mobility on political dynamics, for example by considering several groups, fully studying the case with frequent decisions over the speed of social mobility, and introducing multiple policy levers impacting social mobility. Third, this 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 impacts 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.

References

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin (2010) "Political Selection and Persistence of Bad Governments," *Quarterly Journal of Economics*, 125 (4): 1511-1575.

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

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin (2015) "Political Economy in a Changing World," *Journal of Political Economy*, 123(5): 1038-1086.

Acemoglu, Daron, and James Robinson (2000) "Why Did The West Extend The Franchise? Democracy, Inequality, and Growth In Historical Perspective," *Quarterly Journal of Economics*, 115(4): 1167-1199.

Acemoglu, Daron, and James Robinson (2001) "A Theory of Political Transitions," *American Economic Review*, 91: 938-963.

Alesina, Alberto and Edward Glaeser (2004) *Fighting Poverty in the US and Europe: A World of Difference*, Oxford University Press, Oxford UK.

Alesina, Alberto and Paola Giuliano (2010) "Preferences for Redistribution" in Jess Benhabib, Alberto Bisin, Matthew O. Jackson (eds.): *Handbook of Social Economics*, Vol. 1A, Netherlands: North-Holland: 93-131.

Amador, Manuel (2003) Essays in Macroeconomics and Political Economy, PhD thesis, MIT.

Austen-Smith, David, and Jeffrey S. Banks (2005) *Positive Political Theory II: Strategy and Structure.* Ann Arbor: U. Michigan Press.

Besley, Timothy and Stephen Coate (1998) "Sources of Inefficiency in a Representative Democracy: A Dynamic Analysis," *American Economic Review,* 88(1), 139-56.

Benabou, Roland, and Efe Ok (2001) "Social Mobility and the Demand for Redistribution: The POUM Hypothesis," *Quarterly Journal of Economics*, 116(2), 447-487.

Benabou, Roland and Jean Tirole (2006) "Belief in a Just World and Redistributive Politics," *Quarterly Journal of Economics*, 121(2), 699-746.

Bisin, Alberto, Alessandro Lizzeri, and Leeat Yariv (2015) "Government Policy with Time-Inconsistent Voters," *American Economic Review*, 105(6): 1711-1737.

Blau, Peter, and Otis Duncan (1967) *The American Occupational Structure*, New York: Wiley.

Bourguignon, Francois and Thierry Verdier (2000) "Oligarchy, democracy, inequality and growth," *Journal of Development Economics,* 62(2), 285-313.

De Tocqueville, Alexis (1835). *De la démocratie en Amérique*, Paris: Librairie de Charles Gosselin: English translation, 2000, *Democracy in America*. Chicago: U. Chicago Press.

Erikson, Robert, and John Goldthorpe (1992) *The Constant Flux: A Study of Class Mobility in Industrial Societies,* Oxford: Clarendon Press.

Fearon, James (1995) "Rationalist Explanations for War," *International Organization,* 49(3): 379-414.

Gomes, Armando, and Philippe Jehiel (2005) "Dynamic Processes of Social and Economic Interactions: On the Persistence of Inefficiencies", *Journal of Political Economy*, 113(3), 626-667.

Gul, Faruk and Wolfgang Pesendorfer (2004) "Self Control, Revealed Preferences and Consumption Choice," *Review of Economic Dynamics,* 7(2): 243-264.

Hungerford, Thomas (1974) *Algebra*, New York: Springer.

Jackson, Matthew and Leeat Yariv (2015) Collective Dynamic Choice: The Necessity of Time Inconsistency, *American Economic Journal: Microeconomics*, forthcoming.

Jehiel, Philippe and Suzanne Scotchmer (2001) "Constitutional Rules of Exclusion in Jurisdiction Formation." *Review of Economic Studies*, 68: 393-413.

Lagunoff, Roger (2006) "Markov Equilibrium in Models of Dynamic Endogenous Political Institutions," Georgetown, mimeo.

Lipset, Seymour, and Reinhard Bendix (1959) *Social Mobility in Industrial Society*, Berkeley: U. California Press, 1959).

Lipset, Seymour (1960) *Political Man: The Social Bases of Politics*, Garden City, New York: Anchor Books.

Lizzeri, Alessandro, and Nicola Persico (2004) "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(2): 705-763.

Marshall, Albert W., Ingram Olkin, and Barry C. Arnold (2011). *Inequalities: Theory of Majorization and Its Applications*, Springer, New York.

Moore, Barrington (1966) *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*, Beacon Press, Boston.

Pareto, Vilfredo (1935) *The Mind and Society*, New York: Harcourt, Brace and Company.

Piketty, Thomas, "Social Mobility and Redistributive Politics," *Quarterly Journal of Economics*, 110, 551–583.

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

Sombart, Werner (1906) *Warum gibt es in den Vereinigten Staaten keinen Sozialismus?* Tübingen: English: *Why is there No Socialism in the United States?* New York: Sharpe, 1976.

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

Wright, Randall (1986) "The Redistributive Roles of Unemployment Insurance and the Dynamics of Voting," *Journal of Public Economics*, 377-399.

Appendix A: Proofs of Main Results

In Appendix A, we provide proofs of Theorems 1–5, for which we need a number of lemmas. Proofs of lemmas A4–A8 are relegated to 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:

$$\begin{aligned}\bar{U} &= \max_{j\in G}|A_j| + \max_{j,k\in G}(b_k - b_j)^2, \\ \bar{u} &= \min_{j,k\in G: j\neq k}(b_k - b_j)^2.\end{aligned}$$

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: i.e., 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 Φ_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*

$$W_j(x) = \sum_{k\in G} \mu_{jk} V_k(x). \tag{A1}$$

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 $\{\mu_{j\cdot}\}$ is first-order stochastically dominated by $\{\mu_{l\cdot}\}$, 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 matrices* $Q = \{q_{sz}\}_{s,z\in S}$ *are such that for* $x < y$*, the distribution* $q_{x\cdot}$ *is (weakly) first-order stochastically dominated by* $q_{y\cdot}$*. Then* $\left\{V_j'(x)\right\}_{j\in G}^{x\in S}$*, defined by*

$$V_j'(x) = u_j(b_{d_x}) + \beta \sum_{y\in S} q_{xy} \sum_{k\in G} \mu_{jk} V_k(y), \tag{A2}$$

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

$$\left(V_l'(y) - V_l'(x)\right) - \left(V_j'(y) - V_j'(x)\right) \geq 2\bar{u}. \tag{A3}$$

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

$$V_l'(s) - V_j'(s) = \left(u_l(b_{d_s}) - u_j(b_{d_s})\right) + \beta \sum_{z\in S} q_{sz}\left(W_l(z) - W_j(z)\right).$$

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 $q_{x\cdot}$ is (weakly) first-order stochastically dominated by $q_{y\cdot}$, and thus the expectation of $W_l(z) - W_j(z)$ is weakly smaller when evaluated with the former distribution than with the latter, i.e.,

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

We thus have

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

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 will slightly abuse notation $W_j(x, y, z, \ldots)$ to denote the continuation value of group j when the sequence of states is $x, y, z, \ldots$.

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 will 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: (a) for each s, $s \in F_s$ and (b) 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, i.e., we construct a matrix $\hat{Q}$ such that $\hat{q}_{sx} > 0$ only if $x \in F_s$, and also $\hat{q}_{x\cdot}$ weakly first-order stochastically dominates $\hat{q}_{y\cdot}$ whenever $x > y$. Second, we prove that there is an equilibrium σ 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 also that Π is compact and convex.

Consider the following correspondence Υ from Π 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 (A1). For each state $s \in S$, let p_s be the ideal policy of pivotal group d_s, i.e., $p_s = b_{d_s}$, and Ψ_s be the expected utility of the members of pivotal group d_s from transitioning into state s, i.e., $\Psi_s = \arg\max_{x \in F_s} W_{d_s}(x)$. Furthermore, let λ_s be any probability distribution over S the support of which is a subset of Ψ_s, and let Λ_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 $j \in G$,

$$V_j'(s) = u_j(p_s) + \beta \sum_{x \in S} \lambda_s(x) W_j(x). \tag{A4}$$

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 Λ_s is nonempty), and define $V_j'(s)$ as in (A4). 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 (A3). 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 Λ_s be λ_s' and λ_s'', respectively. For any $\alpha \in (0,1)$, $\alpha\lambda_s' + (1-\alpha)\lambda_s''$ is a

probability distribution in Λ_s, and in particular its support is in F_s, and moreover,

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

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 Π to $2^S \setminus \{\varnothing\}$, the set of nonempty subsets of S (and has a closed graph when $2^S \setminus \{\varnothing\}$ is endowed with discrete topology); (ii) a mapping from $2^S \setminus \{\varnothing\}$ to $\Delta(S)$, where each subset $X \in 2^S \setminus \{\varnothing\}$ 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 Π, which is linear and thus continuous. Since 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 Π 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

$$V_j(s) = u_j(b_{d_s}) + \beta \sum_{x \in S} \hat{q}_{sx} \sum_{k \in G} \mu_{jk} V_k(x). \tag{A5}$$

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 λ_x is (weakly) first-order stochastically dominated by λ_y, which means this is also true for $\hat{q}_{x\cdot}$ and $\hat{q}_{y\cdot}$. This proves that both properties of $\hat{Q}$ are satisfied.

We now construct an equilibrium σ 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 solutions to (A5). We construct strategies of the players as follows. Denote the stage where a representative from group d_s makes 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 represenative from d_s chosen to make a proposal randomizes over proposals in Ψ_s and proposes $x \in \Psi_s$ with probability $\hat{q}_{sx} = \lambda_s(x)$ (and makes any other proposal with probability zero), 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 political decision is made, and since 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 Ψ_s are made with these respective probabilities, and are accepted. Finally, in each of the subsequent stages, no alternative from Ψ_s is ever voted down, even by another alternative from Ψ_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 transition to alternative x takes place with probability $\hat{q}_{sx}$.

Denote the resulting profile of strategies σ_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 σ, which prescribes strategies for all players in the original game Γ. By construction, the corresponding transition mapping is $Q(\sigma) = \hat{Q}$, and if profile σ 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, σ is a SPE: by 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 σ_s is a SPE in the game $\Gamma_{s_t, p_{t-1}}$. Thus, σ is a MPE in Γ. Since in the construction of σ_s, the strategies were defined identically for different players in the same group, the MPE is symmetric, and since $\hat{Q}$ is feasible and monotone, these properties are also retained by σ. Thus, σ is an equilibrium with

the desired properties. This completes the proof of existence 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 1. Take symmetric MPE σ. 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 transition, nor does it affect any actions is subsequent periods, because strategies in σ 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 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 $l+1$, 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. Lastly, 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 rejects (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 σ, $p_t = b_{d_s}$ with probability 1, which completes the proof.

Proof of Part 2. Take an equilibrium σ, 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 σ, then V is the unique solution to (A5), and it may be obtained through infinite iteration of mapping (A2), 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 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, $s^g_{t+1} \in \Xi_s$.

Since $p^g_t \in \Xi_s$ in all subgames, we can prove the following statement by backward induction: if at some stage l, $0 \leq l \leq g$, $s^l_{t+1} = \max \Psi_s$ (which also equals $\max \Xi_s$), then $s^g_{t+1} \in \Psi_s$. 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 $s^{l-1}_{t+1} = \max \Psi_s$. Suppose, to obtain a contradiction, that $s^g_{t+1} \notin \Psi_s$ with a positive probability. By induction, this is only possible if $s^l_{t+1} \neq \max \Psi_s$, which, in turn, is only possible if proposal $\tilde{s}^l_{t+1} \neq \Psi_s$ is made and is accepted. However, we showed that subsequent subgame will result in s^g_{t+1} having some distribution with support in Ξ_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^{J-1}_{t+1} (by proposing $\tilde{s}^J_{t+1} = s^{J-1}_{t+1}$, so $s^J_{t+1} = s^{J-1}_{t+1}$) 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 $s^J_{t+1} = \max \Psi_s$ to $s^J_{t+1} = s^{J-1}_{t+1}$. Consequently, $s_{t+1} \in \Psi_s$ with probability 1, which completes the proof.

Proof of Part 3. Take equilibrium σ, and take $x, y \in S$ such that $x < y$. Suppose that $q_{x,a} > 0$ and $q_{y,b} > 0$; by Part 2, 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, $\{W_j(x)\}^{x \in S}_{j \in G}$ 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 4. To prove this part of the theorem, we will 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 $\left(\{b_k\}_{k \in G}, \{A_k\}_{k \in G}, \{\gamma_k\}_{k \in G}, \{\mu_{jk}\}_{j,k \in G}, \beta\right)$ satisfy a nontrivial polynomial equation with rational coefficients (we will 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 zero. (In fact, we will 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 $\tilde{q}_{y\cdot} = q_{y\cdot}$ 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 Λ_x, α_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 $H_{jx}\left(b_{d_y}\right) = \frac{1}{b_{d_y}}\left(W_j(\Lambda_x) - \sum_{\tau=1}^{\infty} \beta^{\tau-1} \mu_{jd_y}^{(\tau)} b_{d_y}^2\right)$ 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 ∞ (more precisely, there is $C, K > 0$ such that $\left|b_{d_y}\right| > K$ implies $\left|H_{jx}\left(b_{d_y}\right)\right| < C$). 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 $W_{d_x}(x, x, x, \ldots)$ 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 Λ_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 $W_{d_x}(x, x, x, \ldots) - \sum_{\tau=1}^{\infty} \beta^{\tau-1} \mu_{d_x d_x}^{(\tau)} b_x^2$ has quadratic terms in b_x, and the coefficient is nonzero because it is polynomial in other parameters, and it cannot be equal to zero if the parameters are generic). Now, continuation values $\{V_j(s)\}_{j \in G}^{s \geq x}$ solve the system of equations (A5) with $q_{\cdot\cdot}$ as linear functions of α_x, which implies that $\{V_j(s)\}_{j \in G}^{s \geq x}$ can be expressed as ratios of polynomials of α_x; then equating $W_{d_x}(\Lambda_x)$ to $W_{d_x}(x)$ results in a nontrivial polynomial of α_x (it is nontrivial, because it holds for some α_x as there is such an equilibrium, but not for some other, say $\alpha_x = 0$, as in that case $W_{d_x}(x) = W_{d_x}(x, x, x, \ldots) \neq W_{d_x}(\Lambda_x)$, as we just proved). Finally, the function $H_{jx}\left(b_{d_y}\right)$ does not depend on α_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 $W_{d_x}(x, x, x, \ldots)$ 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 $W_{d_x}(x,x,x,\ldots) - \sum_{\tau=1}^{\infty} \beta^{\tau-1} \mu_{d_x d_x}^{(\tau)} b_x^2$ 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 α_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, α_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 $y < \Lambda_x$, $H_{jx}\left(b_{d_y}\right)$ is bounded for $\left|b_{d_y}\right|$ large enough. Notice that $H_{jx}\left(b_{d_y}\right)$ 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 both in the numerator and the denominator. It now suffices to prove that the denominator does not tend to 0 as b_{d_y} tends to ∞. Since each α_y satisfies a polynomial equation with coefficient that are polynomials in the parameters of the model, either α_y does not depend on b_{d_y} explicitly, or there is only a finite number of limit points (including ∞) of the solutions to this equation as b_{d_y} tends to ∞. If for at least one of these limit points, the denominator tends to 0, this yields a polynomial equation on α_z for z being the smallest element in A greater than x. This means that there are two polynomial equations on α_z that have a common root, which is only possible if their resultant equals zero, 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 λ_1 and Λ_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 lemma will be used in several proofs. (Proof of this and all subsequent lemmas are relegated to Appendix B.)

Lemma A4 *Suppose that σ 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 σ' 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 $x, y \in S'$.*

The next lemma will be used in the proof of Theorem 2.

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 $Q' : S \to S$ given by*

$$q'_{sa} = \begin{cases} q_{sa} & \text{if } s \neq x, \\ 1 & \text{if } s = x \text{ and } a = y, \\ 0 & \text{if } s = x \text{ and } a \neq y \end{cases} \tag{A6}$$

is monotone.

Proof of Theorem 2. Part 1. Let β_0 be defined by $\beta_0 = \frac{\zeta}{\zeta + 2\bar{U}}$, where

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

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

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

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 2 of Theorem 1.

Part 2. In this proof, let $Z_s = \arg\min_{z\in S} \left|b_{d_z} - b^{(\infty)}_{d_s}\right|$; this set is either a singleton or consists of two adjacent states. The result follows from the following three steps.

Step 1. Denote

$$\begin{aligned} \xi &= \min_{s,y,z\in S, \left|b_{d_y} - b^{(\infty)}_{d_s}\right| > \left|b_{d_z} - b^{(\infty)}_{d_s}\right|} \left(\left(b_{d_y} - b^{(\infty)}_{d_s}\right)^2 - \left(b_{d_z} - b^{(\infty)}_{d_s}\right)^2 \right), \\ \Xi &= b_m - b_1, \end{aligned}$$

and take $\varepsilon = \frac{\xi}{4\Xi}$. For such ε there exists $T \geq 1$ such that for all $s \in S$ and $t > T$, $\left| b_{d_s}^{(t)} - b_{d_s}^{(\infty)} \right| < \varepsilon$. Let $\tilde{\beta} = \left(1 - \frac{\xi}{4\Xi^2}\right)^{1/T}$. 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:

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

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

Step 2. Suppose that β 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 such 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 $\Phi_s \geq s$, for 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 β, 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 in $z \in Z_s$, 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 $b_{d_x}^{(\infty)}$ and 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 estalished 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 $q' : S \to S$ defined by (A6) (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 $W_{d_z}(y) > W_{d_z}(z)$, 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 $W_{d_z}(y) > W_{d_z}(z)$ 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 σ' 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 2 of the Theorem. ■

We next state three more lemmas that will be used in the remaining proofs.

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 $\arg\min_{z \in S} \left| b_j^{(\infty)} - b_{d_z} \right| = \{s\}$. 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 α, 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 α 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 α, $f(\alpha) = 0$, then $f(\alpha') > 0$ for $\alpha' > \alpha$ and $f(\alpha') < 0$ for $\alpha' < \alpha$.

Proof of Theorem 3. Uniqueness when β is sufficiently small is straightforward: consider the sets $A = \left\{ x \in \mathbb{R} \mid x = b_{d_s}^{(1)} \text{ for some } s \in S \right\}$ and $B = \left\{ y \in \mathbb{R} \mid y = \frac{b_{d_s} + b_{d_{s+1}}}{2} \text{ for some } s \in \{1, \ldots, m-1\} \right\}$. For generic parameter values, $A \cap B = \varnothing$. If so, then there is a unique mapping satisfying the description in Part 1 of Theorem 2, and therefore, the equilibrium is generically unique if $\beta < \beta_0$.

We now turn to generic uniqueness under Assumption 2, which will be proved in several steps.
Step 1. Suppose that there are two equilibria σ_1 and σ_2, and let Q^1 and Q^2 be the corresponding transition matrices. Then, for generic parameter values, if $Q^1 \neq Q^2$, then there are two at least states $x, y \in S$, $x \neq y$, such that the distributions $q_{x\cdot}^1 \neq q_{x\cdot}^2$ and $q_{y\cdot}^1 \neq q_{y\cdot}^2$. In other words, it is impossible that transition probabilities from only one state are different.
Proof of Step 1. Suppose not, so there is a unique state s such that $q_{s\cdot}^1 \neq q_{s\cdot}^2$. Let us first prove that, generically, for set $\Omega = \left(\Phi_s^1 \cup \Phi_s^2\right) \setminus \{s\}$, $|\Omega| = 1$. Indeed, if Ω is empty, $\Phi_s^1 = \Phi_s^2 = \{s\}$, hence

$q^1_{s\cdot} = q^2_{s\cdot}$, 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 $x \in \Phi^1_s$. Then by Part 4 of Theorem 1, for generic parameter values, $y \notin \Phi^1_s$, which means that $y \in \Phi^2_s$, which, again by Part 4 of Theorem 1, implies $x \notin \Phi^2_s$ for generic parameter values. Now, consider three possibilities. If $x < s < y$, then, from Part 3 of Theorem 1, $x \in \Phi^1_s$ implies $\Phi^1_z \leq x$ for $z < s$; moreover, for such z, $q^2_{z\cdot} = q^1_{z\cdot}$. Therefore, if society moves from state s to x, the continuation utilities of group d_s should be the same for both equilibria: $W^1_{d_s}(x) = W^2_{d_s}(x)$. Similarly, from $y \in \Phi^2_s$ implies $\Phi^2_z \geq y$ for all $z > s$; moreover, for such z, $q^1_{z\cdot} = q^2_{z\cdot}$. Thus, if the society moves from state s to y, the continuation utilities again coincide: $W^1_{d_s}(y) = W^2_{d_s}(y)$. But by Part 2 of Theorem 1, we have $W^1_{d_s}(x) \geq W^1_{d_s}(y) = W^2_{d_s}(y) \geq W^2_{d_s}(x) = W^1_{d_s}(x)$, which implies that both inequalities hold with equality, in particular, $W^1_{d_s}(x) = W^1_{d_s}(y)$. This means $x, y \in \Psi^1_s$, which, as shown in the proof of Part 4 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 $q^1_{sx} \neq q^2_{sx}$; otherwise, since $\Phi^1_s \subset \{s, x\}$ and $\Phi^2_s \subset \{s, x\}$ we would have $q^1_{ss} = q^2_{ss}$, again meaning that $q^1_{s\cdot} = q^2_{s\cdot}$ and contradicting the choice of s. Without loss of generality, assume $q^1_{sx} < q^2_{sx}$, so in equilibrium σ_1 the society stays in s longer than in equilibrium σ_2, in expectation; this means, in particular, $q^1_{sx} < 1$ and $q^2_{sx} > 0$. It must be that the sequence $b^{(t)}_{d_s}$ is nondecreasing and, moreover, it is nonstationary, for otherwise $q^2_{sx} > 0$ would contradict Lemma A6.

Let $j = d_s$. The continuation utilities from moving to x are the same in both equilibria: $W^1_j(x) = W^2_j(x)$, because the transition probabilities are identical thereafter. Moreover, in equilibrium σ_2, transiting to x is a best response, so $W^2_j(x) \geq W^2_j(s)$, and in equilibrium σ_1, staying is a best response, so $W^1_j(s) \geq W^1_j(x)$. We thus have $W^1_j(s) \geq W^1_j(x) = W^2_j(x) \geq W^2_j(s)$, meaning that the utility of individuals from group j from staying is at least as high under σ_1 as under σ_2. Denote $W_j(s;\alpha)$ the utility of staying in s if the subsequent equilibrium play has probability α of moving to x; then $W_j\left(s; q^1_{sx}\right) = W^1_j(s)$ and $W_j\left(s; q^2_{sx}\right) = W^2_j(s)$. By Lemma A8, the function $f(\alpha) : [0,1] \to \mathbb{R}$, defined by, $f(\alpha) = W_j(s;\alpha) - W_j(x)$, satisfies the strict single-crossing condition. Now, if $f\left(q^1_{sx}\right) = 0$, then $f\left(q^2_{sx}\right) > 0$, meaning that $W_j\left(s; q^2_{sx}\right) > W_j(x)$, which contradicts that moving to x is a best response in σ_2. Similarly, if $f\left(q^2_{sx}\right) = 0$, then $f\left(q^1_{sx}\right) < 0$, meaning that $W_j\left(s; q^1_{sx}\right) < W_j(x)$, which contradicts that staying at s is a best response in σ_1. If $f\left(q^1_{sx}\right) \neq 0$ and $f\left(q^2_{sx}\right) \neq 0$, then since staying in s is a best response in σ_1, we must have $f\left(q^1_{sx}\right) > 0$; similarly, we must have $f\left(q^2_{sx}\right) < 0$. But then by continuity there is $\alpha \in \left(q^1_{sx}, q^2_{sx}\right)$ such that $f(\alpha) = 0$. In that case, it must be that $f\left(q^1_{sx}\right) < 0 < f\left(q^2_{sx}\right)$. 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, σ_1 and σ_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 the transition matrices in equilibria σ_1 and σ_2. Let $Z \subset S$ be the set of $z \in S$ such that $q^1_{z\cdot}$ and $q^2_{z\cdot}$ are different distributions; from Step 1 it follows that $|Z| \geq 2$. In what follows, let $L = \left\{s \in S : \Phi^1_s \leq s, \Phi^2_s \leq s\right\}$ and $R = \left\{s \in S : \Phi_s \geq s, \Phi^2_s \geq s\right\}$. 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 $\Phi^1_s = \Phi^2_s = \{s\}$. 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 that m is 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 $\Phi^1_x = \Phi^2_x = \{x\}$ 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 $q^1_{z\cdot}$ and $q^2_{z\cdot}$ are different is the whole set S. Without loss of generality, suppose $q^1_{11} > q^2_{11}$. Since $q^2_{11} < 1$, $q^2_{12} > 0$, and in a monotone equilibrium we must have $q^2_{22} = 1$; this means $q^1_{22} < 1$, and thus $q^1_{21} > 0$ and again by monotonicity $q^1_{11} = 1$. From Lemma A6, this implies that the sequence $b^{(t)}_{d_1}$ is nondecreasing (because equilibrium σ_2 exists) and $b^{(t)}_{d_2}$ is nonincreasing (because equilibrium σ_1 exists). Suppose $b^{(\infty)}_{d_1} < \frac{b_{d_1}+b_{d_2}}{2}$, then Lemma A7 implies that $q^1_{11} = q^2_{11} = 1$, which contradicts $q^1_{11} > q^2_{11}$. If $b^{(\infty)}_{d_2} > \frac{b_{d_1}+b_{d_2}}{2}$, then we get a similar contradiction. Since $b^{(\infty)}_{d_1} \leq b^{(\infty)}_{d_2}$ by Assumption 1, we must have $b^{(\infty)}_{d_1} = b^{(\infty)}_{d_2} = \frac{b_{d_1}+b_{d_2}}{2}$, which is nongeneric. This proves that under Assumption 3 for generic parameter values, we have a unique equilibrium. ■

Proof of Corollary 1. By Part 2 of Theorem 2, there exists an equilibrium with the desired properties, and since the equilibrium is unique, the result follows. ■

Proof of Corollary 2. Follows immediately from Corollary 1. ■

Proof of Theorem 4. Let us first prove that there is an equilibrium with democracy stable 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. Since from (10) $\frac{b_{d_{x-1}}+b_{d_x}}{2} \leq b_{d_x}^{\infty} \leq \frac{b_{d_x}+b_{d_{x+1}}}{2}$, and also we have $\frac{b_{d_{x-1}}+b_{d_x}}{2} \leq b_{d_x} \leq \frac{b_{d_x}+b_{d_{x+1}}}{2}$, under Assumption 3, we have $\frac{b_{d_{x-1}}+b_{d_x}}{2} \leq b_{d_x}^{(\tau)} \leq \frac{b_{d_x}+b_{d_{x+1}}}{2}$ for any τ. Therefore, there is no deviation that would make group d_x better off. This implies that there is an equilibrium with transition matrix Q, i.e., 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 σ. 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 3 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 2 of Theorem 1. Now for matrix M, taking into account that $b_{d_y}^{(\tau)} \leq b_{d_y}^{\prime(\tau)}$ for every $\tau \geq 0$, single-crossing implies that $W_{d_y}(y, y, \ldots) \geq W_{d_y}(x)$, 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 such two equilibria would contradict Lemma A8 (which is applicable because strict inequality $b_j^{(t)} < b_j^{\prime(t)}$ 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 footnote 16. 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, $q_{xa}^{(t)} > 0$ and $q_{ya}^{(t)} > 0$ are possible for $a \in \{x, y\}$ only. Therefore, we have (using the same calculus as in the proof of step 1 in Theorem 2):

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

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

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

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

$$f(\alpha) = W_j(y) - W_j(x) = -\left(b_{d_x} - b_{d_y} \right) \sum_{t=1}^{\infty} (\beta(1-\alpha))^{t-1} \left(2b_j^{(t)} - b_{d_x} - b_{d_y} \right).$$

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

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

Since we have $b_j'^{(\tau)} \geq b_j^{(\tau)}$ for all τ 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 $f'(a) < 0$. Now, since $q'_{yx} \in (0,1)$, it must satisfy $f'\left(q'_{yx}\right) = 0$, and then by Lemma A8 we must have $q_{yx} = \alpha < q'_{yx}$. Now consider the second case, where $q_{yx} = 1$. By Part 2 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 $f'\left(q'_{yx}\right) < 0$. 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 (10) 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 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. ■

Appendix B: Additional Proofs, Results, and Examples — Not for Publication

B1 Proofs of Lemmas from Appendix A

Proof of Lemma A4. By construction, transition mapping Q' is feasible under S' and F'. Furthermore, the continuation utilities under transition mapping Q', in particular, $W'_j(x, y, z, \ldots)$ for any group j and any path of states $x, y, z, \ldots$ are the same: $W'_j(x, y, z, \ldots) = W_j(x, y, z, \ldots)$. Therefore, for every state $s \in S'$, if $x \in S'$ is such that $q'_{sx} > 0$, then transition to x maximizes the utility of group d_x among all feasible transitions: $x \in \arg\max_{z \in F'_s} W'_{d_s}(z)$; indeed, if for some $y \in F'_s$ we had $W'_{d_s}(y) > W'_{d_s}(x)$, then, since $y \in F'_s$ implies $y \in F_s$, we would have $y \in F_s$, $q_{sx} > 0$, and $W_{d_s}(y) > W_{d_s}(x)$, which cannot be the case if σ is an equilibrium, by Part 2 of Theorem 1. Given that, we can repeat the argument in the proof of Theorem 1 to construct the strategy profile σ' that gives rise to transition mapping Q' and is an equilibrium. This completes the proof. ■

Proof of Lemma A5. Suppose, to obtain a contradiction, that for any $x, y \in S$ such that $W_{d_x}(y) > \tilde{W}_{d_x}$, Q' given by (A6) is not monotone. Take $x, y, a \in S$ such that $|y - a|$ is minimal among all tuples (x, y, a) such that $W_{d_x}(y) > \tilde{W}_{d_x}$ and $a \in \Phi_x$ (informally, we consider the shortest deviation). By our assertion, the corresponding Q' is not monotone. Since Q is monotone and Q and Q' differ by the distribution $Q_{x\cdot}$ and $Q'_{x\cdot}$ only, there are two possibilities: either for some $z < x$ and some $b \in \Phi_z$, $y < b \leq \Phi_x$, or for some $z > x$ and some $b \in \Phi_z$, $\Phi_x \leq b < y$. Assume the former (the latter case may be considered similarly). Let s be defined by

$$s = \min(z \in S : b > y \text{ for some } b \in \Phi_z) = \min(z \in S : \Phi_z \not\leq y);$$

in the case under consideration, the set of such z is non-empty (e.g., x is its member, and z found earlier is one as well), and hence state s is well-defined. We have $s < x$; since Q is monotone, $\Phi_s \leq \Phi_x$.

Notice that if we redefined $\Phi'(s) = \{y\}$, we would get a monotone correspondence (in other words, a deviation by the society in state s to y is monotone) Indeed, there is no state $\tilde{z}$ such that $\tilde{z} < s$ and $y \not\geq \Phi_{\tilde{z}} \leq \Phi_s$ by construction of s, and there is no state $\tilde{z} > s$ such that $\Phi_s \leq \Phi_{\tilde{z}} \not\geq y$ as this would contradict that $y < b$ for some $b \in \Phi_s$ (indeed, the latter would imply $\Phi_{\tilde{z}} > y$). By hypothesis, $W_{d_s}(y) \leq \tilde{W}_{d_s}$, since this deviation cannot be profitable for d_s. Moreover, by the definition of $\tilde{W}_{d_s}$, for any $b \in \Phi_s$ such that $W_{d_s}(b) = \tilde{W}_{d_s}$ and, moreover, there is $b \in \Phi_s$ such that $y < b$ (such b exists by definition of s). Since W satisfies increasing differences and $d_s < d_x$, for this b we have $W_{d_x}(y) < W_{d_x}(b)$.

On the other hand, recall that $W_{d_x}(y) > \tilde{W}_{d_x}$. We therefore have

$$W_{d_x}(b) > W_{d_x}(y) > \tilde{W}_{d_x},$$

so $W_{d_x}(b) > \tilde{W}_{d_x}$. Notice, however, that $y < b \leq a$ for all $a \in \Phi_x$, and strict inequality $W_{d_x}(b) > \tilde{W}_{d_x} = W_{d_x}(a)$ implies $a \neq b$, so in fact $y < b < a$. This implies that $|b - a| < |y - a|$. This

contradicts the choice of $x, y, a \in S$ such that $|y-a|$ is minimal among tuples (x,y,a) such that $W_{d_x}(y) > \tilde{W}_{d_x}$ and $a \in \Phi_x$, as tuple (x,b,a) satisfies the same properties but has $|b-a| < |y-a|$. This contradiction proves that our initial assertion was wrong, and this proves the lemma. ■

Proof of Lemma A6. Suppose that $b_j^{(t)}$ is nondecreasing (the complementary case is considered similarly). Suppose, to obtain a contradiction, that $\Phi_s \not\geq s$. Denote $x = \min \Phi_s$, then $x < s$. Notice that for any $y \in S$, we have

$$\begin{aligned}
\beta W_j(y) &= \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} q_{ya}^{(t)} \sum_{k \in G} \mu_{jk}^{(t)} u_k(b_{d_a}) \\
&= \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} q_{ya}^{(t)} \sum_{k \in G} \mu_{jk}^{(t)} \left(A_k - (b_k - b_{d_a})^2 \right) \\
&= \sum_{t=1}^{\infty} \sum_{k \in G} \beta^t \mu_{jk}^{(t)} A_k - \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} q_{ya}^{(t)} \mu_{jk}^{(t)} (b_k - b_{d_a})^2 .
\end{aligned}$$

Now take any two states $y < z$ and consider the difference $W_j(z) - W_j(y)$:

$$\begin{aligned}
\beta(W_j(z) - W_j(y)) &= \sum_{t=1}^{\infty} \beta^t \left(\sum_{a \in S} \sum_{k \in G} q_{ya}^{(t)} \mu_{jk}^{(t)} (b_k - b_{d_a})^2 - \sum_{a \in S} \sum_{k \in G} q_{za}^{(t)} \mu_{jk}^{(t)} (b_k - b_{d_a})^2 \right) \\
&= \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \left(q_{ya}^{(t)} - q_{za}^{(t)} \right) \sum_{k \in G} \mu_{jk}^{(t)} (b_k - b_{d_a})^2 \\
&= \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \left(q_{ya}^{(t)} - q_{za}^{(t)} \right) \left(\sum_{k \in G} \mu_{jk}^{(t)} b_k^2 - 2 \sum_{k \in G} \mu_{jk}^{(t)} b_k b_{d_a} + \sum_{k \in G} \mu_{jk}^{(t)} b_{d_a}^2 \right) \\
&= \sum_{t=1}^{\infty} \beta^t \left(\left(\sum_{k \in G} \mu_{jk}^{(t)} b_k^2 \sum_{a \in S} \left(q_{ya}^{(t)} - q_{za}^{(t)} \right) \right) + \sum_{a \in S} \left(q_{ya}^{(t)} - q_{za}^{(t)} \right) \left(-2 b_j^{(t)} b_{d_a} + b_{d_a}^2 \right) \right) \\
&= \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \left(q_{ya}^{(t)} - q_{za}^{(t)} \right) \left(-2 b_j^{(t)} b_{d_a} + b_{d_a}^2 \right) \\
&= \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \left(q_{ya}^{(t)} - q_{za}^{(t)} \right) \left(\left(b_j^{(t)} \right)^2 - 2 b_j^{(t)} b_{d_a} + b_{d_a}^2 \right) \\
&= \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \left(q_{ya}^{(t)} - q_{za}^{(t)} \right) \left(b_j^{(t)} - b_{d_a} \right)^2 .
\end{aligned}$$

Applying this to x and s, we have

$$\beta(W_j(x) - W_j(s)) = \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \left(q_{sa}^{(t)} - q_{xa}^{(t)} \right) \left(b_j^{(t)} - b_{d_a} \right)^2 . \tag{B1}$$

Consider two cases. The first case is where $\Phi_s \leq s$; this holds for generic parameter values, as in Part 4 of Theorem 1 (indeed, $x \in \Phi_s$ and $x < s$). In that case, $b_j^{(t)} \geq b_j \geq b_{d_a}$ for all $a \leq s$, so

$b_j^{(t)} \geq b_{d_a}$ and thus $\left(b_j^{(t)} - b_{d_a}\right)^2$ is decreasing in a for $a \leq s$. Consequently, for each t,

$$\sum_{a \leq s} q_{sa}^{(t)} \left(b_j^{(t)} - b_{d_a}\right)^2 \leq \sum_{a \leq s} q_{xa}^{(t)} \left(b_j^{(t)} - b_{d_a}\right)^2,$$

because the distribution $q_{s\cdot}^{(t)}$ first-order stochastically dominates $q_{x\cdot}^{(t)}$ as the equilibrium is monotone. This implies $W_j(x) \leq W_j(s)$. In fact, this inequality is strict. This can be seen for $t = 1$: the probability distributions $q_{s\cdot}^{(1)}$ and $q_{x\cdot}^{(1)}$ are different, and $\left(b_j^{(t)} - b_{d_a}\right)^2$ is strictly increasing in a. Thus, $W_j(s) > W_j(x)$, which contradicts Part 2 of Theorem 1 in that $x \in \Phi_s$ does not maximize the utility of group $j = d_s$. Notice that for the proof in this case, we did not need that $b_j^{(t)}$ is monotone in t, only that $b_j^{(t)} \geq b_j$ for all t.

Now consider the case where for some $y \in \Phi_s$, $y > s$. This case is nongeneric, but the statement holds here as well. Indeed, consider $W_j(s)$; it is a linear combination of paths where the society stays in s for $\tau \geq 1$ periods (including the current period) and then departs either to lower or higher states. All equilibrium paths $\{s_t\}$ where the eventual departure is to lower states (starting from some z such that $x \leq z \leq s$) satisfy $W_j(z \mid \forall t : s_t \leq z) > W_j(x)$, similarly to the previous case. Now consider some path that eventually departs to higher states, and suppose that it stays in z for exactly τ periods, after which it departs to $y > s$. Let us denote the probability distribution of states in period $t \geq 1$ if an immediate transition to x occurs by $p_{x\cdot}^{(t)}$, and that in the case an immediate transition to y occurs by $q_{y\cdot}^{(t)}$; then these are also distributions of states in period $t + \tau$ if transition occurs in period τ. We know that the individuals in group j are indifferent between transiting to x and to y, meaning that

$$\sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} p_{xa}^{(t)} \mu_{jk}^{(t)} (b_k - b_{d_a})^2 = \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} q_{ya}^{(t)} \mu_{jk}^{(t)} (b_k - b_{d_a})^2,$$

which, by increasing differences, implies

$$\sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} p_{xa}^{(t)} \mu_{jk}^{(t+\tau)} (b_k - b_{d_a})^2 \leq \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} q_{ya}^{(t)} \mu_{jk}^{(t+\tau)} (b_k - b_{d_a})^2;$$

this follows from that $b_j^{(t+\tau)} \geq b_j^{(t)}$ for each t (as earlier in the proof, only the expectations of $\mu_{j\cdot}^{(t+\tau)}$ matter). Now we have

$$\begin{aligned}
\beta W_j \left(\underbrace{s, \ldots, s}_{\tau \text{ times}}, y, \ldots\right) &= \sum_{t=1}^{\tau} \beta^t \sum_{k \in G} \mu_{jk}^{(t)} (b_k - b_j)^2 + \beta^\tau \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} q_{ya}^{(t)} \mu_{jk}^{(t+\tau)} (b_k - b_{d_a})^2 \\
&\geq \sum_{t=1}^{\tau} \beta^t \sum_{k \in G} \mu_{jk}^{(t)} (b_k - b_j)^2 + \beta^\tau \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} p_{xa}^{(t)} \mu_{jk}^{(t+\tau)} (b_k - b_{d_a})^2 \\
&= \beta W_j \left(\underbrace{s, \ldots, s}_{\tau \text{ times}}, x, \ldots\right).
\end{aligned}$$

Consequently, for each such path, we have

$$W_j\left(\underbrace{s,\ldots,s}_{\tau \text{ times}},y,\ldots\right) \geq W_j\left(\underbrace{s,\ldots,s}_{\tau \text{ times}},x,\ldots\right) > W_j(x).$$

Aggregating, we have that $W_j(s) > W_j(x)$ holds in this case as well, and this contradicts Part 2 of Theorem 1. This contradiction completes the proof. ■

Proof of Lemma A7. Suppose that $b_j^{(t)}$ is nondecreasing (the complementary case is considered similarly). By Lemma A6, $\Phi_s \geq s$. Suppose, to obtain a contradiction, that $\Phi_s \neq \{s\}$, then $x \in \Phi_s$ for some $x > s$. Since the calculations from the proof of Lemma A6 are applicable, (B1) implies (since $\Phi_s \geq s$)

$$\beta\left(W_j(s) - W_j(x)\right) = \sum_{t=1}^{\infty}\beta^t \sum_{a\geq s}\left(q_{xa}^{(t)} - q_{sa}^{(t)}\right)\left(b_j^{(t)} - b_{d_a}\right)^2.$$

For any fixed t, consider the sequence $\left|b_j^{(t)} - b_{d_a}\right|$ for $a \geq s$. Since $b_{d_s} = b_j \leq b_j^{(t)} \leq b_j^{(\infty)}$, we have $\arg\min_{z\in S}\left|b_j^{(t)} - b_{d_z}\right| = \{s\}$, so $b_j^{(t)} \in \left[b_{d_s}, \frac{b_{d_s}+b_{d_{s+1}}}{2}\right)$. This implies $\left|b_j^{(t)} - b_{d_a}\right|$ is increasing in a for $a \geq s$, and thus $\left(b_j^{(t)} - b_{d_a}\right)^2$ is also increasing. Similarly to the proof of Lemma A6, this implies

$$\sum_{a\geq s} q_{xa}^{(t)}\left(b_j^{(t)} - b_{d_a}\right)^2 \geq \sum_{a\geq s} q_{sa}^{(t)}\left(b_j^{(t)} - b_{d_a}\right)^2,$$

since the distribution $q_{x\cdot}^{(t)}$ first-order stochastically dominates $q_{s\cdot}^{(t)}$, and for at least one such t (e.g., $t = 1$) the inequality is strict. This implies $W_j(s) > W_j(x)$, but since we assumed $x \in \Phi_s$, this contradicts Part 2 of Theorem 1. This contradiction implies that $\Phi_s = \{s\}$, which completes the proof. ■

Proof of Lemma A8. Suppose that $f(\alpha) = 0$ and $\alpha' > \alpha$ (the case $\alpha' < \alpha$ is analogous). We have $W_j^{(\tau)}(s;\alpha) < W_j^{(\tau)}(x)$ for all $\tau > 1$, because the sequence of expected bliss points $b_j^{(t+\tau)} \geq b_j^{(t)}$ for all τ, and for at least some t the inequality is strict. Therefore, we have

$$\begin{aligned}
f(\alpha') - f(\alpha) &= W_j(s;\alpha') - W_j(s;\alpha) \\
&= \beta\left(\left(1-\alpha'\right)W_j^{(1)}(s;\alpha') + \alpha' W_j^{(1)}(x) - (1-\alpha)W_j^{(1)}(s;\alpha) - \alpha W_j^{(1)}(x)\right) \\
&= \beta\left((1-\alpha)\left(W_j^{(1)}(s;\alpha') - W_j^{(1)}(s;\alpha)\right) + \left(\alpha'-\alpha\right)\left(W_j^{(1)}(x) - W_j^{(1)}(s;\alpha')\right)\right) \\
&> \beta(1-\alpha)\left(W_j^{(1)}(s;\alpha') - W_j^{(1)}(s;\alpha)\right) = \cdots \\
&> \left(\beta(1-\alpha)\right)^2\left(W_j^{(2)}(s;\alpha') - W_j^{(2)}(s;\alpha)\right) = \cdots \\
&> \left(\beta(1-\alpha)\right)^{\tau}\left(W_j^{(\tau)}(s;\alpha') - W_j^{(\tau)}(s;\alpha)\right) \text{ for any } \tau > 2.
\end{aligned}$$

Since $W_j^{(\tau)}(s;\alpha') - W_j^{(\tau)}(s;\alpha)$ is bounded, we must have $f(\alpha') - f(\alpha) > 0$. This proves that $f(\alpha)$ satisfies the single-crossing condition. ■

B2 Omitted Proofs of From the Text

Proof of Theorem 6. Suppose not. Since one can always pick $\beta_0 = 0$, it suffices to prove existence of β_1 with desired properties. Suppose that such β_1 does not exist. Then for some pair of states $s, x \in S$ there are values of β arbitrarily close to 1 for which (12) does not hold. Again, without loss of generality, assume $x > s$. Multiplying both sides by $(1-\beta)$ and taking limit as $\beta \to 1$, we get, again after simplifications, that $(b_{d_x} - b_{d_s})\left(2b_{d_s}^{(\infty)} - b_{d_s} - b_{d_x}\right) \geq 0$, and since $x > s$, $b_{d_s}^{(\infty)} \geq \frac{b_{d_s}+b_{d_x}}{2}$.

Consider two possibilities. If $b_{d_s}^{(\infty)} > \frac{b_{d_s}+b_{d_x}}{2}$, then $\left|b_{d_s} - b_{d_s}^{(\infty)}\right| > \left|b_{d_x} - b_{d_s}^{(\infty)}\right|$, which means that $s \notin \arg\max_{z\in S}\left|b_{d_z} - b_{d_s}^{(\infty)}\right|$. Take $y \in \arg\max_{z\in S}\left|b_{d_z} - b_{d_s}^{(\infty)}\right|$; then by Assumption 2, d_y belongs to the same irreducible component of social mobility matrix M as d_s, which means $b_{d_y}^{(\infty)} = b_{d_s}^{(\infty)}$, and therefore $y \in \arg\max_{z\in S}\left|b_{d_z} - b_{d_y}^{(\infty)}\right|$. By Lemma A7, it must be that y is stable in any equilibrium. But then Step 2 of the proof of Part 2 of Theorem 2 implies that for some $\beta_1 < 0$, for all $\beta > \beta_0$ state s cannot be stable, a contradiction.

The remaining possibility is $b_{d_s}^{(\infty)} = \frac{b_{d_s}+b_{d_x}}{2}$, in which case $b_{d_s}^{(\infty)} - b_{d_s} = b_{d_x} - b_{d_s}^{(\infty)}$; however, Assumption 3 then implies that for all $t \geq 1$, $0 \leq b_{d_s}^{(t)} - b_{d_s} \leq b_{d_x} - b_{d_s}^{(t)}$, and thus $\left|b_{d_s} - b_{d_s}^{(t)}\right| > \left|b_{d_x} - b_{d_s}^{(t)}\right|$. From this it immediately follows that (12) holds, contradicting the assertion that it does not. This contradiction proves that slippery slope is impossible for high β.

Let us now prove that under the extra condition, one can take $\beta_0 > 0$. Suppose not; then it must be that for some pair of states $s, x \in S$ there are values of β arbitrarily close to 0 for which (12) does not hold. Without loss of generality, assume $x > s$. Dividing by β (so that the term for $t = 1$ does not contain β) and taking the limit as $\beta \to 0$, we get, after straightforward simplification, that $(b_{d_x} - b_{d_s})\left(2b_{d_s}^{(1)} - b_{d_s} - b_{d_x}\right) \geq 0$, and since $x > s$, $b_{d_s}^{(1)} \geq \frac{b_{d_s}+b_{d_x}}{2}$. Since we assumed that equality is impossible, it must imply that $b_{d_s}^{(1)} > \frac{b_{d_s}+b_{d_x}}{2}$. However, this implies that $\left|b_{d_s} - b_{d_s}^{(1)}\right| > \left|b_{d_x} - b_{d_s}^{(1)}\right|$, and if so, Part 1 of Theorem 2 implies that for some $\beta_0 > 0$, for all $\beta < \beta_0$ state s cannot be stable. This contradicts the assumption of the theorem, thus proving that slippery slope is impossible for low β.

Finally, Example B1 proves that for some values of β the statement is not necessarily true, which completes the proof of the theorem. ■

The next example illustrates the second part of Theorem 6

Example B1 ***(The non-monotonic effect of beta on slippery slope)*** There are five groups of identical size with political bliss points $b = (-4, -3, 0, 3, 4)'$, all $A_i = 0$, and the social mobility matrix is given by

$$M = \begin{pmatrix} \frac{7}{10} & \frac{1}{5} & \frac{1}{10} & 0 & 0 \\ \frac{1}{10} & \frac{3}{5} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{10} & \frac{1}{10} & \frac{3}{5} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{3}{5} & \frac{1}{10} \\ 0 & 0 & \frac{1}{10} & \frac{1}{5} & \frac{7}{10} \end{pmatrix}.$$

For such M, the equilibrium is generically unique for any discount factor β. Notice that this example satisfies all the assumptions, in particular Assumption 3 (the simplest way to see this is to notice that matrix M satisfies the conditions of Lemma B3).

With this transition matrix, members of the middle group 3 expect, on average, to prefer policy 0 due to symmetry, and thus there is no transition out of state 3. For members of group 4, the preferences of their future selves are the following. The expected political bliss policy of their tomorrow's self is $\frac{3}{2}$, the next day it is $\frac{3}{4}$, then $\frac{3}{8}$, etc. This means that tomorrow's self is indifferent between living under state 3 or 4, whereas all future selves strictly prefer state 3. This implies that in equilibrium, group 4 must move from state 4 to state 3 with probability one. Similarly, group 2 would move out of state 2 to state 3 with probability one.

Consider the incentives of groups 1 and 5 (they are symmetric). For members of group 5, the preferences of their future selves are: $\frac{17}{5} = 3.4, \frac{67}{25} = 2.68, \frac{1013}{500} = 2.026, \frac{3733}{2500} = 1.4932, \ldots$ Thus, ideally, members of this group would prefer to have state 4 in periods $2, 3, 4$, and state 3 thereafter. However, by the argument above, they can only enjoy state 4 in one period, for after that group 4 which is in power in that state would move to state 3.

Thus, members of group 5 effectively compare staying in state 5 versus spending one period in state 4 and moving to 3 thereafter. Not surprisingly, if β is small, then they prefer to move, discounting the disutility from moving to 3 too fast.

The following describes the equilibrium:

If $0 < \beta < 0.0282$, then the equilibrium is $\phi(1, 2, 3, 4, 5) = (2, 3, 3, 3, 4)$ (here, we used ϕ to denote a deterministic transition mapping).

If $0.0282 < \beta < 0.0368$, then the equilibrium involves mixing between transiting from 1 to 2 and staying at 1, and, symmetrically, between transiting from 5 to 4 and staying at 5. Here, the slippery slope effects begin to kick in: members of group 5 are already unhappy about fast transition to 3, and try to mitigate the problem by delaying this transition and staying at 5 with some probability. The best response to staying in 5 is still moving to 4, especially because the third period, where current members of group 5 are most willing to spend in state 4, is given sufficient weight; at the same time, the best response to moving to 4 is now staying in 5, because it is much more preferable to spend the third period in states 5 or 4 rather than 3. This leads to mixing.

If $0.0368 < \beta < 0.5621$, then the equilibrium is $\phi(1, 2, 3, 4, 5) = (1, 3, 3, 3, 5)$. Here, slippery slope considerations are in effect: the decision-maker in state 5 are sufficiently concerned about moving to state 3 too fast, and thus they prefer to stay in state 5. They are willing to stay in state 5 now even if this implies staying there forever.

If $0.5621 < \beta < 1$, then the equilibrium involves mixing between transiting from 1 to 2 and staying at 1, and, symmetrically, between transiting from 5 to 4 and staying at 5 (for example, if $\beta = 0.9$, then they stay with probability 0.69 and move with probability 0.31). For these values of β, distant future is sufficiently important. Decision-makers in state 5 still prefer to stay in state 5 instead of moving to state 4 immediately; however, now the weight given to distant future is high,

and so if the society were to stay in state 5 forever, they would prefer to deviate immediately and move to 4 (followed by 3).

This example illustrates that slippery slope considerations may be important only for intermediate values of β, but not for very low or very high ones.

To prove Theorem 7 and Theorem 8, we need the following stronger version of Lemma A6. It was, essentially, established in the proof of Lemma A6, but not formulated.

Lemma B1 *Suppose that for some j, for all $t > 0$, $b_j^{(t)} \geq b_j$ (respectively, $b_j^{(t)} \leq b_j$). Suppose, furthermore, that in state $s \in S$ such that $j = d_s$, either $\Phi_s \geq s$ or $\Phi_s \leq s$. Then $\Phi_s \geq s$ (respectively, $\Phi_s \leq s$).*

Proof. Suppose that for all $t > 0$, $b_j^{(t)} \geq b_j$ (the complementary case is considered similarly). Suppose, to obtain a contradiction, that $\Phi_s \not\geq s$. In this case, by assumption, we must have $\Phi_s \leq s$. If so, the argument in the proof of Lemma A6 (first case in that proof) goes through as long as $b_j^{(t)} \geq b_j$ for all t. We thus arrive at a contradiction that completes the proof. ■

Proof of Theorem 7. Notice that for all τ, either $b_{d_x} \leq b_{d_x}^{(\tau)} \leq b_{d_x}^{\prime(\tau)} \leq b_{d_x}^{\prime(\infty)} = b_{d_x}^{(\infty)}$ or the opposite inequalities hold. Then $\frac{b_{d_{x-1}}+b_{d_x}}{2} < b_{d_x}^{(\infty)} < \frac{b_{d_x}+b_{d_{x+1}}}{2}$ implies that $\frac{b_{d_{x-1}}+b_{d_x}}{2} < b_{d_x}^{(\tau)} < \frac{b_{d_x}+b_{d_{x+1}}}{2}$ for all τ. We use this to prove that for each equilibrium σ under M, democracy is stable (the same argument would apply to M'). Suppose not, so for democracy x, $\Phi(x) \neq \{x\}$. Let $s \in \Phi \setminus \{x\}$ and, as in the proof of Lemma A6, consider the utility of decision-makers at x, group d_x, if the society stayed in x for τ periods and then transitioned to s. Notice that for any $\tau > 0$,

$$W_{d_x}\left(\underbrace{x,\ldots,x}_{\tau \text{ times}},s,\ldots\right) > W_{d_x}(s).$$

Indeed, if $s > x$, then by monotonicity distribution $q_{s\cdot}^{(t+\tau)}$ (weakly) first-order stochastically dominates $q_{s\cdot}^{(t)}$ for any $t \geq 1$, which in turn (weakly) first-order stochastically dominates the degenerate distribution with an atom at s, which implies that for all $t > \tau$, the future self with the expected ideal point $b_{d_x}^t$ would prefer to have stayed in x for τ times. Since for the first τ selves, such preference is strict (they get their ideal point x as opposed to some other state), the inequality is strict. Similar logic applies to the case $s < x$. This implies that for any $s \in \Phi \setminus \{x\}$, staying in x for any number of periods and eventually transitioning to s is preferred to transitioning to s immediately. Aggregating this inequality over $s \in \Phi \setminus \{x\}$, we find that $W_{d_x}(x) > W_{d_x}(s)$ for any such s. However, this contradicts Part 2 of Theorem 1. This contradiction proves that $\Phi(x) = \{x\}$ in each equilibrium under M and, similarly, under M'.

Denote $y = x - 1$ and $z = x + 1$, and suppose for simplicity that both $y, z \in S$ (if not, then a straightforward simplification of the argument applies). Suppose that under M, x is asymptotically

stable in any equilibrium. Suppose, to obtain a contradiction, that $b_{d_y}^{(\infty)} \leq b_{d_y}$; let us prove that under M, there is some equilibrium σ with $\Phi(y) \leq y$ (this would contradict asymptotic stability). Since Definition 2 applies to matrices M and M', we have $b_{d_y}^{(t)} \leq b_{d_y}$ for all $t \geq 1$. Notice that if the parameter values are generic (in the sense of Part 4 of Theorem 1), then by that result, in any equilibrium, either $\Phi(y) \leq y$ or $\Phi(y) \geq y$, and then Lemma B1 implies that $\Phi(y) \leq y$. In the case of nongeneric parameter values, we need to take a sequence of generic ones that converges to the original parameter values and find equilibria in those cases; taking the limit, we would again find an equilibrium with $\Phi(y) \leq y$. Existence of such equilibrium σ contradicts asymptotic stability of x under M, which proves that $b_{d_y}^{(\infty)} > b_{d_y}$. We can similarly prove that $b_{d_z}^{(\infty)} < b_{d_z}$.

Suppose that there is equilibrium σ' under M' where x is not asymptotically stable. This implies that either $\Phi'(y) = \{y\}$ or there is some $s < y$ with $s \in \Phi'(y)$ or $\Phi'(z) = z$ or there is some $s > z$ with $s \in \Phi'(z)$. We consider the first and the second possibilities, while the third and the fourth are completely analogous.

Consider the case $\Phi'(y) = \{y\}$. Take matrix M, and consider restricted transitions $a \in F_s$ if either $s \geq x$ or $a \leq y$ (i.e., any transitions are possible, except if the origin is y or below and destination is x or above). Consider equilibrium $\tilde{\sigma}$ under these restricted transitions. Since $\Phi(y) \leq y$, Lemma B1 is applicable and implies $\tilde{\Phi}(y) = \{y\}$. Notice that the same logic as above implies that democracy is stable: $\tilde{\Phi}(x) = \{x\}$. Let us remove the restriction on transitions; by Lemma A5, we either get an equilibrium, or group d_y prefers to deviate from staying in y to moving to x. However, if they prefer to do so under M, single-crossing considerations would imply that they would prefer to do so under M', which means that there cannot exist an equilibrium σ' under M' such that $\Phi'(x) = \{x\}$ and $\Phi'(y) = \{y\}$. This contradiction implies that removing the restrictions results in an equilibrium σ under M where y is stable. However, this contradicts that y is asymptotically stable for all equilibria under M.

Now consider the case where there is $s < y$ with $s \in \Phi'(y)$. If $\Phi'(y) \leq y$, we would get an immediate contradiction with Lemma B1, thus, we must have that $x \in \Phi'(y)$ as well, and therefore $W_{d_y}(x) = W_{d_y}(s)$. Since we have $b_{d_y}^{(t)} \geq b_{d_y}$ for all $t \geq 1$, this means that all future selves of group d_y strictly prefer state y over state s or any state below. Thus, we have $W_{d_y}(y, y, y, \ldots) > W_{d_y}(s)$, and therefore $W_{d_y}(y, y, y, \ldots) > W_{d_y}(x) = W_{d_y}(x, x, x, \ldots)$. Consequently, using an argument similar to above, we can prove that for some equilibrium σ' under M', y is stable, as in that case group d_y would get a higher continuation utility than if they deviated to x or s or a state below s. But this reduces this case to the previous one, and we can again get to a contradiction.

The cases where $\Phi'(z) = z$ or there is some $s > z$ with $s \in \Phi'(z)$ are considered similarly and lead to contradictions. This proves that for each equilibrium under M', x is asymptotically stable. ■

Proof of Theorem 8. Suppose that $b_{d_x}^{(\infty)} \leq b_{d_x}$ (the opposite case is analogous). Then we can follow an argument similar to one in the proof of Theorem 7 to show that state $y = x-1$ is stable in all equilibria under both M and M', which implies that x is not asymptotically stable for any equilibrium.

Suppose that democracy is not stable for any equilibrium under M. Under matrix M, restrict

transitions as follows: let $a \in F_s$ if either $s < x$ or $a \geq x$ (i.e., transitions from x to the left are forbidden). Then there is an equilibrium $\tilde{\sigma}$ with transition mapping $\tilde{Q}$. Given the restriction on transitions, it must be that $\tilde{\Phi}_x \geq x$, and then Lemma B1 implies that $\tilde{\Phi}_x = \{x\}$. In addition, as before, it must be that $\tilde{\Phi}_y = \{y\}$, so y is stable. Let us now lift the restriction on transitions; by Lemma A5, either there is an equilibrium under M with transition mapping $\tilde{Q}$, or group d_x would be better off if the society transited to y instead of staying in x. In the first case, however, we would get a contradiction to the assertion that x is not stable under any equilibrium under M. In the second case, since social mobility is faster under M' than under M, we get that there cannot be an equilibrium under M' such that both x and y are stable, since in such equilibrium d_x would be better off if the society transited to y instead of staying in x, which would contradict Part 2 of Theorem 1. Since, as we proved, y is stable for all equilibria under M', it must be that x is unstable for any equilibrium under M'. ■

Proof of Proposition 1. With only two groups affected by social mobility, within-person monotonicity is automatically satisfied, and the equilibrium is (generically) unique. If $\gamma_M > \gamma_P$, then members in M prefer to be in democracy, where it rules, at any point in the future, and thus democracy is stable. Given that, continuation payoffs, starting in democracy, are given by:

$$\begin{aligned} V_R(d) &= A_R - b_R^2 + \beta V_R(d); \\ V_M(d) &= \beta\left((1-\theta) V_M(d) + \theta V_P(d)\right); \\ V_P(d) &= A_P - b_P^2 + \beta\left(\left(1 - \frac{\gamma_M}{\gamma_P}\theta\right) V_P(d) + \theta\frac{\gamma_M}{\gamma_P} V_M(d)\right). \end{aligned}$$

Thus,

$$\begin{aligned} V_R(d) &= \frac{A_R - b_R^2}{1-\beta}; \\ V_M(d) &= \frac{A_P - b_P^2}{1-\beta}\frac{\beta\theta}{1-\beta+\beta\theta\left(1+\frac{\gamma_M}{\gamma_P}\right)}; \\ V_P(d) &= \frac{A_P - b_P^2}{1-\beta}\frac{1-\beta+\beta\theta}{1-\beta+\beta\theta\left(1+\frac{\gamma_M}{\gamma_P}\right)}. \end{aligned}$$

So, $V_R(d)$ does not depend on d, whereas $V_M(d)$ is decreasing and $V_P(d)$ is increasing in θ, since $A_P - b_P^2 < 0$.

Now consider the case $\gamma_M < \gamma_P$. Here, the state with the poor in power (denote it l) is stable, and starting from d, the society can start in d or transition to l, but not to the state where the rich are in power, r (this follows from Lemma A6). The utility of the players from being in state l is given by (similarly to previous)

$$\begin{aligned} V_R(l) &= A_R - (b_R - b_P)^2 + \beta V_R(l); \\ V_M(l) &= -b_P^2 + \beta\left((1-\theta) V_M(l) + \theta V_P(l)\right); \\ V_P(l) &= A_P + \beta\left(\left(1 - \frac{\gamma_M}{\gamma_P}\theta\right) V_P(l) + \theta\frac{\gamma_M}{\gamma_P} V_M(l)\right). \end{aligned}$$

and thus

$$\begin{aligned}
V_R(l) &= \frac{A_R-(b_R-b_P)^2}{1-\beta};\\
V_M(l) &= \frac{1}{1-\beta}\frac{(1-\beta)\left(-b_P^2\right)+\beta\theta\left(A_P-\frac{\gamma_M}{\gamma_P}b_P^2\right)}{1-\beta+\beta\theta\left(1+\frac{\gamma_M}{\gamma_P}\right)};\\
V_P(l) &= \frac{1}{1-\beta}\frac{(1-\beta)A_P+\beta\theta\left(A_P-\frac{\gamma_M}{\gamma_P}b_P^2\right)}{1-\beta+\beta\theta\left(1+\frac{\gamma_M}{\gamma_P}\right)}.
\end{aligned}$$

Suppose the probability of transition from d to l equals $q=q_{dl}$. In that case, the utilities being in d are given as equations

$$\begin{aligned}
V_R(d) &= A_R-b_R^2+\beta(1-q)V_R(d)+\beta qV_R(l);\\
V_M(d) &= \beta(1-\theta)(1-q)V_M(d)+\beta(1-\theta)qV_M(l)+\beta\theta(1-q)V_P(d)+\beta\theta qV_P(l);\\
V_P(d) &= A_P-b_P^2+\beta\left(1-\theta\frac{\gamma_M}{\gamma_P}\right)(1-q)V_P(d)+\beta\left(1-\theta\frac{\gamma_M}{\gamma_P}\right)qV_P(l)\\
&\quad+\beta\theta\frac{\gamma_M}{\gamma_P}(1-q)V_M(d)+\beta\theta\frac{\gamma_M}{\gamma_P}qV_M(l).
\end{aligned}$$

One can check that for $\theta<\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta}$, $w_M(d)>w_M(l)$ for any q; for $\theta>\frac{1}{2}$, $w_M(d)<w_M(l)$ for any q, and for $\theta\in\left[\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta},\frac{1}{2}\right]$, there is a unique $q\in[0,1]$ such that $w_M(d)=w_M(l)$, and this q corresponds to a unique equilibrium; moreover, q is increasing in θ.

This implies the result for preferences of R, as $V_R(d)$ depends on θ only through q, and is decreasing in q. Consider the middle class M. For $\theta<\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta}$, $q=0$, so d is stable, and M prefers a smaller θ. For $\theta\in\left[\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta},\frac{1}{2}\right]$, $w_M(d)=w_M(l)$, or in other words, $V_M(d)=V_M(l)+b_P^2$; similarly, for $\theta>\frac{1}{2}$, there is an immediate transition from d to l, and $V_M(d)=V_M(l)+b_P^2$ also holds. But $V_M(l)$ is decreasing in θ as follows from the formula, and therefore $V_M(d)$ is monotonically decreasing in θ.

Let us now consider P. For $\theta\leq\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta}$, $q=0$, and for $\theta\geq\frac{1}{2}$, $q=1$; in both intervals, a marginal increase in θ only leads to more social mobility, and the poor are better off. If $\theta\in\left(\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta},\frac{1}{2}\right)$, let us rewrite the equation for $V_M(d)$ (by collapsing $\beta((1-\theta)V_M(l)+\theta V_P(l))$ into $V_M(l)+b_P^2$) as

$$V_M(d)=\beta(1-\theta)(1-q)V_M(d)+\beta\theta(1-q)V_P(d)+q\left(V_M(l)+b_P^2\right).$$

Now, we can plug in $V_M(d)=V_M(l)+b_P^2$ to obtain

$$V_M(l)+b_P^2=\beta(1-\theta)(1-q)\left(V_M(l)+b_P^2\right)+\beta\theta(1-q)V_P(d)+q\left(V_M(l)+b_P^2\right);$$

rearranging and dividing by $1-q$ (which is nonzero in the interior of the interval), we get

$$\left(V_M(l)+b_P^2\right)(1-\beta+\beta\theta)=\beta\theta V_P(d),$$

and thus

$$\begin{aligned}
V_P(d) &= \frac{(1-\beta+\beta\theta)}{\beta\theta}\left(V_M(l)+b_P^2\right)\\
&= \frac{1-\beta+\beta\theta}{\beta\theta}\left(\frac{1}{1-\beta}\frac{(1-\beta)\left(-b_P^2\right)+\beta\theta\left(A_P-\frac{\gamma_M}{\gamma_P}b_P^2\right)}{1-\beta+\beta\theta\left(1+\frac{\gamma_M}{\gamma_P}\right)}+b_P^2\right)\\
&= \frac{1-\beta+\beta\theta}{(1-\beta)\theta}\frac{\theta A_P-\left(1-\theta-\beta+\beta\theta\left(1+\frac{\gamma_M}{\gamma_P}\right)\right)b_P^2}{1-\beta+\beta\theta\left(1+\frac{\gamma_M}{\gamma_P}\right)}.
\end{aligned}$$

Differentiating and simplifying, we get

$$\frac{dV_P(d)}{d\theta}=\frac{b_P^2}{\theta^2}-\frac{\beta\frac{\gamma_M}{\gamma_P}\left(A_P+b_P^2\right)}{\left(1-\beta+\beta\theta\left(1+\frac{\gamma_M}{\gamma_P}\right)\right)^2},$$

which is positive, since $A_P+b_P^2<0$ by assumption. Thus, $V_P(d)$ is strictly increasing in θ for all θ, which completes the proof. ■

Proof of Proposition 2. If $\gamma_M>\gamma_R$ then, as in Theorem 1, democracy is stable for any θ. Similarly to the proof there, one can easily show that R prefer a lower θ, M prefer a higher θ, and P are indifferent.

If $\gamma_M<\gamma_R$, then let $q=q_{dr}$ be the probability of transition from d to r in each period. Then, once again as in Theorem 1, there is a unique equilibrium for each θ; moreover, for $\theta\leq\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_R}\right)\beta}$, $q=0$, for $\theta\in\left[\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_R}\right)\beta},\frac{1}{2}\right]$, q is monotonically increasing from 0 to 1, and for $\theta\geq\frac{1}{2}$, $q=1$. Accordingly, $V_P(d)$ is strictly increasing on $\theta\in\left[\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_R}\right)\beta},\frac{1}{2}\right]$ and constant outside of it, and $V_M(d)$ is strictly increasing for all θ (this may be proven analogously to Theorem 1). As for $V_R(d)$, it is strictly decreasing for $\theta<\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_R}\right)\beta}$ or $\theta>\frac{1}{2}$.

To complete the proof, consider $V_R(d)$ for $\theta\in\left[\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_R}\right)\beta},\frac{1}{2}\right]$. Here, $V_R(d)$ is given (similarly to Theorem 1) by

$$V_R(d)=\frac{1-\beta+\beta\theta}{(1-\beta)\theta}\frac{\theta A_R-\left(1-\theta-\beta+\beta\theta\left(1+\frac{\gamma_M}{\gamma_R}\right)\right)b_R^2}{1-\beta+\beta\theta\left(1+\frac{\gamma_M}{\gamma_R}\right)}.$$

Its derivative with respect to θ equals

$$\frac{dV_R(d)}{d\theta}=\frac{b_R^2}{\theta^2}-\frac{\beta\frac{\gamma_M}{\gamma_R}\left(A_R+b_R^2\right)}{\left(1-\beta+\beta\theta\left(1+\frac{\gamma_M}{\gamma_R}\right)\right)^2}.$$

This expression is different, because $A_R + b_R^2 > 0$, and the sign of this expression is potentially ambiguous. More precisely, $V_R(d)$ locally increasing for

$$\theta < \frac{1-\beta}{\sqrt{\beta\frac{\gamma_M}{\gamma_R}\left(\frac{A_R}{b_R^2}+1\right)}-\beta\left(1+\frac{\gamma_M}{\gamma_R}\right)} = \theta^*,$$

and is locally decreasing otherwise.

One can easily check that $\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_R}\right)\beta} < \theta^*$ is equivalent to $\beta\frac{\gamma_M}{\gamma_R}\left(\frac{A_R}{b_R^2}+1\right) < 4$. If the latter condition does not hold, then $V_R(d)$ is monotonically decreasing on $\theta \in \left[\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_R}\right)\beta}, \frac{1}{2}\right]$, and thus for all θ; if it holds, there is an interval up to $\min\left(\theta^*, \frac{1}{2}\right)$ where $V_R(d)$ is increasing. ■

Proof of Proposition 3. Consider the case $\gamma_M > \gamma_P$. By Theorem 1, M prefer a lower θ and P prefer a higher θ. However, since $\gamma_M > \gamma_P$, any $\theta > 0$ will be defeated in a plurality voting by $\hat{\theta} = 0$. Thus, $\hat{\theta} = 0$ is the unique core element.

Now consider the case $\gamma_M < \gamma_P$. Consider $\theta < \frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta}$; such θ will be defeated in a plurality voting by $\hat{\theta} = \frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta}$, because R are indifferent, and the more numerous of the remaining groups, P, prefers θ'. If $\theta > \frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta}$, then it will again be defeated by $\hat{\theta} = \frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta}$, because R and M both prefer a slower social mobility on this interval, and together they constitute a majority. Thus, $\hat{\theta} = \frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta}$ is the unique core element. ■

Proof of Proposition 4. Consider the case $\gamma_M > \gamma_R$. By Theorem 2, R prefer a lower θ and M prefer a higher θ. However, since $\gamma_M > \gamma_R$, all θ, except for the maximum value, will be defeated in a plurality voting. Thus, $\hat{\theta}$ is the maximal admissible value of θ; in our case, Assumption 1 is satisfied whenever $\theta \leq \frac{1}{1+\frac{\gamma_M}{\gamma_R}}$, so $\hat{\theta} = \frac{1}{1+\frac{\gamma_M}{\gamma_R}}$.

Now consider the case $\gamma_M < \gamma_R$. Here, consider the following possibilities. First, if $\beta\frac{\gamma_M}{\gamma_R}\left(\frac{A_R}{b_R^2}+1\right) \geq 4$, then the utility of R is monotonically decreasing in θ. Thus, any $\theta > 0$ will be defeated, in a plurality voting, by $\hat{\theta} = 0$ (M would favor $\theta > 0$, but R, who are more numerous, would vote for $\hat{\theta}$, and sometimes they would be joined by P). Thus, $\hat{\theta} = 0$ is the unique core element in this case.

More generally, it is easy to see that given the conflict of interest between M and P, $\hat{\theta}$ will equal the value that maximizes $V_R(d)$. There are two candidate values for this value of θ: 0 and $\min\left(\theta^*, \frac{1}{2}\right)$. Notice that $\frac{1}{1+\frac{\gamma_M}{\gamma_R}} > \frac{1}{2}$ in this case, so this value is necessarily admissible.

Compute first the values of $V_R(d)$ at $\theta = 0$ and $\theta = \frac{1}{2}$; we have

$$\begin{aligned} V_R^{\theta=0}(d) &= \frac{A_R - b_R^2}{1-\beta}, \\ V_R^{\theta=\frac{1}{2}}(d) &= \frac{2-\beta}{1-\beta}\frac{\left(A_R - b_R^2\right) + \beta b_R^2\left(1-\frac{\gamma_M}{\gamma_R}\right)}{2-\beta\left(1-\frac{\gamma_M}{\gamma_R}\right)}. \end{aligned}$$

We have $V_R^{\theta=\frac{1}{2}}(d) > V_R^{\theta=0}(d)$ if and only if $(2-\beta)\left(\frac{\gamma_R}{\gamma_M}-1\right) > \left(\frac{A_R}{b_R^2}-1\right)$. On the other hand, $\theta^* < \frac{1}{2}$ if and only if $\frac{\left(2-\beta+\beta\frac{\gamma_M}{\gamma_R}\right)^2}{\beta\frac{\gamma_M}{\gamma_R}} < \left(\frac{A_R}{b_R^2}+1\right)$. Let us show that $\theta^* < \frac{1}{2}$ cannot hold if $(2-\beta)\left(\frac{\gamma_R}{\gamma_M}-1\right) > \left(\frac{A_R}{b_R^2}-1\right)$. Indeed, if this is not the case, we must have

$$(2-\beta)\left(\frac{\gamma_R}{\gamma_M}-1\right)+2 > \frac{A_R}{b_R^2}+1 > \frac{\left(2-\beta+\beta\frac{\gamma_M}{\gamma_R}\right)^2}{\beta\frac{\gamma_M}{\gamma_R}}.$$

Simplifying, we would get $\left(2-2\beta+\beta\frac{\gamma_M}{\gamma_R}\right)\left(2-\beta+\beta\frac{\gamma_M}{\gamma_R}\right) < 0$, which is impossible. Therefore, if $(2-\beta)\left(\frac{\gamma_R}{\gamma_M}-1\right) > \left(\frac{A_R}{b_R^2}-1\right)$, then $V_R^{\theta=\frac{1}{2}}(d)$ maximizes $V_R^{\theta}(d)$ for all θ, and is thus the unique core element.

Consider the case $(2-\beta)\left(\frac{\gamma_R}{\gamma_M}-1\right) < \left(\frac{A_R}{b_R^2}-1\right)$. Here, we may get a core element other than 0 only if $\theta^* < \frac{1}{2}$ and $V_R^{\theta=\theta^*}(d) > V_R^{\theta=0}(d)$. We have, after simplification,

$$\begin{aligned} V_R^{\theta=\theta^*}(d) &= \frac{1-\beta+\beta\theta^*}{(1-\beta)\theta^*}\frac{\theta^* A_R - \left(1-\theta^*-\beta+\beta\theta^*\left(1+\frac{\gamma_M}{\gamma_R}\right)\right)b_R^2}{1-\beta+\beta\theta^*\left(1+\frac{\gamma_M}{\gamma_R}\right)} \\ &= \left(\frac{1}{\theta^*}+\frac{\beta}{1-\beta}\right)\left(\frac{A_R+b_R^2}{\frac{1-\beta}{\theta^*}+\beta\left(1+\frac{\gamma_M}{\gamma_R}\right)}-b_R^2\right) \\ &= \frac{A_R+\left(1+\frac{\gamma_M}{\gamma_R}\right)b_R^2-2\sqrt{\beta\frac{\gamma_M}{\gamma_R}\left(A_R+b_R^2\right)}b_R}{1-\beta}. \end{aligned}$$

Then $V_R^{\theta=\theta^*}(d)$ exceeds $V_R^{\theta=0}(d)$ if and only if $\frac{\left(2+\frac{\gamma_M}{\gamma_R}\right)^2}{4\beta\frac{\gamma_M}{\gamma_R}} > \left(\frac{A_R}{b_R^2}+1\right)$. However, this is incompatible with $\theta^* < \frac{1}{2}$. Indeed, if both were true at the same time, we would have

$$\frac{\left(2+\frac{\gamma_M}{\gamma_R}\right)^2}{4\beta\frac{\gamma_M}{\gamma_R}} > \left(\frac{A_R}{b_R^2}+1\right) > \frac{\left(2-\beta+\beta\frac{\gamma_M}{\gamma_R}\right)^2}{\beta\frac{\gamma_M}{\gamma_R}},$$

which, after simplification, implies $(2\beta-1)\left(1-\frac{\gamma_M}{\gamma_R}\right) > 1$. But this is impossible, which means that $V_R^{\theta=\theta^*}(d) > V_R^{\theta=0}(d)$ only if $\theta^* > \frac{1}{2}$. Consequently, if $(2-\beta)\left(\frac{\gamma_R}{\gamma_M}-1\right) < \left(\frac{A_R}{b_R^2}-1\right)$, then the utility of R is maximized for $\theta = 0$. Consequently, $\hat{\theta} = 0$ is the unique core element in this case. ■

B3 Additional Examples

In this part of Appendix B we provide three additional examples.

Example B2 ***(Multiple equilibria)*** There are five groups with political bliss points $b_{1,2,3,4,5} = -\frac{21}{10}, -1, 0, 1, \frac{21}{10}$ (there would be two equilibria even if the extreme political bliss points are ± 2

rather than ± 2.1, but this would be a knife-edge case). All $A_i = 0$, discount factor $\beta = \frac{1}{2}$, and the reshuffling matrix M is given by

$$M = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}.$$

One can show that the following two mappings, $\phi_1(1,2,3,4,5) = (1,2,3,4,4)$ and $\phi_2(1,2,3,4,5) = (1,2,4,4,4)$, form an equilibrium. To see why, consider the incentives of a member of group 3. Today (in period 1), his political bliss point is 0. The next day, he will have political bliss points $-1, 0, 1, \frac{21}{10}$ with equal probabilities. For quadratic utility functions, it is the average that matters, and his expected political bliss point equals $\frac{21}{40}$. Since $\frac{21}{40}$ is closer to $b_4 = 1$ than to $b_3 = 0$, when players only care about the next period, an individual of group 3 would choose $\phi(3) = 4$. For a more patient individual, the situation is more complicated. In period 3, his expected political bliss point would equal $\frac{73}{160} < \frac{1}{2}$, and it would continue to decrease in the subsequent periods, monotonically converging to zero. Thus, ideally, he would prefer state 4 in period 2 and state 3 starting from period 3 on. But this is not feasible: once the society reaches state 4, it will stay there forever, as the decision-makers there are not willing to move to state 3, as one can easily show (more precisely, they would prefer to remain in state 4 for periods 3 through 8 and move to state 3 after that, but given the discount factor, this makes them willing to stay in 4 rather than move to 3). Consequently, he needs to decide whether to stay in 3 or move to 4 taking into account the fact that 4, would be an absorbing state in equilibrium.

This decision is ultimately made by taking the decisions of future members of group 3 into account. If they would opt to stay in state 3, then in period 1 the effective choice is between staying in state 3 forever or moving permanently to state 4. In this case, current members of group 3 would prefer to stay, even if their short-term incentives are different. However, if future members of group 3 would move to state 4, then staying in state 3 is for one period only (period 2), and it so happens that this is the only period where members of group 3 would actually prefer to be in state 4. Consequently, the best response today is to move to state 4 immediately. As a result, both ϕ_1 and ϕ_2 are equilibria (verifying that other groups act as prescribed is straightforward).

One can also verify that equilibrium ϕ_1 is preferred to ϕ_2 by individuals who start in groups $1, 2, 3$, and the opposite is true for those in groups 4 *and* 5. In other words, today's decision-makers (group 3) are in favor of ϕ_1. Given that the decision is made by a representative agent, one could wonder what makes ϕ_2 an equilibrium. One way of interpreting equilibrium mapping ϕ_2 is coordination failure, but not by individuals living in one period, but rather by members of group 3 from different periods. At their respective time, they would all be better off staying in 3. However, if future decision-makers in state 3 move to 4, then it is a best response to do so immediately. (The problem does not disappear if we truncate the future, i.e., consider a finite number of periods: then in the penultimate period, members of group 3 would move to 4, and this would ensure the

survival of the equilibrium corresponding to ϕ_2).

As always, when there are two equilibria, there is also a third one, where starting in state 3, group 3 decides to stay with probability $\alpha \approx 0.5667$ and move to state 4 with probability $1-\alpha$.

Example B3 ***(Mixing between noncontiguous states)*** There are five groups; the weights of the groups are $\frac{3}{100}, \frac{1}{100}, \frac{6}{100}, \frac{50}{100}, \frac{40}{100}$, and their political bliss points are $b = (0, 0.9, 1, 2, 30)'$, respectively. All $A_i = 0$, and the social mobility matrix is given by

$$M = \begin{pmatrix} \frac{70}{100} & \frac{10}{100} & \frac{20}{100} & 0 & 0 \\ \frac{30}{100} & \frac{10}{100} & \frac{60}{100} & 0 & 0 \\ \frac{10}{100} & \frac{10}{100} & \frac{30}{100} & \frac{30}{100} & \frac{20}{100} \\ 0 & 0 & \frac{6}{100} & \frac{54}{100} & \frac{40}{100} \\ 0 & 0 & 0 & \frac{53}{100} & \frac{47}{100} \end{pmatrix}.$$

Suppose that the discount factor $\beta = 0.5$.

The unique equilibrium in the game has the following transition mapping: $\phi(2) = 3$, $\phi(3,4,5) = 4$, and from state 1, the society moves to state 3 with probability $z \approx 0.896$ and stays in state 1 with the complementary probability $1 - z \approx 0.104$.

The intuition for why the society does not find it even better to transit to state 2 is the following. The transition matrix is such that individuals from group 1 prefer the society to stay in 1 tomorrow, and be in state 4 thereafter. They know that from states $3, 4, 5$ there will be an immediate transition to 4, therefore, since staying in 1 forever is a bad idea in the long run, moving to state 3 is a reasonable compromise. On the other hand, if future members of group 1 are sufficiently likely to move to state 3, then the current ones would rather prefer to spend an extra period in state 1, which would lead to mixing between states 1 and 3. This mixing is a compromise between the desires to spend an extra period in state 1 and to reach state 4 sooner rather than later.

It would seem that moving to state 2*,* rather than mixing between states 1 and 3*,* is a reasonable middle ground, enabling the accomplishment of both goals. It turns out, however, that it accomplishes neither. Moving to state 2 does not allow members of group 1 to benefit from being in state 1 for an extra period. At the same time, since from state 2 the society moves to state 3 rather than 4, going to state 2 does not make state 4 any closer. The parameter values, where state 2 is "unimportant" (the group which rules there is small, and its bliss policy is very close to that in state 3) make sure that the immediate utility of members of group 1 from moving to state 2 is only marginally better than that from moving to state 3, but it delays transition to state 4. As a result, the path initiated by moving to state 2 runs in-between the corresponding paths for staying at 1 and moving to 3, but in the important few periods the payoff is closer to the path that yields a lower payoff in that period. As a result, in equilibrium, the mixing is between staying and moving to a non-neighboring state, even though all utility functions are concave and even quadratic.

Example B4 ***(Nonconvergence to long run ideal policy and slippery slope)*** *There are four groups, with weights* $\frac{1}{10}, \frac{2}{5}, \frac{1}{10}, \frac{2}{5}$*, and their political bliss points are* $b = (-5, 0, 1, 6)'$*, respectively. All*

$A_i = 0$, *and the social mobility matrix is given by*

$$M = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{1}{6} & \frac{5}{6} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & \frac{1}{6} & \frac{5}{6} \end{pmatrix}.$$

Suppose, however, that there are only three states: with the first, third, and fourth group ruling (so there is no state where policy 0 *is implemented).*

Notice that Assumption 2 does not hold in this example: in the long run, the ideal policy of individuals from group 1 *is* $\frac{\frac{1}{10}\times(-5)+\frac{2}{5}\times 0}{\frac{1}{10}+\frac{2}{5}} = -1$, *and the closest policy that can be implemented in some state is* 1, *which would happen if group* 3 *rules. However, individuals from group* 1 *can never move to group* 3.

Consider any $\beta \in (0,1)$. *Then the state where group* 4 *rules is stable, whereas from the state where group* 3 *rules, there will be an immediate transition to the state where group* 4 *rules. Consider the problem of group* 1 *in the state where it makes decisions. Its expected ideal point in the next period is* $-\frac{5}{3}$, *in the following period it is* $-\frac{10}{9}$, *etc, converging to* -1; *which implies that all future selves prefer the state where group* 3 *rules. At the same time, they know that a transition to that state would put group* 4 *in power in the following period, which they clearly dislike. Thus, if they are sufficiently forward-looking (namely, if* $\beta > 0.103$*), then they would prefer to stay in the same state where group* 1 *chooses policy, so this state is stable.*

Therefore, this serves as a counterexample both to Theorem 2 Part 2 (since from state 1 *there is no convergence to the state that the future selves prefer, even if* β *is close to* 1*), and to Theorem 6 (since state* 1 *is stable, but very distant future selves would prefer that the society always stayed in the state where group* 3 *is in charge.*

B4 Conditions for mixed strategies

Our next results clarify the conditions under which we should see equilibria in mixed strategies. Consider first the following definition.

Definition 3 *We say that social mobility is slow if the preferred state of each individual's today's and tomorrow's selves coincide. More formally, this property holds if for all states* s,

$$b_{d_s} \in \arg\min_{z\in S} \left| b^{(1)}_{d_s} - b_{d_z} \right|.$$

This property is guaranteed to hold, for example, if M is sufficiently close to diagonal.

Theorem B1 *The following is true for any* M, *any* $\mathbf{b}$ *and* A.

(i) There is $\beta_0 > 0$ *such that for any* $0 < \beta < \beta_0$. *Then there is an equilibrium which involves pure transitions only and, generically, this is true for all equilibria;*

(ii) Suppose that social mobility is slow, but in at least one state $s \in S$, $b_{d_s} \notin \arg\min_{z\in S} \left| b^{(\infty)}_{d_s} - b_{d_z} \right|$ *(this is guaranteed to hold for generic parameter values, provided that there*

are at least two states). Then there is $\beta_1 < 1$ such that for any $\beta_1 < \beta < 1$, the equilibrium mapping involves mixing.

Proof of Theorem B1. To establish (i), let us take generic parameter values, in the sense of Part 4 of Theorem 1. Notice that for such parameter values, for every $x \in S$, the state $\arg\min_{z\in S}\left|b_{d_z} - b^{(1)}_{d_x}\right|$ is a singleton. Thus, by Theorem 2, Part 1, the transition mapping is uniquely defined and involves pure transitions only. Now, to show that there is such equilibrium for non-generic parameter values, we can take a converging sequence of generic parameter values and use upper-hemicontinuity of equilibria.

To establish (ii), suppose not. Then there is a sequence of β, $\{\beta_i\}$ converging to 1 such that for each β there is an equilibrium with deterministic transition mapping. Since there is only a finite number of such mappings, we can take a subsequence $\left\{\beta_{i_k}\right\}$ for which there are equilibria with the same deterministic transition mapping. Denote this mapping by $\phi : S \to S$.

If $\phi(s) = s$ for all s, then take state x that satisfies $b_{d_x} \notin \arg\min_{z\in S}\left|b^{(\infty)}_{d_x} - b_{d_z}\right|$ (i.e., existence of such a state is assumed). For β high enough, group d_x would be better off deviating and moving to a state y that maximizes $\arg\min_{z\in S}\left|b^{(\infty)}_{d_x} - b_{d_z}\right|$, which contradicts that such ϕ occurs in an equilibrium for arbitrarily high β. Now suppose that $\phi(s) \neq s$ for some s. Monotonicity of ϕ implies that there are x and y such that $|x - y| = 1$ and such that $\phi(x) = \phi(y) = y$. In this case, however, the decision-makers at x, d_x, would prefer to deviate and stay in x for an extra period. This contradiction proves the statement of the theorem.

Finally, let us prove that existence of a state with $b_{d_s} \notin \arg\min_{z\in S}\left|b^{(\infty)}_{d_s} - b_{d_z}\right|$ holds for generic parameter values. Indeed, generically, all elements of M are positive, and therefore $b^{(\infty)}_{d_x} = b^{(\infty)}_{d_y}$ for all $x, y \in S$. Denote this value by $b^{(\infty)}$; notice that the only case where $b^{(\infty)} \in \arg\min_{z\in S}\left|b^{(\infty)}_{d_s} - b_{d_z}\right|$ for every $s \in S$ is where S consists of two elements (say x and y), and $b^{(\infty)} = \frac{1}{2}\left(b_{d_x} + b_{d_y}\right)$. This however, is nongeneric, establishing the result. ■

One can also prove that for any fixed β, if M is sufficiently close to identity matrix then the equilibrium is in pure strategies. Interestingly, with a finite number of periods, there would (generically) only be equilibria in pure strategies. A proof is available upon request.

B5 Conditions for monotonicity of MPE

We first provide an example of symmetric nonmonotone MPE.

Example B5 ***(Nonmonotone equilibrium)*** There are four groups, with identical weights. Their political bliss points are $b = (-1, 0, 1, 40)'$, respectively. All $A_i = 0$, and the social mobility matrix is given by

$$M = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{3}{10} & \frac{3}{10} & \frac{2}{5} & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} \end{pmatrix}.$$

Furthermore, suppose that there are only two states: in state 1, the leftmost group (with bliss point -1) is ruling, and in state 2, the second group (with bliss point 0) is ruling.

This example with only two states is deliberately simple. For any $\beta \in (0,1)$ it admits the monotone equilibrium $\phi(1) = \phi(2) = 2$. (Members are group 1 are indifferent between staying at 1 and moving to 2, but staying at 1 with a positive probability is not an equilibrium, since then they would strictly prefer to move to 2 because of a nonzero chance to stay in 1 in subsequent periods.) This is the only monotone equilibrium.

However, there is also a nonmonotone equilibrium, ψ with $\psi(1) = 2$ and $\psi(2) = 1$, for $\beta > \beta^* \approx 0.2174$. It works as follows. Expecting that future decision-makers would alternate between states 1 and 2, the current decision-makers, at both states 1 and 2, effectively choose between the following two paths: $1, 2, 1, 2, 1, 2, \ldots$, and $2, 1, 2, 1, 2, 1, \ldots$. For $\beta > \beta^*$, the immediate considerations are not too important, but what is important is when members of the group get a chance to move to the group with radical preferences (group 4); strategically, members of either of the two groups 1 and 2 would want to be in state 2 at the time of first encounter. For members of group 2, this encounter happens in two periods, hence they prefer the path $1, 2, 1, 2, 1, 2, \ldots$ to $2, 1, 2, 1, 2, 1, \ldots$ and are thus willing to move to state 1, contrary to their immediate preferences. On the other hand, members of group 1 prefer the latter path, which reinforces their incentives to move to state 2. As a result, neither group wants to deviate, and mapping ψ may arise in equilibrium.

It should be noted that in this example, Assumption 3 (within-person monotonicity)is satisfied: the expected bliss points of current members of groups 1, 2, and 3 monotonically converge upwards to $b^{(\infty)} = 10$, and the expected bliss points of current members of group 4 monotonically converge downwards to this value.

The following theorem provides sufficient conditions for when all symmetric MPE are monotone in the sense of Definition 1.

Theorem B2 *Every symmetric MPE is monotone for generic parameter values if either of the following conditions holds:*

(i) The discount factor β is sufficiently low, provided that for any states s and $x \neq y$;

(ii) There is sufficiently little social mobility, in the sense that the matrix M is sufficiently close to the identity matrix.

Proof of Theorem B2. To establish (i), notice that if β is low enough, then in any equilibrium σ, $\{W_j(x)\}_{j\in G}^{x\in S}$ satisfies strict increasing differences. Thus, the result of Part 2 of Theorem 1 holds, and in any state s, the $y \in \Phi_x$ implies that $y \in \arg\max_{z\in S} W_{d_x}(z)$.

Suppose, to obtain a contradiction, that there is a nonmonotone equilibrium. Then for some states $x, y, a, b \in S$ such that $x < y$ and $a > b$, we have $a \in \Phi_x$, $b \in \Phi_b$. This means that $a \in \arg\max_{s\in S} W_{d_x}(s)$ and $b \in \arg\max_{s\in S} W_{d_y}(s)$, in particular, this implies $W_{d_x}(a) \geq W_{d_x}(b)$

and $W_{d_y}(b) \geq W_{d_y}(a)$. Note that we have

$$\left| W_j(a) - W_j(b) - \sum_{k \in G} \mu_{j,k} \left(u_k(a) - u_k(b) \right) \right| \leq \frac{\beta}{1-\beta} 2\bar{U}.$$

Taking $j = d_x$, this implies that $\sum_{k \in G} \mu_{d_x k} \left(u_k(a) - u_k(b) \right)$ cannot be negative, (otherwise the inequality would not hold for β small enough, and taking $j = d_y$, we get that $\sum_{k \in G} \mu_{d_y k} \left(u_k(a) - u_k(b) \right)$ cannot be positive. Since $b_{d_a} > b_{d_b}$, we have $b^{(1)}_{d_x} \geq \frac{b_{d_a} + b_{d_b}}{2} \geq b^{(1)}_{d_y}$. However, by Assumption 1, $x < y$ implies $b^{(1)}_{d_x} \leq b^{(1)}_{d_y}$, consequently, $b^{(1)}_{d_x} = \frac{b_{d_a} + b_{d_b}}{2} = b^{(1)}_{d_y}$. And yet, this equality does not hold for generic parameter values (in the sense of Part 4 of Theorem 1).

To establish (ii), fix β. Suppose that the statement is not true, then there are states $x, y, a, b \in S$ such that $x < y$ and $a > b$, and we have $a \in \Phi_x$, $b \in \Phi_b$ for equilibria for matrices M arbitrarily close to unity matrix. Since we can always choose a sequence of matrices $\{M_i\}$ that converges to unit matrix such that corresponding equilibria matrices $\{Q_i\}$ also converge to some Q (not necessarily satisfying the conditions $q_{xa} > 0$, $q_{yb} > 0$), we find that under M equal to unity matrix, there is an equilibrium where decision-makers in x (group d_x) weakly prefers transition to state a to transition to state b, and group d_y weakly prefer transition to state b. If matrix Q has monotone transitions, then $\{W_j(x)\}_{j \in G}^{x \in S}$ satisfies strict increasing differences by Lemma A2, and thus $a \in \arg\max_{s \in S} W_{d_x}(s)$ and $b \in \arg\max_{s \in S} W_{d_y}(s)$ cannot hold together, which is a contradiction. If matrix Q does not satisfy monotone transitions, then one can easily get a contradiction as in Theorem 7 and 8 of Acemoglu, Egorov, and Sonin (2015); indeed, under M, there is no social mobility, and the argument in that paper straightforwardly generalizes the case of nondeterministic transitions (details available upon request). ■

B6 Some results on social mobility matrices

We first formally state the result that if a social mobility matrix M satisfies (2) and (3), then there is a probability distribution over permutations of individuals that induces transition probabilities given by matrix M. In other words, any such matrix M is implementable, even with a finite number of individuals.

Lemma B2 ***(Corollary of Birkhoff-von Neumann Theorem)*** *If M is a $g \times g$ matrix that satisfies* (2) *and* (3)*, then there exist $n!$ nonnegative coefficients $\{\alpha_\pi\}_{\pi \in S_n}$ that sum to* 1 *such that for any $j, k \in N$,*

$$\mu_{g_j g_k} = \sum\nolimits_{\pi \in S_n : \pi(j) \in g_k} \alpha_\pi, \tag{B2}$$

where g_i is the group containing individual i.

Proof. Consider $n \times n$ matrix $\tilde{M}$, with rows and columns numbered by individuals, and defined by

$$\tilde{\mu}_{jk} = \frac{1}{n_{g_k}} \mu_{jk};$$

in other words, transition matrix $\tilde{M}$ postulates that individual j has an equal chance of taking the place of any individual from group k. Now, (2) implies $\sum_{k=1}^{n} \tilde{\mu}_{jk} = 1$ for all j, and (2) implies $\sum_{j=1}^{n} \tilde{\mu}_{jk} = 1$; consequently, $\tilde{M}$ is a doubly stochastic matrix. By Birkhoff-von Neumann Theorem (see, e.g., Theorem A2 in Marshall et al., 2011), $\tilde{M}$ lies in the convex hull of permutation matrices $\{P_\pi\}_{\pi \in S_n}$. This means that $\tilde{M}$ may be represented as a convex combination of these matrices:

$$\tilde{M} = \sum\nolimits_{\pi \in S_n} \alpha_\pi P_\pi.$$

This implies that $\tilde{\mu}_{jk} = \sum_{\pi \in S_n : \pi(j) = k} \alpha_\pi$ for any $j, k \in N$. Since $\tilde{\mu}_{jk} = \tilde{\mu}_{jl}$ for any k, l such that $g_k = g_l$, we have $n_{g_k} \tilde{\mu}_{jk} = \sum_{\pi \in S_n : \pi(j) \in g_k} \alpha_\pi$, which immediately implies (B2). ■

The next example illustrates that some matrices of social mobility may have multiple representations as a sum of permutation matrices.

Example B6 ***(Multiple representations of a mobility matrix as lottery over permutations)*** For a given A, the distribution μ such that $A = \Omega(\mu)$ need not be unique. E.g., take $n = 3$ and

$$A = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

It may be represented as

$$A = \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

which corresponds to three equally likely permutations id, (123) and (132), and

$$A = \frac{1}{3}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

which corresponds to three equally likely permutations (13), (12), (23).

Note that if a matrix satisfies conditions (2) and (3), then it takes the form of a block-diagonal matrix consisting of one or more blocks $\{K_x\}$. Each K_x is a connected block determining the extent of social mobility. (Assumption 1 requires that the blocks are connected.)

Lemma B3 ***(Characterization of matrices satisfying Assumption 3)*** *Suppose a $m \times m$ matrix M satisfies all the assumptions for all b. Then it satisfies within-person monotonicity if and only if it has the following structure: For each component K_x, corresponding to groups $H_{l_x}, \ldots, H_{r_x}$, there is a number $\kappa_x \in [0, 1]$, such that the transition probabilities for all groups except for the two extreme ones, i.e., for $l_x < j < l_y$, satisfy*

$$\mu_{jk} = \kappa_x \frac{n_k}{\sum_{i=l_x}^{l_y} n_i} + (1 - \kappa_x)\, \mathbf{1}_{j=k}. \tag{B3}$$

Proof. Sufficiency. Straightforward.

Necessity. Take any group H_j such that $l_x < j < l_y$. Let us show that for any $k_1, k_2 \neq j$, the probabilities μ_{jk_1} and μ_{jk_2} are proportional to the sizes of the groups: $\mu_{jk_1} n_{k_2} = \mu_{jk_2} n_{k_1}$. Suppose, to obtain a contradiction, the opposite, i.e., for some k_1 and k_2 this is not true. Without loss of generality, we may assume $k_1 < j < k_2$, and among such pairs, $k_2 - k_1$ is the maximal. For such k_2 and k_1, it is also true that $\left(\sum_{i=l_x}^{k_1} \mu_{ji}\right)\left(\sum_{z=k_2}^{l_y} n_z\right) \neq \left(\sum_{i=k_2}^{l_y} \mu_{ji}\right)\left(\sum_{z=l_x}^{k_1} n_z\right)$ (denote the difference right-hand side and left-hand side by Y).

Consider the following vector $\mathbf{b}^\varepsilon$ for each $\varepsilon > 0$:

$$(\mathbf{b}^\varepsilon)_i = \begin{cases} -\sum_{z=k_2}^{l_y} n_z + \varepsilon (i-j) & \text{if } l_x \leq i \leq k_1 \\ \varepsilon (i-j) & \text{if } k_1 < i < k_2 \\ \sum_{z=l_x}^{k_1} n_z + \varepsilon (i-j) & \text{if } k_2 \leq i \leq l_y \end{cases}$$

(outside of K_x, b_i are defined arbitrarily, subject to monotonicity). We have $(\mathbf{b}^\varepsilon)_j = 0$ for every ε. If we consider the $(M\mathbf{b}^\varepsilon)_j$, then as $\varepsilon \to 0$, we have $(M\mathbf{b}^\varepsilon)_j \to Y \neq 0$. Take δ_1 to be such that $\left|(M\mathbf{b}^\varepsilon)_j\right| > \frac{|Y|}{2}$ for $\varepsilon \leq \delta_1$. Now, observe that the sequence M^z converges, as $z \to \infty$, to a matrix M^∞ such that its elements satisfy

$$\mu_{jk}^\infty = \frac{n_k}{\sum_{i=l_x}^{l_y} n_i}.$$

This means that as $\varepsilon \to 0$, we have $(M^\infty \mathbf{b}^\varepsilon)_j \to -\left(\frac{\sum_{k=l_x}^{k_1} n_k}{\sum_{i=l_x}^{l_y} n_i}\right)\left(\sum_{z=k_2}^{l_y} n_z\right) + \left(\frac{\sum_{z=k_2}^{l_y} n_k}{\sum_{i=l_x}^{l_y} n_i}\right)\left(\sum_{z=l_x}^{k_1} n_z\right) = 0$. Thus, there is δ_2 such that $\left|(M^\infty \mathbf{b}^\varepsilon)_j\right| < \frac{|Y|}{2}$ for $\varepsilon \leq \delta_2$. Consequently, for $\varepsilon = \max(\delta_1, \delta_2)$, we have $0 = (\mathbf{b}^\varepsilon)_j < \left|(M^\infty \mathbf{b}^\varepsilon)_j\right| < \frac{|Y|}{2} < \left|(M\mathbf{b}^\varepsilon)_j\right|$. Since all inequalities are strict, there is $h : 1 < h < \infty$ such that this inequality holds if M^∞ is replaced by M^h. This implies that the subsequence $(\mathbf{b}^\varepsilon)_j, (M\mathbf{b}^\varepsilon)_j, \left(M^h \mathbf{b}^\varepsilon\right)_j$ is not monotone, a contradiction.

We have thus proved that $\mu_{jk_1} n_{k_2} = \mu_{jk_2} n_{k_1}$ for all $k_1, k_2 \neq j$, and thus there is $\kappa_x = \kappa_{x,j}$ such that μ_{jk} are given by (B3). The fact that these numbers are the same for each $j : l_x < j < l_y$ follows from Assumption 1 that M is assumed to satisfy. Indeed, if $\kappa_{x,j_1} < \kappa_{x,j_2}$ for $j_1 < j_2$, we would have $\mu_{j_1 l_x} < \mu_{j_2 l_x}$, and thus (4) would be violated for $q = l_x$; similarly, if $\kappa_{x,j_1} > \kappa_{x,j_2}$ for $j_1 < j_2$, then $\mu_{j_1 l_y} > \mu_{j_2 l_y}$, and thus (4) would be violated for $q = l_y - 1$. ■

Notice that Lemma B3 does not require the extreme groups in a given class to conform to the same formula given by (B3). For example, the following matrices satisfy (2), (3), as well as monotonicity across and within individuals:

$$\begin{pmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{pmatrix}, \begin{pmatrix} 3/5 & 2/5 & 0 & 0 \\ 1/5 & 2/5 & 1/5 & 2/5 \\ 1/5 & 1/5 & 2/5 & 1/5 \\ 0 & 0 & 2/5 & 3/5 \end{pmatrix}.$$