Social Mobility and Stability of Democracy: Re-evaluating De Tocqueville

Daron Acemoglu
MIT

Georgy Egorov
Northwestern University

Konstantin Sonin
University of Chicago

October 2015

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, peripheral coalitions between rich and poor segments of society may form because the rich are opposed to social mobility for economic reasons and the poor for political reasons—because it destabilizes democracy.

Keywords: Social mobility, institutions, stability, democracy, De Tocqueville, dynamics.
JEL Classification: D71, D74.

Preliminary and Incomplete.
Please Do Not Circulate

*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, and seminars at the University of Chicago, Higher School of Economics, MIT, and the University of Zurich for helpful comments.
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 has intuitive appeal. It also 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 this 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 why this idea is intuitive, but also why greater social mobility may actually destabilize democracy.

Example 1 Consider a society with \(n\) individuals, with \(\frac{3}{5}n\), 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-wing 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 at each date 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. Suppose, instead, that there is high social mobility, involving complete reshuffling of all individuals across the three social groups. If so, a rich individual expects to be part of the rich, the middle class, and the poor with probabilities $2/5, 1/5, \text{ and } 2/5$, respectively. His second-period expected utility is then $-\frac{2}{5} (p_2 + 1)^2 - \frac{1}{5} p_2^2 - \frac{2}{5} (p_2 - 1)^2 = -p_2^2 - \frac{4}{5} p_2 + \frac{2}{5}$. Thus, he prefers, in expectation, $p_2 = 0$. To achieve this, he would prefer next period’s political institutions to be democratic.

While this example illustrates a simple form of De Tocqueville’s hypothesis, it can also be used to highlight the other side of the political forces in play.

**Example 1 (continued)** Suppose next that society starts out as a democracy, where the decisive group is the middle class. Social mobility now corresponds to $r$ middle-class agents becoming rich and $r$ rich agents moving 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. (At the time decisions are made, who will stay and who will move is not known, so there is no asymmetry of information or conflict of interest within a group). Now, if sufficiently many middle-class members move upwards (i.e., if $\alpha > 1/2$), then middle-class agents, on average, expect to have the preferences of rich agents tomorrow, and hence prefer elite dictatorship tomorrow to democracy.

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

---

$^1$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 as having certain preferences or making certain decisions.
the politically pivotal middle class expect to change their preferences in a certain direction, they will have an incentive to change the institution in that direction as well.\textsuperscript{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 redistribution), then there would be no need for institutional changes. Second, it 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 most importantly, 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.\textsuperscript{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 $\rho$ rich agents move downwards. Let $\alpha = \frac{5r}{n}$ again denote the share of the middle class moving upwards.

In left-wing 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 political institutions. The middle class’ preferences, however, depend on their expectations of future institutions and of how future middle-class agents will behave. Suppose $1/4 < \alpha < 1/2$. 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, because social mobility makes her long-run preferences very similar to those of the current rich. But if today’s middle class expects a transition to elite dictatorship tomorrow, then it would prefer to remain in democracy. If, on the other hand, it expects the continuation of democracy, then it would prefer an immediate transition. This logic not only illustrates the interplay between the preferences and strategies of current and future ‘selves’ but also shows that there is

\textsuperscript{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 the middle class expected to move upwards or downwards symmetrically, then its members would continue to prefer democracy to other political regimes because they would lose in expectation even more from elite (or left-wing) dictatorship than they would gain.

\textsuperscript{3}The conflict between the different ‘selves’ of an individual, which is a central aspect of our model as discussed further in footnote 4, does not arise because of time-inconsistent preferences, but because of political economy reasons. Because an individual expects to change her social group in the future, she expects her future preferences to be different from the preferences of those who will occupy the same social station as herself at the moment.
no pure-strategy equilibrium in this case because of this same interplay.

Our baseline framework corresponds to a straightforward generalization of the setup discussed in this example. Society consists of a finite number of social groups, each of which comprises a finite number of identical individuals. Individuals (and thus groups) are ordered with respect to their policy preferences. Social mobility results from well-defined stationary probabilities specifying how each individual transitions from one social group to another. There is a finite set of alternative political institutions, which we refer to as ‘states’, and each state is represented by a set of weights assigned to individuals within each social group. 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 of course 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, equilibria may be in mixed strategies, we demonstrate that mixed strategy equilibria take a particular form: there is mixing only between keeping the current institution and transiting to a uniquely defined alternative. This property implies that the equilibrium direction of transition is always well defined, and different mixed strategies simply change the speed of transitions. Similarly, the interplay between different selves of the current pivotal voter can lead to multiple equilibria. Nevertheless, we establish 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 the horizon increases, an individual’s preferences will either gradually shift to the left or to the right, and enables consistent aggregation of the preferences of future selves.4

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

---

4Our analysis thus highlights the importance of two kinds of conflicts of interest: between agents with different economic interests today; and between today’s decision-maker and his future selves (meaning the agents who will occupy the future the same social group as the current decision-maker), which arises from the fact that today’s decision-maker anticipates to be in a different social class in the future. The latter conflict of interest 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 political power to agents with different preferences, whereas with social mobility, even with unchanged institutions, political power will be effectively delegated to agents with different preferences.
social mobility process $M$ than $M'$, if it is stable (respectively, asymptotically stable) under $M$ whenever it is stable (respectively, asymptotically stable) under $M'$. 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 far 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 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 and 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. One notable exception is Leventoglu (2005) who 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 arrange-

---

5 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).
ments, including Roberts (1999), Besley and Coate (1998), Bourguignon and Verdier (2000), Acemoglu and Robinson (2000 and 2001), Lizzeri and Persico (2004), Gomes and Jehiel (2005), Lagunoff (2006), and Acemoglu, Egorov, and Sonin (2010, 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), and Jackson and Yariv (2015), 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 an environment without the within-person monotonicity assumption. Section 6 endogenizes social mobility and studies how concerns about changes in future social mobility constrain equilibrium mobility decisions in a simplified version of our baseline model. It shows, most importantly, that because of social mobility’s impact on the stability of democracy, peripheral coalitions between rich and poor segments may form, with the aim of limiting equilibrium mobility. Section 7 concludes. Appendix A contains the proofs of results presented in the text, while Appendix B, which is not for publication, includes 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 $G_k$, $1 \leq k \leq g$, 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:

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

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).\textsuperscript{6} In what follows, $b = \{b_k\}$ will denote the column vector of political bliss points.

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 any state $s$ — namely, the group $d(s)$ such that $\sum_{k=1}^{d(s)} w_k(s) \frac{n_k}{n} \geq \frac{1}{2}$ and $\sum_{k=d(s)}^g w_k(s) \frac{n_k}{n} \geq \frac{1}{2}$ — is uniquely defined. Since, for our purposes, two states that have the same pivotal group are equivalent, we can without loss of any generality take $S$ to be a subset of $G$, so that each state corresponds to a different social group being pivotal. Without loss of generality, let us 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 individual’s offspring belonging to a different social group than herself (and the individual 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

\textsuperscript{6}For example, if all $A_k = 0$, then members of the middle class would not want to become rich if the political institution is democracy, because this will hurt policy payoff. This is inconsequential if social mobility is exogenous, but would lead to unrealistic predictions once we endogenize social mobility.
exogenous here, this assumption can simply be interpreted as supposing that there exists a stationary aggregate distribution and that we start the analysis once society has reached this stationary distribution.\footnote{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 structure of society, which would be continuously ongoing if the aggregate distribution of population across social groups did not remain constant.}

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

\begin{align*}
\sum_{k=1}^{g} \mu_{jk} &= 1 \text{ for all } j, \text{ and} \\
\sum_{j=1}^{g} n_j \mu_{jk} &= n_k \text{ for all } k,
\end{align*}

where the latter condition imposes the stationarity assumption requiring that the sizes of different groups remain constant. Within each group, there is no heterogeneity, and thus the stochastic process for social mobility is the same for each individual within the same social group.\footnote{Matrix $M$ can be equivalently defined by using permutations $\pi \in S_N$ of all individuals and assuming that in each period, Nature the changes identities of individuals according to $\pi$ with probability $\lambda_\pi$, such that $\sum_{\pi \in S_N} \lambda_\pi = 1$. The symmetry requirement then becomes $\lambda_\pi = \lambda_{\sigma \circ \tau}$ for any $\sigma, \tau \in S_N$ that reshuffle individuals within groups only. In this case,}

\begin{equation}
\mu_{jk} = \frac{1}{n_j} \sum_{i \in G_j} \sum_{\pi \in S_N, \pi(i) \in G_k} \lambda_\pi.
\end{equation}

The converse is also true: for any matrix $M = \{\mu_{jk}\}$ of nonnegative elements satisfying (2)–(3) there is a corresponding distribution $\lambda$ over permutations $\pi$ (this distribution may be not uniquely defined). This relatively minor generalization of Birkhoff-von Neumann theorem for doubly stochastic matrices is proved in Appendix C (see also Budish et al., 2013).
explicitly stating it. 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 $\mu_{j,k} = \frac{n_k}{n}$; 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.

### 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, $\pi_s : \{1, ..., g\} \to G$, determining the sequence in which (representatives of) different groups make proposals. The only requirement we impose is that group $d_s$ gets a chance to propose in state $s$, and given this, none of our results depend on this order.

We start with a given state, $s_0$, and default policy, $p_0$, in the first period. Thereafter, denoting the group that individual $i$ belongs to at time $t$ by $g^t_i$, the timing in each period $t \geq 1$ is as follows.

1. **Policy decision:**
   
   (a) In each state $s_t$, we start with $j = 1$.

   (b) A random agent $i$ from group $\pi_{s_t}(j)$ is chosen as the agenda setter and makes a policy proposal $p^t_j$. (Since all members of social groups have the same preferences, which agent is chosen to do this is immaterial).

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

   (d) If $\sum_{i=1}^n w_{g^t_i}(s_t) \mathbf{1}\{v^p_i(j) = Y\} > \frac{1}{2}$, then the current proposal is implemented: $p_t = p^t_j$, and the game moves to stage 2. Otherwise, the game returns back to stage 1(b) with $j$ increased by 1.

   (e) If for all $j \in \{1, \ldots, g\}$, the proposals are rejected, then the default (previous period’s policy) is implemented: $p_t = p_{t-1}$.

2. **Political decision:**

   (a) In each state $s_t$, we start with $j = 1$.

   (b) A random agent $i$ from group $\pi_{s_t}(j)$ is chosen as the agenda setter and makes a proposal of political transition, $s^t_{t+1}$. 

9
(c) All individuals vote, sequentially, with each individual \( i \) casting vote \( v_i^{s_t} (j) \in \{Y, N\} \).

(d) If \( \sum_{i=1}^{n} w_{g_i} (s) 1 \{v_i^{s_t} (j) = Y\} > \frac{1}{2} \), then next period’s state is the current proposal, \( s_{t+1} = s_{j+1}^t \), and the game moves to stage 3. Otherwise, the game proceeds to stage 2(b) with \( j \) increased by 1.

(e) If for all \( j \in \{1, \ldots, g\} \) the proposals are rejected, then there is no transition, i.e., \( s_{t+1} = s_t \).

3. *Payoffs:* Each individual \( i \) receives time-\( t \) payoff of \( u_{g_i} (p_t) \), 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 \( \mu_{g_i^t, k} \).

### 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 (Example B2 in Appendix B provides an example of non-monotonic equilibrium, and Theorem B1 presented there provides sufficient conditions for all equilibria to be monotone). Since, as shown in Example B3 in the Introduction, pure-strategy equilibria may fail to exist, throughout we allow for mixed strategies.

We next define our equilibrium concept more formally.

**Definition 1** *(Symmetric Monotone Markov Perfect Equilibrium)* A subgame perfect equilibrium \( \tilde{\sigma} \) is a Markov Perfect Equilibrium (MPE) if the strategy of each player \( i \), \( \tilde{\sigma}_i \), 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).

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

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

\[
\text{Pr} (s_\tau \leq l \mid s_t = x) \geq \text{Pr} (s_\tau \leq l \mid s_t = y) .
\]  

(5)

In what follows, we refer to symmetric monotone MPE simply as ‘equilibria’. Although equilibria formally correspond to a complete list of strategies, in what follows it will 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 \textit{stable}, if \( s_t = s \) implies that \( s_{t+1} = s \). We say that a state \( s \) is \textit{asymptotically stable} if \( s_t \in \{s-1, s, s+1\} \cap S \) implies that \( \lim_{t \to \infty} \Pr(s_{t} = s) = 1 \), in other words, if 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 (local) 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 MPE, asymptotic stability of a state implies stability. We also quantify the notion of stability by saying that a state becomes \textit{more stable} under a change in parameters, if it is stable whenever it was stable before the change and it is asymptotically stable whenever it was asymptotically stable before the change. The notion of \textit{less stable} is defined analogously.

3 Analysis

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

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 maximize the discounted utility of the current pivotal group and also that mixed strategies will take a particularly simple form.

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

\[
\arg\max_{x \in S} \sum_{j \in G} \mu_{d_s,j} V_j(x),
\]  

(6)
where \( \{V_j(x)\}_{j \in G}^{x \in S} \) satisfies

\[
V_j(x) = u_j(b_{d_x}) + \beta \sum_{y \in S} q_{x,y} \sum_{k \in G} \mu_{j,k} V_k(y).
\] (7)

3. The transitions induced by the equilibrium are strongly monotonic: if \( x < y \) and \( q_{x,a} > 0 \), \( q_{y,b} > 0 \) (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 \( q_{s,x} > 0 \) and \( q_{s,y} > 0 \) 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, \( d_s \). 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 simply 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 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 simply required first-order stochastic dominance of the equilibrium path when starting from a higher state. The result here instead establishes that starting from a higher state, there is never any transition to a lower state than those to which there are transitions 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: such equilibria involve mixing only between the current state and some other state. Mixed strategies arise only 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 also stay. This was a reflection of the fact that the

---

9 There is an analogous result in Roberts (1999) in a non-strategic environment (and without social mobility), and also in Acemoglu, Egorov and Sonin (2015) in a setting without social mobility.
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.\footnote{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 B2 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).}

One implication of this characterization is that even though there may be mixed strategies, this will not change the direction of transitions, but will only affect its speed.

To formulate our next result, it is useful to introduce additional notation to designate the policy preferences of tomorrow’s self. An agent currently belonging to group \(j\) will belong in \(\tau\) periods to group \(k\) with probability \(\mu^\tau_{jk}\), where \(\mu^\tau_{jk}\) is the corresponding element of the matrix \(M^\tau\). Therefore, the expected stage utility of an agent currently in group \(j\) in \(\tau\) periods if policy \(p\) were to be implemented at that point is

\[
\sum_{k=1}^{g} \mu^\tau_{jk} (A_k - (b_k - p))^2 = \left( \sum_{k=1}^{g} \mu^\tau_{jk} b_k - p \right)^2 + \left( \sum_{k=1}^{g} \mu^\tau_{jk} b_k \right)^2 + \sum_{k=1}^{g} \mu^\tau_{jk} (A_k - b_k^2).
\]

The last two terms 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 \(\tau\) periods given by

\[
b_j^{(\tau)} = \sum_{k=1}^{g} \mu^\tau_{jk} b_k = (M^\tau b)_j.
\]

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

**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 in pure strategies. Moreover, if in period \(t\) the state is \(s\), then the state in period \(t+1\) is \(z \in S\) that minimizes \(b_{dz} - b_{dz}^{(1)}\). 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, \(b_{dz}^{(1)}\).
2. 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 $|b_{ds} - b_{ds}^{(\infty)}|$.

The first result is straightforward: sufficiently myopic players in the pivotal group will choose the institution that maximizes the welfare of their immediate future selves. The second result is a little more subtle: if $\beta$ 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 $|b_{ds} - b_{ds}^{(\infty)}|$ is stable, so once the society gets there, it stays there forever. This can be established from the following intuitive argument: in the long run, the distribution of future selves of individuals from groups $ds$ and $dz$ is the same, and therefore their interests are aligned. So decision-makers from group $dz$ prefer to maintain state $z$, which is exactly what group $ds$, 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 $\beta$ is close to 1 because the agents 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 — the middle class expecting to move up in the near future due to ‘fast social mobility’, but also expecting to be much closer to the poor in the very far future due to ‘slow social mobility’ that affects the whole society.

**Example 4** Consider an infinite-horizon environment as in Example 2, but with the following changes: first, the discount factor, $\beta$, 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, and 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 of the
same ‘irreducible component’, 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 \( \left( \frac{5}{8} \times \frac{11}{20} + \frac{11}{20} \times \frac{5}{8} \right) \times 1 + \left( \frac{1}{10} \times \frac{19}{20} + \frac{11}{20} \times \frac{1}{10} \right) \times (-1) = \frac{1533}{3200} \times \frac{1}{2} \). In fact, thereafter their bliss policies decline monotonically towards zero, which is the preferences of the very distant future self of all agents, given by the likelihood of being in different social groups in the (ergodic) stationary distribution.

It can be verified that for \( \beta < 0.373 \), there is an equilibrium, in which democracy is stable. If \( 0.373 < \beta < 0.830 \), then both dictatorships are stable. If, on the other hand, \( 0.830 < \beta < 0.921 \), then left-wing dictatorship is stable and the rich democratize with a positive probability. Finally, if \( 0.921 < \beta < 1 \), then both dictatorships become democracies with positive probabilities in each period. 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.

Consider, however, an alternative strategy profile, which forms an equilibrium for all \( \beta \): the middle class immediately transitions to the elite dictatorship, and both dictatorships are stable. The reason why this strategy profile constitutes a best response for the middle class is that, because of fast social mobility upwards, the next period is the only one where having elite dictatorship is beneficial for them, but its members also recognize that in two periods, they will end up there anyway, because tomorrow’s middle class will implement a transition to elite dictatorship according to the equilibrium strategy. Therefore, for today’s middle class the issue is not whether to move to elite dictatorship, but when, and their preferences imply that a transition today is preferable to transition tomorrow. Clearly, elite dictatorship is stable in this equilibrium: the rich would benefit from democracy in the long run, but they know that any democratization will last for only one period, and is thus not beneficial. For the poor under left-wing dictatorship, democratization will result in elite dictatorship, which is even worse. Consequently, this strategy profile is an equilibrium for all values of \( \beta \), even though for \( \beta \) sufficiently high, it is Pareto dominated by the equilibrium described in the previous paragraph, where democracy is stable.\(^\text{11}\)

\(^\text{11}\)Pareto ranking is a feature of this example and does not apply in general. For example, for intermediate values of \( \beta \) there are multiple equilibria, but these are not Pareto ranked.

As may be expected, if \( \beta > 0.373 \), there is also a third equilibrium, where the middle
class moves to elite dictatorship with some probability, and both dictatorships are stable. Elite dictatorship is stable because if middle-class agents are willing to transfer power to the rich, then the rich a fortiori wish to stay in power. In left-wing dictatorship, on the other hand, the poor are unwilling to transition to democracy because democracy itself will make way to elite dictatorship.

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 will be in the same social group as herself and those in other social groups. Though this type of multiplicity can occur whenever there are mobilities at different speeds, a simple (though not necessarily weak) assumption is sufficient to rule it out. We present this assumption next.

Assumption 2 (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 \( b_k^{(\tau)} \geq b_k^{(\tau+1)} \) or \( b_k^{(\tau)} \leq b_k^{(\tau+1)} \) for \( \tau = 0, 1, \ldots \) with at least one strict inequality.

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 is more comfortable delegating to them.

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 generically uniquely within the class of symmetric monotone MPE) if either the discount factor \( \beta \) is sufficiently low or if Assumption 2 (within-person monotonicity) is satisfied.
That the equilibrium is generically unique when the players are very myopic (have 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 $\beta$ 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 long-run preferences, not their short-run incentives. As $\beta$ approaches 1, these short-run incentives become less and less important, but the coordination problem does not disappear.

The within-person monotonicity condition and its role in uniqueness become particularly intuitive if we view the problem of dynamics of institutions under social mobility as a problem of aggregation of preferences of all future selves. To understand when such aggregation has a well-defined solution, 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 also implies uniqueness of equilibrium — again because of the uniqueness of the weighted median voter in the presence of such well-defined preferences. This argument also provides a complimentary intuition for why within-person monotonicity is not needed when $\beta$ is sufficiently low: in this case, tomorrow’s self receives almost all of the weight, and the problem of aggregation of preferences of different future selves becomes moot.

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 \( \beta \) takes an arbitrary value). In this subsection, we state these corollaries under Assumption 2 to simplify the exposition. In Section 5, we provide analogs of these results without this assumption.

Theorem 2 implies that when \( \beta \) is very high, the preferences of very distant future selves \( b^{(\infty)} \) 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 other words, \( L_M(j) \) 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 (communicating class) of the Markov chain describing social mobility. Equivalently, the set of groups \( G \) can be partitioned into non-intersecting components: \( L_M(j) \cap L_M(k) \neq \emptyset \) if and only if \( L_M(j) = L_M(k) \). Also from Assumption 1, we have that each component is ‘connected’, that is, whenever \( k_1 < k_2 < k_3 \) and \( k_1, k_3 \in L_M(j) \), we have that \( k_2 \in L_M(j) \). This enables us to write the preferences of the current decision-maker, from group \( d_x \), in the very distant future as the average preferences of all agents within the same component:

\[
b^{(\infty)}_{d_x} = \frac{\sum_{k \in L_M(d_x)} n_k b_k}{\sum_{k \in L_M(d_x)} n_k}.
\]

Corollary 1 (Farsighted stability of institutions) Suppose that Assumption 2 holds. State \( s \in S \) is stable for sufficiently high \( \beta \) (formally, there exists \( \tilde{\beta} < 1 \) such that for any \( \beta \in (\tilde{\beta}, 1) \), \( q_{s,s} = 1 \)) if and only if

\[
s \in \arg\min_{z \in S} \left| b_{d_z} - b^{(\infty)}_{d_z} \right|.
\]

This result states that when players are sufficiently farsighted (the discount factor is sufficiently close to 1), a state is stable if and only if it guarantees the policy outcome closer to the average of the political bliss points of groups which the current decisions can move to, weighted by the sizes of those groups, than policy choice in any other state. Applying this result to the democratic institution, we get 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 any other policy which may be implemented under any institution. 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:\[^{12} \]

\[^{12} \text{To formally cover the cases in which the political institutions are the lowest and highest feasible ones, i.e., 1 and } g \text{ respectively, in what follows we set } b_0 = -\infty \text{ and } b_{g+1} = +\infty, \text{ which ensures that condition (10) is satisfied automatically for these lowest and highest political institutions.} \]
Corollary 2 (Farsighted stability of democracy) Suppose that Assumption 2 holds. Suppose also that group $x$ contains the median voter. Democracy is stable for sufficiently high $\beta$ if and only if
\[
\frac{b_{d_{x-1}} + b_{d_x}}{2} \leq b_{d_x}^{(\infty)} \leq \frac{b_{d_x} + b_{d_{x+1}}}{2}.
\] (10)

This corollary provides a simple, and as it will turn out powerful, characterization of the stability of democracy when the discount factor, $\beta$, is sufficiently close to 1. Intuitively, it requires that the preferences of the current median voter in the very distant future is closer to his own current preferences than the preferences of either neighboring group.\(^{13}\) When this is the case, the current median voter prefers to delegate future decisions to the future median voter. When it is not, he would like to empower a group other than the one containing the future 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 $\beta$, but also shapes the comparative statics of democracy with respect to the speed of social mobility (for any value of $\beta$).

A complementary interpretation of conditions (9) and (10) also helps with the intuition. Note from (8) that $b_{d_x}^{(\infty)}$ is the average bliss point within the component to which group $x$ belongs to. 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 $x \in \arg\min_{z \in S} |b_{d_x} - b_{d_x}^{(\infty)}|$ 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.

Definition 2 Suppose we have two matrices of social mobility $M$ and $M'$ with the same components (which implies that $b^{(\infty)} = b'^{(\infty)}$). Then, we say that social mobility is (weakly) faster under $M'$ than under $M$ if for each $j \in N$ and each $t \geq 1$, either $b_j \leq b_j^{(t)} \leq b_j^{(\infty)} \leq b_j^{(t)} = b_j^{(\infty)}$ or $b_j \geq b_j^{(t)} \geq b_j^{(\infty)} = b_j^{(\infty)}$, with at least one strict inequality.

\(^{13}\)This condition is equivalent to $|b_{d_x} - b_{d_x}^{(\infty)}| \leq |b_{d_{x-1}} - b_{d_x}^{(\infty)}|$ and $|b_{d_x} - b_{d_x}^{(\infty)}| \leq |b_{d_x+1} - b_{d_x}^{(\infty)}|$. 


This definition simply means that two matrices $M$ and $M'$ are comparable in terms of the speed of social mobility 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 emphasizes 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 $M$ (and assume that it satisfies 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 $\gamma' \geq \gamma$.

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 stability of democracy) Suppose that Assumption 2 holds. Suppose also that social mobility under $M'$ is faster than under $M$, and the 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 $M'$.

In the case where (10) holds, this theorem thus supports De Tocqueville’s hypothesis that social mobility contributes to the stability of democracy. Intuitively, when this condition holds, the long-run future self of the pivotal voter in democracy has higher utility in democracy than in any other political institution (and this was the reason why Corollary 1 applied). The same condition also ensures that faster social mobility increases the size of the basin of attraction of democracy. To understand this result, recall that faster social mobility implies that, for any $\beta$, 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 is the same for all groups (within the component), this also implies that the preferences of all social groups approach each other (in the limit where the matrix $M$ involves complete ‘reshuffling’, which is the fastest feasible level of social mobility, all groups will have the same preferences). Since, from condition (10),
the very distant self of the current decision-maker (and of all other groups) prefers democracy to any other political system, faster social mobility makes neighboring groups also more willing to prefer democracy to any other political system.

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, $\beta$, 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 stability of democracy)** Suppose that Assumption 2 holds. Suppose that social mobility under $M'$ is faster than under $M$. Suppose also that for $M$, the inequality (10) does not hold, but we have

$$\frac{b_d(x-2) + b_d(x-1)}{2} \leq b_d^{(\infty)}(x) \leq \frac{b_d(x+1) + b_d(x+2)}{2}. \quad (11)$$

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; faster social mobility makes the preferences of all agents closer to the preferences of their very distant future self. But because now (10) does not hold, these very distant preferences would induce a move away from democracy. So the faster is social mobility, the more likely are the preferences of the very distant self to dominate and induce the abandonment of 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 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.
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) prefers to move to state \( x \neq s \), 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.

**Theorem 6 (Slippery slopes)** Suppose that Assumption 2 holds. There exist \( 0 < \beta_0 < \beta_1 < 1 \) 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 always being in \( x \) cannot exceed their equilibrium continuation utility:

\[
\sum_{t=1}^{\infty} \sum_{k=1}^{g} \mu_{d_s}^k u_{d_s} (b_{d_s}) \geq \sum_{t=1}^{\infty} \sum_{k=1}^{g} \mu_{d_s}^k u_{d_s} (b_{d_s}).
\]

If, on the other hand, \( \beta \in (\beta_0, \beta_1) \), this need not be the case.

In other words, this result suggests that for both high and low \( \beta \), all stable states give higher expected utility to the current decision-maker (with the expectation taken with respect to the social mobility process) than any other state. When slippery slope considerations are important, this need not be the case: there may be a state providing a higher expected utility to the current decision-maker than the current state, but moving to this state would unleash another set of transitions that reduce the discounted continuation payoff of the current decision-maker. Theorem 6 shows that such slippery slope considerations arise only for intermediate values of \( \beta \). (Example A1 in Appendix A establishes the second part of the theorem.)
The intuition for why slippery slope considerations do not play a role for myopic players (with low $\beta$) is straightforward: such players care only about the next period’s state, so the subsequent moves do not modify their rankings over states. That these considerations do not arise for very farsighted players (with high $\beta$) is more interesting and perhaps surprising. Suppose a situation in which the current-decision-maker, who is pivotal in the current state $s$, prefers a different state, $x$, where by definition he will not belong to the pivotal group unless his preferences change due to social mobility. Such preferences are possible only when members of the current pivotal group have a positive probability of joining the group that is pivotal in state $x$ (and conversely, those in the group pivotal in state $x$ could move to the group that is pivotal in state $s$). An implication is that even though the distribution of political power in states $s$ and $x$ have a conflict of interest today, because of social mobility their preferences in the distant future will be similar. Thus 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 $x$, obviating slippery slope considerations. In contrast, with intermediate discount factors, the loss of control in the near future can trigger concerns about slippery slopes, encouraging the current decision-maker not to move in the direction of states that increased 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 Corollaries 1 and 2 and Theorems 4 and 5 under Assumption 2 (within-person monotonicity). We next provide direct generalizations of all four results when this assumption is relaxed. The substantive and intuitive economic content of these results are essentially identical. The analogs of Corollaries 1 and 2 are particularly simple.

**Corollary 3 (Farsighted stability of institutions without within-person monotonicity)** A state $s \in S$ is stable, in some equilibrium, for sufficiently high $\beta$ (formally, there exists $\bar{\beta} < 1$ such that for any $\beta \in (\bar{\beta}, 1)$ there exists an equilibrium $\sigma$ such that $q_s^\sigma = 1$) only if

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

Moreover, if $s$ is the unique state satisfying (9) (which holds generically), then condition (9) is also sufficient for the existence of an equilibrium where $s$ is stable (for $\beta$ sufficiently high).
Corollary 4 (Farsighted stability of democracy without within-person monotonicity)

Suppose that group $x$ contains the median voter. Then there exists $\beta < 1$ such that for $\beta \in (\beta, 1)$ there is an equilibrium where democracy is stable only if

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

The analogs of Theorems 4 and 5 are a little more involved because the language has to be adjusted for possible multiplicity of equilibria. Nevertheless, our next two results show that economic contents of these two theorems are still essentially the same.

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 inequality (10) holds for $M$ or $M'$. Then democracy is more stable for $M'$ than for $M$. More precisely, there are equilibria $\sigma$ (for $M$) and $\sigma'$ (for $M'$) such that democracy is stable in both these equilibria, and, furthermore, if it is asymptotically stable in some equilibrium under $M$, then there is an equilibrium under $M'$ where it is asymptotically stable.

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 that for $M$, the inequality (10) does not hold, but

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

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

6 Endogenous social mobility

In this extension, we allow the political choices to impact 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 $\gamma_P$, $\gamma_M$ and $\gamma_R$, 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. Finally, we further simplify the analysis by assuming that collective decisions about social mobility are made only once, at the beginning of the game.
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.

Finally, we normalize the preferences of the middle class, \( b_M = 0 \), and let \( b_P < 0 \) and \( b_R > 0 \) be the political bliss points of the poor and the rich, respectively, and also set \( A_M = 0 \) and assume that \( A_P < -b_P^2 \) and \( A_R > b_R^2 \); the latter assumptions merely say that even if the poor rule, it is better to be in the middle class than to be poor, and if the middle class rules, being rich is still better than being middle-class. 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.

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. Finally, in the last subsection we study the case in which there are frequent choices over social mobility (modeled as the speed of social mobility being chosen at the beginning of each period).

### 6.1 Preferences for social mobility

The next two propositions characterize the preferences of the three social groups over the speed of social mobility (assumed to be chosen once and for all).

We start with social mobility at the bottom (between the poor and the middle class). Let \( \theta^b \) 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 \( M_P b^b \). The values of \( \theta^b \) consistent with Assumption 1 are \( \theta^b \in [0, \theta^b_{\text{max}}] \), where \( \theta^b_{\text{max}} = \frac{1}{1 + \gamma_P} \).

**Proposition 1** (Endogenous mobility at the bottom) If \( \gamma_M > \gamma_P \), then a higher \( \theta^b \) makes the poor better off and the middle class worse off, while the rich are indifferent.

If \( \gamma_M < \gamma_P \), then a higher \( \theta^b \) makes the poor better off and the middle class worse off. The rich become weakly worse off as \( \theta^b \) increases: strictly worse off if \( \theta^b \) increases within the interval \[ \frac{1 - \beta}{2(1 + \gamma_P)} \beta \frac{1}{2} \] (the probability of transiting from \( d \) to \( p \) increases from 0 to 1 on this interval), and their utility is locally constant outside of this interval.
The poor always value social mobility at the bottom, both for economic reasons (this enables them to transition to a richer group, middle class) and for political reasons (it can lead to institutional change from democracy to dictatorship of the poor, provided that the middle class has fewer people than the poor). In contrast, the middle class, which stands to transition to a lower social class, dislikes of 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 $\theta^b$ (corresponding to higher social mobility), democracy becomes unstable, making way to a left-wing dictatorship. In this case, the rich lose out indirectly from greater social mobility — because it destabilizes democracy in favor of a left-wing dictatorship.

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

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

If $\gamma_M < \gamma_R$, then a higher $\theta^t$ makes the middle class better off, and the utility of the poor is monotonically decreasing on the interval $\left[\frac{1-\beta}{2-\beta\left(1+\frac{\gamma_M}{\gamma_R}\right)}, \frac{1}{2}\right]$, where the probability of transition to $r$ increases from 0 to 1, and is locally constant outside of the interval. The utility of the rich is monotonically decreasing in $\theta^t$ if $\beta^2 \frac{\gamma_M}{\gamma_R} \left(\frac{R_{\theta^t}}{b_{\theta^t}} + 1\right) \geq 4$, and is nonmonotone in $\theta$ otherwise.\(^{14}\)

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 if it makes democracy less stable, making way to an elite dictatorship, it makes the poor worse off indirectly. In contrast, social mobility at the top always benefits the middle class. By contrast, for the rich, there is now a trade-off: 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 institution in their favor. Theorem 2 describes how this trade-off is resolved: a marginal increase in the speed 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 $\theta^t$. The rich are more likely to

\[^{14}\text{More precisely, the utility of the rich is increasing on the interval}\]

$$\left\{\frac{1-\beta}{2-\beta\left(1+\frac{\gamma_M}{\gamma_R}\right)}, \min\left(\left(1-\beta\right)\left(\frac{\gamma_M}{\gamma_R}\left(\frac{R_{\theta^t}}{b_{\theta^t}} + 1\right) - \beta\left(1 + \frac{\gamma_M}{\gamma_R}\right)\right)^{-1}, \frac{1}{2}\right)\right\}, \text{ which is in this case nonempty.}\]

At the lower end of the interval, the middle class is indifferent between staying in democracy forever and transiting to $r$. The interval may extend all the way up to $\frac{1}{2}$, where transition to $r$ becomes immediate, or to an interior point, in which case the rich do not benefit from a faster transition as it is achieved by too fast social mobility.
benefit from social mobility if $A_R$ is small, i.e., when “inequality” between $M$ and $R$ is limited. 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, the inequality is high, it is more important for the rich to stay rich than to influence institutional change. A lower $b^2_R$, meaning less conflict of interest between middle class and rich, decreases the chance that the rich will benefit from mobility, because institutional (and policy) consequences are less important. The rich are also more likely to benefit from mobility if $\gamma_M/\gamma_R$ is low (so most rich agents will remain rich longer even with a higher 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 and for all times at the beginning of the game, and society starts in democracy. More specifically, the exact game in this case is as follows (where once again the exact sequence in which different groups make proposals is immaterial, and the default level of social mobility $\theta_0$ also does not matter for equilibrium):

0. Social mobility decision:

(a) A random agent $i$ from the rich is chosen as the agenda setter and makes a policy proposal $\theta^i \in [0, \theta_{max}]$.

(b) All individuals vote sequentially, with each agent $i$ casting vote $v^\theta_i(j) \in \{Y, N, A\}$.

(c) If $\sum_{i=1}^n w_{g_i}(d) \{v^\theta_i(j) = Y\} > \sum_{i=1}^n w_{g_i}(d) \{v^\theta_i(j) = N\}$, then the current proposal is implemented: $\theta = \theta^i$, and the game moves to stage 1 (described in Section 2). Otherwise, the game returns back to stage 0(a), where instead of the rich, a middle-class agent proposes, and if this is done, then a poor citizen proposes.

(d) If for all $j \in \{R, M, P\}$, the proposals are rejected, then some default social mobility ($\theta = \theta_0$ is implemented).

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 produces a unique winner (which is the analog of the Condorcet winner for plurality) and thus leads to a unique equilibrium as we show next.

**Proposition 3** If $\gamma_M > \gamma_P$, the unique equilibrium choice of social mobility at the bottom $\theta^b$ is 0. If $\gamma_M < \gamma_P$, then the unique equilibrium choice of social mobility at the bottom is...
\[ \hat{\theta}^b = \frac{1 - \beta}{2 - \left(1 + \frac{\gamma_M}{\gamma_P}\right)\beta}, \] which is higher for low \( \beta \) or for higher share 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. The larger of the two groups gets their preferred policy, as long as it is within the interests of the third group, the rich. The rich here are indifferent between any social mobility, as long as democracy remains stable. Thus, if the middle class is more numerous than the poor, there is no mobility, whereas if the middle class is thin, then the poor are able to ensure some mobility, up to a point where the middle class is indifferent between preserving democracy and abandoning it.

To understand this result and its comparative statics, observe that, when \( \gamma_M < \gamma_P \), social mobility is limited by a coalition of the middle class and rich. These forces also make the speed of social mobility decreasing in \( \beta \). The rich are concerned with the possibility of institutional change. For any fixed speed of social mobility, the middle class is more likely to deviate from democracy if it is sufficiently forward-looking. Then, the rich will put a tighter bound on the speed of social mobility that they will accept, and the poor will have to settle for that lower speed in order to get the support of the rich, which they need as no group constitutes a majority. The comparative statics with respect to \( \frac{\gamma_M}{\gamma_P} \) are also intuitive: once \( \gamma_M < \gamma_P \), an even smaller middle class means that current middle class members would spend quite a lot of time being poor for any given \( \theta^b \), and thus they are more willing to delegate to the poor. To prevent this, the rich would have to insist on lower mobility. Consequently, both very large middle class and very small middle class are bad for social mobility at the bottom, which is greatest if the middle class is sufficiently large to value democracy in the long run, but not too powerful to stop social mobility.

We next turn to the mobility at the top, which shows how peripheral coalitions (between the rich and the poor) can form because of the endogeneity of social mobility.

**Proposition 4** If \( \gamma_M > \gamma_R \), the equilibrium choice of social mobility at the top \( \hat{\theta}^t \) is \( \theta^t_{\text{max}} \). If \( \gamma_M < \gamma_R \) and

\[
\frac{A_R}{b_R^2} > 2 - \beta \left( \frac{\gamma_R}{\gamma_M} - 1 \right) + 1 \quad (12)
\]

holds, then \( \hat{\theta}^t = 0 \), and if (12) does not hold, then \( \hat{\theta}^t = \frac{1}{2} \). In the first two cases, democracy is stable, whereas in the third case it is immediately abandoned in favor of elite dictatorship. The condition (12) is more likely to hold if \( A_R \) is high, \( b_R \) is low, \( \beta \) is high, or \( \frac{\gamma_M}{\gamma_R} \) is high.

In other words, if the size of the middle class exceeds the size of the rich, the society will choose the highest social mobility possible, and yet democracy will be stable. If the middle class
is relatively small, the middle class will wish to have more rapid mobility, but the preferences of the rich are potentially nonmonotone. Their bliss point is either no mobility or just sufficient mobility to persuade the middle class to abandon democracy.

Condition (12) ensures that democracy remains stable. This condition is very intuitive. If $A_R$ is sufficiently high, which means that the rich-middle class inequality is high, then society will choose no mobility at the top, and the reason is that the rich value their position in the society high enough to risk social mobility. Similarly, if $b^2_R$ is low, then the conflict of interest between the middle class and the rich over policy is low enough, and changing the institution for a minor change in policy is not worthwhile. A higher share of the middle class makes social mobility less attractive for the rich because it corresponds to more rapid transitions to the middle class. (In contrast and by the same reasoning, when the rich are numerous and the middle class relatively few, the rich would be willing to undergo social mobility, because the time they will spend, on average, in the middle class will be limited.)

### 6.3 Joint dynamics of institutions and social mobility

In the previous subsection, we assumed that social mobility is decided once and for all, via a sequence of proposals and votings in the beginning of the game. Here, we turn to environment with more frequent decisions over social mobility, and assume instead that the speed of social mobility is decided at the beginning of every period.

Formally, we assume that the game of the previous subsection is played at the beginning of every period. We make a few small modifications to remove uninteresting multiplicities. First, while in democracy all the groups have equal weight, in elite dictatorship the rich have almost all the weight: they have $1 - 2\varepsilon$, and the middle class and the poor have $\varepsilon$ each (and, similarly, for the dictatorship of the poor), and $\varepsilon$ is taken to be small. The presence of this small “perturbation” $\varepsilon$ is used to break indifferences when the more powerful group is indifferent, and thus ensures uniqueness in the presence of plurality voting as described above. Second, the default social mobility in every period is taken to be the same $\theta_0$. (This assumption is adopted, since the alternative where the default is the previous period’s social mobility would make this variable part of the payoff-relevant state variable, complicating the analysis). We also continue to focus on in MPE in monotone strategies, which implies that the equilibrium choice of social mobility may only depend on the current institution.

While the complete analysis is involved, we can use a similar analysis to that developed so far to characterize the equilibria in this case.

As before, we start with social mobility at the bottom, and then characterize equilibrium social mobility starting from each possible state (rather than just focusing on democracy which
was feasible in social mobility decisions were made once and for all). We use the notation \( \tilde{\theta} \) for equilibrium social mobility choices in this case with frequent decisions to contrast with the equilibrium 1 social mobility decisions were made once and for all, \( \theta^* \).

**Proposition 5 (Joint dynamics of institutions and mobility at the bottom)** When in power, the poor always choose maximum mobility \( \tilde{\theta}_p^b = \theta_{max}^b \). When the rich are in power, they are indifferent, and the equilibrium social mobility is \( \tilde{\theta}_r^b = \theta_{max}^b \) if \( \gamma_M < \gamma_P \) and \( \tilde{\theta}_r^b = 0 \) otherwise. In democracy, on the other hand, we have: if \( \gamma_M > \gamma_P \), then \( \tilde{\theta}_d^b = 0 \) (and thus no equilibrium social mobility); otherwise, then \( \tilde{\theta}_d^b > \tilde{\theta}_d^b > 0 \). In either case, democracy is stable.\(^{16}\)

In left dictatorship, the poor would impose maximum mobility. In elites dictatorship, the rich are indifferent across different levels of social mobility at the bottom provided that this does not cause them to relinquish power. This implies that, in this case, equilibrium social mobility will be chosen by the more numerous of the two lower classes. More importantly for our focus here, the characterization of equilibrium social mobility in democracy is very similar to the case where mobility is decided once and for all. In particular, a large middle class will be able to enforce zero social mobility. In contrast, when the middle class is small (which is the last case considered in the proposition), there will again be equilibrium social mobility, which will be exactly at the speed that makes the middle class just indifferent between preserving and abandoning democracy. This last case also emphasizes one difference from the environment in which social mobility choices were infrequent: now the current middle class expects the equilibrium speed of social mobility to increase even further once transition to left dictatorship takes place, and this further encourages them to preserve democracy (and thus they are now willing to preserve democracy for more rapid levels of social mobility, encouraging the rich to tolerate more rapid social mobility at the bottom than in Proposition 3).

We next turn to the analysis of mobility at the top in the presence of frequent decisions over social mobility.

**Proposition 6 (Joint dynamics of institutions and mobility at the top)** When in power, the rich choose no mobility \( \tilde{\theta}_r^t = 0 \). When the poor are in power, they are indifferent, and the society chooses \( \tilde{\theta}_p^t = 0 \) if \( \gamma_M < \gamma_R \) and \( \tilde{\theta}_p^t = \theta_{max}^t \) otherwise.

In democracy, the choices are the following. If \( \gamma_M > \gamma_R \), then \( \tilde{\theta}_d^t = \tilde{\theta}_d^t = \theta_{max}^t \). If \( \gamma_M < \gamma_R \),

\(^{15}\)The proofs of the next two propositions are available upon request.

\(^{16}\)The exact value is \( \tilde{\theta}_d^t = \frac{1}{1+\frac{\gamma_M}{\gamma_P}} - \frac{2(1-\frac{\gamma_M}{\gamma_P})}{2-\beta\left(1+\frac{\gamma_M}{\gamma_P}\right)+\left(2-\beta\left(1+\frac{\gamma_M}{\gamma_P}\right)\right)^{-1}} \left(1+\frac{\gamma_M}{\gamma_P}\right)\left(\frac{\gamma_P}{\gamma_P}\right)\right)\right)\right)\right)\right).
but condition
\[
\frac{A_R}{b^2_R} > \frac{\gamma_R}{\gamma_M} + \frac{1}{\beta + (1 - \beta) \frac{\gamma_M}{\gamma_R}} - 1
\] (13)
(which is weaker than (12)) holds, then \(\tilde{\theta}_d^T = 0\), and if it does not hold, then \(\tilde{\theta}_d^T > \frac{1}{2}\). In the first two cases, democracy is stable, whereas in the third case it is immediately abandoned in favor of elite dictatorship.\(^{17}\)

This characterization is also very similar to the one in the presence of infrequent social mobility decisions (Proposition 4). The main difference is that when the middle class is relatively small, social mobility is less likely to be preferred by the rich (but if the rich do prefer social mobility, it will be faster than otherwise). Intuitively, the middle class is more reluctant to delegate to the rich, since such delegation will reduce future social mobility. Thus, it will only do so for a higher speed of social mobility. The rich are, therefore, less likely to choose this over no social mobility.\(^{18}\)

The results above imply the following takeaway. If the middle class is in power in democracy, then democracy is always stable against transition to left dictatorship. Transition to elite dictatorship may happen if the middle class is sufficiently small (\(\gamma_M/\gamma_R\) low); if inequality at the top is low (\(A_R\) low); and if there is substantial conflict of interest between the rich and the middle class (\(b_R\) high).

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 these issues using a dynamic model of political economy. Our framework provides a natural formalization of De Tocqueville’s hypothesis, showing that greater social mobility can further 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

\(^{17}\)More precisely, in this case \(\tilde{\theta}_d^T = \frac{1}{1 - \beta (\frac{A_R + v_M}{\gamma_R} + \frac{v_M}{\gamma_R}) + \frac{\frac{1}{\beta (\frac{A_R + v_M}{\gamma_R} + \frac{v_M}{\gamma_R}) + \frac{\beta (A_R + v_M)}{(1 - \beta) \gamma_M} (\gamma_M R + 1))}{1 - \beta (\frac{A_R + v_M}{\gamma_R} + \frac{v_M}{\gamma_R}) + \frac{\beta (A_R + v_M)}{(1 - \beta) \gamma_M} (\gamma_M R + 1))}}.

\(^{18}\)In addition, now the rich will be keener to obtain power in order to reduce future social mobility, though this effect is dominated by the one we emphasize in the text.
up (respectively down), she would prefer to give less voice to poorer (respectively richer) social
groups. We provided a tight characterization of these two competing forces, demonstrating 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 is 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 far apart, 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 institution may be disliked by the current decision-makers not only because
their future selves prefer another institution, 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.

We also characterized the conditions for slippery slope considerations — which prevent cer-
tain policy and institutional choices because of the further series of changes that these 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 potentially more
important.

Finally, we also showed how our results can be extended when society decides the extent
of social mobility. Our results here suggest the possibility of peripheral coalitions between the
poor and the rich. For example, when there is social mobility at the top (between the middle
class and the rich), the rich may dislike the prospect of moving down the social hierarchy, while
the poor may be concerned about the middle class abandoning democracy for elite dictatorship.
This paves the way for a poor-rich coalition aimed at decreasing social mobility at the top.

There are many fruitful areas of research related to the political implications of social mobil-
ity. First, there is a 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 and control over different aspects of social mobility among these groups). Third, this framework can also be enriched to include individual decisions, such as on the quantity or quality of education, which will affect a dynasty’s own future mobility and will impact both this individual’s and others’ political attitudes. Fourth, the framework we present here can also be generalized to include other 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 which often accompanies periods of rapid social mobility and changes the sizes of different social groups. An extension in this direction would be particularly interesting to see what types of structural changes contribute to the emergence and consolidation of democracy.
Appendix A: Proofs of Main Results

Proof of Theorem 1. We first prove Parts 1–4, and then use them and the intermediate results proved therein to show existence of a monotone equilibrium. (Notice that showing existence of any MPE is straightforward, as this is a direct application of Kakutani theorem.)

Proof of Part 1. In an MPE, the society’s decision on today’s policy will not affect the strategies in the continuation game or continuation utilities, apart from the utility from today’s policy decision. The preferences of each group and each individual over this decision are single-peaked (quadratic), and there is a unique policy in the core, namely, the policy \( b_{d_s} \) preferred by the effective median voter \( d_s \). Standard backward induction arguments (see, e.g., Acemoglu, Egorov, and Sonin, 2012) imply that no policy other than \( b_{d_s} \) may be chosen in equilibrium, which means that in this MPE, the society chooses \( b_{d_s} \) with probability 1. This proves Part 1.

Proof of Part 2. To prove this, we first establish an auxiliary result, which we will later use again.

Auxiliary result. Any monotone correspondence \( Q \) induces continuation utilities \( V_j(x) \), defined by (7), that satisfy the increasing differences condition: if \( j_1 < j_2 \) and \( x_1 < x_2 \), then \( V_{j_2}(x_2) - V_{j_2}(x_1) > V_{j_1}(x_2) - V_{j_1}(x_1) \).

Proof of auxiliary result. Consider the mapping from the set of continuation utilities \( \{V_j(x)\} \) onto itself, with the true continuation utilities \( \{V_j(x)\} \) being a unique fixed point (we use that in state \( s \), policy \( b_{d_s} \) is chosen, per Part 1):

\[
V_j^t(x) = u_j(b_{d_s}) + \sum_{y \in S} q_{x,y} \sum_{k \in G} \mu_{j,k} V_k(y). \tag{A1}
\]

To show that \( \{V_j^t(x)\}_{j \in G}^{x \in S} \) satisfies increasing differences, we use the following argument. The mapping defined by (A1) is a contraction. Thus, it suffices to prove that if \( \{V_j^t(x)\}_{j \in G}^{x \in S} \) satisfy increasing differences, then \( \{V_j'(x)\}_{j \in G}^{x \in S} \), defined by (A1), also do.

Take two states \( x, z \in S \) with \( x < z \). Monotonicity implies that \( \max \{y \in S : q_{x,y} > 0\} \leq \min \{y \in S : q_{z,y} > 0\} \). Thus, there is \( s \in S \) such that \( q_{x,y} > 0 \) implies \( y \leq s \) and \( q_{z,y} > 0 \) implies \( s \leq y \). Consider the following difference:

\[
V_j'(z) - V_j'(x) = (u_j(b_{d_s}) - u_j(b_{d_s})) + \beta \sum_{k \in H} \mu_{j,k} \left[ \sum_{y \in S} q_{x,y} V_k(y) - \sum_{y \in S} q_{x,y} V_k(y) \right] \\
= \left( (b_j - b_{d_s})^2 - (b_j - b_{d_s})^2 \right) + \beta \sum_{k \in H} \mu_{j,k} Z_k \\
= (b_{d_s} - b_{d_s}) (2b_j - b_{d_s} - b_{d_s}) + \beta \sum_{k \in H} \mu_{j,k} Z_k, \tag{A2}
\]
where we denoted
\[ Z_k = \sum_{y \in S} (q_{z,y} - q_{x,y}) V_k(y). \]

Let us prove that \( Z_k \) is weakly increasing in \( k \). Indeed, if we take two groups \( k, l \) with \( k < l \), then
\[
Z_l - Z_k = \sum_{y \in S} (q_{z,y} - q_{x,y}) (V_l(y) - V_k(y))
\]
\[ = \sum_{y \in S} q_{z,y} (V_l(y) - V_k(y)) - \sum_{y \in S} q_{x,y} (V_l(y) - V_k(y)) = 0,
\]
but \( V_l(y) - V_k(y) \) is monotonically increasing in \( y \) due to the assumption of increasing differences. Therefore, the expectation of this function under the probability distribution \( \{q_{z,}\} \) (the first term) is at least as high as that under the probability distribution \( \{q_{x,}\}, \) which it weakly first-order stochastically dominates (the second term). Consequently, \( Z_l - Z_k \geq 0 \), meaning that \( Z_k \) is weakly increasing in \( k \).

Going back to (A2), observe that the first term \( (b_d \delta - b_d \delta) (2b_l - b_d \delta - b_d \delta) \) is increasing in \( \delta \), because \( b_l \) is increasing in \( \delta \) and the difference \( b_d \delta - b_d \delta \) is positive. To show that the second term is non-decreasing in \( \delta \), take two groups, \( j, l \in G \) such that \( j < l \). By assumption, the probability distribution \( \{\mu_{j,}\} \) is first-order stochastically dominated by \( \{\mu_{l,}\} \). Then the expected values of a monotone sequence \( \{Z_k\} \) satisfy the inequality
\[
\sum_{k \in H} \mu_{j,k} Z_k \leq \sum_{k \in H} \mu_{l,k} Z_k.
\]
This proves that the second term \( \sum_{k \in H} \mu_{j,k} Z_k \) is non-decreasing in \( \delta \). Therefore, (A2) is increasing in \( \delta \), which implies that \( \{V_{j'}(x)\}_{x \in S}^{x \in G} \) satisfies increasing differences. Given that (A1) is a contraction, this completes the proof that \( \{V_{j'}(x)\}_{x \in S}^{x \in G} \) satisfies increasing differences.

**Finishing the proof of Part 2.** When deciding on the next state, an individual from group \( j \) acts as to maximize the expected continuation value, which in this case depends on both this decision and on the social mobility transformation that occurs in the end of the period. Thus, this individual maximizes
\[
w_j(x) = \sum_{k \in H} \mu_{j,k} V_k(x).
\]
These \( \{w_j(x)\}_{x \in S}^{x \in G} \) satisfy (weak) increasing differences. Indeed, for \( y > x \),
\[
w_j(y) - w_j(x) = \sum_{k \in H} \mu_{j,k} (V_k(y) - V_k(x)),
\]
and since \( V_k(y) - V_k(x) \) is monotonically increasing in \( k \), its expectation with respect to distribution \( \{\mu_{l,}\} \) is at least as high as that with respect to distribution \( \{\mu_{j,}\} \) if \( l > j \). In this
case, standard backward induction arguments imply that \( x \) will be chosen so as to maximize the expected continuation utility of the effective median voter, \( w_{ds} (x) \). This completes the proof of Part 2.

**Part 3.** This result holds for all parameter values if matrix of social mobility \( M \) satisfies strict first-order stochastic dominance; otherwise it holds for generic parameter values. Suppose, to obtain a contradiction, that \( x < y \), \( a > b \), and yet \( q_{x,a} > 0 \), \( q_{y,b} > 0 \). This means that \( a \in \arg\max_{s \in S} w_{ds} (s) \) and \( b \in \arg\max_{s \in S} w_{dy} (s) \), in particular, this implies \( w_{ds} (a) \geq w_{ds} (b) \) and \( w_{dy} (b) \geq w_{dy} (a) \). But if \( M \) satisfies strict first-order stochastic dominance, then \( \{ w_{j} (x) \}_{j \in G} \) satisfy strict increasing differences. We get a contradiction that completes the proof.

**Part 4.** This easily follows from genericity considerations, and the proof is omitted.

**Proof of existence of a monotone equilibrium.** Consider the following mapping from the set of continuation utilities \( \{ V_{j} (x) \}_{j \in G} \) that satisfy (weak) increasing differences onto itself. We restrict attention to a sufficiently large compact, where \( V_{j} (x) \leq \frac{1}{1-\beta} M \), with \( M \) defined as

\[
M = \max_{k \in G, y \in S} |u_{k} (b_{dy})|.
\]  

(3.3)

For each state \( s \in S \), consider a one-period game described in Section 2 and consider the set of all its Nash equilibria in mixed strategies, if the utility of an agent in group \( j \) is given by

\[
u_{j} (p) + \beta \sum_{k \in G} \mu_{j,k} V_{k} (y), \]

where \( p \) is the policy they agree upon and \( y \) is the next period’s state they decide on. From the proof of Part 1 it follows that \( p = b_{ds} \) for all such equilibria. From the proof of Part 2 it follows that \( y \) maximizes the expected continuation utility \( w_{j} (s) = \sum_{k \in G} \mu_{ds,k} V_{k} (y) \), and from the proof of Part 3 it follows that the transition mapping is monotone in each equilibrium. From the auxiliary result proved in the proof of Part 2 it now follows that for every combination of Nash equilibria (for different \( s \in S \)) of this one-period game, the continuation utility satisfies increasing differences. Now, Kakutani’s theorem implies that there exists a vector \( \{ V_{j} (x) \}_{j \in G} \) and Nash equilibria (for which correspond to a fixed point of the correspondence from continuation utilities to itself (it is standard to verify that other requirements are satisfied as well). Clearly, this set of Nash equilibria corresponds to a monotone Markov Perfect equilibrium of the original dynamic game. This completes the proof.

**Lemma A1** Theorem 1 continues to hold if we impose the following constraints on equilibrium transitions: \( q_{s,x} > 0 \) only if \( (x,s) \in F \), where \( F \subset S \times S \) satisfies: (a) for each \( s \), \( (s,s) \in F \) and (b) if \( a < b < c \) or \( a > b > c \), \( (a,c) \in F \) implies \( (a,b) \in F \) and \( (b,c) \in F \).
Proof of Lemma A1. The proof almost literally repeats the proof of Theorem 1 and is omitted. See also Acemoglu, Egorov, and Sonin (2015) for explicit proofs under a similar restriction on transitions. ■

Lemma A2 Suppose that for a monotone transition correspondence \( \{q_{x,y}\}_{x,y \in S} \) there exist \( x, y \in S \) such that for any \( a \) satisfying \( q_{x,a} > 0 \), \( w_{dx}(y) > w_{dx}(a) \). Then there exist \( x, y, a \in S \) such that \( q_{x,a} > 0 \), \( w_{dx}(y) > w_{dx}(a) \) and, in addition, the correspondence \( q': S \to S \) given by

\[
q'_{s,b} = \begin{cases} 
q_{s,b} & \text{if } s \neq x, \\
1 & \text{if } s = x \text{ and } b = y, \\
0 & \text{if } s = x \text{ and } b = y 
\end{cases} \tag{A4}
\]

is monotone.

Proof of Lemma A2. In this proof, \( A < b \) means \( a < b \) for all \( a \in A \), and \( A < B \) means \( a < b \) for all \( a \in A, b \in B \). Suppose, to obtain a contradiction, that for any \( x, y, a \in S \) such that \( q_{x,a} > 0 \) and \( w_{dx}(y) > w_{dx}(a) \), \( q' \) given by (A4) is not monotone. Take \( x, y \in S \) such that \( |y - a| \) is minimal among all tuples \((x, y, a)\) such that \( a \in \Phi(x) \) and \( w_{dx}(y) > w_{dx}(a) \) (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

\[
s = \min \{z \in S : b > y \text{ for some } b \in \Phi(z)\} = \min \{z \in S : \Phi(z) \nsubseteq y\};
\]

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 a deviation in state \( s \) to \( y \) is monotone: indeed, there is no state \( \tilde{z} \) such that \( \tilde{z} < s \) and \( y \nsubseteq \Phi(\tilde{z}) \leq \Phi(s) \) by construction of \( s \), and there is no state \( \tilde{z} > s \) such that \( \Phi(s) \leq \Phi(\tilde{z}) \nsubseteq y \) as this would contradict that \( y < b \) for some \( b \in \Phi(s) \) (indeed, the latter would imply \( \Phi(\tilde{z}) > y \)). By assertion, this deviation cannot be profitable for decision-makers in \( s \), meaning that \( w_{dx}(y) \leq w_{dx}(b) \) for any \( b \in \Phi(s) \), and in particular, for some \( b \) such that \( y < b \) (such \( b \) exists by definition of \( s \)). Since \( \{w(\cdot)\} \) satisfies weak increasing differences and \( d_s < d_x \), for this \( b \) we have \( w_{dx}(y) \leq w_{dx}(b) \).

On the other hand, deviation from \( x \) to \( y \) is profitable, so \( w_{dx}(y) > w_{dx}(a) \) for any \( a \in \Phi(x) \). We therefore have

\[
w_{dx}(b) \geq w_{dx}(y) > w_{dx}(a),
\]

37
so \( w_{ds}(b) > w_{ds}(a) \). Notice, however, that \( y < b \leq a \) for all \( a \in \Phi(x) \), and strict inequality \( w_{ds}(b) > w_{ds}(a) \) implies \( a \neq b \), so in fact \( y < b < a \). This implies that \( |b - a| < |y - a| \). This contradicts the choice of \( y \) such that \( |y - a| \) is minimal among tuples \((x, y, a)\) such that \( a \in \Phi(x) \) and \( w_{ds}(y) > w_{ds}(a) \). This contradiction proves that our initial assertion was wrong, and this proves the lemma. \( \blacksquare \)

**Proof of Theorem 2.** Part 1. Let \( \beta_0 \) be defined by \( \beta_0 = \frac{\zeta}{\zeta + 2M} \), where

\[
\zeta = \min_{s, y, z \in S, |b_{dy} - b_{dy}^{(1)}| > |b_{dz} - b_{dz}^{(1)}|} \left( \left( b_{dy} - b_{dy}^{(1)} \right)^2 - \left( b_{dz} - b_{dz}^{(1)} \right)^2 \right),
\]

where \( M \) is defined by (A3). Suppose that this is not the case, i.e., for some \( s \in S \), a transition to a state \( z \) which does not minimize \( |b_{dz} - b_{dz}^{(1)}| \) occurs. This means that for some \( y \in S \), \( |b_{dy} - b_{dy}^{(1)}| < |b_{dz} - b_{dz}^{(1)}| \). Now consider the utility of individuals from group \( d_s \) if they transited to \( y \) instead. Their gain in utility (after factor \( \beta \)) would be

\[
w_{ds}(y) - w_{ds}(z) = \sum_{k \in H} \mu_{d_s,k} (V_k(y) - V_k(z))
\]

\[
= \sum_{k \in H} \mu_{d_s,k} \left( A_k - (b_k - b_{dy})^2 - A_k + (b_k - b_{dz})^2 \right) + \beta (\ldots)
\]

\[
\geq \sum_{k \in H} \mu_{d_s,k} \left( (b_k - b_{dz})^2 - (b_k - b_{dy})^2 \right) + \frac{\beta}{1 - \beta} 2M
\]

\[
= (b_{dy} - b_{dz}) \sum_{k \in H} \mu_{d_s,k} \left( 2b_k - b_{dy} - b_{dz} \right) + \frac{\beta}{1 - \beta} 2M
\]

\[
= (b_{dy} - b_{dz}) \left( 2b_{dy}^{(1)} - b_{dy} - b_{dz} \right) + \frac{\beta}{1 - \beta} 2M
\]

\[
= \left( b_{dy}^{(1)} - b_{dz} \right)^2 - \left( b_{dy}^{(1)} - b_{dz} \right)^2 + \frac{\beta}{1 - \beta} 2M > 0,
\]

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 \( Z_s = \arg \min_{z \in S} \left| b_{dz} - b_{dz}^{(\infty)} \right| \); this set is either a singleton or consists of two adjacent states. The result follows from the following three steps.

**Step 1.** Denote

\[
\xi = \min_{s, y, z \in S, |b_{dy} - b_{dy}^{(\infty)}| > |b_{dz} - b_{dz}^{(\infty)}|} \left( \left( b_{dy} - b_{dy}^{(\infty)} \right)^2 - \left( b_{dz} - b_{dz}^{(\infty)} \right)^2 \right),
\]

\[
\Xi = b_m - b_1,
\]

and take \( \varepsilon = \frac{\xi}{\Xi} \). For such \( \varepsilon \) there exists \( T \geq 1 \) such that for all \( s \in S \) and \( t > T \), \( \left| b_{dz}^{(t)} - b_{dz}^{(\infty)} \right| < \varepsilon \). Let \( \tilde{\beta} = \left( 1 - \frac{\xi}{\Xi} \right)^{1/T} \). Then for \( \beta \in (\tilde{\beta}, 1) \), if for \( s \in S \), state \( z \in Z_s \) is stable, 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:

$$w_{d_s}(z) - w_{d_s}(x) = \sum_{k \in H} \mu_{d_s,k} (V_k(z) - V_k(x))$$

$$= \sum_{t \geq 1} \sum_{k \in H} \sum_{y \in S} \beta^{t-1} \mu_{d_s,k} \Pr(s_t = y) \left(A_k - (b_k - b_{d_s})^2 - A_k + (b_k - b_{d_y})^2\right)$$

$$= \sum_{t \geq 1} \sum_{k \in H} \sum_{y \in S \setminus Z_s} \beta^{t-1} \mu_{d_s,k} \beta^{t-1} M_{d_s,k} \Pr(s_t = y) \left((b_k - b_{d_y})^2 - (b_k - b_{d_z})^2\right)$$

$$= \sum_{t \geq 1} \sum_{k \in H} \sum_{y \in S \setminus Z_s} \beta^{t-1} \Pr(s_t = y) \left(b_{d_s} - b_{d_y}\right) \left(2b_{d_y} - b_{d_z}\right)$$

$$= \sum_{t \geq 1} \sum_{k \in H} \sum_{y \in S \setminus Z_s} \beta^{t-1} \Pr(s_t = y) \left(b_{d_s} - b_{d_y}\right) \left(2b_{d_y} - b_{d_z}\right)$$

$$= \sum_{t \geq 1} \sum_{k \in H} \sum_{y \in S \setminus Z_s} \beta^{t-1} \Pr(s_t = y) \left(b_{d_s} - b_{d_y}\right) \left(2b_{d_y} - b_{d_z}\right)$$

$$\geq \frac{\beta}{1 - \beta} \xi - 2\frac{\beta(1 - \beta T)}{1 - \beta} \Xi^2 - 2\frac{\beta T + 1}{1 - \beta} \Xi$$

$$= \frac{\beta}{1 - \beta} \left(\xi - 2(1 - \beta T) \Xi^2 - 2\beta T \Xi \xi\right)$$

$$> \frac{\beta}{1 - \beta} \left(\xi - 2(1 - \beta T) \Xi^2 - 2\Xi \xi \frac{\xi}{4 \Xi}\right) = 0.$$ 

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

**Step 2.** Suppose that $\beta$ is sufficiently high, and in some equilibrium, for state $s \in S$, at least one of the states $z \in Z_s$ is stable. Then such state is reached from $s$ with probability 1.

**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 $q_{s,y} = 0$ for $y > z$ due to monotonicity. On the other hand, from Step 1 it follows that $q_{s,x} = 0$ for $x < s$: 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, $s$ is less than all states in $Z_s$. From Step 1 it follows that $q_{s,x} = 0$ for $x < s$. If $q_{s,y} > 0$ for $y > s$, then by monotonicity $y \leq z$, where $z \in Z_s$ and is stable, and the same is true for all possible paths starting 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 remaining possibility is, $q_{s,s} = 1$, so $s$ is stable. But this impossible per Step 1. This proves that a stable state from $Z_s$ is reached with probability 1.

**Step 3.** For sufficiently high $\beta$, there exists an equilibrium such that for each state $s \in S$, at least one of the states $z \in Z_s$ is stable: $q_{z,z} = 1$.

**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, $b_{d_s}^{(\infty)} = b_{d_s}^{(\infty)}$, and thus $Z_s = Z_s$. 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.

Let us now make use of Lemma A1. Suppose first that $Z_s$ is a singleton \{z\}. Then define 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 $x = y = z$ (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). By Lemma A1, this game has an equilibrium, with a corresponding transition matrix $\tilde{Q}$. By construction, $\tilde{q}_{z,z} = 1$. If there exists an 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 A2 there must exist a monotone deviation, namely, states $x, y, a \in S$ such that $\tilde{q}_{x,a} > 0$, $w_{d_x} (y) > w_{d_x} (a)$ and, in addition, the correspondence $q' : S \rightarrow S$ defined by (A4) (replacing $q$ with $\tilde{q}$) is monotone.

Notice that it must be that $x = z$. Indeed, if not, then without loss of generality assume $x > z$, then 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 an MPE. Thus, $x = z$, and then $a = x = z$ ($\tilde{q}_{z,a} > 0$ implies $a = z$). Now, $w_{d_x} (y) > w_{d_x} (a)$ implies $w_{d_x} (y) > w_{d_x} (z)$, so $y \neq z$. Without loss of generality, assume $y > z$. But by monotonicity of this deviation, we must have $\tilde{\Phi} (y) \geq y$, and therefore all paths that start from $y$ never reach $z$. But then $w_{d_x} (y) > w_{d_x} (z)$ contradicts Step 1, because, as argued above, $Z_s = Z_s = \{z\}$. 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 $x, y \in Z_s$. By Lemma A1 there is an equilibrium $\sigma'$ that gives rise to a transition mapping $Q'$. Since from $z$ and $z'$ it is possible to transit onto this set only, 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 MPE in the original game, the result is proven. Otherwise, as before, by Lemma A2 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 however, $x \in Z_s = \{z, z'\}$. A deviation within $Z_s$ (i.e., $y \in Z_s$) cannot be profitable 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. ■

**Lemma A3** Suppose that for some $j$, the sequence $b_j^{(t)}$ is non-decreasing (respectively, non-increasing). Then in state $s$ where $d_s = j$, $x < s$ (respectively, $x > s$) implies $q_{s,x} = 0$.

**Proof.** Suppose that $b_j^{(t)}$ is non-decreasing (the complementary case is considered similarly). Suppose, to obtain a contradiction, that for some $x < s$, $q_{s,x} > 0$. Without loss of generality, assume that $x$ is the minimal state with $q_{s,x} > 0$. Notice that for any $y \in S$, we have

$$
\beta V_j(y) = \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} q_{y,a}^{(t)} \sum_{k \in G} \mu_{j,k}^{(t)} u_k(b_{da})
$$

$$
= \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} q_{y,a}^{(t)} \sum_{k \in G} \mu_{j,k}^{(t)} \left(A_k - (b_k - b_{da})^2 \right)
$$

$$
= \sum_{t=1}^{\infty} \sum_{k \in G} \beta^t \mu_{j,k}^{(t)} A_k - \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} q_{y,a}^{(t)} \mu_{j,k}^{(t)} (b_k - b_{da})^2.
$$
Now take any two states $y < z$ and consider the difference $V_j(z) - V_j(y)$:

$$
\beta (V_j(z) - V_j(y)) = \sum_{t=1}^{\infty} \beta^t \left( \sum_{a \in S} \sum_{k \in G} q_{y,a}^{(t)} \mu_{j,k}^{(t)} (b_k - b_{da})^2 - \sum_{a \in S} \sum_{k \in G} q_{z,a}^{(t)} \mu_{j,k}^{(t)} (b_k - b_{da})^2 \right)
$$

$$
= \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \left( q_{y,a}^{(t)} - q_{z,a}^{(t)} \right) \sum_{k \in G} \mu_{j,k}^{(t)} (b_k - b_{da})^2
$$

$$
= \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \left( q_{y,a}^{(t)} - q_{z,a}^{(t)} \right) \left( \sum_{k \in G} \mu_{j,k}^{(t)} b_k^2 - 2 \sum_{k \in G} \mu_{j,k}^{(t)} b_k b_{da} + \sum_{k \in G} \mu_{j,k}^{(t)} b_{da}^2 \right)
$$

$$
= \sum_{t=1}^{\infty} \beta^t \left( \sum_{k \in G} \mu_{j,k}^{(t)} b_k^2 \sum_{a \in S} \left( q_{y,a}^{(t)} - q_{z,a}^{(t)} \right) + \sum_{a \in S} \left( q_{y,a}^{(t)} - q_{z,a}^{(t)} \right) \left(-2b_{j}^{(t)} b_{da} + b_{da}^2\right) \right)
$$

$$
= \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \left( q_{y,a}^{(t)} - q_{z,a}^{(t)} \right) \left(-2b_{j}^{(t)} b_{da} + b_{da}^2\right)
$$

$$
= \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \left( q_{y,a}^{(t)} - q_{z,a}^{(t)} \right) \left(b_{j}^{(t)} - b_{da}\right)^2.
$$

Applying this to $x$ and $s$, we have

$$
\beta (V_j(s) - V_j(x)) = \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \left( q_{s,a}^{(t)} - q_{x,a}^{(t)} \right) \left(b_{j}^{(t)} - b_{da}\right)^2.
$$

Consider two cases. The first case is where $q_{s,a} > 0$ implies $a \leq s$; this holds generically by Part 4 of Theorem 1 (indeed, $q_{s,a} > 0$ and $x < s$). In that case, $b_{j}^{(t)} \geq b_j \geq b_{da}$ for all $a \leq s$, so $b_{j}^{(t)} \geq b_{da}$ and thus $\left(b_{j}^{(t)} - b_{da}\right)^2$ is increasing in $a$ for $a \leq s$; consequently, for each $t$,

$$
\sum_{a \leq s} q_{s,a}^{(t)} \left(b_{j}^{(t)} - b_{da}\right)^2 \geq \sum_{a \leq s} q_{x,a}^{(t)} \left(b_{j}^{(t)} - b_{da}\right)^2,
$$

because the distribution $q_{s,a}^{(t)}$ first-order stochastically dominates $q_{x,a}^{(t)}$ as the equilibrium is monotone. This implies $V_j(s) \geq V_j(x)$. A closer inspection suggests that the inequality is strict: e.g., for $t = 1$, the probability distributions $q_s$ and $q_x$ are different, and $\left(b_{j}^{(t)} - b_{da}\right)^2$ is strictly increasing in $a$. Thus, $V_j(s) > V_j(x)$, which contradicts Part 2 of Theorem 1 in that $x$ does not maximize the utility of group $j = d_s$. Notice that for 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 $y > s$, $q_{s,y} > 0$. This case is nongeneric, but the statement holds here as well. Consider $V_j(s)$; it is a linear combination of paths where the society stays in $s$ for $\tau \geq 1$ periods and then departs either to lower or higher states. All
equilibrium paths \( \{s_t\} \) where the departure is to lower states satisfy \( V_j(z \mid s_t \leq z) > V_j(x) \), similarly to the previous case. Now consider some path which departs to higher states, and suppose that it stays in \( z \) for exactly \( \tau \) periods, after which it departs to \( y > s \). Let us denote the probability distribution of states if an immediate transition to \( x \) occurs by \( p_{x,y}^{(t)} \), and that in the case an immediate transition to \( y \) occurs by \( q_{y,z}^{(t)} \). We know that the individuals in group \( j \) are indifferent between transiting to \( x \) and to \( y \), meaning that

\[
\sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} p_{x,a}^{(t)} \mu_{j,k}^{(t)} (b_k - b_{da})^2 = \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} q_{y,a}^{(t)} \mu_{j,k}^{(t)} (b_k - b_{da})^2,
\]

which, by increasing differences, implies

\[
\sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} p_{x,a}^{(t)} \mu_{j,k}^{(t+\tau)} (b_k - b_{da})^2 \leq \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} q_{y,a}^{(t)} \mu_{j,k}^{(t+\tau)} (b_k - b_{da})^2;
\]

this follows from that \( b_{y}^{(t+\tau)} \geq b_{y}^{(t)} \) for each \( t \) (transformations similar to the ones earlier in the proof would clearly show that only the expectation of \( \mu_{j,k}^{(t+\tau)} \). Now we have

\[
\beta V_j \left( s, \ldots, s, y, \ldots \right) = \sum_{t=1}^{\tau} \beta^t \sum_{k \in G} \mu_{j,k}^{(t)} (b_k - b_j)^2 + \beta^t \sum_{a \in S} \sum_{k \in G} q_{y,a}^{(t)} \mu_{j,k}^{(t+\tau)} (b_k - b_{da})^2 \geq \sum_{t=1}^{\tau} \beta^t \sum_{k \in G} \mu_{j,k}^{(t)} (b_k - b_j)^2 + \beta^t \sum_{a \in S} \sum_{k \in G} p_{x,a}^{(t)} \mu_{j,k}^{(t+\tau)} (b_k - b_{da})^2\]

Consequently, for each such path, we have

\[
V_j \left( s, \ldots, s, y, \ldots \right) \geq V_j \left( s, \ldots, s, x, \ldots \right) > V_j(x).
\]

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

**Proof of Theorem 3.** The proof for the case where \( \beta \) is sufficiently small. Consider the set \( A = \{ x \in \mathbb{R} \mid x = b_{d(s)}^{(1)} \text{ for some } s \in S \} \) and \( B = \{ y \in \mathbb{R} \mid y = \frac{b_s + b_{s+1}}{2} \text{ for some } s \in \{1, \ldots, m-1\} \} \). For generic parameter values, \( A \cap B = \emptyset \). If so, then there is a unique mapping satisfying the description in Part 1 of Theorem 2, and therefore, by that Theorem, there MPE is essentially unique if \( \beta < \beta_0 \).

The proof for the case where within-person monotonicity is satisfied is in several steps.
**Step 1.** Suppose that there are two monotone MPEs $\sigma_1$ and $\sigma_2$, 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 states $x, y \in S$, $x \neq y$, such that the distributions $q^1_{x,y} \neq q^2_{x,y}$ and $q^1_{y,x} \neq q^2_{y,x}$. In other words, it is impossible that transition probabilities from only one state are different.

**Proof.** Suppose not, so there is a unique state $s$ such that $q^1_{s,s} = q^2_{s,s}$. Let us first prove that, generically, $\left| x \in S \setminus \{s\} : q^1_{s,x} + q^2_{s,x} > 0 \right| = 1$. Indeed, if there is no such $x$, then $q^1_{s,x} = q^2_{s,x} = 0$ for all $x \neq s$, meaning that $q^1_{s,s} = q^2_{s,s} = 1$ and $q^1_{y,s} = q^2_{y,s}$, which contradicts the choice of $s$. On the other hand, suppose that there are $x \neq y$ that satisfy this property; without loss of generality, $x < y$. Without loss of generality, suppose $q^1_{s,x} > 0$. Then by Part 4 of Theorem 1, for generic parameter values, $q^1_{s,y} = 0$, which means that $q^2_{s,y} > 0$, which, again by Part 4 of Theorem 1, implies $q^2_{s,x} = 0$ for generic parameter values. Now, consider three possibilities. If $x < s < y$, then, from Part 3 of Theorem 1, from $q^1_{s,x} > 0$ it follows that for $z < s$, $q^1_{z,x} > 0$ implies $a \leq x$; moreover, for such $z$, $q^2_{z,x} = q^1_{z,x}$. Therefore, if the society moves from state $s$ to $x$, the continuation utilities of the current decision-maker should be the same for both equilibria: $w^1_{d_1}(x) = w^2_{d_1}(x)$. Similarly, from $q^2_{s,y} > 0$ it follows that for $z > s$, $q^2_{z,y} > 0$ implies $a \geq y$; moreover, for such $z$, $q^2_{z,x} = q^1_{z,x}$. Thus, if the society moves from state $s$ to $y$, the continuation utilities of the current decision-maker again coincide: $w^1_{d_2}(y) = w^2_{d_2}(y)$. But by Part 2 of Theorem 1, we have $w^1_{d_2}(x) \geq w^1_{d_1}(y) = w^2_{d_1}(y) \geq w^2_{d_2}(x) = w^1_{d_2}(x)$, which implies that both inequalities hold with equality, in particular, $w^1_{d_1}(x) = w^1_{d_2}(y)$. But from the proof of Part 4 of Theorem 1 this is impossible for generic parameter values. The remaining possibilities are $x < y < s$ and $s < x < y$; they are considered similarly (and even simpler).

Thus, we proved that there is a unique $x \neq s$ such that $q^1_{s,x} + q^2_{s,x} > 0$. Without loss of generality, assume $x > s$. Clearly, it must be that $q^1_{s,x} \neq q^2_{s,x}$; otherwise, since the supports of $q^1_{s}$ and $q^2_{s}$ are subsets of $\{s,x\}$, we would have $q^1_{s,s} = q^2_{s,s}$, again meaning that $q^1_{s,s} = q^2_{s,s}$, and contradicting the choice of $s$. Without loss of generality, assume $q^1_{s,x} < q^2_{s,x}$, so in equilibrium $\sigma_1$ the society stays in $s$ longer than in equilibrium $\sigma_2$, in expectation; this means, in particular, $q^1_{s,s} < 1$ and $q^2_{s,s} > 0$. It must be that the sequence $b^{(t)}_{d_1}$ is non-decreasing and, moreover, it is nonstationary, for otherwise $q^2_{s,s} > 0$ would contradict Lemma A3.

Let $j = d_s$. The continuation utilities from moving to $x$ are the same in both equilibria: $V^1_j(x) = V^2_j(x)$, because the transition probabilities are identical thereafter. Moreover, in equilibrium $\sigma_2$, transiting is a best response, so $V^2_j(x) \geq V^2_j(s)$, and in equilibrium $\sigma_1$, staying is a best response, so $V^1_j(s) \geq V^1_j(x)$. We thus have $V^1_j(s) \geq V^1_j(x) = V^2_j(x) \geq V^2_j(s)$, meaning that the utility of individuals from group $j$ from staying is at least as high under $\sigma_1$ as under $\sigma_2$. Denote $V_j(s; \alpha)$ the utility of staying in $s$ if the subsequent equilibrium play has probability $\alpha$ of moving to $x$; then $V_j(s; q^1_{s,s}) = V^1_j(s)$ and $V_j(s; q^2_{s,s}) = V^2_j(s)$. 44
Consider the function \( f(\alpha) : [0, 1] \to \mathbb{R} \), defined by
\[
f(\alpha) = V_j(s; \alpha) - V_j(x).
\]
Let us prove that it satisfies the following strict single-crossing property: if for some \( \alpha, f(\alpha) = 0 \), then \( f(\alpha') > 0 \) for \( \alpha' > \alpha \) and \( f(\alpha') < 0 \) for \( \alpha' < \alpha \). Suppose that \( f(\alpha) = 0 \) and \( \alpha' > \alpha \) (the case \( \alpha' < \alpha \) is analogous). Let us denote the continuation utility of individuals from current group \( j \) after the society spent \( \tau \) periods in state \( s \) and stays there for an extra period with transition probability is \( \alpha \) thereafter by \( V_j^{(\tau)}(s; \alpha) \), and if it departs to state \( x \), by \( V_j^{(\tau)}(x) \).

We have \( V_j^{(\tau)}(s; \alpha) < V_j^{(\tau)}(x) \) for all \( \tau > 1 \), because the sequence of expected bliss points \( b_j^{(t+\tau)} \geq b_j^{(t)} \) for all \( \tau \), and for at least some \( t \) the inequality is strict. Therefore, we have
\[
f(\alpha') - f(\alpha) = V_j(s; \alpha') - V_j(s; \alpha) = \beta \left( (1 - \alpha') V_j^{(1)}(s; \alpha') + \alpha' V_j^{(1)}(x) - (1 - \alpha) V_j^{(1)}(s; \alpha) - \alpha V_j^{(1)}(x) \right) = \beta \left( (1 - \alpha) V_j^{(1)}(s; \alpha') - V_j^{(1)}(s; \alpha) \right) + (\alpha' - \alpha) \left( V_j^{(1)}(x) - V_j^{(1)}(s; \alpha') \right) \]
\[> \beta (1 - \alpha) \left( V_j^{(1)}(s; \alpha') - V_j^{(1)}(s; \alpha) \right) = \cdots \]
\[> (\beta (1 - \alpha))^2 \left( V_j^{(2)}(s; \alpha') - V_j^{(2)}(s; \alpha) \right) = \cdots \]
\[> (\beta (1 - \alpha))^\tau \left( V_j^{(\tau)}(s; \alpha') - V_j^{(\tau)}(s; \alpha) \right) \text{ for any } \tau > 2.
\]
Since \( V_j^{(\tau)}(s; \alpha') - V_j^{(\tau)}(s; \alpha) \) is bounded, we must have that \( f(\alpha') - f(\alpha) > 0 \). This proves that \( f(\alpha) \) satisfies the single-crossing condition.

Now, if \( f(q_{s,x}^1) = 0 \), then \( f(q_{s,x}^2) > 0 \), meaning that \( V_j(s; q_{s,x}^2) > V_j(x) \), which contradicts that moving to \( x \) is a best response in \( \sigma_2 \). Similarly, if \( f(q_{s,x}^2) = 0 \), then \( f(q_{s,x}^1) < 0 \), meaning that \( V_j(s; q_{s,x}^1) < V_j(x) \), which contradicts that staying at \( s \) is a best response in \( \sigma_1 \). If \( f(q_{s,x}^1) \neq 0 \) and \( f(q_{s,x}^2) \neq 0 \), then, since staying in \( s \) is a best response in \( \sigma_1 \), we must have \( f(q_{s,x}^1) > 0 \); similarly, we must have \( f(q_{s,x}^2) < 0 \). But then by continuity there is \( \alpha \in (q_{s,x}^1, q_{s,x}^2) \) such that \( f(\alpha) = 0 \). In that case, it must be that \( f(q_{s,x}^1) < 0 < f(q_{s,x}^2) \) and not the other way around, a contradiction. This completes the proof of Step 1.

**Step 2.** Let \( m \) be the minimal number of states for which there are two monotone MPEs \( \sigma_1 \) and \( \sigma_2 \). Then \( m = 2 \).

**Proof.** Suppose not, then either \( m = 1 \) or \( m \geq 3 \). If \( m = 1 \), there is only one possible transition mapping: \( Q \) with \( q_{1,1} = 1 \). Suppose \( m > 3 \) and let \( Q^1 \) and \( Q^2 \) the transition matrices in equilibria \( \sigma_1 \) and \( \sigma_2 \). Let \( Z \subset S \) be the set of \( z \in S \) such that \( q_{z,s}^1 \) and \( q_{z,s}^2 \) are different distributions; from Step 1 it follows that \( |Z| \geq 2 \). In what follows, let \( L = \{ s \in S : \forall x > s : q_{s,x}^1 = q_{s,x}^2 = 0 \} \) and \( R = \{ s \in S : \forall x < s : q_{s,x}^1 = q_{s,x}^2 = 0 \} \). By Lemma A3, \( L \cup R = S \); let us denote \( I = L \cap R \).

45
First, we show that if \( s \in S \) and \( 1 < s < m \), then \( s \notin I \). Indeed, otherwise, we would have \( q^1_{s,s} = q^2_{s,s} = 1 \). If there is \( x < s \) such that \( x \in Z \), then there are two equilibria \( \sigma_1|_{[1,s]} \) and \( \sigma_2|_{[1,s]} \) in the game with the set of states \( S' = S \cap [1,s] \); otherwise there must be \( x > s \) such that \( x \in Z \), and 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 \( x = \min (s : s \in Z, s > 1) \) and \( y = \max (s : s \in Z, s < m) \) (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 \( q^1_{x,x} = q^2_{x,x} = 1 \) by definition of \( R \), and then \( x \notin Z \), a contradiction. If, on the other hand, \( x \in R \) and \( x < m \), then, similar to the proof of Claim 1, \( \sigma_1|_{[s,m]} \) and \( \sigma_2|_{[s,m]} \) are different equilibria in the game with the set of states \( S' = S \cap [m,s] \), a contradiction. We can similarly prove that \( y \in R \).

There are two possibilities. If \( Z \neq \{1,m\} \), then \( x = \min (s : s \in Z, s > 1) = \min (s : s \in Z \cap [2,m-1]) \leq \max (s : s \in Z \cap [2,m-1]) = \max (s : s \in Z, s < m) = y \). In that case, we have that \( \sigma_1|_{[x,y]} \) and \( \sigma_2|_{[x,y]} \) are two different equilibria on \( [x,y] \), which again contradicts with 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 \( m = 2 \).

**Finishing 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 \( q^1_z \) and \( q^2_z \) are different is the whole set \( S \). Without loss of generality, suppose \( q^1_{1,1} > q^2_{1,1} \). Since \( q^2_{1,1} < 1 \), \( q^2_{1,2} > 0 \), and in a monotone equilibrium we must have \( q^2_{2,1} = 1 \); this means \( q^1_{2,1} < 1 \), and thus \( q^1_{2,1} > 0 \) and again by monotonicity \( q^1_{1,1} = 1 \). This implies (by Lemma A3) that the sequence \( b^{(t)}_{d_1} \) is non-decreasing (because equilibrium \( \sigma_2 \) exists) and \( b^{(t)}_{d_2} \) is non-increasing (because equilibrium \( \sigma_1 \) exists). Suppose \( b^{(\infty)}_{d_1} < \frac{b^{(1)}_{d_1} + b^{(2)}_{d_1}}{2} \), then one can easily prove (similar to the proof of Lemma A3) that \( \sigma_2 \) cannot be an equilibrium, as the group \( d_1 \) would strictly prefer to stay in 1 under \( \sigma_2 \), while \( q^2_{1,1} < 1 \). Similarly, if \( b^{(\infty)}_{d_2} > \frac{b^{(1)}_{d_2} + b^{(2)}_{d_2}}{2} \), then \( \sigma_1 \) cannot be an equilibrium. The only remaining possibility (since \( b^{(\infty)}_{d_1} \leq b^{(\infty)}_{d_2} \) by Assumption 1) is where \( b^{(\infty)}_{d_1} = b^{(\infty)}_{d_2} = \frac{b^{(1)}_{d_1} + b^{(2)}_{d_2}}{2} \), which is nongeneric (moreover, for both \( \sigma_1 \) and \( \sigma_2 \) to exist, it must be that \( b^{(t)}_{d_1} = b^{(t)}_{d_2} = \frac{b^{(1)}_{d_1} + b^{(2)}_{d_2}}{2} \) starting from \( t = 1 \)). This proves that for generic parameter values, if within-person monotonicity condition holds, the equilibrium is unique. ■

**Proof of Corollary 1.** By Part 2 of Theorem 2, there exists an equilibrium with the desired properties. For generic parameter values it is unique by Theorem 3, and the result follows. For
other parameter values, the result may be proved directly but we omit the proof. ■

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

**Proof of Theorem 4.** To be completed. ■

**Proof of Theorem 5.** To be completed. ■

**Proof of Theorem 6.** To be completed. ■

The next example illustrates the second part of Theorem 6

**Example A1** *(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

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

For such \( M \), the equilibrium is generically unique for any discount factor \( \beta \).

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 \( \beta \) 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 \( \phi(1, 2, 3, 4, 5) = (2, 3, 3, 3, 4) \).
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 by 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 $\beta$, 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 $\beta$, but not for very low or very high ones.

**Proof of Theorem 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:

\[
V_R (d) = A_R - b^2_R + \beta V_R (d);
\]

\[
V_M (d) = \beta ((1 - \theta) V_M (d) + \theta V_P (d));
\]

\[
V_P (d) = A_P - b^2_P + \beta \left( \left( 1 - \frac{\gamma_M \theta}{\gamma_P} \right) V_P (d) + \theta \frac{\gamma_M}{\gamma_P} V_M (d) \right).
\]
Thus,

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

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

Now consider the case \(\gamma_M < \gamma_P\). Here, the state with poor in power (denote it \(l\)) is stable, and starting from \(d\), the society can start in \(d\) or transit to \(l\), but not to the state with rich in power, \(r\) (this follows from Lemma A3). The utility of the players from being in state \(l\) are given by (similarly to previous)

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

and thus

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

Suppose the probability of transition from \(d\) to \(l\) equals \(q = q_{dl}\). In that case, the utility of individuals from different groups from being in \(d\) are given by equations

\[
V_R (d) = A_R - b_R^2 + \beta (1 - q) V_R (d) + \beta q V_R (l); \\
V_M (d) = \beta (1 - \theta) (1 - q) V_M (d) + \beta (1 - \theta) q V_M (l) + \beta \theta (1 - q) V_P (d) + \beta \theta q V_P (l); \\
V_P (d) = A_P - b_P^2 + \beta \left(1 - \theta \frac{\gamma_M}{\gamma_P}\right) (1 - q) V_P (d) + \beta \left(1 - \theta \frac{\gamma_M}{\gamma_P}\right) q V_P (l) + \beta \theta \frac{\gamma_M}{\gamma_P} (1 - q) V_M (d) + \beta \theta \frac{\gamma_M}{\gamma_P} q V_M (l).
\]

One can check that for \(\theta < \frac{1 - \beta}{2 \left(1 + \frac{\gamma_M}{\gamma_P}\right)}\), \(w_M (d) > w_M (l)\) for any \(q\); for \(\theta > \frac{1}{2}\), \(w_M (d) < w_M (l)\) for any \(q\), and for \(\theta \in \left[\frac{1 - \beta}{2 \left(1 + \frac{\gamma_M}{\gamma_P}\right)}, \frac{1}{2}\right]\) 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 \(\theta\).

49
This implies the result for preferences of $R$, as $V_R(d)$ depends on $\theta$ only through $q$, and is decreasing in $q$. Consider middle class $M$. For $\theta < \frac{1-\beta}{2-\left(1+\frac{2M}{\gamma p}\right)\beta}$, $q = 0$, so $d$ is stable, and $M$ prefers a smaller $\theta$. For $\theta \in \left[\frac{1-\beta}{2-\left(1+\frac{2M}{\gamma p}\right)\beta}, \frac{1}{2}\right]$, $w_M(d) = w_M(l)$, 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 $\theta$ as follows from the formula, and therefore $V_M(d)$ is monotonically decreasing in $\theta$.

Let us now consider $P$. For $\theta \leq \frac{1-\beta}{2-\left(1+\frac{2M}{\gamma p}\right)\beta}$, $q = 0$, and for $\theta \geq \frac{1}{2}$, $q = 1$; in both intervals, an marginal increase in $\theta$ only leads to more social mobility, and the poor are better off. If $\theta \in \left(\frac{1-\beta}{2-\left(1+\frac{2M}{\gamma p}\right)\beta}, \frac{1}{2}\right)$, 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

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

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

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

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

$$\frac{(V_M(l) + b_P^2)(1 - \beta + \beta \theta)}{(1 - \beta) \theta} = \beta \theta V_P(d),$$

and thus

$$V_P(d) = \frac{(1 - \beta + \beta \theta)}{\beta \theta} (V_M(l) + b_P^2)$$

$$= \frac{1 - \beta + \beta \theta}{\beta \theta} \left( \frac{1}{1 - \beta} (1 - \beta) \left(-b_P^2\right) + \beta \theta \left(A_P - \frac{\gamma M}{\gamma p} b_P^2\right) + b_P^2 \right)$$

$$= \frac{1 - \beta + \beta \theta}{(1 - \beta) \theta} \left( \frac{1}{1 - \beta} \left(1 - \beta + \beta \theta \left(1 + \frac{\gamma M}{\gamma p}\right)\right) b_P^2\right).$$

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 $\theta$ for all $\theta$, which completes the proof. ■
Proof of Corollary 3. Consider the case $\gamma_M > \gamma_P$. By Theorem 1, if $M$ prefer a lower $\theta$ and $P$ prefer a higher $\theta$. 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 $\theta < \frac{1-\beta}{2 - (1+\gamma_M \gamma_P)\beta}$; such $\theta$ will be defeated in a plurality voting by $\hat{\theta} = \frac{1-\beta}{2 - (1+\gamma_M \gamma_P)\beta}$, because $R$ are indifferent, and the more numerous (than $M$) $P$ prefer $\theta'$. If $\theta > \frac{1-\beta}{2 - (1+\gamma_M \gamma_P)\beta}$, then it will again be defeated by $\hat{\theta} = \frac{1-\beta}{2 - (1+\gamma_M \gamma_P)\beta}$, because $R$ and $M$ both prefer a slower social mobility on this interval, and they together constitute a majority. Thus, $\hat{\theta} = \frac{1-\beta}{2 - (1+\gamma_M \gamma_P)\beta}$ is the unique core element. This completes the proof. $\blacksquare$

Proof of Theorem 2. If $\gamma_M > \gamma_R$ then, similarly to Theorem 1, democracy is stable for any $\theta$. Similarly to the proof there, one can easily show that $R$ prefer a lower $\theta$, $M$ prefer a higher $\theta$, 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, similarly, to Theorem 1, there is a unique equilibrium for each $\theta$; moreover, for $\theta \leq \frac{1-\beta}{2 - (1+\gamma_M \gamma_R)\beta}$, $q = 0$, for $\theta \in \left[\frac{1-\beta}{2 - (1+\gamma_M \gamma_R)\beta}, \frac{1}{2}\right]$, $q$ is monotonically increasing from 0 to 1, and for $\theta \geq \frac{1}{2}$, $q = 1$.

Accordingly, $V_P(d)$ is strictly increasing on $\theta \in \left[\frac{1-\beta}{2 - (1+\gamma_M \gamma_R)\beta}, \frac{1}{2}\right]$ and constant outside of it, and $V_M(d)$ is strictly increasing for all $\theta$ (this may be proven analogously to Theorem 1). As for $V_R(d)$, it is strictly decreasing for $\theta < \frac{1-\beta}{2 - (1+\gamma_M \gamma_R)\beta}$ or $\theta > \frac{1}{2}$.

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

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

Its derivative with respect to $\theta$ equals

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

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 $\theta < \frac{1 - \beta}{\sqrt{\beta \frac{\gamma_M}{\gamma_R} \left(A_R b_R^2 + 1\right)} - \beta \left(1 + \frac{\gamma_M}{\gamma_R}\right)} = \theta^*$. 51
and is locally decreasing otherwise.

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

**Proof of Corollary 4.** Consider the case \( \gamma_M > \gamma_R \). By Theorem 2, \( R \) prefer a lower \( \theta \) and \( M \) prefer a higher \( \theta \). However, since \( \gamma_M > \gamma_R \), all \( \theta \), except for the maximum value, will be defeated in a plurality voting. Thus, \( \hat{\theta} \) is the maximal admissible value of \( \theta \); in our case, Assumption 1 is satisfied whenever \( \theta \leq \frac{1}{1 + \frac{\gamma_M}{\gamma_R}} \). Now consider the case \( \gamma_M < \gamma_R \). Here, consider the following possibilities. First, if \( \beta \frac{\gamma_M}{\gamma_R} \left( \frac{A_R}{b_R} + 1 \right) \geq 4 \), then the utility of \( R \) is monotonically decreasing in \( \theta \). 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 to the value that maximizes \( V_R(d) \). There are two candidate values for this value of \( \theta \): 0 and \( \min (\theta^\ast, \frac{1}{2}) \). 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

\[
V_R^{\theta=0}(d) = \frac{A_R - b_R^2}{1 - \beta},
\]

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

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

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

Simplifying, we would get \( (2 - 2\beta + \beta \frac{\gamma_M}{\gamma_R}) (2 - \beta + \beta \frac{\gamma_M}{\gamma_R}) < 0 \), which is impossible, if \( (2 - \beta) \left( \frac{2R}{\gamma_M} - 1 \right) > \left( \frac{A_R}{b_R} - 1 \right) \), then \( V_R^{\theta=\frac{1}{2}}(d) \) maximizes \( V_R^{\theta}(d) \) for all \( \theta \), 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 \(V_{R}^{\theta^*} (d) > V_{R}^{\theta=0} (d)\). We have, after simplification,

\[
V_{R}^{\theta=0} (d) = \frac{1 - \beta + \beta \theta^* \theta^* A_R - \left( 1 - \theta^* - \beta + \beta \theta^* \left( 1 + \frac{\gamma_M}{\gamma_R} \right) \right) b_R^2}{1 - \beta + \beta \theta^* \left( 1 + \frac{\gamma_M}{\gamma_R} \right)}
\]

\[
= \left( \frac{1}{\theta^*} + \frac{\beta}{1 - \beta} \right) \frac{A_R + b_R^2}{\frac{1 - \beta}{\theta^*} + \beta \left( 1 + \frac{\gamma_M}{\gamma_R} \right) - b_R^2}.
\]

\[
= A_R + \left( 1 + \frac{\gamma_M}{\gamma_R} \right) \frac{b_R^2}{1 - \beta} - 2 \sqrt{\beta \frac{\gamma_M}{\gamma_R} (A_R + b_R^2) b_R}
\]

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

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

which, after simplification, implies \((2\beta - 1) \left( 1 - \frac{\gamma_M}{\gamma_R} \right) > 1\). But this is impossible, which means that \(V_{R}^{\theta^*} (d) > V_{R}^{\theta=0} (d)\) 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. This completes the proof. ■
References


Appendix B: Additional results and examples — Not for publication

B1 Examples

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

Example B1 (Multiple equilibria) There are five groups with political bliss points $b_{1,2,3,4,5} = -\frac{21}{16}, -1, 0, 1, \frac{21}{16}$ (there would be two equilibria even if the extreme political bliss points are $\pm 2$ rather than $\pm 2.1$, 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

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

(e.g., if each of the following permutations: (12), (23), (345), (354) happen with probability $\frac{1}{4}$ each, the result would be matrix $M$).

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 $-1, 0, 1, \frac{21}{16}$ 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 $b_3 = 0$, then an individual of group 3 who cared only about the next period (i.e., very myopic one) 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. Unfortunately, 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 \( \phi_1 \) and \( \phi_2 \) are equilibria (verifying that other groups act as prescribed is straightforward).

One can also verify that equilibrium \( \phi_1 \) is preferred to \( \phi_2 \) by individuals who start in groups 1, 2, 3, and the opposite is true for those in groups 4, 5. In other words, today’s decision-makers (group 3) are in favor of \( \phi_1 \). Given that the decision is made by a representative agent, one could wonder what makes \( \phi_2 \) an equilibrium. One way of interpreting equilibrium mapping \( \phi_2 \) 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. (Remarkably, the problem does not disappear if we truncate the future, i.e., consider a finite number of periods: then in the last but one period, members of group 3 would move to 4, and actually the equilibrium corresponding to \( \phi_2 \) will survive.)

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 B2 \( (\text{Mixing between noncontiguous states}) \) 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

\[
M = \begin{pmatrix}
70 & 10 & 20 & 0 & 0 \\
10 & 10 & 10 & 0 & 0 \\
10 & 10 & 10 & 30 & 20 \\
0 & 0 & 6 & 34 & 10 \\
0 & 0 & 0 & 53 & 10 \\
\end{pmatrix}.
\]

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

B-2
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 which allows to accomplish 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.

B2 Conditions for mixed strategies

Our next results are theoretical, highlighting cases where 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 matrix $M$ satisfies

$$ b_j = \arg \min_{b \in \{b_{i(s)}\}_{k=1}^{s}} \left| (M b)_j - b \right|. $$

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

Proposition B1 The following is true for any $M$, any $b$ and $A$:

(i) There is $\beta_0 > 0$ such that for any $0 < \beta < \beta_0$, the equilibrium mapping involves pure transitions only.

(ii) Suppose that social mobility is slow, but $M$ is not an identity matrix. Then there is $\beta_1 < 1$ such that for any $\beta_1 < \beta < 1$, the equilibrium mapping involves mixing.
Interestingly, with a finite number of periods, there would generically be only equilibria in pure strategies. A proof is available upon request. At the same time, for any fixed $\beta$, if $M$ is sufficiently close to identity matrix then the equilibrium is in pure strategies.

### B3 Conditions for monotonicity of MPE

We first provide an example of nonmonotone MPE.

**Example B3 (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

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

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 if 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, $\psi$ 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, ..., and 2, 1, 2, 1, 2, 1, .... 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 the prefer the path 1, 2, 1, 2, 1, 2, ... to 2, 1, 2, 1, 2, 1, ... 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 $\psi$ may arise in equilibrium.

It should be noted that in this Example, Within-Person Monotonicity property 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 MPE are monotone in the sense of Definition 1.

**Theorem B1** Every MPE is monotone if either of the following conditions holds:

(i) The discount factor \(\beta\) is sufficiently low;
(ii) There is sufficiently little social mobility, in the sense that the matrix \(M\) is sufficiently close to the identity matrix.
(iii) In each state, the decisions are made by a single decisive agent; more precisely, in each state \(s\), the weight of the pivotal group \(d_s\) have weight 

\[ w_s(d_s) > \frac{1}{2} . \]

The latter condition suggest that nonmonotone equilibria would be ruled out if every state were a dictatorship of its respective pivotal group. Since such equilibria do not depend on the exact weights, provided that the pivotal groups remain the same, monotone equilibria are also the most ‘robust’.

**B4 Some results on social mobility matrices**

We start with an example illustrating the different representations of mobility matrices.

**Example B4** *(Multiple representations of a mobility matrix as lottery over permutations)* For a given \(A\), the distribution \(\mu\) such that \(A = \Omega(\mu)\) need not be unique. E.g., take \(n = 3\) and

\[ A = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} . \]

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 of more blocks \( \{K_x\} \). Each \( K_x \) is a connected block determine the extent of social mobility. (Assumption 1 requires that the blocks are connected.)

**Lemma B4 (Characterization of matrices, satisfying Assumption 2)** 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

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

(B1)

**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 \( \mu_{j,k_1} \) and \( \mu_{j,k_2} \) are proportional to the sizes of the groups: \( \mu_{j,k_1} n_{k_1} = \mu_{j,k_2} n_{k_2} \).

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 \( \left( \sum_{i=l_x}^{l_y} \mu_{ji} \right) \left( \sum_{z=k_2}^{l_y} n_z \right) \neq \left( \sum_{i=k_1}^{l_y} \mu_{ji} \right) \left( \sum_{z=l_x}^{l_y} n_z \right) \) (denote the difference right-hand side and left-hand side by \( Y \)).

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

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

(outside of \( K_x \), \( b_i \) are defined arbitrarily, subject to monotonicity). We have \( (\mathbf{b}^\varepsilon)_j = 0 \) for every \( \varepsilon \). 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 \( \delta_1 \) to be such that \( |(M\mathbf{b}^\varepsilon)_j| > |Y|/2 \) for \( \varepsilon \leq \delta_1 \). Now, observe that the sequence \( M^z \) converges, as \( z \to \infty \), to a matrix \( M^\infty \) 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 \( (M^\infty \mathbf{b}^\varepsilon)_j \to -\left( \frac{\sum_{i=l_x}^{l_y} n_k}{\sum_{i=l_x}^{l_y} n_i} \right) \left( \sum_{z=k_2}^{l_y} n_z \right) + \left( \sum_{z=k_2}^{l_y} n_k \right) \left( \sum_{z=l_x}^{l_y} n_z \right) = 0 \). Thus, there is \( \delta_2 \) such that \( |(M^\infty \mathbf{b}^\varepsilon)_j| < |Y|/2 \) for \( \varepsilon \leq \delta_2 \). Consequently, for \( \varepsilon = \max(\delta_1, \delta_2) \), we have \( 0 = (\mathbf{b}^\varepsilon)_j < |(M^\infty \mathbf{b}^\varepsilon)_j| < |Y|/2 < |(M\mathbf{b}^\varepsilon)_j| \). Since all inequalities are strict, there is \( h : 1 < h < \infty \) such that this inequality holds if \( M^\infty \) is replaced by \( M^h \). This implies that the subsequence \( (\mathbf{b}^\varepsilon)_j, (M\mathbf{b}^\varepsilon)_j, (M^h\mathbf{b}^\varepsilon)_j \) 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 $\mu_{jk}$ are given by (B1). 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 $q = l_y - 1$. This completes the proof. ■

Remark 1 Lemma B4 does not require anything the extreme groups in a given class to conform to the same formula given by (B1). For example, the following matrices satisfy (2), (3), as well as monotonicity across and within individuals:

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