Social Mobility and Stability of Democracy: Re-evaluating de Tocqueville*

Daron Acemoglu — **MIT**

Georgy Egorov — **Northwestern University**

Konstantin Sonin — **Higher School of Economics**

March 2015

Abstract

An influential thesis going back to 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. In this paper, we investigate this hypothesis using a dynamic model of political economy in the presence of social mobility. 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 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.

Keywords: Social mobility, institutions, stability, democracy, de Tocqueville, dynamics.

JEL Classification: D71, D74.

Comments Are Welcome!

*We thank participants of Stanford Institute for Theoretical Economics (SITE) conference on Dynamics of Collective Decision-making 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.[1] This perspective has intuitive appeal: if the rich expect to become middle class in the near future, then it would be counterproductive for them to try to exclude the middle class from the political process.[2] It also has major implications for the health of democracy. For example, it suggests that a greater social mobility, caused by improvements in the educational system, the dismemberment of barriers against occupational mobility, or technological changes, inexorably improves the prospects of democracy's survival.

Despite its important role in major social theories and its continued relevance in modern debates on democracy, there has been little systematic formalization or critical investigation of this idea. The next example shows why this idea is intuitive, but also how inherent forces present in dynamic political economy models, which link the anticipation of future economic and political changes to current conflict (e.g., Acemoglu and Robinson, 2000, 2001, Acemoglu, Egorov and Sonin, 2008, 2012), imply that greater social mobility may actually undermine support for democracy.

[1] De Tocqueville, for example, argued: "Among aristocratic nations, as families remain for centuries in the same condition, often on the same spot, all generations become, as it were, contemporaneous . . . . Among democratic nations [as the United States], new families are constantly springing up, others are constantly falling away, and all that remain change their condition. . . ." and articulated the importance of social mobility as follows "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 (1960) and Barrington Moore (1966), among others, built on De Tocqueville's ideas. Pareto (1935) emphasized the related notion that political stability depends on social mobility into and out of the elite, while Sombart (1906) linked the weakness of socialist ideology in the United States to social mobility (see also Thernstorm, 1984). Blau and Duncan's famous (1967) study concludes that "the stability of American democracy is undoubtedly related to the superior chances of the upward mobility in this country". See also Erikson and Goldthorpe (1992).

[2] This might, for example, contribute to an explanation for why members of the landed aristocracy have typically been less likely to support the extension of political rights to the poorer segments of society than have merchants and professional classes.

Example 1 *Consider a society with n individuals, with $\frac{2}{5}n$, or 40 percent of them, being rich, $\frac{1}{5}n$ or 20 percent, being middle class, and $\frac{2}{5}n$, or 40 percent, being poor. There are three possible political institutions: democracy, where decisions are made by the median voter who is a member of the middle class; the left-wing dictatorship, where the rich and the middle class are excluded from political participation and all political decisions are made by a member of the poor group; and the elite dictatorship, where the poor and the middle class are excluded and all decisions are the preview of the rich. Suppose that the economy lasts for two periods, at each date society adopts a single policy, p_t, and there is no discounting between the two periods. (Alternatively, the two periods may correspond to two generations of the same dynasty, with parents having utility defined over their and their offsprings' stage payoffs). All agents have stage payoffs given by $-(p_t - b_i)^2$, with political bliss points, b_i, for the poor, middle-class, and rich social groups being, respectively, $-1, 0$, and 1. Society starts out with one of the three political institutions described above. Then in the first period, a member of the politically decisive social group (who has exactly the same preferences as all other members of his social group) decides both the current policy and the political institution for the second period, and in the second period, the group in power chooses policy.*

Imagine, to start with, that the rich are in power (i.e., we start with the elite dictatorship). Without social mobility, the rich would prefer to keep their dictatorship so as to be able to set the policy in the second period as well.[3] *Suppose, however, that there is high social mobility, namely, the group identity of an individual's second period self does not depend on who he is in the first period. If so, a rich individual expects to be part of the rich, the middle class, and the poor with probabilities $2/5, 1/5$, and $2/5$, respectively. His second-period expected utility is then $-\frac{2}{5}(p_2+1)^2 - \frac{1}{5}p_2^2 - \frac{2}{5}(p_2-1)^2 = -p_2^2 - \frac{4}{5}$. Thus, he prefers, in expectation, policy $p_2 = 0$ to be implemented in the second period. To achieve this, he would prefer next period's political institutions to be democratic, as they would benefit his expected future self. This illustrates a simple form of de Tocqueville's argument.*

However, the same simple setup can also be used to illustrate the limits of this argument. Suppose that the society starts out as a democracy. Now the decisive group is the middle class,

[3] 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.

and it prefers to preserve democracy and enjoy its political bliss policy 0 *in both periods, as there is no incentive to do the contrary. Suppose next that there is social mobility: between periods* 1 *and* 2*, a certain number of the middle class members* r *become rich (let* $\alpha = \frac{r}{5n}$ *denote the share of the middle class that moves upwards) and to keep the aggregate distribution across social groups stationary,* r *rich agents become middle class. (At the time of decision-making, the middle class members do not know who will stay and who will move, and thus there is still 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 the 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 support for democracy. In particular, if social mobility means that members of the politically pivotal middle class expect to change their preferences in a certain direction, they are more willing to change the institution in that direction as well.[4] The example not only delineates the potential limitation of de Tocqueville's hypothesis, it also introduces the bare-bones structure we will use in our analysis. In particular, instead of a two-period setup, we will consider an infinite horizon economy. This is not for the mere sake of generality, but because several important strategic considerations emerge when the political horizon is longer than two periods.

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, and more importantly, what matters for the political equilibrium is not simply mobility next period, but the evolution of the preferences of an agent's "future selves" (because of evolving social mobility) and her expectation of others' behavior in the future. This is illustrated in the next example, which extends Example 1 to an infinite-horizon setting.

[4]The fact that the social mobility in this example makes middle-class agents more likely to move upwards than downwards is important as we will see in our analysis. If the middle class expected to move upwards and upwards symmetrically, then its members would continue to prefer democracy to other political regimes because they would lose in expectation even more from an elite (or a left-wing) dictatorship as they would gain.

Example 2 *Consider the same setting as in Example 1, but now with infinite number of periods. In each period, the current decision-maker determines the next period's institution, and in-between the periods r people move upwards from the middle class, while exactly the same number of people move down from the ranks of the rich to the middle class. We also assume that agents care about the discounted sum of their stage payoffs, with a discount factor β. For illustration purposes, set $\beta = 4/5$ in this example, and also suppose that the society starts out as a democracy, i.e. with the middle class in power.*

If the society ever became a left-wing dictatorship, then 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. If, conversely, society ever became the elite dictatorship, then it can be shown that the rich, who would then be in power, would also have no incentive to change this institution. What about the middle class? Suppose that mobility is sufficiently high (in particular, $\alpha = \frac{r}{5n} > 1/2$ as in Example 1). Then members of the middle class would prefer the elite dictatorship to democracy starting next period. However, $\alpha < 1/2$ is not sufficient to make democracy stable in this case: middle-class agents prefer democracy to survive next period so that they are still pivotal then, but may still prefer the elite dictatorship in subsequent periods. Intuitively, these middle class agents expect that in the long run they will be rich $2/3$ of the time and in the middle class $1/3$ of the time (because of the ergodic property of the corresponding Markov chain). This implies that, when they do not discount too heavily, they prefer the rich to be pivotal rather than the middle class in the future.

Given $\beta = 4/5$, the critical threshold for social mobility for the middle class to prefer the rich to be pivotal in the long run turns out to be $\alpha = 1/4$. When $\alpha < 1/4$, there is too little social mobility relative to the discount factor and middle-class agents prefer to remain in power and thus maintain democracy. If, on the other hand, $1/4 < \alpha < 1/2$, then a member of the middle class prefers her group to remain in power in the next period, but the rich to be pivotal after a few periods, leading to non-existence of pure strategy Markov Perfect Equilibria: if today's middle-class agents expect a transition from democracy to the elite dictatorship in the future, they would then prefer to remain in democracy today. But if they expect future middle-class agents to maintain democracy, then they would rather move to the elite dictatorship today. This implies that any Markov Perfect Equilibrium must be in mixed strategies in this range. It can also be shown that as α increases within this range (towards the $1/2$), then the probability of

a switch from democracy to the elite dictatorship increases (one can verify that this probability equals $\frac{4\alpha-1}{4-6\alpha}$*).*

Our baseline framework generalizes the setup considered in the previous two examples. Because our focus is on the impact of social mobility on political equilibria, we focus on an exogenous social mobility process (though we analyze political preferences over different rates of social mobility and allow the society to choose it in a particular case of three social groups in Section 5.2). Society consists of a finite number of social groups, each of which consists of a finite number of identical individuals. Individuals (and thus groups) are ordered with respect to their wealth and their policy preferences. Social mobility is modeled as in the previous examples: in each period, there are well-defined and stationary probabilities for individuals from each social group to be reallocated to one of the other social groups, and we also assume that the aggregate distribution of individuals across social groups remains stationary.[5]

There is a finite set of alternative political institutions (political states), each represented by a set of weights assigned to individuals within each social group. Given these weights, a group, or alternatively, an individual within that group, is the pivotal voter. This pivotal voter chooses the current policy as well as next period's political state, or alternatively, next period's pivotal voter. As highlighted by our examples, if the current pivotal voter expects to be in a different social group (because of social mobility) in the next period, she might prefer to shift the political power to another social group.

Our main results are of two sorts. First, we establish the existence and certain basic characterization results for Markov Perfect Equilibria in this economy. We show that under some mild conditions, such equilibria are "monotone" in that if the equilibrium path starts from a state to the further right, then the equilibrium distribution of future states will be to the further right than had the economy started with a state to the left. We also demonstrate that such monotone equilibria can be equivalently characterized as maximizing the discounted utility of the current pivotal voter at each date. As Example 2 suggests, equilibria are often in mixed strategies. Nev-

[5] Assuming that the aggregate distribution remains unchanged allows us to focus on social mobility rather than rise of some social groups or decline of others, which is an interesting but distinct phenomenon. From a more technical standpoint, our assumption allows us to focus on a well-defined notion of Markov Perfect Equilibrium. Without an assumption guaranteeing stationarity of distribution of people across social groups over time, Markovian restriction would have little bite, as strategies would have to depend on the current (and future) distributions of people.

ertheless, we show that even such mixed strategy equilibria take a particular form, namely, the society only mixes between keeping the current institution and transiting to a uniquely defined alternative. This, in particular, implies that the direction of transition is always well defined within an equilibrium, and different mixed strategies simply change the speed of transitions. This enables us to determine how political institutions change under various conditions, and in particular as a function of social mobility. We also show that monotone equilibria are unique under a simple *within-person monotonicity* condition, which requires that the preferences of the future selves of an individual change monotonically (i.e., an individual expects that as the horizon increases, his preferences will either gradually shift to the left or to the right). Intuitively, this condition implies that the preferences of the future selves can be consistently aggregated. Without this condition, there may be multiple equilibria whenever individuals are not too myopic, because an individual's — in particular, the pivotal voter's — preferences over different political institutions depends on how future members of the group she belongs to now will vote. Our analysis highlights the importance of two kinds of conflicts of interest: that between agents with different economic interests today, and that between today's decision-maker and his future selves, which arises from the fact that today's decision-maker anticipates to be in a different social class and the future. The latter conflict of interest is not only essential for understanding the political implications of social mobility, but it 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.

Second and more substantively, we provide a comprehensive analysis of the relationship between social mobility and stability of democracy. We quantify the stability of democracy with the size of its basin of attraction along the equilibrium path. So greater social mobility makes democracy more stable if its basin of attraction becomes larger, meaning that democracy will arise and persist starting from a broader set of initial political institutions. As in Example 1, social mobility may make democracy asymptotically unstable—even starting in democracy, society will not stay there (implying that its basin of attraction is empty). Theorem 4 shows that, when agents are sufficiently forward-looking, the (asymptotic) stability of democracy turns on the difference between median preferences (which represent the preferences of the pivotal voter in democracy) and mean preferences (which represent the "average" preferences of individual

under very high social mobility). If median preferences are close to mean preferences, then starting with any distribution of political rights, society will transition to full democracy. This is because democracy ensures that future policies will not be too far from the preferences of individuals that are currently powerful — regardless of how their preferences change. This result thus provides a simple formalization of the "de Tocqueville hypothesis".

However, generalizing the examples above, we also characterize how social mobility can weaken the support for democracy. Specifically, the same simple condition mentioned in the previous paragraph, comparing median and mean preferences, is sufficient to determine the other side of the relationship between social mobility and the support for democracy: when the mean and the median of the preferences are far apart (in a sense that is made precise below), an increase in social mobility reduces the basin of attraction of democracy and can make democracy asymptotically unstable, so that even starting from democracy society will not stay there.

We further consider two additional issues. The first is the study of slippery slope in dynamic political economy, whereby potentially efficiency-enhancing institutional or policy changes are not adopted because they will create a "slippery slope," paving the way for further changes that will benefit the currently politically powerful individuals or groups. In Acemoglu, Egorov and Sonin (2012), we emphasized the importance of slippery slope arguments in the dynamics of constitutions, coalitions and political institutions in cases in which agents have discount factors close to 1—since this is when they especially care about future payoffs. We show here that, in the presence of social mobility, slippery slope arguments are also important, but not when the discount factor is high, but when it is intermediate. The reason for this is that with a discount factor very close to 1, agents care about the preferences of their "long-run selves," not their current preferences, thus dulling the slippery slope considerations. In contrast, with intermediate discount factors, agents care most about their preferences in the near future, which tend to be similar to the current preferences, which heightens the slippery slope considerations. (When agents have very small discount factors, then they do not care about the future, so the slippery slope issues do not arise). Second, we also consider political preferences over social mobility. Focusing on settings with the three social groups (poor, middle-class and rich), we show that depending on whether social mobility is upwards or downwards (respectively between the middle class and the poor or between the middle class and the rich), it tends to directly benefit groups that are moving upwards, but at the same time, through the political equilibrium responses,

it may hurt both these groups and those that are not part of the social mobility process. The reason for this is that, for example, when social mobility is upwards and becomes more rapid, the middle class will increasingly prefer the elite dictatorship to democracy, ultimately changing equilibrium policies in a direction that is harmful to the poor. This observation points to new endogenous limits to the extent of social mobility — this time coming from groups that will lose from the political responses to social mobility.

Our paper is related to several branches of the political economy literature. Most closely related is 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 related to the de Tocqueville 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), and do not link these ideas to endogenous institutional change or to the stability of democracy. Benabou and Tirole (2006) construct a political model, where people's 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, and does not feature the dynamic political trade-offs that are at the heart of our paper. Most importantly, neither this literature nor any other that we are aware of has noticed or studied the destabilizing role of social mobility for democratic systems.

Some of the modeling approach here overlaps with dynamic political economy models studying democratization, constitutional change, the dynamics of repression and the efficiency of long-run institutional arrangements. Roberts (1999) is one of the first contributions in this line, analyzing the dynamics of political clubs, while Besley and Coate (1998) discuss similar issues in a two-period democratic setting. Acemoglu and Robinson (2000 and 2001), Lizzeri and Persico (2004), Lagunoff (2006), and Acemoglu, Egorov, and Sonin (2015) consider

dynamic models of democratization and transitions back from democracy to dictatorship, while Bourguinon and Verdier (2000) study the choice of educational policy which affects political participation in the future, creating a role for inefficient educational choices. Barbera and Jackson (2004) consider the stability of different constitutional arrangements in a two-period setting, while Gomes and Jehiel (2005) analyze the efficiency of long-run institutional change. Fearon (2005), Powell (2005) and Acemoglu, Ticchi and Vindigni (2010) use related models to study the determinants of civil and international wars. Our previous work, Acemoglu, Egorov and Sonin (2012, 2015), provides a general framework delineating some of the key forces present in this class of models and showing how these ideas could be applied to general models of institutional change and also to the study of the dynamics of repression. None of these papers discuss social mobility or political preferences when individual's economic or social status changes over time.

Finally, the role of the implicit conflict between the current self and the future selves of the pivotal voter, induced by social mobility in our model, relates to a handful of papers considering time-inconsistency of collective or political decisions, in particular, Amador (2003), Gul and Pesendorfer (2004), Strulovici (2010), Bisin, Lizzeri, and Yariv (2015), and Jackson and Yariv (2015), though none of these works notes the conflict between current and future cells resulting from social mobility or studies implications of such 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 stability of democracy. Section 5 contains further results, which study when slippery slope is possible in a model with social mobility and which also endogenize social mobility and study the preferences of different groups over future social mobility. Section 6 concludes.

2 Model

In this section, we introduce the basic model and define 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, $1 \leq k \leq g$, consisting of $n_k > 0$ agents (so $\sum_{k=1}^{g} n_k = n$). The groups are ordered, and the order reflects their "economic" preferences (e.g., groups with lower numbers could be those that are richer and prefer lower taxes). All individuals share a common discount factor $\beta \in (0, 1)$.

Individuals have preferences over a policy space represented by the real line, $\mathbb{R}$. We assume that individuals in each group have the same stage payoffs represented by the following quadratic function of the distance between policy and their bliss point:

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

where p_t is the policy at time t, b_k is the political bliss point of agents in group k, and A_k is an arbitrary constant, allowing for the possibility that some groups are better off than others (e.g., because they are richer).[6] In what follows, $\mathbf{b} = \{b_k\}$ will denote the column vector of political bliss points.

Decision-making power in the society depends on the current state, or social arrangement; in each period, the 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 particular, we assume that 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).

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 very mild assumption adopted for technical convenience, and in particular, it holds generically within the class of weights. In our model, 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 the same as G, so that each state corresponds to a different

[6] For example, if all $A_k = 0$, then members of the middle class would not want to become rich if the 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.

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 each individual 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 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 (or has come close to) this stationary distribution.

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

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

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

where the latter condition imposes the stationarity assumption (or equivalently 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.[7] Throughout the paper, we also impose the following assumption:

[7] An alternative way to define matrix M is to start with with permutations $\pi \in S_N$ of all individuals, and assume that in each period, Nature changes identities of individuals according to π with probability λ_π, such that $\sum_{\pi \in S_N} \lambda_\pi = 1$. The symmetry requirement then becomes $\lambda_\pi = \lambda_{\sigma \circ \pi \circ \tau}$ for any $\sigma, \tau \in S_N$ that reshuffle individuals within groups only. In this case, we will have

$$\mu_{jk} = \frac{1}{n_j} \sum_{i \in G_j} \sum_{\pi \in S_N : \pi(i) \in G_k} \lambda_\pi.$$

The converse is also true: for any matrix $M = \left\{\mu_{jk}\right\}$ of nonnegative elements satisfying (2)–(3) there is a

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

$$\sum_{k=1}^{l} \mu_{j_1,k} \geq \sum_{k=1}^{l} \mu_{j_2,k}. \tag{4}$$

This assumption essentially rules out 'deterministic reversals of fortune', where poorer people become (in expectation) richer than the current rich. In other words, the current relative ranking of two individuals persists, at least in the weak sense, or put differently, the distribution of a richer individual's future selves first-order stochastically dominates the distribution of the poorer individual's future selves.

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 correspondes 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 identitiy 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. None of our results depend on this order (provided that group d_s has a chance to make an offer in state s, as we have imposed).

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_i^t, the timing in each period $t \geq 1$ is as follows.

1. *Policy decision:*

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

corresponding distribution λ over permutations π (this distribution may be not uniquely defined). This is simple generalization of Birkhoff-von Neumann theorem for doubly stochastic matries is proved in the Appendix (see also Budish et al., 2013).

(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 individual i casting vote $v_i^p(j) \in \{Y, N\}$.

(d) If $\sum_{i=1}^{n} w_{g_i^t}(s) \mathbf{1}\{v_i^p(j) = Y\} > \frac{1}{2}$, 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 1b 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, 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+1}^j.

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

(d) If $\sum_{i=1}^{n} w_{g_i^t}(s) \mathbf{1}\{v_i^s(j) = Y\} > \frac{1}{2}$, then next period's state is the current proposal, $s_{t+1} = s_{t+1}^j$, and the game moves to stage 3. Otherwise, the game proceeds to stage 2b 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 gets payoff $u_{g_i^t}(p_t)$, given by (1).

4. *Social mobility:* At the end of the period, there is social mobility, so that individual i who belonged to group g_i^t in period t will start period $t+1$ in group k with probability $\mu_{g_i^t,k}$.

2.4 Definition of equilibrium

We focus on"symmetric" Markov Perfect Equilibrium (MPE for short). In particular, we required that equilibria are symmetric with respect to individuals' actions (i.e., individuals i and i' that happen to be in group k, perhaps in different periods, play identically We allow for mixed strategies: Example 2 in the Introduction illustrates that a pure strategy equilibrium may fail to

exist in even simple settings. (Proposition C1 in Appendix C provides more general conditions under which there are no equilibria in pure strategies.) Furthermore, for most of the paper we focus on monotone MPE, which we simply refer to as "equilibria," as defined next.

Definition 1 ***(Monotone Markov Perfect equilibrium)*** *MPE σ is called 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 $s_t = y$. Formally, for any $l \in [1, m]$,*

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

For most of our analysis, we focus on monotone MPE. Theorem 9 presented in the Appendix provides conditions under which all MPE are monotone.

Note that MPE are formally defined in terms of a complete list of strategies. However, 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 MPE. In particular, even though MPEs will not be unique in terms of a complete list of strategies, they will often all correspond to the same policy choices in equilibrium transitions, and we will, with a slight abuse of terminology, refer to a unique MPE in such circumstances. (See the Appendix).

Finally, we say that a (political) state s is *stable*, if $s_t = s$ implies that $s_{t+1} = s$. We say that a state s is *asymptotically stable* if $s_t \in \{s-1, s, s+1\}$ implies that $\lim_{\tau \to \infty} \Pr(s_\tau = s) = 1$, in other words, if the sequence of states induced in equilibrium converges to s with probability 1. This last definition captures (in discrete state space) the usual idea of asymptotic stability: starting with a small enough deviation from a stable state, the equilibrium path will take the state arbitrarily close to the original 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 'more stable' under a change in parameters, if it is (asymptotically) stable whenever it was (asymptotically) stable before the change. The notion of 'less stable' is defined analogously.

3 Analysis

In this section, we establish existence, 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 (monotone MPE) and characterizes certain important properties of this equilibrium. In particular, such an equilibrium can be represented by a sequence of policies and transitions that maximize the discounted utility of the current pivotal group. It also induces a set of transitions across states that are also monotonic.

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

1. *The equilibrium policy coincides with the bliss policy of the current pivotal group at each t: i.e., if the current state at time t is s, then the policy is $p_t = b_{d_s}$.*

2. *The next state is always the one that maximizes the expected continuation utility of current members of the current pivotal group: i.e., if we define the transition correspondence Q by $q_{s,z} = \Pr(s_{t+1} = z \mid s_t = s)$, then $q_{s,z} > 0$ implies*

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

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

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

3. *The transitions induced by the equilibrium are 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 current policies and transitions that maximize the discounted utility of the pivotal group, d_s.[8] 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 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. Thus it implies — but is also much stronger than — the equilibrium distribution of states starting from a higher state first-order statistically dominating the distribution of states starting from a lower state.

Finally, Part 4 shows that even if the equilibrium involves mixed strategies, these are neither arbitrary nor very complicated. In particular, a mixed strategy must involve mixing only between the current state and some other state. Loosely speaking, mixed strategies arise only as a way of slowing down the transition from today's state to some unique 'target' state. Intuitively, 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 this next decision-maker to also stay. This was a reflection of the fact that the current decision-maker prefers the current state but would like to be in a different state because he expects his preferences to change in the near future due to social mobility. Mixed strategies resolve this problem by slowing down transitions: when he expects the next decision maker to slowly move away (i.e., more way with some probability), the current decision maker is indifferent between moving towards his 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.[9]

[8] There is an equivalent of this result in Roberts (1999) in a non-strategic setting and also in Acemoglu, Egorov and Sonin (2014) in a setting without social mobility or this type of stochasticity.

[9] 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 3 in Appendix B demonstrates shows that it may take place as a result

One implication of this characterization is that even though the 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 notation to designate the policy preferences of tomorrow's self. An agent currently belonging to group j will belong in τ periods to group k with probability μ_{jk}^{τ}, where μ_{jk}^{τ} is the corresponding element of the matrix M^{τ}. Therefore, the expected utility of an agent currently in group j will have stage utility in τ periods if policy p is implemented at that point given by:

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

The last two terms are constants (reflecting, after rearranging, the expectation A_k and the variance of b_k), and thus policy preferences are given by the square of the distance between the policy and the political bliss point of the self in τ periods,

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

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

Theorem 2 ***(Very myopic or very patient players)*** *The following is true for the limit values of discount factor:*

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 $z \in S$ which minimizes $\left| b_{d_z} - b_{d_s}^{(1)} \right|$; in other words, if agents are myopic, then society immediately moves to a state where policy closest to the political bliss point of tomorrow's self of the current pivotal group, $b_{d_s}^{(1)}$, will be chosen.*

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

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).

The first result is straightforward: 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 β is high, agents are patient, and are willing to act in a way that will eventually lead to a state where the utilities of their distant future selves are maximized. Thus, if the equilibrium evolution were to fail to take the society there, the current decision-maker would have an incentive to move there immediately. Intuitively, these players are willing to care about their current and near-future selves only inasmuch as this does not compromise the well-being of distant future selves, because of farsightedness. To complete the argument, one needs to show that the state z that minimizes $\left|b_{d_z} - b_{d_s}^{(\infty)}\right|$ is stable, so once the society gets there, it stays there forever. This follows from the following intuitive argument: in the long run, the distribution of future selves of individuals from groups d_s and d_z is the same, and therefore their interests are aligned. Therefore, in the long run, decision-makers from group d_z would be intrested in preserving state z, which is exactly what group d_s wants in the beginning of the game. Theorem 2 does not imply immediate transition to the long-run stable state even when β is close to 1 because the agents might still prefer to spend the next several periods in the current state.

3.2 Uniqueness

The results stated so far apply to any (monotone) MPE. We next provide conditions for uniqueness, which will turn on whether the preferences of future selves can be consistently aggregated.

Theorem 3 ***(Uniqueness)*** *The monotone MPE is generically (essentially) unique (meaning that decisions on current policy and transitions in each state are determined generically uniquely within the class of monotone MPE) if either the discount factor β is sufficiently low, or there is* ***within-person monotonicity****, meaning that the sequence of political bliss points of a player's selves in all future periods is monotone (increasing or decreasing): for any $j \in H$, the sequence $b_j, b_j^{(1)}, b_j^{(2)}, \ldots$ is weakly monotone.*

Recall that uniqueness here refers not to the complete list of strategies but to the behavior induced along the equilibrium path (as captured by the qualifier 'essentially').

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 maximized their next period utility, which generically has a unique solution. The within-person

monotonicity condition plays a central role beyond this special case. In each period, the current decision-maker chooses the state tomorrow, and hence indirectly the sequence of states at all future dates. Imagine a situation in which the current decision-maker expects his preferences to first move to the right and then to the left (thus violating the within-person monotonicity). He might be happy to stay in the original state in order to balance the interests of all future selves. However, if he expects future decision-makers to move right in the next period, he would prefer to do so immediately, because tomorrow's self is the only one that benefits from such a move. This paves way for mutiliplicity. If, on the other hand, the within-person monotonicity condition is satisfied, this sort of multiplicity is not possible: his tomorrow's self wants a move to the right more than his current self, and if tomorrow's decision-makers are more likely to move to the right, the current one is more comfortable delegating to them.

It is worth noting that in the absence of within-person monotonicity, multiplicity of equilibria does not disappear as β approaches 1. The reason is that even if the current and long-run selves have similar preferences, they still need to ensure that individuals who are in charge are willing and able to resist the temptation to pursue their short-run incentives instead. While these short-run incentives are not an important part of the utility function, there is a problem of coordination leading to multiplicity. The following example illustrates the problem of coordination between individuals that occupy the same niche in the society at different times.

Example 4 *Consider an infinite-horizon environment as in Example 2, but introduce the following changes to the social mobility (and make the discount factor arbitrary). As before, suppose that in each period, x members from the middle class become rich and an equal number of rich become middle class. In addition to that, y other members from the middle class become poor, while an equal number of poor become middle class. Furthermore, assume that $x = \frac{5}{8}k$ and $y = \frac{1}{10}k$ (these are all integers provided that k is divisible by 40, which simplifies the example by ensuring that transitions are not probabilistic).*

The person who is currently in the middle class prefers policy 0; however, the next period he will prefer $\frac{5}{8} \times 1 + \frac{1}{10} \times (-1) = \frac{21}{40}$ on average. As a result, he prefers the rich to rule in the next period. However, his preferences in the subsequent periods are such that he again prefers democracy: for example, in the period after next he prefers $\left(\frac{5}{8} \times \frac{11}{16} + \frac{11}{40} \times \frac{5}{8}\right) \times 1 + \left(\frac{1}{10} \times \frac{19}{20} + \frac{11}{40} \times \frac{1}{10}\right) \times (-1) = \frac{1533}{3200} < \frac{1}{2}$, and his expected ideal policy decreases further down the road, monotonically converging to zero. Intuitively, there is substantial mobility between the

middle class and the rich but little mobility between the middle class and the poor, so in the short run, a member of the middle class expects to move upward, while in the long run the fact that the entire society is mobile becomes relevant, and the distribution of one's long-run future selves converges to the (ergodic) stationary distribution, which is the original distribution of citizens across classes.

Can be verified that for $\beta > 0.373$*, there is an equilibrium where democracy is stable. If* $0.373 < \beta < 0.830$*, then both dictatorships are stable; if* $0.830 < \beta < 0.921$*, then the 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 resists 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 β*: the middle class immediately transitions to the elite dictatorship, and the two dictatorships are stable. The reason why this strategy profile constitutes a best response for the middle class is that the next period is the only one where having the 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 the elite dictatorship according to the equilibrium strategy. Therefore, for today's middle class the issue is not whether to move to the elite dictatorship, but when, and their preferences imply that a transition today is preferable to transition tomorrow. Clearly, the elite dictatorship is stable: the rich would benefit from democracy in the long run, but they know that any democratization will last for one period only and thus does not make sense. For the poor under left-wing dictatorship, democratization will result in the elite dictatorship, which is even worse. As a result, this strategy profile is an equilibrium for all values of* β*, even though if* β *is sufficiently high, it is Pareto dominated by the equilibrium where democracy is stable. (Notice that Pareto ranking is a feature of this example and may not be expected in general; for example, for intermediate values of* β *there are multiple equilibria, but they are not Pareto ordered.)*

As may be expected, if $\beta > 0.373$*, there is also a third equilibrium, where the middle class moves to the elite dictatorship with some probability, and both dictatorships are stable. The elite dictatorship because if middle-class agents are willing to transfer power to the rich, and the rich*

want to stay in power even more, while in the left-wing dictatorship, the poor are unwilling to transition to democracy because democracy itself will make way to the elite dictatorship.

The within-person monotonicity condition becomes particularly intuitive if one views the problem of dynamics of institutions under social mobility as a problem of aggregation of preferences of all future selves of all current agents. To understand when such aggregation will have 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. The within-person monotonicity condition 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 for 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 (i.e., a member of the current pivotal group). 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 reveals why we do not need the extra condition when β is sufficiently low because in this case, tomorrow's self receives almost all of the weight, and thus the problem of aggregation of preferences of different future selves becomes moot.

3.3 Farsighted stability of institutions

In what follows, we will assume that the within-person monotonicity condition holds. Under this condition, Theorem 2 yields two corollaries, which we will use in the rest of the analysis. If β is high then, as follows from Theorem 2, the preferences of very distant future selves $\mathbf{b}^{(\infty)}$ play a key role. Fortunately, they are straightforward to characterize and 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 become (trivially, it suffices to consider $\tau = g - 1$). Under Assumption 1, a member of group j may eventually move to group k if and only if members of group k can

move to group j. This implies that the set of groups G can be partitioned into non-intersecting components (i.e., $L_M(j) \cap L_M(k) \neq \varnothing$ 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 $k_2 \in L_M(j)$.[10] This enables us to write the preferences of the current decision maker, from group d_x, in the very distant future as

$$b_{d_x}^{(\infty)} = \frac{\sum_{k \in L_M(x)} n_k b_k}{\sum_{k \in L_M(x)} n_k}.$$

Corollary 1 ***(Farsighted stability of institutions)*** *Take a state $s \in S$. It is stable for sufficiently high β (formally, there exists $\tilde{\beta} < 1$ such that for any $\beta \in \left(\tilde{\beta}, 1\right)$, $q_{s,s} = 1$) if and only if*

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

This result states that a state is stable, when players are sufficiently farsighted (the discount factor is sufficiently close to 1), if this state 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 (to make it valid for the cases where the median voter is in one of the extreme groups, i.e., if either of these groups have a majority, we denote $b_0 = -\infty$ and $b_{g+1} = +\infty$).

Corollary 2 ***(Farsighted stablity of democracy)*** *Suppose that group x contains the median voter. Democracy is stable for sufficiently high β if and only if*

$$\frac{b_{d_x - 1} + b_{d_x}}{2} \leq b_{d_x}^{(\infty)} \leq \frac{b_{d_x} + b_{d_x + 1}}{2}. \tag{8}$$

[10] Technically, $L_M(j)$ denotes the irreducible component of the Markov chain given by M that contains group j.

This corollary provides a simple, and as it will turn out powerful, characterization of the stability of democracy when the discount factor, β, is sufficiently high (sufficiently close to 1). Intuitively, it requires that the preferences of the current decision- maker in the very distant future (which is the median group in democracy) be closer to his own current preferences than the preferences of either neighboring group.[11] When this is the case, the current median voter prefers to delegate future decisions to the future median voter. When it is not the case, 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 democracy is stable or not, but also provides us the comparative statics of democracy would respect to social mobility.

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. Given this focus, in what follows, we fix all other parameters of the model, and vary the matrix of social mobility.

Definition 2 *Suppose we have two matrices of social mobility M and M' such that $\mathbf{b}^{(\infty)} = \mathbf{b}'^{(\infty)}$. We say that social mobility is (weakly)* faster *under M' than under M if for each $j \in N$, if $b_j'^{(\infty)} \geq b_j^{(\infty)}$, then $b_j'^{(t)} \geq b_j^{(t)}$ for all $t \geq 1$, and if $b_j'^{(\infty)} \leq b_j^{(\infty)}$, then $b_j'^{(t)} \leq b_j^{(t)}$ for all $t \geq 1$.*

For example, take a baseline matrix M, which satisfies the within-person monotonicity property, and consider a family of matrices of social mobility $M(\gamma) = \gamma M + (1-\gamma) I$, where I is the identity matrix and $\gamma \in [0,1]$ is a parameter. Then social mobility for $M(\gamma')$ is faster than that in $M(\gamma)$ if and only if $\gamma' > \gamma$.

The next theorem shows that the relationship between social mobility and the stability of democracy turns on condition (8) introduced in Corollary 1.

Theorem 4 *Suppose that social mobility under M' is faster than under M, and the inequality* (8) *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 also is under M'.*

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

In other words, de Tocqueville's prediction about social mobility and support for democracy was correct, as long as condition (8) holds. Intuitively, when this condition holds, Corollary 1 established that, because the long-run future self of the pivotal voter in democracy has higher utility in democracy than in any other political institution. This ensures the stability of democracy. Interestingly, however, the same condition also ensures that greater social mobility increases the size of the "basin of attraction of democracy". This is because with greater social mobility, the preferences of other social groups are also becoming closer and closer to that of the group containing the median voter. For example, in the special case where M involves complete 'reshuffling' and when $\gamma = 1$, all groups will have the same preferences.

What if (8) does not hold? In this case, the current median voter expects to prefer another state in the very long run. This does not necessarily bar democracy from being stable for $\beta < 1$ (provided that it is not too close to 1). Therefore, the 'only if' part of Corollary 1 does not have a bite. But at the same time, an increase in social mobility, corresponding to an increase in γ, makes democracy less stable rather than more stable.

Theorem 5 *Suppose that for M, the inequality* (8) *does not hold. Then democracy becomes 'less stable' as γ increases. More precisely, suppose $\gamma' < \gamma$; then if democracy is stable under $M(\gamma)$ for some parameters, then it is also stable under $M(\gamma')$ for the same parameters.*

The intuition for this result is related to the intuition for Theorem 4: now as the speed of mobility γ increases, both the current median voter and neighboring groups prefer to be in a different state than democracy in the future, and this implies that, for a given β, the basin of attraction of democracy shrinks as γ increases.

5 Further Results and Extensions

In this section, we discuss some additional results of our baseline model concerning slippery slope considerations, and an extension in which the political system determines the extent of social mobility.

5.1 Slippery slopes

In Acemoglu, Egorov, and Sonin (2012), we emphasized how slippery slope considerations — whereby one round of change in political institutions or states is expected to lead to a series of

further changes — can lock in an inefficient outcome because of the fear that this will be the first step that in a slippery slope leading to a series of further changes entailing lower payoffs. 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 instead. In models without social mobility, slippery slope considerations are more powerful when the discount factor is closer to 1 because in this case agents 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 (since these have the same dynamic political economy forces), 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 *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 in a monotone equilibrium, 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:*

$$\sum_{t=1}^{\infty} \sum_{k=1}^{g} \mu_{d_s k}^{t} u_{d_s} \left(b_{d_s} \right) \geq \sum_{t=1}^{\infty} \sum_{k=1}^{g} \mu_{d_s k}^{t} u_{d_s} \left(b_{d_x} \right).$$

In other words, this result suggests that for both high and low β, all stable states give higher expected utility to the current decision-maker (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. In particular, there may be a state providing a higher expected utility to the current decision-maker than the current state, but moving to this state would unleash another set of transitions that reduce the discounted continuation payoff of the current decision-maker. Theorem 6 shows that such slippery slope considerations arise only for intermediate values of β. Example 4 in the Appendix illustrates that slippery slope considerations are actually possible.

The intuition for why slippery slope considerations do not play a role for myopic players (with low β) is straightforward: such players care only about the next period's state, so the subsequence sequence of changes does not modify their rankings over states. That these considerations do not arise for very farsighted players (with high β) is more interesting and perhaps surprising. Suppose a situation in which the current-decision-maker, who is pivotal in the current state s, prefers a different state, x, where by definition he will not belong to the pivotal group unless his preferences change due to social mobility. Therefore, such a ranking is feasible 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 while the decision-makers in states s and x have a conflict of interest today, their preferences in the distant future will be similar because of social mobility. Thus with a sufficiently high discount factor, the current decision-maker will not be worried about decision rights shifting to those in the group that is pivotal in state x, obliviating 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 a result in Acemoglu, Egorov and Sonin, 2012).

5.2 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 would only three social groups, the poor (P), the middle class (M), and the rich (R), with shares γ_P, γ_M, and γ_R, respectively; $\gamma_P+\gamma_M+\gamma_R = 1$. We also assume that $\gamma_P, \gamma_R < \frac{1}{2}$, so that the median voter belongs to the middle class. 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 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 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.

We start with social mobility at the bottom. Let θ 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 to the middle class is $\frac{\gamma_M}{\gamma_P}\theta$.

Theorem 7 *If $\gamma_M > \gamma_P$, then mobility at the bottom does not make democracy unstable. A higher θ makes the poor better off and the middle class worse off; the rich are indifferent.*

If $\gamma_M < \gamma_P$, then mobility at the bottom makes the poor better off and the middle class worse off. The rich become weakly worse off as θ increases, and strictly worse off if θ increases within the interval $\left(\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta}, \frac{1}{2}\right)$.

The poor always value social mobility at the bottom, both because this enables them to transition to a richer group (middle class) and because 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 — provided that democracy remains stable. This stability is guaranteed when $\gamma_M > \gamma_P$, and also holds when $\gamma_M < \gamma_P$ provided that social mobility is not very high. For higher θ (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.

Let us next turn to the collective choices over θ. Let us assume that when choosing between any two possible levels of mobility θ, those who are indifferent abstain (or their vote is split fifty-fifty). In this case, it is easy to see there is a unique Condorcet winner: if $\gamma_M > \gamma_P$, $\theta = 0$ is the Condorcet winner and thus there will be no social mobility at the bottom; if $\gamma_M < \gamma_P$, then $\theta = \frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta}$ is the Condorcet winner. In other words, if the middle class is large enough, there will be too little support for social mobility at the bottom; if it is small enough, then there will be support for some social mobility, but only to the extent that it does not undermine stability of democracy.

Corollary 3 *The unique Condorcet winner $\hat{s}$ is:*

(i) if $\gamma_M > \gamma_P$, then $\hat{s} = 0$;

(ii) if $\gamma_M < \gamma_P$, then $\hat{s} = \frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_P}\right)\beta}$.

The most important implication of this result is that the political implications of social mobility introduce another endogenous limit on the extent of social mobility. In particular, 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 β this range of parameters. The intuition for this result is as follows. 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. Thus, if people are forward-looking, the rich will put a tighter bound on the speeds of social mobility that they will accept, and the poor will have to settle for that lower speed in order to get support of the rich, which they need as no group constitutes a majority.

We next turn to social mobility at the top. Let us now denote by θ the share of middle class who become rich, and then the share of the rich that move to the middle class is $\frac{\gamma_M}{\gamma_R}\theta$. In what follows, we denote $\varphi = \frac{\gamma_M}{\gamma_R} \in (0,1)$, and $\Delta = \frac{A_R}{b_R^2} - 1 > 0$. The equilibrium can be characterized as follows.

Theorem 8 *If $\gamma_M > \gamma_R$, then mobility at the top does not make democracy unstable. A higher θ makes the middle class better off and the rich worse off; the poor are indifferent.*

If $\gamma_M < \gamma_R$, then mobility at the top makes the middle class better off and a marginal increase in the extent of social mobility makes the poor weakly worse off (strictly worse off if $\theta \in \left(\frac{1-\beta}{2-\left(1+\frac{\gamma_M}{\gamma_R}\right)\beta}, \frac{1}{2}\right)$. The utility of the rich is locally increasing in θ if and only if

$$\theta \in \left(\frac{1-\beta}{2-(1+\varphi)\beta}, \min\left(\frac{1-\beta}{\sqrt{r\Delta\beta}-(1+\varphi)\beta}, \frac{1}{2}\right)\right)$$

and is locally decreasing in θ otherwise. This interval is nonempty if and only if $\varphi\Delta\beta < 4$.

Now conversely, the poor do not directly care about social mobility at the top, and they also do not care about it at all provided that 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. On the other hand, 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, which makes them better off. Theorem 8 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 θ. The rich are more likely to benefit from social mobility if Δ is small, i.e., when "inequality" 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 (Δ is high), it is more important for the rich to stay rich than to influence institutional change.

As before, we can use the notion of Condorcet winner to understand the choice of the society in this case, but the structure of the coalition supporting such a Condorcet winner is more interesting in this case. If $\gamma_M > \gamma_R$, democracy is stable and the poor are indifferent, and the middle class, which is pivotal in this case, chooses the maximal speed of social mobility possible (the maximal speed that satisfies Assumption 1 is $\theta = \frac{1}{1+\varphi}$). If $\gamma_M < \gamma_R$, the poor are either indifferent or against social mobility and the middle class is in favor it. If the rich have much to lose from social mobility economically, there will effectively be a peripheral coalition against social mobility consisting of the poor and the rich. If, on the other hand, the rich only suffer a small economic loss from social mobility, they may also support greater social mobility because this will destabilize democracy in favor of a dictatorship favoring their preferences. The next corollary gives a complete characterization.

Corollary 4 *The unique Condorcet winner is:*

(i) if $\gamma_M > \gamma_R$, *then* $\widehat{\theta} = \theta$, *where* $\theta = \frac{1}{1+\frac{\gamma_M}{\gamma_R}}$ *is the maximum speed possible;*

(ii) if $\gamma_M < \gamma_R$ *and* $\Delta > (2-\beta)\left(\frac{1}{\varphi} - 1\right)$, *then* $\widehat{\theta} = 0$;

(iii) if $\gamma_M < \gamma_R$ *and* $\Delta < (2-\beta)\left(\frac{1}{\varphi} - 1\right)$, *then* $\widehat{\theta} = \frac{1}{2}$;

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. If the middle class is relatively small, the choice will depend on Δ, which measures the extra utility of being rich (a proxy for inequality). If it is sufficiently high, then then the society will choose no mobility at the top; if it is sufficiently low, the society will ensure that a half of the middle class becomes rich, which is the lowest speed of

mobility that guarantees immediate transition to the institution ruled by the elite. Thus, social mobility is higher if Δ is higher, or if β is lower. Here again, myopic rich are more likely to favor social mobility; for them, switching to their preferred institution is more important than long-run considerations about the possibility of becoming middle class.

6 Conclusion

An influential thesis going back to 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 role in many institutional discussions, it has received little attention from the political economy literature. With the exception of a few papers investigating the link between social mobility and preferences for redistribution, which essentially confirmed de Tocqueville's intuition, even if he did not focus on the link between social mobility and the support for democracy, there are no modern works modeling the links between social mobility and dynamic political preferences.

This paper has investigated these issues using a simple but fairly general model of dynamic political economy. Our framework provides a natural formalization of de Tocqueville's ideas, showing that greater social mobility can increase the support for democracy for reasons anticipated by de Tocqueville. However, more importantly, it also shows the limits of this hypothesis. There is a robust reason why greater social mobility can undermine support for democracy: when the median voter expects to move 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 the society is mobile, the current institution may be disliked by the current decision makers not only because their future selves will like some other institution more, but also because should the institution stay, future pivotal players will make decisions detrimental to the future selves of the current ones. To the best extent of our knowledge, this link between preferences of future selves and the nature of dynamic political equilibria has not been noted in the literature.

We also characterized the conditions under which 'slippery slope' considerations—which prevent certain policy and institutional choices because of the further series of changes that these would induce—arise centrally in this framework, but differently from other dynamic political economy settings, they 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 similar preferences 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 (e.g., 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 the elite dictatorship. This paves the way for a poor-rich coalition aimed at decreasing social mobility at the top.

There are many important topics for research related to the political implications of social mobility. Most important are systematic empirical analyses of the impact of social mobility (perception) on political attitudes. In addition, a fruitful area for future research is the development of dynamic theoretical models in which individuals make investments that impact of social mobility as well as voting on policies and institutional choices impacting social mobility.

References

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

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

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin (2015) "Political Economy in a Changing World," *Journal of Political Economy*, forthcoming.

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

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

Acemoglu, Daron, and James Robinson (2006) *Economic Origins of Dictatorship and Democracy*, Cambridge University Press, New York.

Acemoglu, Daron, Davide Ticchi, and Andrea Vindigni (2010). "Persistence of Civil Wars," *Journal of the European Economic Association,* 8(2-3), pages 664-676

Alesina, Alberto, Ignazio Angeloni, and Federico Etro. (2005) "International Unions." *American Economic Review*, 95(3): 602-615.

Alesina, Alberto and Eliana La Ferrara (2005) "Preferences for redistribution in the land of opportunities" *Journal of Public Economics,* 89(5–6), 897–931.

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

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

Austen-Smith, David, and Jeffrey S. Banks 1999 *Positive Political Theory I: Collective Preference.* Ann Arbor: The University of Michigan Press.

Barberà, Salvador, Michael Maschler and Jonathan Shalev (2001) Voting for Voters: A Model of the Electoral Evolution, *Games and Economic Behavior*, 37: 40-78.

Besley, Timothy and Stephen Coate (1998) "Sources of Inefficiency in a Representative Democ-

racy: A Dynamic Analysis," *American Economic Review,* 88(1), 139-56.

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

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

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

Bordignon, Massimo and Sandro Brusco (2006) "On Enhanced Cooperation." *Journal of Public Economics*, 90(10-11): 2063-2090.

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

Budish, Eric, Yeon-koo Che, Fuhito Kojima, and Paul Milgrom (2013) "Designing Random Allocation Mechanisms: Theory and Applications," *American Economic Review*, 103(2), 585-623.

Dolmas, James and Gregory Huffman (2004), "On the Political Economy of Immigration and Income Redistribution," *International Economic Review*, 45(4):1129-1168.

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

Glaeser, Edward and Andrei Shleifer (2005), "The Curley Effect," *Journal of Law, Economics, and Organization*, April 2005.

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

Gregory, Paul R., Philipp J.H. Schroeder, and Konstantin Sonin (2011) "Rational Dictators and the Killing of Innocents: Data from Stalin's Archives" *Journal of Comparative Economics.* 39(1), 34-42.

Hirshleifer, Jack, Michele Boldrin, and David K Levine (2009) "The Slippery Slope Of Concession," *Economic Inquiry*, vol. 47(2), pages 197-205.

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

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

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

Lipset, Seymour

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

Lizzeri, Alessandro, and Nicola Persico (2004) "Why Did the Elites Extend the Suffrage? Democracy and the Scope of Government, With an Application to Britain's 'Age of Reform'." *Quarterly Journal of Economics*, 119(2): 705-763.

Milgrom, Paul, and Christine Shannon (1994) "Monotone Comparative Statics," *Econometrica*, 62(1): 157-180.

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

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

Powell, Robert (2006) War as a Commitment Problem, *International Organization,* 60(1), 169-203.

Roberts, Kevin (1999) "Dynamic Voting in Clubs," unpublished manuscript.

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

Wolitzky, Alexander (2011) "A Theory of Repression, Extremism, and Political Succession," mimeo.

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

Appendix A: Proofs of Main Results

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 MPE 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):

$$V_j'(x) = u_j(b_{d_x}) + \beta \sum_{y \in S} q_{x,y} \sum_{k \in G} \mu_{j,k} V_k(y). \tag{A1}$$

To show that $\{V_j(x)\}_{j \in G}^{x \in S}$ safisfies increasing differences, we use the following argument. The mapping defined by (A1) is a contraction. Thus, it suffices to prove that if $\{V_j(x)\}_{j \in G}^{x \in S}$ satisfy increasing differences, then $\left\{V_j'(x)\right\}_{j \in G}^{x \in S}$, 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:

$$\begin{aligned} V_j'(z) - V_j'(x) &= (u_j(b_{d_z}) - u_j(b_{d_x})) + \beta \sum_{k \in H} \mu_{j,k} \left[\sum_{y \in S} q_{z,y} V_k(y) - \sum_{y \in S} q_{x,y} V_k(y) \right] \\ &= \left((b_j - b_{d_x})^2 - (b_j - b_{d_z})^2 \right) + \beta \sum_{k \in H} \mu_{j,k} Z_k \\ &= (b_{d_z} - b_{d_x})(2b_j - b_{d_x} - b_{d_z}) + \beta \sum_{k \in H} \mu_{j,k} Z_k, \end{aligned} \tag{A2}$$

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

$$\begin{aligned} Z_l - Z_k &= \sum_{y \in S} (q_{z,y} - q_{x,y})(V_l(y) - V_k(y)) \\ &= \sum_{y \in S} q_{z,y}(V_l(y) - V_k(y)) - \sum_{y \in S} q_{x,y}(V_l(y) - V_k(y)). \end{aligned}$$

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 $\{q_{z,\cdot}\}$ (the first term) is at least as high as that under the probability distribution $\{q_{x,\cdot}\}$, 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 $(b_{d_z} - b_{d_x})(2b_j - b_{d_x} - b_{d_z})$ is increasing in j, because b_j is increasing in j and the difference $b_{d_z} - b_{d_x}$ is positive. To show that the second term is nondecreasing in j, take two groups, $j, l \in G$ such that $j < l$. By assumption, the probability distribution $\{\mu_{j,\cdot}\}$ is first-order stochastically dominated by $\{\mu_{l,\cdot}\}$. 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 $\beta \sum_{k \in H} \mu_{j,k} Z_k$ is nondecreasing in j. Therefore, (A2) is increasing in j, which implies that $\left\{ V_j'(x) \right\}_{j \in G}^{x \in S}$ satisfies increasing differences. Given that (A1) is a contraction, this completes the proof that $\{V_j(x)\}_{j \in G}^{x \in S}$ 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 $\{w_j(x)\}_{j \in G}^{x \in S}$ satisfy (weak) increasing differences. Indeed, for $y > x$,

$$w_j(y) - w_j(x) = \sum_{k \in H} \mu_{j,k} \left(V_k(y) - V_k(x) \right),$$

and since $V_k(y) - V_k(x)$ is monotonically increasing in k, its expectation with respect to distribution $\left\{\mu_{l,\cdot}\right\}$ is at least as high as that with respect to distribution $\left\{\mu_{j,\cdot}\right\}$ 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, $w_{d_s}(x)$. 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 $a \in \arg\max_{s \in S} w_{d_x}(s)$ and $b \in \arg\max_{s \in S} w_{d_y}(s)$, in particular, this implies $w_{d_x}(a) \geq w_{d_x}(b)$ and $w_{d_y}(b) \geq w_{d_y}(a)$. But if M satisfies strict first-order stochastic dominance, then $\{w_j(x)\}_{j \in G}^{x \in S}$ 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 $\{V_j(x)\}_{j \in G}^{x \in S}$ 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

$$M = \max_{k \in G, y \in S} \left| u_k\left(b_{d_y}\right) \right|. \tag{A3}$$

For each state $s \in S$, consider a one-period game desribed 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 $p = b_{d_s}$ for all such equilibria. From the proof of Part 2 it follows that y maximizes the expected continuation utility $w_j(s) = \sum_{k \in G} \mu_{d_s,k} V_k(y)$, 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 $\{V_j(x)\}_{j\in G}^{x\in S}$ 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. ■

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

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

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 $\left|b_{d_z}-b_{d_s}^{(1)}\right|$ occurs. This means that for some $y \in S$, $\left|b_{d_y}-b_{d_s}^{(1)}\right| < \left|b_{d_z}-b_{d_s}^{(1)}\right|$. Now consider the utility of individuals from group d_s if they transited to y instead. Their gain in utility (after factor β) would be

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

provided that $\beta \in (0, \beta_0)$. Therefore, a transition to z does not maximize the continuation utility of the pivotal group d_s (they would be better off moving to y), which contradicts Part 2 of Theorem 1.

Part 2. This result follows from the following two steps.

Step 1. For sufficiently high β, there exists an equilibrium such that for each state $s \in S$, at least one of the states $z \in S$ that minimize $\left|b_{d_z} - b_{d_s}^{(\infty)}\right|$ is stable: $q_{z,z} = 1$.

Proof. First, we notice that generically (for almost all parameter values) such state z is unique for each s; the only case where if is non-unique is when $b_{d_s}^{(\infty)}$ is exactly halfway between two bliss points for some s.

To prove this result, TO BE COMPLETED.

Suppose not, i.e., from z, the society may transit to some other state x with a positive probability. If it does, then by monotonicity (Part 3 in Theorem 1) the path of states will never return to z.

Step 2. Suppose that β is sufficiently high, and in some equilibrium, for any state $s \in S$, at least one of the states $z \in S$ that minimize $\left|b_{d_z} - b_{d_s}^{(\infty)}\right|$ is stable. Then with probability 1 the sequence of states starting from s converges to such a state.

Proof. To be completed. ■

Lemma A1 *Suppose that for some j, the sequence $b_j^{(t)}$ is nondecreasing (respectively, nonincreasing). 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 nondecreasing (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

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

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

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

Applying this to x and s, we have

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

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

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

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 $q_{s,\cdot}$ and $q_{x,\cdot}$ are different, and $\left(b_j^{(t)} - b_{d_a}\right)^2$ 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 τ periods, after which it departs to $y > s$. Let us denote the probability distribution of states if an immediate transition to x occurs by $p^{(t)}_{x,\cdot}$, and that in the case an immediate transition to y occurs by $q^{(t)}_{y,\cdot}$. We know that the individuals in group j are indifferent between transiting to x and to y, meaning that

$$\sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} p^{(t)}_{x,a} \mu^{(t)}_{j,k} (b_k - b_{d_a})^2 = \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} q^{(t)}_{y,a} \mu^{(t)}_{j,k} (b_k - b_{d_a})^2 ,$$

which, by increasing differences, implies

$$\sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} p^{(t)}_{x,a} \mu^{(t+\tau)}_{j,k} (b_k - b_{d_a})^2 \leq \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} q^{(t)}_{y,a} \mu^{(t+\tau)}_{j,k} (b_k - b_{d_a})^2 ;$$

this follows from that $b^{(t+\tau)}_j \geq b^{(t)}_j$ for each t (transformations similar to the ones earlier in the proof would clearly show that only the expecteation of $\mu^{(t+\tau)}_{j,\cdot}$. Now we have

$$\begin{aligned}
\beta V_j \left(\underbrace{s, \ldots, s}_{\tau \text{ times}}, y, \ldots \right) &= \sum_{t=1}^{\tau} \beta^t \sum_{k \in G} \mu^{(t)}_{j,k} (b_k - b_j)^2 + \beta^{\tau} \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} q^{(t)}_{y,a} \mu^{(t+\tau)}_{j,k} (b_k - b_{d_a})^2 \\
&\geq \sum_{t=1}^{\tau} \beta^t \sum_{k \in G} \mu^{(t)}_{j,k} (b_k - b_j)^2 + \beta^{\tau} \sum_{t=1}^{\infty} \beta^t \sum_{a \in S} \sum_{k \in G} p^{(t)}_{x,a} \mu^{(t+\tau)}_{j,k} (b_k - b_{d_a})^2 \\
&= \beta V_j \left(\underbrace{s, \ldots, s}_{\tau \text{ times}}, x, \ldots \right) .
\end{aligned}$$

Consequently, for each such path, we have

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

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 β is sufficiently small. Consider the set $A = \left\{ x \in \mathbb{R} \mid x = b^{(1)}_{d(s)} \text{ for some } s \in S \right\}$ and $B = \left\{ y \in \mathbb{R} \mid y = \frac{b_s + b_{s+1}}{2} \text{ for some } s \in \{1, \ldots, m-1\} \right\}$. For generic parameter values, $A \cap B = \varnothing$.

If so, then there is a unique mapping satisfying the description in Part 1 of Theorem 2, and therefore, 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 σ_1 and σ_2, and let Q^1 and Q^2 be the corresponding transition matrices. Then, for generic parameter values, if $Q^1 \neq Q^2$, then there are two states $x, y \in S$, $x \neq y$, such that the distributions $q^1_{x,\cdot} \neq q^2_{x,\cdot}$ and $q^1_{y,\cdot} \neq q^2_{y,\cdot}$. 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 $q^1_{s,\cdot} \neq q^2_{s,\cdot}$. 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 $q^1_{s,\cdot} = q^2_{s,\cdot}$, 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 $z < s$, $q^1_{z,a} > 0$ implies $a \leq x$; moreover, for such z, $q^2_{z,\cdot} = q^1_{z,\cdot}$. 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: $w^1_{d_s}(x) = w^2_{d_s}(x)$. Similarly, from $q^2_{s,y} > 0$ it follows that for $z > s$, $q^2_{z,a} > 0$ implies $a \geq y$; moreover, for such z, $q^1_{z,\cdot} = q^2_{z,\cdot}$. Thus, if the society moves from state s to y, the continuation utilities of the current decision-maker again coincide: $w^1_{d_s}(y) = w^2_{d_s}(y)$. But by Part 2 of Theorem 1, we have $w^1_{d_s}(x) \geq w^1_{d_s}(y) = w^2_{d_s}(y) \geq w^2_{d_s}(x) = w^1_{d_s}(x)$, which implies that both inequalities hold with equality, in particular, $w^1_{d_s}(x) = w^1_{d_s}(y)$. 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 $q^1_{s,\cdot}$ and $q^2_{s,\cdot}$ are subsets of $\{s, x\}$, we would have $q^1_{s,s} = q^2_{s,s}$, again meaning that $q^1_{s,\cdot} = q^2_{s,\cdot}$ and contradicting the choice of s. Without loss of generality, assume $q^1_{s,x} < q^2_{s,x}$, so in equilibrium σ_1 the society stays in s longer than in equilibrium σ_2, in expectation; this means, in particular, $q^1_{s,x} < 1$ and $q^2_{s,x} > 0$. It must be that the sequence $b^{(t)}_{d_s}$ is nondecreasing and, moreover, it is nonstationary, for otherwise $q^2_{s,x} > 0$ would contradict Lemma C2.

Let $j = d_s$. The continuation utilities from moving to x are the same in both equilibria: $V_j^1(x) = V_j^2(x)$, because the transition probabilities are identical thereafter. Moreover, in equilibrium σ_2, transiting is a best response, so $V_j^2(x) \geq V_j^2(s)$, and in equilibrium σ_1, staying is a best response, so $V_j^1(s) \geq V_j^1(x)$. We thus have $V_j^1(s) \geq V_j^1(x) = V_j^2(x) \geq V_j^2(s)$, meaning that the utility of individuals from group j from staying is at least as high under σ_1 as under σ_2. Denote $V_j(s;\alpha)$ the utility of staying in s if the subsequent equilibrium play has probability α of moving to x; then $V_j\left(s;q_{s,x}^1\right) = V_j^1(s)$ and $V_j\left(s;q_{s,x}^2\right) = V_j^2(s)$.

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 α, $f(\alpha) = 0$, 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 τ periods in state s and stays there for an extra period with transition probability is α 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 ideal points $b_j^{(t+\tau)} \geq b_j^{(t)}$ for all τ, and for at least some t the inequality is strict. Therefore, we have

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

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 $f\left(q_{s,x}^1\right) = 0$, then $f\left(q_{s,x}^2\right) > 0$, meaning that $V_j\left(s;q_{s,x}^2\right) > V_j(x)$, which contradicts that moving to x is a best response in σ_2. Similarly, if $f\left(q_{s,x}^2\right) = 0$, then $f\left(q_{s,x}^1\right) < 0$, meaning that $V_j\left(s;q_{s,x}^1\right) < V_j(x)$, which contradicts that staying at s is a best response in σ_1. If $f\left(q_{s,x}^1\right) \neq 0$ and $f\left(q_{s,x}^2\right) \neq 0$, then, since staying in s is a best response in σ_1, we must have $f\left(q_{s,x}^1\right) > 0$; similarly, we must have $f\left(q_{s,x}^2\right) < 0$. But then by continuity there is $\alpha \in \left(q_{s,x}^1, q_{s,x}^2\right)$

such that $f(\alpha) = 0$. In that case, it must be that $f\left(q^1_{s,x}\right) < 0 < f\left(q^2_{s,x}\right)$ and not the other way around, a contradiction. This completes the proof of Step 1.

Step 2. Without loss of generality, suppose that m is the minimal number of states for which there are two monotone MPEs σ_1 and σ_2. Then $m = 2$.

Proof. Suppose not, then either $m = 1$ or $m \geq 3$. If $m = 1$, there is only one possible transition mapping: Q with $q_{1,1} = 1$. Suppose $m > 3$ and let Q^1 and Q^2 the transition matrices in equilibria σ_1 and σ_2. Let $Z \subset S$ be the set of $z \in S$ such that $q^1_{z,\cdot}$ and $q^2_{z,\cdot}$ are different distributions; from Step 1 it follows that $|Z| \geq 2$. In what follows, let $L = \left\{s \in S : \forall x > s : q^1_{s,x} = q^2_{s,x} = 0\right\}$ and $R = \left\{s \in S : \forall x < s : q^1_{s,x} = q^2_{s,x} = 0\right\}$. By Lemma A1, $L \cup R = S$; let us denote $I = L \cap R$.

First, we show that if $s \in S$ and $1 < s < m$, then $s \notin I$. Indeed, otherwise, we would have $q^1_{s,s} = q^2_{s,s} = 1$. 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 $q^1_{z,\cdot}$ and $q^2_{z,\cdot}$ 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 $q^2_{2,2} = 1$; this means $q^1_{2,2} < 1$, and thus $q^1_{2,1} > 0$ and again by monotonicity $q^1_{1,1} = 1$. This implies (by Lemma A1) that the sequence $b^{(t)}_{d_1}$ is nondecreasing (because equilibrium σ_2 exists) and $b^{(t)}_{d_2}$ is nonincreasing (because equilibrium σ_1 exists). Suppose $b^{(\infty)}_{d_1} < \frac{b_{d(1)}+b_{d(2)}}{2}$, then one can easily prove (similar to the proof of Lemma A1) that σ_2 cannot be an equilibrium, as the group d_1 would strictly prefer to stay in 1 under σ_2, while $q^2_{1,1} < 1$. Similarly, if $b^{(\infty)}_{d_2} > \frac{b_{d(1)}+b_{d(2)}}{2}$, then σ_1 cannot be an equilibrium. The only remaining possibility (since $b^{(\infty)}_{d_1} \leq b^{(\infty)}_{d_2}$ by Assumption 1) is where $b^{(\infty)}_{d_1} = b^{(\infty)}_{d_2} = \frac{b_{d(1)}+b_{d(2)}}{2}$, which is nongeneric (moreover, for both σ_1 and σ_2 to exist, it must be that $b^{(t)}_{d_1} = b^{(t)}_{d_2} = \frac{b_{d(1)}+b_{d(2)}}{2}$ 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. ■

Proof of Theorem 7. To be completed. ■

Proof of Corollary 3. To be completed. ■

Proof of Theorem 8. To be completed. ■

Proof of Corollary 4. To be completed. ■

Appendix B: Examples

Example 1 ***(Multiple representations of transition matrix as lottery over permutations)*** *For a given A, the distribution μ such that $A = \Omega(\mu)$ need not be unique. E.g., take $n = 3$ and*

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

It may be represented as

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

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

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

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

Example 2 ***(Multiple equilibria)*** *There are five groups with political bliss points $b_{1,2,3,4,5} = -\frac{21}{10}, -1, 0, 1, \frac{21}{10}$ (there would be two equilibria even if the extreme political bliss points are ± 2 rather than ± 2.1, but this would be a knife-edge case). All $A_i = 0$, discount factor $\beta = \frac{1}{2}$, and the reshuffling matrix M is given by*

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

(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* $-1, 0, 1, \frac{21}{10}$ *with equal probabilities. For quadratic utility functions, it is the average that matters, and his expected political bliss point equals* $\frac{21}{40}$*. Since* $\frac{21}{40}$ *is closer to* $b_4 = 1$ *than to* $b_3 = 0$*, 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* ϕ_1 *and* ϕ_2 *are equilibria (verifying that other groups act as prescribed is straightforward).*

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

so immediately. (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* ϕ_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 3 ***(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* $\mathbf{b} = (0, 0.9, 1, 2, 30)'$*, respectively. All* $A_i = 0$*, and the social mobility matrix is given by*

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

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

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

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

It would seem that moving to state 2 *rather than mixing between states* 1 *and* 3 *is a reasonable middle ground 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 ideal policy is very close to that in state* 3*, make sure that the immediate utility of members of group* 1 *from moving to state* 2 *is only marginally better than that from moving to state* 3*, but it delays transition to state* 4*. As a result, the path initiated by moving to state* 2 *runs in-between the corresponding paths for staying at* 1 *and moving to* 3*, but in the important few periods the payoff is closer to the path that yields a lower payoff in that period. As a result, in equilibrium, the mixing is between staying and moving to a non-neighboring state, even though all utility functions are concave and even quadratic.*

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

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

For such M*, the equilibrium is generically unique for any discount factor* β*.*

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

Consider the incentives of groups 1 *and* 5 *(they are symmetric). For members of group* 5*, the preferences of their future selves are:* $\frac{17}{5} = 3.4, \frac{67}{25} = 2.68, \frac{1013}{500} = 2.026, \frac{3733}{2500} = 1.4932, \ldots$ *Thus, ideally, members of this group would prefer to have state* 4 *in periods* $2, 3, 4$*, and state* 3

thereafter. However, by the argument above, they can only enjoy state 4 *in one period, for after that group* 4 *which is in power in that state would move to state* 3.

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

The following describes the equilibrium:

If $0 < \beta < 0.0282$*, then the equilibrium is* $\phi(1,2,3,4,5) = (2,3,3,3,4)$.

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 preferrable to spend the third period in states* 5 *or* 4 *rather than* 3*. This leads to mixing.*

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

If $0.5621 < \beta < 1$*, then the equilibrium involves mixing between transiting from* 1 *to* 2 *and staying at* 1*, and, symmetrically, between transiting from* 5 *to* 4 *and staying at* 5 *(for example, if* $\beta = 0.9$*, then they stay with probability* 0.69 *and move with probability* 0.31*). For these values of* β*, distant future is sufficiently important. Decision-makers in state* 5 *still prefer to stay in state* 5 *instead of moving to state* 4 *immediately; however, now the weight given to distant future is high, and so if the society were to stay in state* 5 *forever, they would prefer to deviate immediately and move to* 4 *(followed by* 3*).*

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

Appendix C: Additional Results

C1 Conditions for mixed strategies

Our next results are theoretical, highlighting cases where we should see equilibria in mixed strategies. <These imply "slow" convergence to the steady state.> Consider first the following definition.

Definition 3 *We say that social mobility is slow [or, alternatively, that periods are frequent enough?] if the preferred state of each individual's today's and tomorrow's selves coincide. More formally, this property holds if matrix M satisfies*

$$b_j = \arg \min_{b \in \{b_k\}_{k=1}^n} \left| (M\mathbf{b})_j - b \right|.$$

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

Proposition C1 *The following is true for any M, any $\mathbf{b}$ and A:*

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

(ii) Suppose that social mobility is slow, but existing (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. <In fact, for high β, it is either no move or mixing.>

Moreover, for any fixed β, the equilibrium is in pure strategies if M is sufficiently close to identity matrix. <Perhaps not so interesting: in the limit, actually, the equilibrium involves no transitions at all.>

Interestingly, with a finite number of periods, there would generically be only equilibria in pure strategies. A proof is available upon request. For example, in Example **??**, there would be no transition in the last period, but there would be in the last-but-one <ELABORATE AFTER CHECKING>.

(mixed strategies are allowed, and we will see below that this is important). Namely, we consider strategies that map any state j to a probability distribution $\Delta(\mathbb{R} \times S)$ over pairs of policy p and next period's state s, which depend only on the current state j, but neither on

history nor on the identities of individuals that compose different groups. More precisely, the definition is the following.

Definition 4 *A mapping $\sigma : S \to \Delta(\mathbb{R} \times S)$ mapping every s into probability distribution over (p, s) is a MPE if and only if for every $j \in S$ and every (p^*, s^*) in the support of $\sigma(j)$,*

$$(p^*, s^*) \in \arg \max_{(p,s) \in \mathbb{R} \times S} \left\{ u_j(p) + \beta \sum_{k \in N} \mu_{j,k} V_k^{\sigma}(s) \right\}, \tag{C1}$$

where $\{V_\cdot^{\sigma}(\cdot)\}$ are expected continuation utilities of individuals of group H_k if the current period starts in state s, if strategies in σ are played ever after, i.e., $\{V_\cdot^{\sigma}(\cdot)\}$ are the unique solution to the system of equations

$$V_j^{\sigma}(z) = \int u_j(p)\, dF^{\sigma(j)}(p) + \beta \sum_{s \in S} q_s^{\sigma(j)} \sum_{k \in N} \mu_{j,k} V_k^{\sigma}(s), \tag{C2}$$

where $F^{\sigma(j)}$ and $q^{\sigma(j)}$ are the distributions of the two components of $\sigma(j)$, respectively.

C2 Conditions for monotonicity of MPE

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

Theorem 9 *Every MPE is monotone if either of the following conditions holds:*

(i) The discount factor β is sufficiently low;

(ii) The discount factor β is sufficiently high;

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

(iv) MAYBE SOMETHING ELSE - NEED TO THINK.

C3 Some results on social mobility matrices

Fact. 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 C2 ***(Characterization of matrices, satisfying within-person monotonicity-for any*** *b) 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 irreducible component K_x, corresponding to groups $H_{l_x}, \ldots, H_{r_x}$, there is a number $\kappa_x \in [0,1]$, such that the transition probabilities for all groups except for the two extreme ones, i.e., for $l_x < j < l_y$, satisfy*

$$\mu_{jk} = \kappa_x \frac{n_k}{\sum_{i=l_x}^{l_y} n_i} + (1-\kappa_x)\,\mathbf{1}_{j=k}. \tag{C3}$$

Proof. Sufficiency. Straightforward.

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

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

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

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

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

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

We have thus proved that $\mu_{jk_1} n_{k_2} = \mu_{jk_2} n_{k_1}$ for all $k_1, k_2 \neq j$, and thus there is $\kappa_x = \kappa_{x,j}$ such that μ_{jk} are given by (C3). 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 C2 does not require anything the extreme groups in a given class to conform to the same formula given by (C3). For example, the following matrices satisfy* (2), (3), *as well as monotonicity across and within individuals:*

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

C4 Some examples with shocks

Example 1 ***(Mixed equilibrium with mixing over two transitions)*** *THIS IS CASE WITH SHOCKS. There are five groups with political bliss points $b_{1,2,3,4,5} = -2, -1, 0, 1, 2$. All $A_i = 0$, discount factor $\beta = \frac{1}{2}$, and the reshuffling matrix M is given by*

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

Without exogenous shocks, the unique transition mapping would be $\phi^*(1, 2, 3, 4, 5) = (2, 3, 3, 3, 4)$.

Suppose, however, that if the current state is 2, then with probability $\frac{3}{10}$ the society moves to state 1. [[[IF WITH PROB $\frac{3}{10}$ IT MOVES TO 3 AND ONLY WITH PROB $\frac{2}{5}$ STAYS, THIS IS TOTALLY FINE.]]] In this case, moving from state 1 to state 3 might take a longer time than the current decision-makers in state 1 would want, so if future decision-makers move to 2, then current ones are better off moving to 3. However, if future decision-makers move to 3,

then current ones will be happy to move to 2 *instead, as in this case, an exogenous shock would affect them at most once. As a result, there is no pure-strategy equilibrium. DETAILS ON HOW MIXED STRATEGY EQUILIBRIUM LOOKS LIKE.*

Example 2 ***(Multiple equilibria in the presence of exogenous shocks)*** *THIS IS CASE WITH SHOCKS. There are three groups with political bliss points* $b_{1,2,3} = -1, 0, 1$*. All* $A_i = 0$*, and the weights of groups are* $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$*, respectively. The discount factor is* $\beta = 0.9$*, and the reshuffling matrix* M *is given by*

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

In the absence of an exogenous shock, the unique equilibrium transition mapping would be $\phi^*(1,2,3) = (2,2,2)$*.*

Suppose, however, that if the current state is 2*, then with probability* $\frac{1}{5}$*, the society moves to state* 1*, and with probability* $\frac{1}{5}$*, it moves to state* 3 *(i.e., the society stays in state* 2 *with probability* $\frac{3}{5}$*. One can easily verify that there is still an equilibrium where the transition mapping is* $\phi^*(1,2,3) = (2,2,2)$*. However, in this case, there is also another equilibrium, where the transition mapping is* $\phi'(1,2,3) = (1,2,3)$*. The intuition for this equilibrium is the following: if the future decision-makers are expected to play* ϕ^**, the current decision-makers in states* 1 *and* 3 *are marginally in favor of moving to state* 2*. However, if members of group* 1 *know that with some probability they will end up in state* 3 *because of an exogenous shock, and when they do, group* 3 *would never relinquish power, then they would prefer not to move to state* 2 *in the first place. In other words, the decisions of groups* 1 *and* 3 *to move to state* 2 *from the states where they are in power are strategic complements, and this results in multiplicity of equilibria.*

C5 Extra Proofs

Sufficiency. Straightforward.

$$
\begin{aligned}
V_j(x) &= \sum_{t=0}^{\infty}\beta^t \sum_{y\in S} q_{x,y}^{(t)} \sum_{k\in H}\mu_{j,k}^{(t)} u_k\left(b_{d_y}\right) \\
&= \sum_{t=0}^{\infty}\beta^t \sum_{y\in S} q_{x,y}^{(t)} \sum_{k\in H}\mu_{j,k}^{(t)} \left(A_k - \left(b_k - b_{d_y}\right)^2\right) \\
&= \sum_{t=0}^{\infty}\sum_{k\in H}\beta^t \mu_{j,k}^{(t)} A_k - \sum_{t=0}^{\infty}\beta^t \sum_{y\in S}\sum_{k\in H} q_{x,y}^{(t)}\mu_{j,k}^{(t)}\left(b_k - b_{d_y}\right)^2
\end{aligned}
$$

The first term depends on j but not on x; denote it by B_j. As for the second term, we have

$$
\begin{aligned}
\sum_{y\in S}\sum_{k\in H} q_{x,y}^{(t)}\mu_{j,k}^{(t)}\left(b_k - b_{d_y}\right)^2 &= \left(\sum_{k\in H}\mu_{j,k}^{(t)} b_k - \sum_{y\in S} q_{x,y}^{(t)} b_{d_y}\right)^2 \\
&\quad + \left(\sum_{k\in H}\mu_{j,k}^{(t)} b_k^2 - \left(\sum_{k\in H}\mu_{j,k}^{(t)} b_k\right)^2\right) + \left(\sum_{y\in S} q_{x,y}^{(t)} b_{d_y}^2 - \left(\sum_{y\in S} q_{x,y}^{(t)} b_{d_y}\right)^2\right) \\
&\quad C_j + D_k.
\end{aligned}
$$

(this is just a variant of $\mathbb{E}\left(X-Y\right)^2 = \left(\mathbb{E}X - \mathbb{E}Y\right)^2 + \mathbb{V}x + \mathbb{V}y$ for appropriately defined X and Y). Denote

$$
\begin{aligned}
C_j &= \sum_{t=0}^{\infty}\beta^t\left(\sum_{k\in H}\mu_{j,k}^{(t)} b_k^2 - \left(\sum_{k\in H}\mu_{j,k}^{(t)} b_k\right)^2\right), \\
D_x &= \sum_{t=0}^{\infty}\beta^t\left(\sum_{y\in S} q_{x,y}^{(t)} b_{d_y}^2 - \left(\sum_{y\in S} q_{x,y}^{(t)} b_{d_y}\right)^2\right),
\end{aligned}
$$

we then have

$$
V_j(x) = B_j - C_j - D_x - \sum_{t=0}^{\infty}\beta^t\left(\sum_{k\in H}\mu_{j,k}^{(t)} b_k - \sum_{y\in S} q_{x,y}^{(t)} b_{d_y}\right)^2.
$$

Now take $z > x$ and consider the difference

$$
\begin{aligned}
V_j(z) - V_j(x) &= D_x - D_z \\
&\quad - \sum_{t=0}^{\infty}\beta^t\left[\left(\sum_{k\in H}\mu_{j,k}^{(t)} b_k - \sum_{y\in S} q_{z,y}^{(t)} b_{d_y}\right)^2 - \left(\sum_{k\in H}\mu_{j,k}^{(t)} b_k - \sum_{y\in S} q_{x,y}^{(t)} b_{d_y}\right)^2\right] \\
&= D_x - D_z \\
&\quad + \sum_{t=0}^{\infty}\beta^t\left(\sum_{y\in S} q_{z,y}^{(t)} b_{d_y} - \sum_{y\in S} q_{x,y}^{(t)} b_{d_y}\right)\left(2\sum_{k\in H}\mu_{j,k}^{(t)} b_k - \sum_{y\in S} q_{z,y}^{(t)} b_{d_y} - \sum_{y\in S} q_{x,y}^{(t)} b_{d_y}\right).
\end{aligned}
$$

This is linear in $\sum_{k\in H}\mu_{j,k}^{(t)}b_k = b_j^{(t)}$.

Now take $l > j$ and consider the difference

$$\begin{aligned}
V_l(x) - V_j(x) &= (B_l - C_l) - (B_j - C_j) \\
&\quad -\sum_{t=0}^{\infty}\beta^t\left[\left(\sum_{k\in H}\mu_{l,k}^{(t)}b_k - \sum_{y\in S}q_{x,y}^{(t)}b_{d_y}\right)^2 - \left(\sum_{k\in H}\mu_{j,k}^{(t)}b_k - \sum_{y\in S}q_{x,y}^{(t)}b_{d_y}\right)^2\right] \\
&= (B_l - C_l) - (B_j - C_j) \\
&\quad +\sum_{t=0}^{\infty}\beta^t\left(\sum_{k\in H}\mu_{l,k}^{(t)}b_k - \sum_{k\in H}\mu_{j,k}^{(t)}b_k\right)\left(2\sum_{y\in S}q_{x,y}^{(t)}b_{d_y} - \sum_{k\in H}\mu_{l,k}^{(t)}b_k - \sum_{k\in H}\mu_{j,k}^{(t)}b_k\right) \\
&= (B_l - C_l) - (B_j - C_j) - \sum_{t=0}^{\infty}\beta^t\left(\sum_{k\in H}\mu_{l,k}^{(t)}b_k - \sum_{k\in H}\mu_{j,k}^{(t)}b_k\right)\left(\sum_{k\in H}\mu_{l,k}^{(t)}b_k + \sum_{k\in H}\mu_{j,k}^{(t)}b_k\right) \\
&\quad +2\sum_{t=0}^{\infty}\beta^t\left(\sum_{k\in H}\mu_{l,k}^{(t)}b_k - \sum_{k\in H}\mu_{j,k}^{(t)}b_k\right)\sum_{y\in S}q_{x,y}^{(t)}b_{d_y}
\end{aligned}$$

Now, the term $(B_l - C_l) - (B_j - C_j)$ is constant in x.

$$\begin{aligned}
V_j(x) &= \sum_{t=0}^{\infty}\beta^t\sum_{y\in S}q_{x,y}^{(t)}\sum_{k\in H}\mu_{j,k}^{(t)}u_k\left(b_{d_y}\right) \\
&= \sum_{t=0}^{\infty}\beta^t\sum_{y\in S}q_{x,y}^{(t)}\sum_{k\in H}\mu_{j,k}^{(t)}\left(A_k - \left(b_k - b_{d_y}\right)^2\right) \\
&= \sum_{t=0}^{\infty}\sum_{k\in H}\beta^t\mu_{j,k}^{(t)}A_k - \sum_{t=0}^{\infty}\beta^t\sum_{y\in S}\sum_{k\in H}q_{x,y}^{(t)}\mu_{j,k}^{(t)}\left(b_k - b_{d_y}\right)^2
\end{aligned}$$

The first term depends on j but not on x; denote it by B_j. As for the second term, we have

$$\begin{aligned}
\sum_{y\in S}\sum_{k\in H}q_{x,y}^{(t)}\mu_{j,k}^{(t)}\left(b_k - b_{d_y}\right)^2 &= \left(\sum_{k\in H}\mu_{j,k}^{(t)}b_k - \sum_{y\in S}q_{x,y}^{(t)}b_{d_y}\right)^2 \\
&\quad + \left(\sum_{k\in H}\mu_{j,k}^{(t)}b_k^2 - \left(\sum_{k\in H}\mu_{j,k}^{(t)}b_k\right)^2\right) + \left(\sum_{y\in S}q_{x,y}^{(t)}b_{d_y}^2 - \left(\sum_{y\in S}q_{x,y}^{(t)}b_{d_y}\right)^2\right) \\
&\quad C_j + D_k.
\end{aligned}$$

(this is just a variant of $\mathbb{E}(X-Y)^2 = (\mathbb{E}X - \mathbb{E}Y)^2 + \mathbb{V}x + \mathbb{V}y$ for appropriately defined X and

Y). Denote

$$\begin{aligned} C_j &= \sum_{t=0}^{\infty} \beta^t \left(\sum_{k \in H} \mu_{j,k}^{(t)} b_k^2 - \left(\sum_{k \in H} \mu_{j,k}^{(t)} b_k \right)^2 \right), \\ D_x &= \sum_{t=0}^{\infty} \beta^t \left(\sum_{y \in S} q_{x,y}^{(t)} b_{d_y}^2 - \left(\sum_{y \in S} q_{x,y}^{(t)} b_{d_y} \right)^2 \right), \end{aligned}$$

we then have

$$V_j(x) = B_j - C_j - D_x - \sum_{t=0}^{\infty} \beta^t \left(\sum_{k \in H} \mu_{j,k}^{(t)} b_k - \sum_{y \in S} q_{x,y}^{(t)} b_{d_y} \right)^2.$$

Now take $z > x$ and consider the difference

$$\begin{aligned} V_j(z) - V_j(x) &= D_x - D_z \\ &\quad - \sum_{t=0}^{\infty} \beta^t \left[\left(\sum_{k \in H} \mu_{j,k}^{(t)} b_k - \sum_{y \in S} q_{z,y}^{(t)} b_{d_y} \right)^2 - \left(\sum_{k \in H} \mu_{j,k}^{(t)} b_k - \sum_{y \in S} q_{x,y}^{(t)} b_{d_y} \right)^2 \right] \\ &= D_x - D_z \\ &\quad + \sum_{t=0}^{\infty} \beta^t \left(\sum_{y \in S} q_{z,y}^{(t)} b_{d_y} - \sum_{y \in S} q_{x,y}^{(t)} b_{d_y} \right) \left(2 \sum_{k \in H} \mu_{j,k}^{(t)} b_k - \sum_{y \in S} q_{z,y}^{(t)} b_{d_y} - \sum_{y \in S} q_{x,y}^{(t)} b_{d_y} \right). \end{aligned}$$

This is linear in $\sum_{k \in H} \mu_{j,k}^{(t)} b_k = b_j^{(t)}$.