Policy Persistence and Drift in Organizations*

Germán Gieczewski†

January 2019

Abstract

I analyze the evolution of organizations that allow free entry and exit of members, such as cities and trade unions. Current members choose a policy for the organization. In turn, policy changes can attract newcomers or drive away dissatisfied members, affecting the set of future policymakers. The resulting feedback effects may take the organization down a "slippery slope", which agents allow in equilibrium despite being forward-looking and patient, a result that contrasts with existing models of elite clubs. The model explains how quickly the organization approaches a steady state; how this limit depends on the distribution of agents' preferences and the initial policy; and when a population of mostly moderate agents might support extremist organizations.

Keywords: dynamic policy choice, median voter, slippery slope, endogenous population, extremism

1 Introduction

This paper studies the dynamic behavior of organizations that are member-owned—that is, whose members choose policies through a collective decision-making process—and which allow for the free entry and exit of members. In this context, policy and

*I am deeply indebted to Daron Acemoglu and Juuso Toikka for their continuous support and numerous suggestions that substantially improved the paper. I would also like to thank Alessandro Bonatti, Glenn Ellison, Robert Gibbons, Matías Iaryczower, Mihai Manea, Robert Townsend, Alex Wolitzky, Muhamet Yildiz, and seminar participants at MIT, Columbia University, the University of Chicago and the Warwick-Princeton-Utah Political Economy Conference for helpful comments. All remaining errors are my own.

†Department of Politics, Princeton University.

membership decisions affect one another: different policies will appeal to or drive away different prospective members, and different groups will make different choices when in charge of the organization. As a result, the equilibrium policy path may exhibit substantial drift over time (an initial policy may attract a set of members wanting a different policy, which in turn attracts other agents, and so on) as well as path-dependence (two organizations with identical fundamentals but different initial policies may exhibit divergent behavior in the long run).

A prominent example where these issues arise is that of cities or localities. Conceptualize each city as an organization and its inhabitants as members. Cities allow people to move in and out freely, and their inhabitants vote for local authorities who implement policies, such as the level of property taxes, the quality of public schools and housing regulations. The interplay between policy changes and migration can lead to demographic and socioeconomic shifts interpreted as urban decay and revitalization, suburbanization, and so on.

The relationship between local taxes and migration has been studied since Tiebout (1956), which considers a population able to move across a collection of communities with fixed policies. Epple and Romer (1991) allow for both redistributive policies and location decisions to respond to one another, but study the problem in static equilibrium, i.e., under the assumption that whatever temporary imbalance might exist between the people living in a community and the policies they want has already resolved itself.

A natural follow-up question to this literature is whether, in a dynamic model, communities will converge to such a static equilibrium—in other words, to a steady state—quickly, slowly or not at all. The main reason convergence may fail to occur is a fear of *slippery slopes*. Here is a simple example: suppose that a community that has set a local tax rate $x_0 = 0.2$ attracts a population whose median voter, m_0, prefers a tax rate $x_1 = 0.18$. m_0 knows that lowering the tax rate to x_1 would attract a different population whose median voter, m_1, has bliss point $x_2 = 0.16$. In turn, if the tax rate is lowered to x_2, a population shift will beget a median voter wanting a tax rate $x_3 = 0.15$, and so on. If agents vote myopically, the tax rate will quickly move not to x_1 but to a much lower steady state, say $x_\infty = 0.1$. Foreseeing this, m_0 might prefer not to change the tax rate after all.

This paper shows that, in a dynamic model where both policies and membership are endogenous, communities *will*, in fact, converge to a steady state, and the set of

steady states is independent of the agents' common discount factor. However, the speed of convergence is slow relative to the discount factor, as slippery slope concerns induce them to make smaller policy changes than their myopic preferences would dictate. In practical terms, communities we observe in the world at any given time may well fail to be in static equilibrium. Even then, the model yields a tractable characterization of the transition dynamics along the equilibrium path with which we can predict their future behavior.

The model yields other substantively valuable insights. For instance, it allows us to predict when organizations will cater to mainstream tastes and when they might be captured by extremists. Broadly speaking, policy drift leads organizations towards high-density areas of the preference distribution, which favors centrism if the distribution is unimodal with a mode near the center. However, a pocket of agents concentrated at an extreme can also support a steady state; whether the organization becomes 'trapped' there may depend on its initial policy and other historical accidents. More importantly, extremism is much more likely when agents' willingness to join is asymmetric across moderates and extremists (i.e., extremists are more willing to be in a moderate organization than vice versa). In general, steady states are more sensitive to the shape of the preference distribution than in models with a fixed population, as they tend to be located close to modes of the distribution rather than near the population median. In particular, when the distribution is close to uniform, small changes to its density can result in dramatic swings in the set of steady states. Hence, a slow, continuous demographic change may at some point trigger a fundamental shift in the organization's policy path which would appear sudden to an outside observer.

The paper is connected to several strands of literature. First, as noted previously, it can be seen as a study of dynamic Tiebout competition. There is a large literature on the Tiebout hypothesis (see, e.g., Cremer and Pestieau (2004) for a review), but most papers in it assume that policies and location decisions must be in static equilibrium, and hence are silent on the transition dynamics that we focus on.

Second, our model can be applied to other organizations with open membership, such as trade unions, nonprofits, and many neighborhood and sports clubs, and it is therefore relevant to applied theory papers about such organizations. For instance, Grossman (1984) provides an explanation for why wages may not go down in a unionized sector after an increase in international competition: layoffs selectively affect less senior workers, so the median voter within the union becomes more senior and more

securely employed, hence prone to making more aggressive wage demands. As in the Tiebout literature, Grossman (1984) assumes that policy and membership are always in static equilibrium, i.e., that they adjust immediately after an external shock; this paper can be seen as providing a model of the transition dynamics.

Finally, the paper makes several contributions to the literature on "elite clubs" (i.e., clubs which can strategically restrict the entry of newcomers or even remove existing members—relevant applications include immigration and enfranchisement). First, we show that, despite the apparent substantive differences, our model of clubs with free entry is closely related to several existing models of elite clubs. Indeed, the tension in both types of models derives from a coupling of policy and decision-making power (i.e., agents cannot choose policies and future distributions of power independently) which introduces the same intertemporal trade-offs in both cases. Second, we extend and clarify existing results regarding the transition dynamics and long-run behavior of these models. Most papers in this literature—for instance, Roberts (2015), Barbera, Maschler and Shalev (2001), Acemoglu, Egorov and Sonin (2008, 2012, 2015) and related work—assume a discrete policy space and obtain the result that "intrinsic" steady states can arise, and are likely to do so when agents are patient;[1] the equilibrium path and the set of steady states can be characterized recursively, but not easily solved for. We show that if a continuous policy space is assumed, these results are overturned: all steady states are myopically stable, they are therefore independent of agents' discount factors, and the equilibrium path can be characterized explicitly (in some cases, in closed form) when agents are patient. Indeed, these results are still true if the policy space is discrete but fine enough; when the policy space is fine *and* agents are patient, which set of results we get depends on the order of limits. A closely related literature on dynamic policy selection (Jack and Lagunoff, 2006; Bai and Lagunoff, 2011) does consider continuous policy spaces, but their results are restricted to smooth equilibria, which are not guaranteed to exist generically;[2] the results in this paper are derived either for all equilibria or for classes of equilibria that exist generically. On a technical note, this paper is also the first in this literature to obtain tractable results in a setting that violates the single-crossing

[1] In the language of Roberts (2015), an intrinsic steady state is a policy that is chosen forever on the equilibrium path, but which would not be an equilibrium in a static version of the game—in other words, agents choose not to change it because they fear the subsequent policy changes that would follow.

[2] See Appendix D for details.

assumption on preferences—a necessary complication that arises in our free-entry world, stemming from the fact that agents too far from the chosen policy can cut their losses by leaving the organization.

The paper is structured as follows. The basic model is presented in Section 2. Section 3 proves some fundamental properties of all equilibria which pin down the organization's policy in the long run. Section 4 characterizes the transition dynamics. Section 5 translates the model to the 'elite clubs' setting, and Section 6 discusses the practical implications of the results. Section 7 is a conclusion. All proofs can be found in the Appendices.

2 The Model

There is a club existing in discrete time $t = 0, 1, \ldots$. Each agent is characterized by a type $\alpha \in [-1, 1]$; the population of agents is given by a continuous density f with support $[-1, 1]$. All agents are potential members.

At each integer time $t \geq 1$, two events take place. First there is a *voting stage*, in which a set of *existing members* $I_{t-1} \subseteq [-1, 1]$ votes on a *policy* $x_t \in [-1, 1]$ to be implemented during the period $[t, t + 1)$. Immediately after, in the *membership stage*, all agents observe x_t and decide whether to be members during the upcoming period $[t, t + 1)$. Agents can freely enter and leave the club as many times as desired at no cost.[3] The set of agents who choose to be members at time t constitutes I_t, the set of incumbent members at the $t + 1$ voting stage.[4] At $t = 0$ the game starts with a membership stage; the club's initial policy is exogenously given.

The essential feature of this setup is that membership affects both an agent's utility and her right to vote. Agents decide whether to be in the club based on their private utility, since their voting power is diluted by the high number of voters, but aggregate membership decisions determine future policies.

[3]The model can be extended to add entry and exit costs; see Appendix E.2 for details.

[4]The assumption that agents vote the period *after* joining prevents self-fulfilling equilibria, in which agents who dislike the club's current policy might join because they expect to immediately change the policy to their liking. Alternatively, we could assume that agents can enter or leave at any time $t \in \mathbb{R}_{\geq 0}$ but gain voting rights after being members for a short time $1 > \epsilon > 0$, with identical results.

Preferences

An agent α has utility

$$U_\alpha\left((x_t)_t, I_\alpha\right) = \sum_{t=0}^{\infty} \delta^t I_{\alpha t} u_\alpha(x_t),$$

where $I_{\alpha t} = \mathbb{1}_{\{\alpha \in I_t\}}$ denotes whether α is a member at time t. In other words, the agent can obtain a flow payoff $u_\alpha(x_t)$ from joining the organization, or leave and obtain a payoff of zero. We assume that:

A1 $u_\alpha(x) : [-1,1]^2 \to \mathbb{R}$ is C^2.

A2 There are $0 < M' < M$ such that $M' \leq \dfrac{\partial^2}{\partial\alpha\partial x} u_\alpha(x) \leq M$ for all α, x.

A3 $u_\alpha(\alpha) > 0$ for all $\alpha \in [-1,1]$.

A4 For a fixed α_0, $u_{\alpha_0}(x)$ is strictly concave in x with peak $x = \alpha_0$.

A5 For a fixed x_0, $\frac{\partial u_\alpha(x_0)}{\partial\alpha} > 0$ if $\alpha < x_0$ and $\frac{\partial u_\alpha(x_0)}{\partial\alpha} < 0$ if $\alpha > x_0$.

Taken together, assumptions A2-A5 say roughly that agent α has bliss point α and wants to be in the club if the policy x_t is close enough to α; higher agents prefer higher policies; and the set of agents desiring membership is always an interval. To fix ideas, the reader may focus on the case $u_\alpha(x) = C - (\alpha - x)^2$, where $C > 0$.

Finally, we assume that:

A6 An agent α's preferences are as defined by U_α when comparing any two paths $S = (s_0, s_1, \ldots)$, $T = (t_0, t_1, \ldots)$ with membership rules I_α^S, I_α^T that are not both zero. However, if $I_\alpha^S = I_\alpha^T = (0, 0, \ldots)$, then α prefers (S, I_α^S) to (T, I_α^T) iff $u_\alpha(s_0) \geq u_\alpha(t_0)$.

In other words, given two paths that both induce the agent to quit forever, the agent breaks ties in favor of the path with the better current policy. This prevents members who intend to quit immediately after the voting stage from making arbitrary choices out of indifference.[5]

[5]This tie-breaking rule would be uniquely selected if we modified the game to add a small time gap between the voting and membership stages, so that outgoing members at time t receive a residual payoff $\varepsilon u_\alpha(x_t)$ from the new policy x_t.

$0 < \delta < 1$ is a common discount factor. That players are forward-looking is crucial: when choosing a policy, they must take into account how it will drift in the future according to the equilibrum path.

Equilibrium Concept

We will impose two important simplifying assumptions on our equilibrium analysis. First, we will abstract from the details of the voting process and assume simply that Condorcet-winning policies are chosen on the equilibrium path.[6] Second, we will focus on Markov strategies. That is, when votes are cast at time t, the only state variable that voters condition on is the set of current members, I_{t-1}; similarly, when entry and exit decisions are made, the only state variable is the chosen policy, x_t.[7] Formally

Definition 1. A Markov strategy profile $(\tilde{s}, I)$ is given by a policy function $\tilde{s} : \mathcal{L}([-1,1]) \to [-1,1]$ and a membership function $I : [-1,1] \to \mathcal{L}([-1,1])$, where $\mathcal{L}([-1,1])$ is the set of Lebesgue-measurable sets contained in $[-1,1]$.

We denote by $s = \tilde{s} \circ I$ the *successor function*. A policy x induces a set of members $I(x)$, who will vote for policy $\tilde{s}(I(x)) = s(x)$ in the next period. Hence, an initial policy y leads to a policy path $S(y) = (y, s(y), s^2(y), \ldots)$.

Definition 2. A Markov Perfect Equilibrium (MPE) is a Markov strategy profile $(\tilde{s}, I)$ such that:

1. Given a policy x, $\alpha \in I(x)$ iff $u_\alpha(x) \geq 0$.

2. Given a set of voters I, the policy path $S(\tilde{s}(I))$ is a Condorcet winner among the available policy paths. That is, for each $y \neq \tilde{s}(I)$, a weak majority of I weakly prefers $S(\tilde{s}(I))$ to $S(y)$.[8] Moreover, if I and I' differ by a set of measure zero, then $\tilde{s}(I) = \tilde{s}(I')$.

[6] This can be rationalized as the result of two office-motivated politicians offering policies to the voters in Downsian competition. For the details, see Appendix E.5.

[7] For an analysis of non-Markovian equilibria, see Appendix E.1.

[8] Note that only one-shot deviations are considered: if voters in $I(x)$ choose a policy $y \neq s(x)$, they expect that after this deviation the MPE will be followed otherwise, i.e., the policy path will be $(y, s(y), \ldots)$ instead of $(s(x), s^2(x), \ldots)$.

From here on we will describe equilibria by the functions I and s rather than I and $\tilde{s}$. This is without loss of detail, as the set of voters is always of the form $I(x)$ on the equilibrium path.[9]

It will be useful to define two additional objects. First, when studying the long-run behavior of the policy path, we will use the following notion of steady states:

Definition 3. $x \in [-1, 1]$ is a *steady state* of a successor function s if $s(x) = x$. x is *stable* if there is a neighborhood $(a, b) \ni x$ such that $s^t(y) \xrightarrow[t\to\infty]{} x$ for all $y \in (a, b)$. We refer to the largest such neighborhood as the *basin of attraction* of x.

Second, we define the *median voter function* m as follows: given a policy x, $m(x)$ is the median member of the induced voter set $I(x)$, i.e., $\int_{-1}^{m(x)} f(y)\mathbb{1}_{\{\alpha\in I(x)\}}dy = \int_{m(x)}^{1} f(y)\mathbb{1}_{\{\alpha\in I(x)\}}dy$. Finally, an important question regarding the structure of the equilibrium is whether the Median Voter Theorem holds, i.e., whether the policy path chosen by $I(x)$ is $m(x)$'s optimal choice. Formally:

Definition 4. Given a successor function s and a set $X \subseteq [-1, 1]$, the MVT holds in X if, for each $x \in X$ and all $y \in [-1, 1]$, $m(x)$ weakly prefers $S(s(x))$ to $S(y)$.

Examples

As an illustration, we map the model to two concrete examples.

The first one is Tiebout-style policy competition between cities. Assume that there is a universe of "normal" cities $c \in [-1, 1]$, and a "special" city which we will call c^*. Cities differ in two ways. First, each city has a *policy*, which is a one-dimensional variable $x_t(c) \in [-1, 1]$ that denotes a certain level of taxation and public goods in city c at time t. For example, a higher x means that the city in question has higher local taxes which finance better public schools and amenities. Second, c^* has an intrinsic attribute that makes it a more desirable location compared to normal cities (uniquely good weather, a strong economy, etc.). For simplicity, suppose that each normal city has a large population of immobile voters tied to it, and the median immobile voter in city c has bliss point c, so that $x_t(c) = c$ for all t, c. In addition, there is a unit mass of mobile agents in the model, whose bliss points are distributed according to a density f. $x_0(c^*)$ is given.

[9] If, given a current policy x, an agent deviates from her equilibrium membership decision, the set of members I will differ from $I(x)$ by a set of measure zero, so the policy chosen tomorrow, $s(x)$, will be unchanged.

We are interested in the policy path of c^* over time and the behavior of mobile agents. At each time t, each mobile agent α chooses a city to live in. If she chooses a normal city c, she gets a flow payoff $u_{\alpha t}(c) = -(c-\alpha)^2$; if she lives in c^*, her payoff is $u_{\alpha t}(c^*) = C - (x_t - \alpha)^2$, where x_t is the policy of city c^* at time t and $C > 0$ is the intrinsic value of c^*.

α's decision boils down to a binary choice: she can either live in c^* or move to one of the normal cities, in which case she should always choose $c = \alpha$, yielding a flow payoff of zero. Living anywhere but in c^* is equivalent to leaving the organization in the general model.

The second example we discuss is that of trade unions. Assume an economy with a unionized firm and a larger competitive (non-union) sector. Firms offer employment contracts (w, l) comprising a wage w and a family leave policy l. The marginal productivity of all workers is normalized to 1, and a leave policy l signifies that the worker only works a fraction $1 - l$ of the time. In equilibrium, competitive firms are willing to offer any contract of the form $(1 - l, l)$; we assume the competitive sector is large enough that all such contracts are available. The union, through collective bargaining, extracts a wage $w_u > 1$ from the unionized firm in the union sector, so that its leadership can bargain for any contract of the form $(w_u - l, l)$, but the same contract will apply to all unionized workers. As in Grossman (1983), assume that the union bargains on behalf of its median voter.

Workers differ in their taste for family leave. A worker of type α has utility function $v_\alpha(w, l) = w + \alpha\sqrt{l}$. Workers are allowed to move freely between firms, including to the union sector; upon joining the union sector they automatically become union members.[10]

Of all the competitive firms, a worker α would prefer to join one offering $l = \frac{\alpha^2}{4}$, and obtain utility $v^*_\alpha = 1 + \frac{\alpha^2}{4}$ from it. Let $u_\alpha(l) = w_1 - l + \alpha\sqrt{l} - v^*_\alpha$ be the net utility of joining the union sector when the union has bargained for a contract $(w_u - l, l)$. We can verify that u satisfies A1-5 and hence the model applies without changes.

3 Equilibrium Characterization

In this Section we show some common properties of all MPEs, which in particular pin down the long-run behavior of any equilibrium. We start by solving for the optimal

[10] This is a common arrangement is known as a 'union shop'.

membership strategy, which is simple:

Lemma 1. *In any MPE, $I(x) = [x - d_x, x + e_x]$ is an interval, and $d_x, e_x > 0$ are given by the condition that $u_{x-d_x}(x) = u_{x+e_x}(x) = 0$.*

Since members can enter or leave at any time, it is optimal for α to join whenever the flow payoff of the current policy, $u_\alpha(x)$, is positive, and leave when it is negative; the Lemma then follows from Assumptions A3 and A5. Armed with this observation, we can describe MPEs solely in terms of successor functions.

Before characterizing s in general, it is instructive to consider two extreme cases with simple solutions. First, suppose that $I(x) = I$ is independent of x (for instance, $I(x) \equiv [-1, 1]$, i.e., everyone always prefers to be in the organization). In this case, regardless of the current policy x, the Condorcet-winner policy is the bliss point of the median member of I. Second, suppose that $\delta = 0$, i.e., agents are completely myopic. Given a current policy x and set of members $I(x)$, the Condorcet-winner policy is the bliss point of $m(x)$, and the policy path will be $(x, m(x), m^2(x), \ldots)$ in equilibrium, which converges to a myopically stable policy $m^*(x) = \lim_{k\to\infty} m^k(y)$. In both scenarios, the simplicity of the solution stems from the lack of tension between the current policy and future control: in the former case there is no link between them, while in the latter case they are linked but voters do not care.

As a first step towards solving the general case, we show that equilibrium paths are always monotonic:

Proposition 1. *In any MPE, for any y, $S(y)$ is monotonic: if $s(y) \geq y$ then $s^k(y) \geq s^{k-1}(y)$ for all k, and analogously if $s(y) \leq y$.*

To see why this must be the case, imagine an equilibrium path $(s_0, s_1, \ldots)$ which increases up to s_k ($k > 0$) and decreases afterwards. For this to work, $S(s_k)$ should be a Condorcet winner in $I(s_{k-1})$, and $S(s_{k+1})$ should be a Condorcet winner in $I(s_k)$. In particular, a majority in $I(s_{k-1})$ must prefer $S(s_k)$ over $S(s_{k+1})$ but a majority in $I(s_k)$ must prefer the opposite. However, $S(s_{k+1})$ has a lower average policy than $S(s_k)$, since it skips s_k which is the highest policy in either path, but the group $I(s_k)$ should have preferences more biased to the right than $I(s_{k-1})$, since $s_k > s_{k-1}$.[11]

We can now pin down the general shape and long-run behavior of any MPE:

[11] A similar result is shown in Acemoglu et al. (2015). The proof here is more involved because, owing to the infinite policy space, we have to rule out cases where the policy path doubles back on itself infinitely many times.

Proposition 2. *In any MPE s and for any y:*

(i) If $m(y) = y$ then $s(y) = y$.

(ii) If $m(y) > y$ then $m^*(y) > s(y) \geq y$. Moreover, if the MVT holds in $[y, m^*(y)]$ then $s(y) > y$ and $s^k(y) \xrightarrow[k\to\infty]{} m^*(y)$.

(iii) If $m(y) < y$ then $m^*(y) < s(y) \leq y$. Moreover, if the MVT holds in $[m^*(y), y]$ then $s(y) < y$ and $s^k(y) \xrightarrow[k\to\infty]{} m^*(y)$.

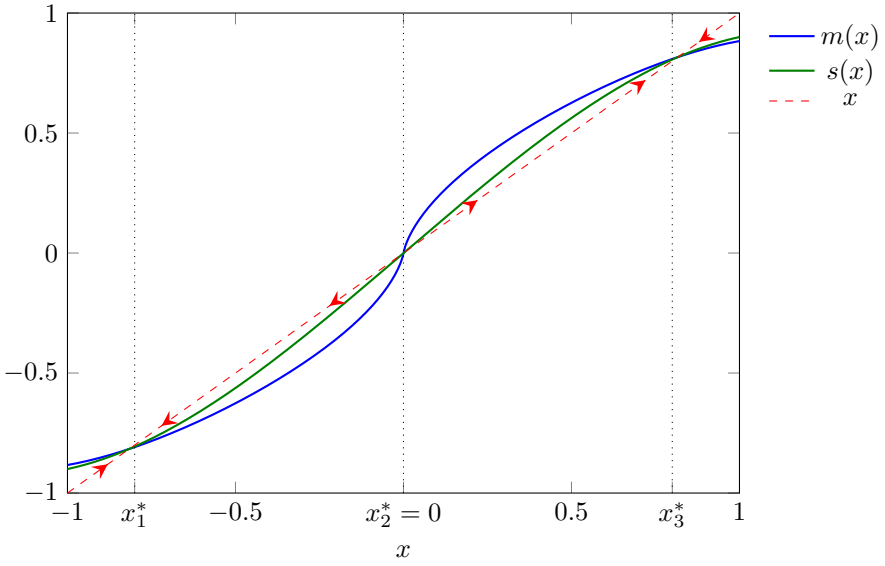


Figure 1: Convergence to steady states in MPE

In other words, when the MVT is satisfied, the steady states of s are simply the fixed points of the mapping $y \mapsto m(y)$. Moreover, stable (unstable) steady states of s are also stable (unstable) fixed points of m, and their basins of attraction coincide.

The intuition for why we should observe $s(x) \leq x$ if $m(x) < x$ and vice versa is straightforward: if $m(x) < x$ in an interval (x^*, x^{**}), any policy in that interval attracts a set of voters whose median wants a lower policy. However, Proposition 2 also implies that slippery slope concerns cannot create myopically unstable steady states—it never happens that $s(x) = x$ despite there being a myopic incentive to change the policy. The logic behind the proof is as follows: suppose $m(x) < x$, but $m(x)$ is afraid of further policy changes if she moves to some $y \in [m(x), x]$. If $m(x)$ chooses a *slightly* better policy $y = x - \epsilon$, her flow payoff tomorrow will be higher by

roughly $\epsilon\left(-\frac{\partial u}{\partial x}\right)$. In exchange, she will relinquish the choice of the continuation path to a slightly different voter, $m(x-\epsilon)$, next period. Due to their preferences being similar, the cost of this loss of control turns out to be of the order of ϵ^2, so such a deviation is always profitable for ϵ small enough.

Figure 1 illustrates the result in an example with three steady states: x_1^* and x_3^* are stable, while x_2^* is unstable. This alternation of stable and unstable steady states occurs in general as long as m is well-behaved. Formally, in the rest of the paper we will impose the following condition:

B1 The equation $m(y) = y$ has finitely many solutions $x_1^* < x_2^* < \ldots < x_N^*$. In addition, $m'(x_i^*) \neq 1$ for all i.[12]

It follows that:

Corollary 1. *The number of steady states is odd. For odd i, $m'(x_i^*) < 1$ and x_i^* is stable; for even i, $m'(x_i^*) > 1$ and x_i^* is unstable.*

Finally, we address the question of whether s is monotonic—a stronger property than path monotonicity—and whether it must satisfy the Median Voter Theorem. It turns out that both properties must hold in a sizable neighborhood of each stable steady state:

Proposition 3. *Let x^* be such that $m(x^*) = x^*$ and $m'(x^*) < 1$; let $x^{***} < x^* < x^{**}$ be the closest unstable steady states to the left and right of x^*; and $J = (x^{***}, x^{**})$. Then, in any MPE, s is weakly increasing and satisfies the MVT in $I(x^*) \cap J$.*

The reason this result may fail outside of $I(x^*) \cap J$ is that it relies on pivotal voters not leaving the organization on the equilibrium path; when pivotal voters quit the club at different times, it is not always possible to cleanly compare policy paths, as the logic that voters with higher bliss points should like higher paths (as per Assumption A2) no longer applies.[13] However, as we will see next, under reasonable conditions there are equilibria that are well-behaved everywhere, in the sense that Proposition 3 holds everywhere.

[12] m is differentiable since f is continuous. For a discussion of this assumption, see Appendix C.

[13] For instance, let $x = 0.6$, $y = 0.5$, and $S = (0.7, 0.1)$, $T = (0.65, 0)$ be two-period policy paths. If x, y never leave under either path, $U_x(S) - U_x(T) > U_y(S) - U_y(T)$ by A2; but if $u_x(0) < 0$ it is possible that $U_x(S) - U_x(T) < U_y(S) - U_y(T)$ instead, as x's distaste for T having a low policy in period 2 is dampened by her intent to leave. By the same logic, the set of voters preferring S over T may be a collection of disjoint intervals, so a winning coalition need not contain the median voter. See Appendix E.4 for an illustration of pathological equilibria stemming from these issues.

4 Transition Dynamics

In this Section we characterize the transition dynamics of the model. Understanding them can be important in practice because, as we will see, convergence to a steady state is far from instant.

We propose a natural class of equilibria, which we call *1-equilibria*, and provide a detailed analysis of them. There are two major reasons to focus on 1-equilibria. The first is tractability: they satisfy the Median Voter Theorem globally, they therefore always feature convergence to a myopically stable policy, and they have other analytically attractive properties. Though there is a continuum of them, they converge to a common limit that can be easily interpreted as policy and membership decisions are allowed to take place more often. The second is robustness: 1-equilibria exist generically, while other types of equilibria (including smooth equilibria, which have been studied in the literature) do not.[14]

Without loss of generality, henceforth we restrict our analysis to the right side of the basin of attraction of a stable steady state. That is, let $x^* < x^{**}$ such that $m(x^*) = x^*$, $m(x^{**}) = x^{**}$ and $m(y) < y$ for all $y \in (x^*, x^{**})$. Then we study s restricted to $[x^*, x^{**}]$.[15]

Formally, this is the class of equilibria we are interested in:

Definition 5. Let s be a successor function on $[x^*, x^{**}]$. s is a *1-function* if there is a sequence $(x_n)_{n\in\mathbb{Z}}$ such that $x_{n+1} < x_n$ for all n, $x_n \xrightarrow[n\to-\infty]{} x^{**}$, $x_n \xrightarrow[n\to\infty]{} x^*$, and $s(x) = x_{n+1}$ if $x \in [x_n, x_{n-1})$. We call $(x_n)_n$ the *recognized sequence* of s.

s is a *1-equilibrium* (henceforth 1E) if it is a 1-function and an MPE.

s is a *quasi-1-equilibrium* (henceforth Q1E) if it is a 1-function and $(1-\delta)U_{m(x_n)}(S(x_{n+1})) = u_{m(x_n)}(x_{n+1})$ for all n.

In a 1E, only policies that are part of the recognized sequence $(x_n)_n$ are chosen in equilibrium. x_n today leads to x_{n+1} tomorrow; if the initial policy x is not part of the recognized sequence, but it is between x_k and x_{k-1}, then x_{k+1} is chosen next, and the path follows along the recognized sequence thereafter. An illustration can be found in Figure 2.[16] The closely related notion of Q1E is useful because Q1Es are

[14] For a discussion of other equilibria, see Appendix D.

[15] For any $y \in (x^*, x^{**})$, $I(y)$ will never want to choose a policy outside of (x^*, x^{**}), so $s|_{(x^*,x^{**})}$ can be studied in isolation. Note also that results are analogous for a basin of attraction of the form $[x^*, 1]$ or to the left of a stable steady state.

[16] 1-equilibria can be found explicitly when m is linear and u is quadratic; see Appendix D.

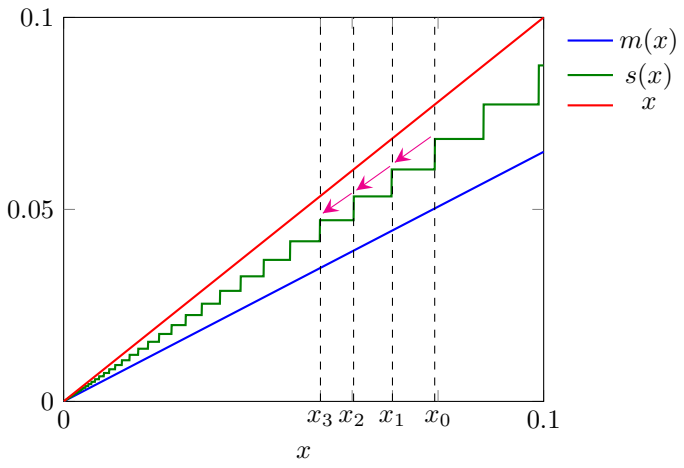


Figure 2: 1-equilibrium for $u_\alpha(x) = C - (\alpha - x)^2$, $m(x) = 0.7x$, $\delta = 0.7$, $C = 1$.

often 1Es and vice versa, but it is easier to check whether a 1-function is a Q1E. The following Proposition states two basic but important properties of 1-equilibria:

Proposition 4. *Let s be a 1E on $[x^*, x^{**}]$. Then s satisfies the Median Voter Theorem for all $x \in [x^*, x^{**}]$. In addition, $(1-\delta)U_{m(x_n)}(S(s(x_n))) = u_{m(x_n)}(x_{n+1})$, i.e., every 1E is a Q1E.*

The intuition behind the second statement is the following. Consider a 1E s given by a recognized sequence $(x_n)_n$, and suppose it satisfies the MVT. Then $S(x_{n+1})$ is $m(x_n)$'s most-preferred policy path from the available choices, but $S(x_{n+2})$ is $m(y)$'s most-preferred path for all $y < x_n$ just under x_n. By continuity, $m(x_n)$ is indifferent between $S(x_{n+1})$ and $S(x_{n+2})$ and prefers these two paths to all others. Equivalently, $m(x_n)$ is indifferent between the static path $(x_{n+1}, x_{n+1}, \ldots)$ and $S(x_{n+1})$. This property relates $m(x_n)$'s equilibrium utility to the distance between x_n and x_{n+1}—the higher $m(x_n)$'s utility, the higher the distance, and hence the faster the policy changes—which is a key driver of later results.

Our next result is that 1-equilibria exist under fairly general conditions:

Proposition 5. *Let $\underline{x} \in (x^*, x^{**})$. Then there is a Q1E $s_{\underline{x}}$ defined on $[x^*, x^{**}]$ such that $x_0 = \underline{x}$.*

Moreover, $s_{\underline{x}}$ is a 1E in $[x^*, m^{-1}(x^* + d)]$ iff $m(x_n) < x_{n+2}$ for all n such that $m(x_n) < x^* + d$.

In other words, Q1Es *always* exist, and there is a continuum of them. The second part of the proposition guarantees that the Q1Es we find are real equilibria, unless

they feature very fast policy changes. A more tractable and powerful condition to determine if Q1Es are equilibria is available if we take the model to continuous time, which we do next.

Continuous Time Limit

A discrete time model is the natural choice for our problem, as in most organizations voting happens periodically (e.g., at monthly meetings, annual elections, etc.). However, the continuous time limit is worth studying, as it is analytically more tractable and offers important insights that apply to the discrete time model.

Assume that decisions are made increasingly often: each period $[t, t+1]$ is broken into j periods of length $\frac{1}{j}$, with discount factor between sub-periods $\delta^{\frac{1}{j}} = e^{-\frac{r}{j}}$, and decisions now happen at times $t, t+\frac{1}{j}, t+\frac{2}{j}, \ldots$ Call this the j-refined game. We are interested in the limit as $j \to \infty$.

A successor function for the continuous time game is a function $s^t(x)$ denoting the successor policy starting from x after a length of time t has passed. We require s to be such that $s^{t+t'}(x) = s^t(s^{t'}(x))$; $s^0(x) = x$; and s is weakly decreasing in t. We say s is *continuous* if $s^t(x) \xrightarrow[t\to 0^+]{} x$ for all x, and we say s is a *continuous solution* if it is continuous; $s^t(x) \xrightarrow[t\to+\infty]{} x^*$; and $U_{m(x)}(S(x)) = u_{m(x)}(x)$ for all $x \in (x^*, x^{**})$. (In general, we define $U_\alpha(S) = \int_0^\infty re^{-rt}u_\alpha(s_t)dt$.)

The following Proposition states the fundamental properties of the continuous solution, if it exists, and relates it to the 1-equilibria of the discrete time game:

Proposition 6. *Suppose that $m \in C^2$ and the continuous time game has a continuous solution s. Then:*

(i) s is the unique continuous solution, and s is C^2 as a function of t;

(ii) given any sequence $(s_j)_j$, where s_j is a Q1E of the j-refined game, $s_j^t(x) \xrightarrow[j\to\infty]{} s^t(x)$ a.e.;

(iii) there is $\overline{\delta} < 1$ such that, for all $\delta > \overline{\delta}$, all Q1Es of the discrete time game with discount factor δ are 1Es.

Figure 3 illustrates a continuous solution for the continuous time game. $s^1(x)$, the policy after a delay of length 1, serves as a direct point of comparison with discrete

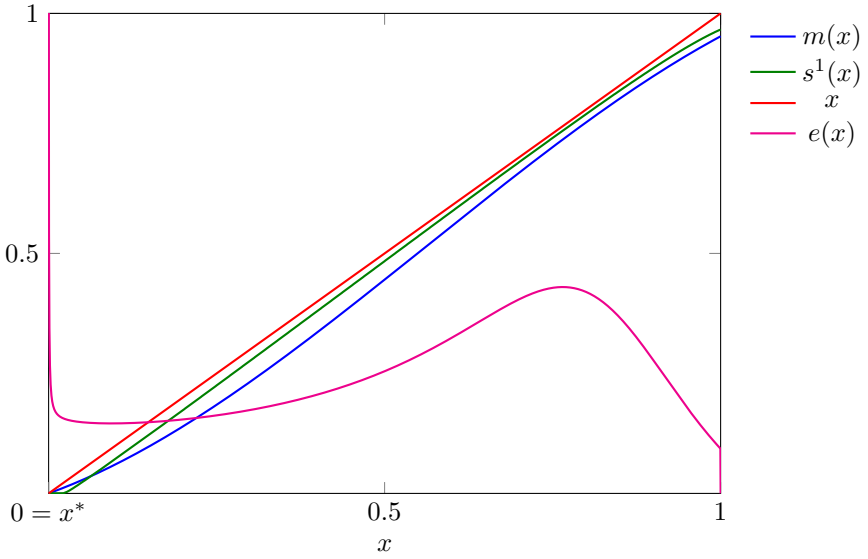


Figure 3: A continuous solution

time equilibria, and $e(x) = -\frac{1}{\frac{\partial s^t(x)}{\partial t}|_{t=0}}$ is the instantaneous delay of the policy path at x. The Figure shows the expected behavior: $e(x)$ is higher—the policy moves more slowly—in regions where $x - m(x)$ is close to zero.

The intuition behind Proposition 6 is as follows. Say we are interested in taking the limit of a sequence of Q1Es as $j \to \infty$, and recall that $U_{m(x_{jn})}(S_j(x_{j(n+1)})) = u_{m(x_{jn})}(x_{j(n+1)})$ for any Q1E s_j. As $j \to \infty$, the discount factor between sub-periods goes to 1, i.e., it is as if agents were more and more patient. This might drive them to choose progressively slower policy paths, i.e., the differences $x_{jn} - x_{j(n+1)}$ might go to zero as $j \to \infty$, so in the limit we should have $U_{m(x)}(S(x)) = u_{m(x)}(x)$ for $x = \lim_{j\to\infty} x_{jn} = \lim_{j\to\infty} x_{j(n+1)}$. Indeed, if this limit is well-behaved, we should have $\frac{x_{jn} - x_{j(n+1)}}{j} \xrightarrow[j\to+\infty]{} \frac{1}{e(x)}$.

It can be shown that $e(x)$ must satisfy the following integral equation:[17]

$$\tilde{e}(x) = \frac{1}{r}\left(-\frac{\frac{\partial^2 u}{\partial x^2}}{\frac{\partial u}{\partial x}} - 2\frac{\frac{\partial^2 u}{\partial \alpha \partial x}}{\frac{\partial u}{\partial x}}m'(x) - \frac{\frac{\partial^2 u}{\partial \alpha^2}}{\frac{\partial u}{\partial x}}m'(x)^2 + \frac{\frac{\partial^2 U_{m(x)}(S(x))}{\partial \alpha^2}}{\frac{\partial u}{\partial x}}m'(x)^2 + \frac{m''(x)}{m'(x)}\right). \quad (*)$$

If we find a candidate solution $\tilde{e}(x)$ that is positive everywhere, a continuous solution

[17] The derivatives of u are all evaluated at $(m(x), x)$. Note that this is an integral equation because $\frac{\partial^2 U_{m(x)}(S(x))}{\partial \alpha^2}$ depends on $S(x)$, which depends on $s^t(x)$ for $t > 0$, i.e, it depends on $e(x')$ for $x' < x$. In the case of quadratic utility, $\frac{\partial^2 U_{m(x)}(S(x))}{\partial \alpha^2} \equiv -2$, so we can obtain a closed-form solution.

can be reconstructed from it, and the claims in Proposition 6 follow. Moreover, Equation (*) gets us as close to an explicit solution as we could hope to get. On the other hand, if $\tilde{e}(x)$ is negative for some x, a continuous solution cannot exist for the continuous time game; in this case, any equilibrium path must involve instantaneous policy jumps that do not vanish as we take the limit to continuous time. To see why this might happen in an example, assume that $u_\alpha(x)$ is quadratic and $m(x) = 0.4x$. Then $m(x)$ prefers policies $y \in [0, 0.8x]$ to policies $z \geq 0.8x$, whence we should expect $\lim_{t\to 0^+} s^t(x) \leq 0.8x < x$.

An important observation on the relationship between patience and the speed of policy change follows from Equation (*). In the continuous time limit, the discount rate acts as a pure rescaling factor: when r is lower, delay becomes proportionally higher so that the effective delay (measured against patience) remains constant. Since discrete time equilibria approach this limit as $\delta \to 1$, it follows that an increase in δ also slows down convergence at a roughly proportional rate (of course, changes in δ do have some "real" effects in discrete time, as they also affect the effective frequency of voting).

Proposition 6 can be extended to the case in which there is no continuous solution. In that case, the continuous time game *still* has a natural solution to which all Q1Es of the discrete time game must converge, as long as m satisfies a genericity condition. This solution is harder to describe succintly, as it involves a series of instantaneous policy jumps separated by temporary stops (intervals where $s^t(x)$ is constant in t). For the details, see Appendix B.

5 A Model of Political Power

We now discuss an important variant of the main model which overturns the assumption of free entry and exit. Consider a polity governed by a *ruling coalition*. At each time $t = 1, 2, \ldots$ the ruling coalition chooses a policy x_t; the policy at time $t = 0$ is given exogenously.

The policy now plays a dual role: it determines the set of agents granted political power at time $t+1$, as well as the flow payoffs of all agents during the period $[t, t+1)$.[18]

[18]For example, x_t determines the set of enfranchised voters, and in each period the median voter chooses an economic policy y_t which affects flow payoffs. Then $u_\alpha(y_t) = u_\alpha(y(I(x_t))$ is indirectly a function of x_t.

In other words, x determines $I(x)$ directly; the mapping $x \mapsto I(x)$ is now taken as a primitive of the model, instead of resulting from equilibrium behavior. Assume for simplicity that $I(x)$ is still an interval of the form $(x - d_x, x + e_x)$. In this model, *all* agents are impacted by the policy, regardless of whether they have political power or not. In other words,

$$U_\alpha\left((x_t)_t, I_\alpha\right) = \sum_{t=0}^{\infty} \delta^t u_\alpha(x_t),$$

and we assume that u_α satisfies A1, A2 and A4. This setting, which is similar to those discussed in Jack and Lagunoff (2006), Bai and Lagunoff (2011) and Acemoglu et al. (2015), is clearly another one where slippery slope concerns apply: a ruling coalition may want to expand the franchise (for example, to lower unrest) but fear that the new voters will choose to expand it even further.

Define an MPE given by a successor function s as before. Then:

Proposition 7. *Propositions 1, 2, 4, 5, 6 and 10 hold in this model. Moreover, any MPE s satisfies Proposition 3 for all $y \in [-1, 1]$, and any Q1E is a 1E globally as long as $m(x_n) < x_{n+2}$ for all n.*

Additionally, let x^* be a stable steady state, and let $I(x^*) \cap J$ be as in Proposition 3. Modify the free entry model and the model of political power by restricting the policy space to $I(x^*) \cap J$ in both cases. Then the sets of equilibria of both models are identical.

In other words, all the main results of the paper hold for this variant of the model. Indeed, some results, namely Propositions 2, 3 and 5, are strengthened, because the MVT is now guaranteed to hold everywhere for all equilibria.[19]

It may appear surprising that these two models have such similar equilibria, despite representing different organizations with different causal relationships between political power, membership, and flow payoffs. However, there is a mechanical equivalence between the components of both models. Indeed, in both cases, members care about two variables, namely, the identity of the future median voter and the current policy, but they only have one 'lever' to pull; hence their choice of tomorrow's policy must optimally trade-off control over the future policy for present flow payoffs.

[19] In the main model, the MVT may fail away from stable steady states because agents' payoffs are harder to compare when they leave the club at different times; this is not a problem here because losing political power does not affect agents' flow payoffs.

Other variants, allowing the model to fit new examples, are possible. For instance, the set of members, I_t, can affect affect payoffs directly: $v_\alpha(x) = u_\alpha(x) + w_\alpha(I(x))$. So long as $v_\alpha(x)$ satisfies A1-4, Proposition 7 still applies. A natural example is immigration: if x_t is a country's immigration policy and I_t is its current set of citizens, x_t does not affect the payoffs of current citizens directly, but the entry of immigrants does, as it has economic and social effects.[20]

6 Discussion

In this Section we highlight some important takeaways from the model.

Myopic Stability of Steady States

We have shown that, although the policy path depends on the selected equilibrium and the agents' discount factor, its limit as $t \to \infty$ depends on neither—in fact, the limit is the same as it would be if $\delta = 0$. In the language of Roberts (2015) all steady states are "extrinsic", i.e., pinned down by the preference distribution. This contrasts with other papers in this literature such as Roberts (2015) and Acemoglu et al. (2008, 2012, 2015), where "intrinsic" steady states exist: that is, policies considered suboptimal by current voters can be sustained permanently for fear that, if a line is crossed, future agents will move towards a different policy too quickly.

Their results hinge on two important assumptions: a discrete policy space and patient agents. In effect, a finite policy space forces agents to choose between moving too fast or not at all, whereas the continuous policy space in our model offers them the option to move slowly. If we simultaneously take δ to 1 and make the (finite) policy space increasingly fine, whether we get intrinsic steady states or not depends on the order of limits.

The upshot is that, in a practical setting, whether slippery slope concerns can stall policy change will depend not just on agents' foresight but also on institutional details, namely, whether incremental changes are possible. For example, take a polity with a limited franchise considering a franchise extension on the basis of income.

[20]This example has been studied in the literature, often in overlapping-generations models: for example, in Ortega (2005) the entry of immigrants affects wages due to labor market complementarities, while in Suwankiri, Razin and Sadka (2016) it affects the net transfers from the welfare system.

Suppose high-income voters prefer a limited franchise, but larger than the smallest one they would be in (a voter at the 90th income percentile wants to extend it to the top 15%, a voter at the 80th percentile wants to extend it to the top 25%, etc.). Then, if it is possible to grant voting rights to the top $x\%$ of the income distribution for any x, slippery slope concerns would not prevent full democracy from being eventually reached through a series of small changes. However, if voting rights can only be extended based on a few criteria (e.g., only to men who can read, to landowners, to taxpayers, etc.), indefinite stalling is more likely.

Distribution of Steady States

A natural question that the model answers concerns the impact of policy drift on extremism: does the tendency to converge to steady states lead to moderate policies in the long run? Or can small factions "capture" the organization indefinitely?

In the quadratic case, or more generally whenever $I(x) = (x-d, x+d)$ is symmetric around x, the distribution of steady states reflects the following intuition: if f is increasing in $I(x)$ then $m(x) > x$ so the policy will drift upward, and vice versa. Hence, stable steady states correspond roughly to maxima of the density function:

Proposition 8. *If $I(x) = (x-d, x+d)$ for all x, and x^* is a stable (unstable) steady state, then $I(x^*)$ contains a local maximum (minimum) of f.*

In particular, if f is increasing (decreasing) everywhere, there is a unique steady state close to 1 (-1); if f is symmetric and single-peaked, 0 is the unique steady state. This suggests that policy drift encourages moderate policies: the organization always moves to the center if the distribution of preferences is bell-shaped. Yet there are three reasons why extremism may be sustainable, in the sense that the organization may converge to a policy more extreme than the bliss points of most voters.

First, even if most voters are moderates, a local maximum near the extreme may support a stable steady state.[21] This is especially likely if d is low, i.e., if the organization attracts a narrow niche, so that an extreme policy would bring in the nearby extremists (who are locally strong) and not be disrupted by a large mass of moderates. Formally:

[21] When there are multiple steady states, whether a minority manages to capture the organization depends on the initial policy, as the solution exhibits path-dependence.

Proposition 9. *If $f'(x^*) = 0$ and $f''(x^*) < 0$, then for all $d > 0$ small enough, $(x - d, x + d)$ contains a stable fixed point of $m((x - d, x + d))$.*

Second, even when the distribution has a single steady state, its location may be unstable when f is close to uniform. For example, consider the densities $f_1(x) = \frac{1}{2} + \epsilon x$, $f_2(x) = \frac{1}{2} - \epsilon x$ and $f_3(x) = \frac{1+\epsilon}{2} - \epsilon|x|$, for $\epsilon > 0$ small. These are all close to each other ($||f_i - f_j||_\infty \leq 2\epsilon$ for all i, j) but f_1 has a unique steady state near -1, f_2 has one near 1, and f_3 has one at 0. Hence, the long-run policy is potentially much more sensitive to demographic changes than in models of voting with a fixed population.

Third, the tendency towards moderate policies can be easily overturned when the voter sets $I(x)$ are asymmetric. That is, if agents with extreme preferences are disproportionately more willing to join the organization, they can often capture it despite being a minority, even locally.

For an example where such asymmetric preferences are natural, suppose that the organization is a nationalist club. There are two types of agents, moderates and extremists. Moderate agents want to engage in benign activities, such as enjoying traditional meals and music, publishing a local newspaper for their community, etc., while hard-liners want to organize attacks against immigrants. Hard-liners would still join the club even if it was too moderate for their tastes, whereas moderates would leave the club if turned xenophobic.

Formally, let the policy space be $[-1, 1]$, where -1 is the most moderate agent and 1 the most radical, and assume $u_\alpha(x) = -|\alpha - x| + (1 + \alpha)$.[22] Then α wants to be a member whenever $x \in [-1, 2\alpha + 1]$, whence $I(x) = [\frac{-1+x}{2}, 1]$.

Suppose f is as follows: moderates constitute 60% of the population and have bliss points uniformly distributed in [-1,-0.9]; radicals, the remaining 40%, have bliss points uniformly distributed in [-0.9,1]. It can then be shown that the *unique* stable steady state is $x^* = \frac{1}{3} > 0$. At the steady state policy, the set of members $I(\frac{1}{3}) = [-\frac{1}{3}, 1]$ is only 28% of the population, all of them radicals.

An important corollary of this example is that social or legal discouragement meant to dissuade extremist behavior may backfire when organizations are involved. Consider a variant where the moderates are initially 90% of the population, and the organization is initially at a stable steady state x^* in $[-1, -0.9]$. Suppose that a

[22] The example is degenerate in that $\frac{\partial^2 u}{\partial \alpha \partial x}$ is only weakly positive and u is only weakly concave in x; this is needed only for simplicity.

media backlash against nationalist organizations results in a reputation cost of 0.1 per period being imposed on all members. This prices out all moderates from the organization no matter the policy chosen, and only one steady state is left: $x^{*'} = \frac{1.1}{3}$. Thus, while the backlash does succeed in shrinking the organization, it also radicalizes it; the social cost of its activities may well increase.

Welfare Implications

A prominent debate in the Tiebout literature concerns whether competition between districts will lead to a socially efficient outcome. Tiebout (1956) conjectured an affirmative answer in a setting with many districts, where agents can all sort into districts offering their respective most-preferred policies. This idea has been formalized (see, e.g., Wooders (1989)) as well as criticized (Bewley (1981) constructs several Tiebout models with inefficient equilibria, largely based on coordination failures which are not possible in our model).

Since in our model there is only one district (or, as discussed in Section 2, a single district with a competitive advantage over the rest), the Tiebout argument cannot apply, and it is not surprising that the solution is typically inefficient. However, it is worth pointing out that two distinct kinds of inefficiency can arise.

The first concerns the long-run behavior of the organization. In the long run, the policy converges to a steady state, which is necessarily Pareto optimal, as every policy is some agent's bliss point. However, if we define welfare as aggregate utility, i.e., $W(S) = \int_{-1}^{1} U_\alpha(S) d\alpha$, then most or even all steady states may be inefficient, asthe welfare-maximizing policy need not be a steady state.

The second, more interesting inefficiency arises on the transition path. Transition paths do not just fail to maximize aggregate utility—they may be Pareto dominated. For a clean illustration, consider the model of political power from Section 5 with utility function $u_\alpha(x) = -(\alpha - x)^2$. In the quadratic case, the payoff from a policy path may be decomposed in terms of its mean and variance, i.e., $U_\alpha(S) = \frac{1}{1-\delta}\left[-(\alpha - E(S))^2 - V(S)\right]$, where $E(S) = (1-\delta)\sum_{t\geq 0}\delta^t s_t$, $V(S) = (1-\delta)\sum_{t\geq 0}\delta^t (s_t - E(S))^2$. It follows that any non-constant equilibrium path $S(x) = (x, s(x), s^2(x), \ldots)$ is inefficient, as all agents would strictly prefer the constant path $(E(S(x)), E(S(x)), \ldots)$, but the latter cannot be attained in equilibrium.

7 Conclusion

We conclude with a brief discussion of some important issues that the model leaves out, and possible extensions that lie beyond the scope of this paper.

The first point concerns multiple organizations. Our model focuses on the behavior of a single organization; it can serve as a model of a setting with many organizations only under rather restrictive assumptions (see the examples from Section 2). A full model of dynamic Tiebout competition would have to solve the problem of multiple organizations choosing policies simultaneously, with an eye not just on internal membership changes but also on their impact on adjacent organizations with which they compete for members. Such a model might also allow for the endogenous creation of new organizations. Our model suggests some tentative insights: an agent able to found an organization would be less likely to do so if her bliss point was far from any steady state, as she would expect faster policy drift away from her initial policy; or, if she did create an organization, she would have stronger incentives to make it non-democratic.

The second point concerns hierarchies and structure within organizations. The organizations in this paper are quite simple: all agents are treated equally, they have the same voting power regardless of seniority or individual characteristics, and they cast their votes independently. In some examples, such as cities holding local elections, these are reasonable assumptions. In others, such as trade unions or various types of elite clubs, it is likely that some agents hold more clout than others. (See Appendix E.3 for a partial extension in this direction.) In the same vein, agents may try to collectively coordinate their behavior. Our model allows agents to form coalitions in the mechanical sense that a group of voters supporting a winning policy can be considered a coalition, but it does not model the notion that agents might group up based on some criteria (e.g., demographic characteristics) and then vote as a bloc. It also does not allow for groups of agents to coordinate a hostile takeover by joining simultaneously, or to make demands and threaten to leave en masse if they are not met. Furthermore, history-dependent or reputation-based strategies are ruled out by the assumption of Markov equilibria (see Appendix E.1 for a discussion of non-Markovian equilibria.)

The third point concerns partial membership restrictions. In the main model we study an organization with free entry and exit, while Section 5 discusses the opposite

case, where the organization can choose the set of future members at will. In addition, Appendix E.2 sketches an extension that allows for positive entry and exit costs. While it is encouraging that the results are similar in all cases, real organizations often wield various kinds of incomplete control over the entry of members, which we do not model fully. For instance, the organization may charge fees to drive some members out, or it may impose various requirements on new members.[23] It is not clear what the net effect of such policies would be: fewer new members translates into slower policy changes, but this, in turn, makes agents less fearful of going down the slippery slope, which can *accelerate* policy changes.

A related point is that, for the results to apply at all, there must be a tension between current payoffs and future control. This is the case when entry and exit are at least partly free, or when the organization has full control over the set of members but has preferences over who its members are, as in the case of immigration. However, there are scenarios in which organizations can shut down these tensions by means of institutional changes: a supporter-owned sports club may switch to private ownership; a country where immigration is poised to reshape the electorate could fashion a policy whereby immigrants are allowed to live and work in the country, but are granted voting rights only after a long wait; or an important policy might be permanently enshrined in a special document (e.g., a constitution). If there is no tension, the pivotal decisionmaker can have her optimal policy *and* never lose control, and there is no drift over time.

The last point concerns the complexity of real political processes. Our reduced-form assumption about the collective decisionmaking process is the simplest possible: that chosen policies should be Condorcet winners. In some ways, this is a strength—the results apply whenever the political process would lead to Condorcet winner policies being chosen, regardless of the institutional details—and it can be micro-founded as the outcome of Downsian competition with two candidates, as shown in Appendix E.5. However, outcomes may differ in settings where the political process is more complicated. For instance, if the organization has more than two candidates running for office, or if it is run by a deliberative decision-making process, then the Condorcet winner policy may not win.[24] Another possibility is for the political pro-

[23]For example, Barcelona FC—a supporter-owned club with about 150,000 *socios*—requires prospective members over 15 to have a close relative who is already a member.

[24]For example, Bouton and Gratton (2015) shows that Condorcet winners may lose in majority runoff elections with three candidates.

cess to be biased towards inaction, for instance due to a supermajority requirement; we show in Appendix E.3 that this dampens policy drift but does not qualitatively change the results. Lastly, leaders have more agency in practice than they are given in this model. If candidates have policy preferences and differ in quality, a high-quality candidate championing a certain policy may affect the policy path, possibly permanently. Moreover, a politician may strategically push for policies that will attract a set of members predisposed to like her. A famous example, discussed in Glaeser and Shleifer (2005), is that of Mayor Curley of Boston, who used wasteful policies in an effort drive out rich citizens of English descent, as he was most popular among the poor Irish population.

A Proofs

Lemma 2. *Let $S = (s_0, s_1, \ldots)$, $T = (t_0, t_1, \ldots)$ be two policy paths, and let $I(S) = \cup_{n=0}^{\infty} I(s_n)$, $I(T) = \cup_{n=0}^{\infty} I(t_n)$. Suppose that $\sup s_n < \inf t_n$; $I(S)$, $I(T)$ are intervals; and $I(S) \cap I(T) \neq \emptyset$. Then there is α_0 such that agents in $[-1, \alpha_0)$ strictly prefer S to T, and agents in $(\alpha_0, 1]$ strictly prefer T to S.*

Proof. Let $\underline{s} = \inf s_n$, $\overline{s} = \sup s_n$, $\underline{t} = \inf t_n$, $\overline{t} = \sup t_n$. By assumption, $\underline{s} \leq \overline{s} < \underline{t} \leq \overline{t}$. Note that all agents $x < \underline{s}$ strictly prefer S to T by A5 and A6; likewise, all $x > \overline{t}$ strictly prefer T to S.

Let $W(\alpha) = U_\alpha(T) - U_\alpha(S)$. Note that W is continuous and $W(\underline{s}) < 0 < W(\overline{t})$. Hence there is some $\alpha_0 \in [\underline{s}, \overline{t}]$ for which $W(\alpha_0) = 0$.

For any $z \in [\underline{s}, \overline{t}]$, let $I_S^-(z) = \sum_{t=0}^{\infty} \delta^t \mathbb{1}_{s_t \leq z, u_z(s_t) > 0}$, $I_S^+(z) = \sum_{t=0}^{\infty} \delta^t \mathbb{1}_{s_t > z, u_z(s_t) > 0}$ and define $I_T^-(z)$, $I_T^+(z)$ analogously.

If $z \in [\overline{s}, \underline{t}]$, then $I_S^-(z) \geq I_T^-(z) = 0 = I_S^+(z) \leq I_T^+(z)$, and $I_S^-(z) + I_T^+(z) > 0$ by the assumption that $I(S) \cap I(T) \neq \emptyset$. Then $W'(z) > 0$ by A5.

If $z \in [\underline{t}, \overline{t}]$, then $I_S^+(z) = 0$, and one of the following must be true. If $I_S^-(z) \leq I_T^-(z)$, then $W(z) > 0$ by A4, and moreover $W(z') > 0$ for all $z' \in [z, \overline{t}]$. If $I_S^-(z) > I_T^-(z)$, then $W'(z) > 0$ by A2 and A5. Similarly, for each $z \in [\underline{s}, \overline{s}]$, either $W'(z) > 0$ or $W(z') < 0$ for all $z' \in [\underline{s}, z]$.

In general, then, we can find thresholds z_0, z_1 such that $\underline{s} \leq z_0 \leq \overline{s} < \underline{t} \leq z_1 \leq \overline{s}$; $W'(z) > 0$ for all $z \in [z_0, z_1]$; $W(z) < 0$ for all $z \in [\underline{s}, z_0)$; and $W(z) > 0$ for all $z \in (z_1, \overline{t}]$. Hence W can vanish at most at one point, so α_0 is unique.

□

Lemma 3. *In any MPE, for any y, $I(y) \cap I(s(y))$ has positive measure. Hence $I(S(y))$ is an interval for all y.*

Proof. Suppose WLOG $y < s(y)$. We argue that $y + e_y > s(y) - d_{s(y)}$. If not, then it would mean that all voters in $I(y)$ get utility 0 from policy $s(y)$. Note that, if $E = \{\alpha \in I(y) : U_\alpha(S(s(y))) > 0\}$ is a *strict* majority of $I(y)$, this leads to a contradiction, as all agents in E would strictly prefer $S(s^2(y))$ to $S(s(y))$. Let $D = \{\alpha \in I(y) : U_\alpha(S(s(y))) \geq U_\alpha(S(y))\} \subseteq E$. Since $S(s(y))$ is a Condorcet winner in $I(y)$, D is a majority in $I(y)$. If $(y - d_y, y + e_y) \subseteq D$ we are done. If not, $\exists \alpha_0 \in D - (y - d_y, y + e_y)$. By continuity of $U_\alpha(S(s(y)))$, $\exists \alpha_1$ such that $0 < U_{\alpha_1}(S(s(y))) < U_{\alpha_1}(S(y))$, and this inequality holds in some neighborhood $(\alpha_1 - \varepsilon, \alpha_1 + \varepsilon)$. Hence $E - D$ has positive measure and E is a strict majority of $I(y)$. □

Corollary 2. *In any MPE, let $S = S(y)$ for some $y < x$ and $T = (x, x, \ldots)$, with $\sup(S) \leq x$. Then there is $\alpha_0 \leq x$ such that agents in $[-1, \alpha_0)$ strictly prefer S to T, and agents in $(\alpha_0, 1]$ strictly prefer T to S.*

Proof. If $I(S) \cap I(x) \neq \emptyset$, this follows directly from Lemma 2 and Lemma 3. If not, it is clear that all $y \geq x - d_x$ strictly prefer T to S and all $y \leq \overline{s}$ strictly prefer S to T. Let α_0' be such that $u_{\alpha_0'}(y) = u_{\alpha_0'}(x) < 0$. If $\alpha_0' \in (\overline{s}, x - d_x)$ then take $\alpha_0 = \alpha_0'$. If $\alpha_0' \leq \overline{s}$ then take $\alpha_0 = \overline{s}$. If $\alpha_0' \geq x - d_x$ take $\alpha_0 = x - d_x$. □

Proof of Proposition 1. Suppose that some $S(y)$ is not monotonic. Denote $s_k = s^k(y)$, $S_k = S(s_k)$ and $I_k = I(s_k)$. Let $\underline{y} = \inf(S(y))$ and $\overline{y} = \sup(S(y))$. We consider two cases:

Case 1: $S(y)$ attains $\underline{y}$ or $\overline{y}$, i.e., $\exists k \in \mathbb{N}$ such that $s_k = \overline{y}$ or $s_k = \underline{y}$. Suppose WLOG the former. Then there is a $k \in \mathbb{N}$ such that $s_{k-1} < \overline{y}$, $s_k = \overline{y}$ and $s_{k+1} < \overline{y}$.[25]

Consider the decision made by voters in I_{k-1} and in I_k. Since S_k is the Condorcet winner in I_{k-1}, at least half of the voters in I_{k-1} must prefer it to S_{k+1}. At the same time, S_{k+1} is Condorcet-winning in I_k, so at least half of the voters there prefer it to S_k. Let $A = (s_{k-1} - d_{k-1}, s_k - d_k)$, $B = (s_k - d_k, s_{k-1} + e_{k-1})$, $C = (s_{k-1} + e_{k-1}, s_k + d_k)$.[26] Note that α prefers S_k to S_{k+1} iff he prefers $(s_k, s_k, \ldots)$ to S_{k+1}. Apply Corollary 2. If $\alpha_0 \in C$, all voters in $A \cup B$ strictly prefer S_{k+1} to S_k, a contradiction. If $\alpha_0 \in B$, all voters in A strictly prefer S_{k+1} to S_k and all voters in C strictly prefer S_k to S_{k+1}, a contradiction.

Case 2: $S(y)$ never attains its infimum nor its supremum. Then there must be a subsequence s_{k_i} (with increasing k_i) such that $s_{k_i} \underset{i\to\infty}{\longrightarrow} \overline{y}$. Construct a sub-subsequence $s_{k_{i_j}}$ such that $s_{k_{i_j}} \underset{j\to\infty}{\longrightarrow} \overline{y}$ and $s_{k_{i_j}-1} \underset{j\to\infty}{\longrightarrow} s_*^{-1}$ for some limit s_*^{-1}, which may or may not be $\overline{y}$. (We can do this because all the s_k are in $[-1, 1]$, which is compact.) Iterating this, construct a nested list of subsequences $((s_{k_{im}})_i)_m$ such that k_{im} is increasing in i for each m; $K_m = \{k_{im} : i \geq 0\} \supseteq K_{m'}$ for $m < m'$; and, for each m, $s_{k_{im}+r} \underset{i\to\infty}{\longrightarrow} s_*^r$ for any $r \in \{-m, \ldots, m\}$, where $s_*^0 = \overline{y}$. Finally let $g_i = k_{ii}$. Then $(s_{g_i})_i$ is a subsequence of $(s_k)_k$ such that $s_{g_i+r} \underset{i\to\infty}{\longrightarrow} s_*^r$ for any $r \in \mathbb{Z}$. We now consider four sub-cases.

Case 2.1: Suppose that $s_*^r < \overline{y}$ for some $r < 0$ and for some $r' > 0$, and let $\underline{r} < 0 < \overline{r}$ be the numbers closest to 0 satisfying these two conditions. Consider the

[25] If we relax the definition of s, there could be paths where $s^k(y) = \ldots = s^{k+m}(y) > s^{k-1}(y), s^{k+m+1}(y)$, but a similar argument would work in this case.

[26] $s_{k-1} + e_{k-1} > s_k - d_k$ by Lemma 3.

decision made by $I_{g_i+\underline{r}}$ vs. the decision made by $I_{g_i+\overline{r}-1}$, for high i. In the limit, they imply that a (weak) majority in $I(s_*^{\underline{r}})$ (weakly) prefers $\overline{y}$ to $\tilde{S}(s_*^{\overline{r}})$, while a majority in $I(\overline{y})$ prefers $\tilde{S}(s_*^{\overline{r}})$ to $\overline{y}$ (here $\tilde{S}(s_*^{\overline{r}}) = (s_*^{\overline{r}}, s_*^{\overline{r}+1}, \ldots)$). This leads to a contradiction by Corollary 2.

Case 2.2: Suppose that $s_*^r < \overline{y}$ for some $r < 0$ but never for $r > 0$. Let $\underline{r}$ be the number closest to 0 satisfying this, so $s_*^r = \overline{y}\ \forall r \geq \underline{r}$. Fix $0 < \nu < \overline{y} - s_*^{\underline{r}}$. For each i, let $r'(i)$ be such that $s_{g_i+r'(i)}$ is the first element of the sequence after $s_{g_i+\underline{r}}$ that is weakly smaller than $s_*^{\underline{r}} + \nu$. Construct a subsequence $(s_{g_{i_j}})_j$ such that $s_{g_{i_j}+r'(i_j)+l} \to s_{**}^l$ for $l \geq -1$ (in particular $s_{**}^{-1} \geq s_*^{\underline{r}} + \nu \geq s_{**}^0$). Now compare the decisions made by $I(s_{g_i+\underline{r}})$ and $I(s_{g_i+r'(i)-1})$. In the limit, they imply that a weak majority in $I(s_*^{\underline{r}})$ prefers $\overline{y}$ to $\tilde{S}(s_{**}^0)$, while a weak majority in $I(s_{**}^{-1})$ prefers the opposite (here $\tilde{S}(s_{**}^0) = (s_{**}^0, s_{**}^1, \ldots)$). This contradicts Corollary 2.

Case 2.3: Suppose that $s_*^r < \overline{y}$ for some $r > 0$ but never for $r < 0$, and let $\overline{r}$ be the smallest r satisfying this. Let $0 < \nu < \overline{y} - s_*^{\overline{r}}$. Let $r'_\nu(i)$ be such that $s_{g_i+r'_\nu(i)}$ is the last element of the sequence before s_{g_i} that is weakly smaller than $\overline{y} - \nu$. Clearly $r'_\nu(i) \underset{i\to\infty}{\longrightarrow} -\infty$.

Consider the choice made by $I(s_{g_i+\overline{r}-1})$. In the limit, a majority in $I(\overline{y})$ prefers a path $(s_*^{\overline{r}}, s_*^{\overline{r}+1}, \ldots)$—a path to the left of $\overline{y}$—over $\overline{y}$. Apply Corollary 2. Clearly $\alpha_0 < \overline{y}$, so $m(\overline{y}) \leq \alpha_0 < \overline{y}$. As m is strictly increasing, $m^k(\overline{y})$ is strictly decreasing in k and converges to a limit $\tilde{m}$; moreover, $m(y) < y$ for all $y \in (\tilde{m}, \overline{y}]$.

Call $g_i + r'_\nu(i) = h_{i\nu}$ and let $s_*^\nu = \liminf_{i\to\infty} s_{h_i\nu}$. Let $s_{**} = \liminf_{\nu\to 0} s_*^\nu$. If $s_{**} < \overline{y}$, take a sequence of $\nu, h_{i\nu}$ such that $s_{h_{i\nu}} \to s_{**}$. By construction $s_{h_{i\nu}+l} \geq \overline{y} - \nu$ for $l = 1, \ldots, L$ for L arbitrarily large as $\nu \to 0$, $h_{i\nu} \to \infty$. Then, in the limit, $\overline{y}$ is a Condorcet winner in $I(s_{**})$; in particular, a majority prefers $\overline{y}$ to $(s_*^{\overline{r}}, s_*^{\overline{r}+1}, \ldots)$, which contradicts Corollary 2.

If $s_{**} = \overline{y}$ we must work away from the limit. Take $\varepsilon > 0$ such that $(\overline{y} - d_{\overline{y}} + \varepsilon, \overline{y} - \varepsilon)$ is a subset of $I(\overline{y} - \upsilon)$ constituting a strict majority of $I(\overline{y} - \upsilon)$ for all $0 < \upsilon \leq \varepsilon$.[27] Take a fixed $\nu' < \varepsilon$; a $\nu < \nu'$ such that $s_*^\nu \geq \overline{y} - \nu'$; and a subsequence s_{f_i} of $s_{h_{i\nu}}$ such that $s_{f_i} \to s_*^\nu$. Let M_i be the largest integer such that $s_{f_i+l} \in (\overline{y} - \nu, \overline{y})$ for $l = 1, \ldots, M_i$ and K_i the set of $l \in 1, \ldots, M_i$ such that $s_{f_i+l} \in (\overline{y} - \frac{\nu}{2}, \overline{y})$. Let

[27] That such ε exists follows from $m(\overline{y}) < \overline{y}$ and the fact that $y - d_y$, $y + e_y$ are continuous in y by A1, A4, A5 and the implicit function theorem.

$k_i = \min(K_i)$. By construction, $M_i, |K_i| \to \infty$. Then, for $\alpha \in (\overline{y} - d_{\overline{y}} + \varepsilon, \overline{y} - \varepsilon)$,

$$\frac{1}{1-\delta} u_\alpha(s_{f_i}) - U_\alpha(S(s_{f_i+1})) = \sum_{t \in K_i} \delta^{t-1} \left(u_\alpha(s_{f_i}) - \mathbb{1}_{\alpha \in I(s_{f_i+t})} u_\alpha(s_{f_i+t}) \right) +$$
$$+ \sum_{M_i \geq t \notin K_i} \delta^{t-1} \left(u_\alpha(s_{f_i}) - \mathbb{1}_{\alpha \in I(s_{f_i+t})} u_\alpha(s_{f_i+t}) \right) + \sum_{t > M_i} \delta^{t-1} \left(u_\alpha(s_{f_i}) - \mathbb{1}_{\alpha \in I(s_{f_i+t})} u_\alpha(s_{f_i+t}) \right)$$
$$\geq \delta^{k_i - 1} \left(u_\alpha(s_{f_i}) - \mathbb{1}_{\alpha \in I(s_{f_i+k_i})} u_\alpha(s_{f_i+k_i}) \right) + 0 - \frac{\delta^{M_i}}{1-\delta} C =$$
$$= \delta^{k_i - 1} \left(\left(u_\alpha(s_{f_i}) - \mathbb{1}_{\alpha \in I(s_{f_i+k_i})} u_\alpha(s_{f_i+k_i}) \right) - \frac{\delta^{|K_i|}}{1-\delta} C \right)$$

where $C = \max_\alpha u_\alpha(\alpha)$. Note that $u_\alpha(s_{f_i}) - u_\alpha(s_{f_i+k_i}) \geq u_{s_{f_i}}(s_{f_i}) - u_{s_{f_i}}(s_{f_i+k_i}) + M'(s_{f_i} - \alpha)(s_{f_i+k_i} - s_{f_i}) \geq M'(s_{f_i} - \overline{y} + \varepsilon)\frac{\nu}{2}$, which converges to $M'(s_*^\nu - \overline{y} + \varepsilon)\frac{\nu}{2} \geq M'\frac{\nu}{2}(\varepsilon - \nu') > 0$ as $i \to \infty$. On the other hand, $u_\alpha(s_{f_i}) \geq u_{\overline{y} - d_{\overline{y}} + \varepsilon}(s_{f_i})$ which converges to $u_{\overline{y} - d_{\overline{y}} + \varepsilon}(s_*^\nu) \geq u_{\overline{y} - d_{\overline{y}} + \varepsilon}(\overline{y} - \nu) > 0$ in the limit. On the other hand, $\delta^{|K_i|} \to 0$ as $i \to \infty$. Hence all $\alpha \in (\overline{y} - d_{\overline{y}} + \varepsilon, \overline{y} - \varepsilon)$ prefer s_{f_i} to $S(s_{f_i+1})$ for high i, so $S(s_{f_i+1})$ is not a Condorcet winner in $I(s_{f_i})$, a contradiction.

Case 2.4: $s_*^r = \overline{y}$ for all r. In other words, the sequence spends arbitrarily long times near $\underline{y}$ and $\overline{y}$ (if not true for both boundaries, one of the former cases applies). We first prove the following claim: $m(y) = y$ for all $y \in [\underline{y}, \overline{y}]$.

Take any $y_0 \in (\underline{y}, \overline{y})$. Take a sequence $(h_i)_i$ such that, for each i, s_{h_i} is the last element of the sequence $(s_k)_k$ before s_{g_i} such that $s_k \leq y_0$. Intuitively, s_{h_i} is the last element of the sequence below y_0 before the sequence goes near $\overline{y}$ for a long time. Take a diagonal subsequence (s_{k_i}) of (s_{h_i}) such that s_{k_i+l} has a limit s_{**}^l for all i. Clearly $s_{**}^0 \leq y_0$ and $s_{**}^l \geq y_0$ for all $l > 0$.

Consider the choice made by $I(s_{k_i})$. If $s_*^0 < y_0$, in the limit, a majority in $I(s_*^0)$ prefers $(s_*^1, s_*^2, \ldots)$ over s_*^0. Apply Corollary 2. Clearly $m(s_*^0) \geq \alpha_0 > s_*^0$; moreover, $u_{m(s_*^0)}(s_*^0) \leq u_{m(s_*^0)}(y_0)$. If $s_*^0 = y_0$, then $m(y_0) < y_0$ leads to a contradiction by an analogous argument as in Case 2.3, so we must have $m(y_0) \geq y_0$. Conversely, considering sequences going near $\underline{y}$ for arbitrarily long, we obtain that either $m(y_0) \geq y_0$ or there is $\tilde{s}_*^0 > y_0$ such that $m(\tilde{s}_*^0) < \tilde{s}_*^0$ and $u_{m(\tilde{s}_*^0)}(\tilde{s}_*^0) \leq u_{m(\tilde{s}_*^0)}(y_0)$.

Given $y \in (\underline{y}, \overline{y})$ such that $y \neq m(y)$, define $\hat{y} \neq y$ to be such that $u_{m(y)}(y) = u_{m(y)}(\hat{y})$. Now suppose that $m(y) \neq y$ for some $y \in (\underline{y}, \overline{y})$, and let y_0 be such that $|y - \hat{y}|$ is maximal. WLOG $m(y_0) < y_0$, so there is $s_*^0 < y_0$ such that $m(s_*^0) > s_*^0$ and $u_{m(s_*^0)}(s_*^0) \leq u_{m(s_*^0)}(y_0)$. Since $m(y_0) > m(s_*^0)$, $u_{m(y_0)}(s_*^0) < u_{m(y_0)}(y_0)$, so $\hat{y_0} > s_*^0$; but

$\hat{s^0_*} \geq y_0$. Hence $|s^0_* - \hat{s^0_*}| \geq |s^0_* - y_0| > |\hat{y_0} - y_0|$, a contradiction.

For the case where $m(y) = y$ for all $y \in [\underline{y}, \overline{y}]$, we use the following

Lemma 4. *Let $S = (y, y, \ldots)$, and let T be a path not identical to S. If x and x' both prefer T to S, and $x < y < x'$, then $x < x' - d_{x'}$ or $x' > x + e_x$.*

Proof. It is enough to check the case where T is contained in $[x, x']$: if not, create a new T' such that $T'_n = x'$ if $T_n > x'$, $T'_n = x$ if $T_n < x$ and $T'_n = T_n$ otherwise. Then T' is contained in $[x, x']$ and it is weakly better for both x and x' than T.

Now, if $x \geq x' - d_{x'}$ and $x' \leq x + e_x$, both x and x' derive non-negative utility from all elements of T'. Let $t = (1-\delta)\sum_s \delta^s T_s$ be the weighted average of policies in T, and $T'' = (t, t, \ldots)$. If $T'' \neq T$, both x and x' strictly prefer T'' over T by Jensen's inequality and A4, but they cannot both strictly prefer T'' over S. If $T'' = T$, $t \neq y$ and at least one agent prefers y over t. □

Take $\epsilon > 0$, $\nu > 0$ small and $y_0 = \underline{y} + \epsilon$. Construct s_{k_i} as before. It follows from previous arguments that $s^0_{**} = y_0$. For all i, a majority in $I(s_{k_i})$ must prefer $S(s_{k_i+1})$ over s_{k_i}. Since s_{k_i} strictly prefers s_{k_i} over $S(s_{k_i+1})$ and $s_{k_i} = m(s_{k_i})$, this can only happen if there are voters both above and below s_{k_i} who prefer $S(s_{k_i+1})$. Let $y'_i < s_{k_i} < y''_i$ be the closest voters to s_{k_i} who weakly prefer the continuation and denote $y'_i - (s_{k_i} - d_{s_{k_i}}) = \eta'_i$, $y''_i - s_{k_i} = \eta''_i$. Note that $\eta'_i, \eta''_i \underset{i\to\infty}{\longrightarrow} 0$.[28] In addition, $y''_i - d_{y''_i} > y'_i$; otherwise we obtain a contradiction as in Lemma 4. Let $\tilde{y}_i$ be such that $\tilde{y}_i - d_{\tilde{y}_i} = y'_i$.

Given the path $T^i = S(s_{k_i+1})$ construct T'^i as follows. If $T^i_j \geq y''_i + \nu$, $T'^i_j = y''_i + \nu$. If $y''_i + \nu > T^i_j \geq \tilde{y}_i$, $T'^i_j = y''_i$. If $\tilde{y}_i > T^i_j \geq s_{k_i}$, $T'^i_j = z_i = \frac{\sum_{\tilde{y}_i > T^i_j \geq s_{k_i}} \delta^j T^i_j}{\sum_{\tilde{y}_i > T^i_j \geq s_{k_i}} \delta^j}$. If $s_{k_i} > T^i_j$, $T'^i_j = v_i = \frac{\sum_{s_{k_i} > T^i_j} \delta^j T^i_j}{\sum_{s_{k_i} > T^i_j} \delta^j}$. Then both y'_i and y''_i weakly prefer T'^i over T^i.[29] Moreover, T'^i is a linear combination of at most four policies; by an abuse of notation, $T'^i = \omega^i_1[y''_i] + \omega^i_2[y''_i + \nu] + \omega^i_3[z_i] + \omega^i_4[v_i]$ with $\sum_j \omega^i_j = 1$. In addition, since $(s_{k_i}, s_{k_i+1}, \ldots)$ spends a long time near $\overline{y}$ (hence above $y''_i + \nu$) before going back under s_{k_i}, $\frac{\omega^i_4}{\omega^i_2} \underset{i\to\infty}{\longrightarrow} 0$.[30]

Finally, take $0 < \omega^i_5 \leq \omega^i_4$ such that $\omega^i_3 z_i + \omega^i_5 w_i = (\omega^i_3 + \omega^i_5) s_{k_i}$,[31] and construct $T'''^i = \omega^i_1[y''_i] + \omega^i_2[y''_i + \nu] + (\omega^i_3 + \omega^i_5)[s_{k_i}] + (\omega^i_4 - \omega^i_5)[v_i]$, $T''^i = w^i_1[y''_i] + w^i_2[y''_i + \nu] + w^i_3[v_i]$

[28] $\eta'_i \to 0$ by a similar argument to Case 2.3. Then, if η''_i did not converge to zero, (y'_i, y''_i) would be a strict majority in $I(s_{k_i})$ for some large i, and we are done.

[29] For this to work, we take ϵ small enough that $u_{y''_i}(\overline{y}) > 0$ for large i.

[30] We take ν small enough that $y_0 + \nu < \overline{y}$ for this to work.

[31] If this is not possible then y'_i could not have preferred T'^i over s_{k_i}, a contradiction.

where $w_1^i = \frac{\omega_1^i}{\omega_1^i+\omega_2^i+\omega_4^i-\omega_5^i}$, $w_2^i = \frac{\omega_2^i}{\omega_1^i+\omega_2^i+\omega_4^i-\omega_5^i}$, $w_3^i = \frac{\omega_4^i-\omega_5^i}{\omega_1^i+\omega_2^i+\omega_4^i-\omega_5^i}$ and $\frac{w_3^i}{w_2^i} \underset{i\to\infty}{\longrightarrow} 0$. Then both y_i' and y_i'' weakly prefer T'''^i over T^i and hence over s_{k_i}. Then

$$Cw_3^i \geq w_3^i u_{y_i'}(v_i) \geq u_{y_i'}(s_{k_i}) = u_{y_i'}(s_{k_i}) - u_{y_i'-\eta_i'}(s_{k_i}) = \eta_i' \frac{\partial u_{\tilde{\alpha}}(s_{k_i})}{\partial \alpha} \geq c\eta_i'$$

$$u_{y_i''}(s_{k_i}) \leq w_1^i u_{y_i''}(y_i'') + w_2^i u_{y_i''}(y_i''+\nu) + w_3^i u_{y_i''}(v_i) \leq (w_1^i + w_2^i) u_{y_i''}\left(y_i'' + \frac{w_2^i \nu}{w_1^i + w_2^i}\right) + w_3^i u_{y_i''}(s_{k_i})$$

$$u_{y_i''}(y_i'' - \eta_i'') = u_{y_i''}(s_{k_i}) \leq u_{y_i''}\left(y_i'' + \frac{w_2^i \nu}{w_1^i + w_2^i}\right) \leq u_{y_i''}\left(y_i'' + w_2^i \nu\right)$$

As A2 and A4 imply $u_\alpha(\alpha) - u_\alpha(\alpha - x) \in \left[\frac{M'}{2}x^2, \frac{M}{2}x^2\right]$, this means

$$\frac{M}{2}(\eta_i'')^2 \geq \frac{M'}{2}(w_2^i)^2 \Longrightarrow \frac{\eta_i''}{\eta_i'} \geq \sqrt{\frac{M'}{M}} \frac{c}{C} \frac{w_2^i}{w_3^i}$$

Since (y_i', y_i'') cannot be a strict majority in $I(s_{k_i})$, we must have $F(y_i'') - F(s_{k_i}) \leq F(y_i') - F(s_{k_i} - d_{s_{k_i}})$ for all i. But this is impossible as $\frac{f(x)}{f(x')}$ is bounded and $\frac{\eta_i''}{\eta_i'} \underset{i\to\infty}{\longrightarrow} \infty$, a contradiction. □

Proof of Proposition 2. Suppose $m(y) = y$ and $s(y) \neq y$; WLOG $s(y) < y$. A majority in $I(y)$ must prefer $S(s(y))$ over $S(y)$, i.e., they must prefer $S(s(y))$ over y. By Proposition 1, $s^k(y) \leq s(y)$ for all k. But then, for small enough $\epsilon > 0$, all agents in $(y - \epsilon, y + e_y)$ will strictly prefer y over $S(s(y))$, a contradiction.

If $m(y) \neq y$, suppose WLOG that $m(y) < y$. If $s(y) > y$ then $s^k(y) \geq s(y) > y$ for all k, so all voters in $(y - d_y, y)$—a strict majority in $I(y)$—strictly prefer y over $S(s(y))$, a contradiction. Hence $s(y) \leq y$. On the other hand, suppose $s(y) < m^*(y)$. Note that $m^*(y) < m(y)$; $m(m^*(y)) = m^*(y)$; $s^k(y) \leq s(y)$ for all k; and choosing $m^*(y)$ leads to the policy path $(m^*(y), m^*(y), \ldots)$ by the previous case. Then, for small enough $\epsilon > 0$, all voters in $(m(y) - \epsilon, y + e_y)$ prefer $S(m^*(y))$ over $S(s(y))$, a contradiction. Hence $s(y) \geq m^*(y)$. Next, suppose $s(y) = m^*(y)$ and consider $T = (m(y), s(m(y)), \ldots)$. Since T is contained in $[m^*(y), m(y)]$ and $T_1 = m(y) > m^*(y)$, all voters in $(m(y) - \epsilon, y + e_y)$ for small $\epsilon > 0$ strictly prefer T over $S(s(y))$, a contradiction. Hence $s(y) > m^*(y)$.

We now show that, if the MVT holds on $[m^*(y), y]$, then $s(y) < y$. Suppose that $s(y) = y$. There must be ϵ_0 such that $s(y - \epsilon) < y - \epsilon$ for all $0 < \epsilon < \epsilon_0$ (otherwise, $m(y)$ would prefer the stable path $(y - \epsilon, y - \epsilon, \ldots)$ over $(y, y, \ldots)$ for ϵ small enough).

Let $s_-(y) = \liminf_{\epsilon\to 0} s(y - \epsilon)$. By our previous results, $s_-(y) \in [m^*(y), y]$. There

are two cases: either $s_-(y) = y$ or $s_-(y) < y$.

If $s_-(y) = y$, then $s^k(z) \to y$ as $z \to y$ for all k. For all $z \in (y - \epsilon_0, y)$, $m(z)$ must prefer $S(s(z))$ to z (that is, to the constant path $(z, z, \ldots)$). Equivalently

$$V(z) = (1-\delta)U_{m(z)}(S(s(z))) - u_{m(z)}(z) = (1-\delta)\sum_{t=0}^{k_z} \delta^t u_{m(z)}\left(s^{t+1}(z)\right) - u_{m(z)}(z) \geq 0$$

where k_z is such that $s^{k_z+1}(z) + e_{s^{k_z+1}(z)} > m(z) > s^{k_z+2}(z) + e_{s^{k_z+2}(z)}$. (Note that $k_z \underset{z \to y}{\longrightarrow} \infty$.) By the envelope theorem,

$$\begin{aligned}
&V'(z) = (1-\delta)\frac{\partial}{\partial \alpha}U_{m(z)}(S(s(z)))m'(z) - \frac{d}{dz}u_{m(z)}(z) \\
&= \sum_{t=0}^{k_z}(1-\delta)\delta^t\left(\frac{\partial}{\partial \alpha}u_{m(z)}\left(s^{t+1}(z)\right) - \frac{\partial}{\partial \alpha}u_{m(z)}(z)\right)m'(z) - (1-\delta^{k_z+1})\frac{\partial}{\partial x}u_{m(z)}(z) - \delta^{k_z+1}\frac{d}{dz}u_{m(z)}(z) \\
&\geq \sum_{t=0}^{k_z}\delta^t\left(-M(z - s^{t+1}(z))\right)m'(z) - (1-\delta^{k_z+1})\frac{\partial}{\partial x}u_{m(z)}(z) - \delta^{k_z+1}\frac{d}{dz}u_{m(z)}(z) \\
&\underset{z \to y}{\longrightarrow} -\frac{\partial}{\partial x}u_{m(z)}(z) > 0.
\end{aligned}$$

Thus $V(z) \geq 0$ and $V'(z) > 0$ for all $z \in (y - \epsilon_1, y)$, whence $V(y) > 0$, which contradicts the assumption that $s(y) = y$.

If $s_-(y) < y$, let (y_n) be a sequence such that $y_n < y \ \forall n$, $y_n \to y$ and $s^k(y_n) \to s_k$ as $n \to \infty$, where $s_1 = s_-(y)$.

$m(y)$ must prefer y over $S(s(y_n))$ for all n, but $m(y_n)$ must prefer $S(s(y_n))$ over y. By continuity, $m(y)$ must be indifferent between y and $(s_k)_k$. Moreover, $m(y_n)$ prefers $S(s(y_n))$ to all other $S(s(y_{n'}))$, hence to (s_k). All this implies

$$\begin{aligned}
&0 \geq U_{m(y)}(S(s(y_n))) - \frac{1}{1-\delta}u_{m(y)}(y) = U_{m(y)}(S(s(y_n))) - U_{m(y)}((s_k)_k) \geq \\
&\geq U_{m(y)}(S(s(y_n))) - U_{m(y)}((s_k)_k) + U_{m(y_n)}((s_k)_k) - U_{m(y_n)}(S(s(y_n))) = \\
&= \sum_{t=0}^{k}\delta^t u_{m(y)}(s^{t+1}(y_n)) - \sum_{t=0}^{k'}\delta^t u_{m(y)}(s_{t+1}) + \sum_{t=0}^{k''}\delta^t u_{m(y_n)}(s_{t+1}) - \sum_{t=0}^{k'''}\delta^t u_{m(y_n)}(s^{t+1}(y_n))
\end{aligned}$$

If $k = k' = k'' = k'''$ for n high enough, then this last expression equals

$$\tilde{M}_n \sum_{t=0}^{k'} \delta^t \left(s^{t+1}(y_n) - s_{t+1}\right)\left(m(y) - m(y_n)\right)$$

for some $\tilde{M}_n \in [M', M]$. Now, crucially,

$$\frac{\tilde{M}_n \sum_{t=0}^{k'} \delta^t \left(s^{t+1}(y_n) - s_{t+1}\right)\left(m(y) - m(y_n)\right)}{y - y_n} \xrightarrow[n\to\infty]{} 0,$$

in other words, $U_{m(y)}(S(s(y_n))) - \frac{1}{1-\delta} u_{m(y)}(y) \in o(y - y_n)$. It can be shown that the same result holds even when k, k', k'' and k''' do not all become equal.[32]

Consider now the possibility of $m(y)$ choosing $S(y_n)$ instead. We can see that

$$\begin{aligned}
&(1-\delta)U_{m(y)}(S(y_n)) - u_{m(y)}(y) = \\
&= (1-\delta)\left(u_{m(y)}(y_n) - u_{m(y)}(y)\right) + \delta\left((1-\delta)U_{m(y)}(S(s(y_n))) - u_{m(y)}(y)\right) \\
&= (1-\delta)(y_n - y)\frac{\partial u_{m(y)}(\tilde{y})}{\partial x} + \delta\left(U_{m(y)}(S(s(y_n))) - U_{m(y)}(y)\right)
\end{aligned}$$

for some $\tilde{y} \in [y_n, y]$. Since $\frac{\partial u_{m(y)}(\tilde{y})}{\partial x} \xrightarrow[n\to\infty]{} \frac{\partial u_{m(y)}(y)}{\partial x} < 0$ and $U_{m(y)}(S(s(y_n))) - \frac{1}{1-\delta} u_{m(y)}(y) \in o(y - y_n)$, the above expression is positive for high n, which contradicts the assumption that $s(y) = y$ was optimal for $m(y)$.

Finally, we show that $s^k(y)$ converges to $m^*(y)$. Since $s^k(y) \in [m^*(y), y]$ for all y and the sequence is monotonically decreasing, it must have a limit $s^* \in [m^*(y), y)$. Suppose $s^* > m^*(y)$. By construction $m(s^*) < s^*$, so there is k_0 such that $m(s^k(y)) < s^*$ for all $k \geq k_0$. Such $m(s^k(y))$ would strictly prefer every element of the path $S(s^{k+2}(y))$ over $s^{k+1}(y)$, and hence would not choose $S(s^{k+1}(y))$, a contradiction. □

Proof of Corollary 1. Let $x_i^* < x_{i+1}^*$ be consecutive fixed points of m. Since m is continuous, either $m(y) > y \ \forall y \in (x_i^*, x_{i+1}^*)$ or $m(y) < y$ for all such y. The first case implies $m'(x_i^*) \geq 1$ and $m'(x_{i+1}^*) \leq 1$, and vice versa; since $m'(x_j^*) \neq 1$, these inequalities are strict, which implies that the intervals must alternate.

A fixed point of m is stable if $m'(x^*) < 1$ and unstable if $m'(x^*) > 1$ (see e.g.

[32]Note that k' is independent of n. If $u_{m(y)}(s_{t+1}) \neq 0$ for all t, then for high enough n, $k = k' = k'' = k'''$ because $y_n \to y$ and $s^{t+1}(y_n) \to s_{t+1}$. If $u_{m(y)}(s_{t+1}) \neq 0$ for some t, some negative terms are replaced by 0, but this can only reduce the difference between terms, i.e., $|\max(a,0) - \max(b,0)| \leq |a - b|$.

Elaydi (2005), Chapter 1.5). Since $m(-1) > -1$ and $m(1) < 1$, x_1^* and x_n^* are both stable, and stable and unstable fixed points alternate in between. □

Proof of Proposition 3. We first prove the monotonicity. Fix $\epsilon > 0$ small. Let $x < y \in [x^*, x^* + \epsilon]$, where $m(x^*) = x^*$, $m(x^{**}) = x^{**}$, $m(y) < y$ for all $y \in (x^*, x^{**})$ and $\epsilon < x^{**} - x^*$. Call $s^k(x) = x_k$, $s^k(y) = y_k$ and suppose $x_1 > y_1$. As usual $S(x_1)$ must be preferred to $S(y_1)$ by a weak majority in $I(x)$, and the opposite must happen in $I(y)$. Since all agents in $I(y) - I(x) = (x + e_x, y + e_y]$ prefer $S(x_1)$ due to A6, there must also be $z_0 \in I(x) - I(y) = [x - d_x, y - d_y)$ that prefers $S(x_1)$ (in fact there must be enough of them, but we only need one).

Let l be such that $x_i > y_i$ for $i = 1, 2, \ldots, l$ but not for $i = l + 1$. If $x_{l+1} = y_{l+1}$ (and hence $S(x_{l+1}) = S(y_{l+1})$) we have a contradiction, as any $z_0 \in I(x) - I(y)$ would prefer $S(y_1)$ over $S(x_1)$ pointwise. By a similar argument, there must be $z_l \in I(y_l) - I(x_l)$ that prefers $S(y_{l+1})$ over $S(x_{l+1})$. If $x_i < y_i$ for *all* $i \geq l+1$ this also yields a contradiction, as any such z_l would prefer $S(x_{l+1})$ over $S(y_{l+1})$ pointwise. More generally, if the ordering between x_l and y_l only changes a finite number of times, we can obtain a contradiction by looking at the last time it happens. The only case left to consider is if there are arbitrarily high i's and j's for which $x_i > y_i$ and $x_j < y_j$. If so, note that

$$0 \leq U_{z_0}(S(x_1)) - U_{z_0}(S(y_1)) = \sum_{t \geq 0} \delta^t \left(u_{z_0}(x_{1+t}) - u_{z_0}(y_{1+t}) \right)$$

$$0 \leq U_{z_0}(S(x_1)) - U_{z_0}(S(y_1)) = \sum_{t=0}^{l-1} \delta^t \left(u_{z_0}(x_{1+t}) - u_{z_0}(y_{1+t}) \right) + \sum_{t \geq l} \delta^t \left(u_{z_0}(x_{1+t}) - u_{z_0}(y_{1+t}) \right)$$

$$-\frac{\partial u_{z_0}(x)}{\partial x}|_{x^*} \sum_{t=0}^{l-1} \delta^t \left(x_{1+t} - y_{1+t} \right) \leq \sum_{t=0}^{l-1} \delta^t \left(u_{z_0}(y_{1+t}) - u_{z_0}(x_{1+t}) \right) \leq \sum_{t \geq l} \delta^t \left(u_{z_0}(x_{1+t}) - u_{z_0}(y_{1+t}) \right)$$

(Note that z_0 gets positive utility from all policies in $S(x_1)$ and $S(y_1)$.) Then

$$0 \leq U_{z_l}(S(y_{l+1})) - U_{z_l}(S(x_{l+1})) = \sum_{t\geq 0} \delta^t \left(u_{z_l}(y_{l+1+t}) - u_{z_l}(x_{l+1+t})\right) = \sum_{t\geq l} \delta^t \left(u_{z_l}(y_{1+t}) - u_{z_l}(x_{1+t})\right)$$

$$-\frac{\partial u_{z_0}(x)}{\partial x}|_{x^*} \sum_{t=0}^{l-1} \delta^t \left(x_{1+t} - y_{1+t}\right) \leq \sum_{t\geq l} \delta^t \left(u_{z_0}(x_{1+t}) - u_{z_l}(x_{1+t}) - u_{z_0}(y_{1+t}) + u_{z_l}(y_{1+t})\right)$$

$$-\frac{\partial u_{z_0}(x)}{\partial x}|_{x^*} \sum_{t=0}^{l-1} \delta^t \left(x_{1+t} - y_{1+t}\right) \leq \sum_{t\geq l} \delta^t M(z_0 - z_l)|x_{1+t} - y_{1+t}|$$

$$-\frac{\partial u_{z_0}(x)}{\partial x}|_{x^*} \max_{0\leq t\leq l-1}\{|x_{1+t} - y_{1+t}|\} \leq \frac{1}{1-\delta} M(z_0 - z_l) \sup_{t\geq l}\{|x_{1+t} - y_{1+t}|\}$$

$$\left(-\frac{\partial u_{z_0}(x)}{\partial x}|_{x^*}\right) \frac{1-\delta}{M(z_0 - z_l)} \max_{0\leq t\leq l-1}\{|x_{1+t} - y_{1+t}|\} \leq \sup_{t\geq l}\{|x_{1+t} - y_{1+t}|\}$$

Since $z_0 - z_l \leq x^* + \epsilon - d_{x^*+\epsilon} - x^* + d_{x^*} \xrightarrow[\epsilon\to 0]{} 0$, by taking ϵ small enough, we can guarantee that

$$D \max_{0\leq t\leq l-1}\{|x_{1+t} - y_{1+t}|\} \leq \sup_{t\geq l}\{|x_{1+t} - y_{1+t}|\}$$

for some fixed $D > 2$ (in fact we can take D arbitrarily large). Take $t_0 = \arg\max_{0\leq t\leq l-1}\{|x_{1+t} - y_{1+t}|\}$ and t_1 the smallest $t \geq l$ for which $|x_{1+t_1} - y_{1+t_1}| \geq 2|x_{1+t_0} - y_{1+t_0}|$. We can apply the same argument to that part of the sequence to obtain t_2 such that $|x_{1+t_2} - y_{1+t_2}| \geq 2|x_{1+t_1} - y_{1+t_1}|$, and so on for t_3, etc. This implies that for large enough j, $|x_{1+t_2} - y_{1+t_2}| > x^{**} - x^*$, a contradiction.

This argument proves (i) for an interval $(x^* - \epsilon, x^* + \epsilon)$. Now let

$$\hat{x} = \inf\{\tilde{x} : s \text{ is not monotonic on } [x^*, \tilde{x}]\}.$$

By the previous argument we know that $\hat{x} \geq x^* + \epsilon$. Suppose WLOG that $x^{**} > m^{-1}(x^* + e_{x^*})$; we will now show that $\hat{x} \geq m^{-1}(x^* + e_{x^*}) > x^* + e_{x^*}$. Suppose $\hat{x} < m^{-1}(x^* + e_{x^*})$.

By construction, for any $\epsilon > 0$, there must be pairs x^ϵ, y^ϵ such that $x^\epsilon < y^\epsilon$, $s(x^\epsilon) > s(y^\epsilon)$ and $x^\epsilon, y^\epsilon \in (\hat{x} - \epsilon, \hat{x} + \epsilon)$. There are two cases. If there are arbitrarily small ϵ for which $s^k(x^\epsilon), s^k(y^\epsilon) \geq \hat{x}$ for all k, and moreover $\lim_{k\to\infty} s^k(x^\epsilon) = \lim_{k\to\infty} s^k(y^\epsilon) \geq \hat{x}$, then we obtain a contradiction by repeating our previous argument. Else we are able to pick x^ϵ and y^ϵ so that, in the addition to the above conditions, $s^k(x^\epsilon) > s^k(y^\epsilon)$ for

all $k \geq 1$.[33] Label $s^k(x^\epsilon) = x_k$, $s^k(y^\epsilon) = y_k$.

Now, a majority in $I(y_0)$ must prefer $S(y_1)$ to $S(x_1)$, and a majority in $I(x_0)$ must prefer $S(x_1)$ to $S(y_1)$. In particular, some $z_0 \in (x_0 - d_{x_0}, y_0 - d_{y_0})$ must prefer $S(x_1)$ over $S(y_1)$. However, A2 implies that all agents $z \in [z_0, x^* + e_{x^*}]$ would then strictly prefer $S(x_1)$ over $S(y_1)$. For ϵ small enough, $m(I(y_0)) < x^* + e_{x^*}$, so a strict majority in $I(y_0)$ would prefer $S(x_1)$ to $S(y_1)$, a contradiction.

Finally we prove that the Median Voter Theorem must hold. Let $y \in I \cap J$ and suppose $m(y)$ strictly prefers $S(y')$ to $S(s(y))$, where $y' < s(y)$. Then, since s is increasing in $I \cap J$, $s^k(y') \leq s^k(s(y))$ for all k, so by A2 all voters $x < m(y)$ prefer $S(y')$ to $S(s(y))$. Some voters $x > m(y)$ close enough to $m(y)$ also prefer $S(y')$ by continuity. Hence $S(s(y))$ is not the Condorcet winner in $I(y)$, a contradiction. On the other hand, suppose $s(y) < y' \leq y$. Then all voters in $[m(y), x^* + e_{x^*}]$ prefer $S(y')$ by A2, and some voters $x < m(y)$ prefer $S(y')$ by continuity. On the other hand, voters $x \in (x^* + e_{x^*}, y + e_y]$ prefer $S(y')$ to $S(s(y))$ because $x > x^* + e_{x^*} \geq y$ and $s^k(y') \geq s^k(s(y))$ for all k. Hence $s(y)$ is not the Condorcet winner in $I(y)$, a contradiction. □

Proof of Proposition 8. If there are three points $x_1 < x_2 < x_3 \in (x-d, x+d)$ such that $f(x_1), f(x_3) < f(x_2)$, then there is a local maximum of f in $(x_1, x_3) \subseteq (x - d, x + d)$, as desired. Hence, if there is no local maximum, there must be $x* \in (x - d, x + d)$ such that f is decreasing in $(x - d, x^*]$ and increasing in $[x^*, x + d)$. Suppose WLOG that $f(x - d) \leq f(x + d)$. Remember that, by definition, $F(m(x)) - F(x - d) = \frac{F(x+d)-F(x-d)}{2}$; this implies

$$f'(x)m'(x) = \frac{f(x + d) + f(x - d)}{2}$$

given that $m(x) = x$. Since x is a stable steady state, $m'(x) < 1$, so $f'(x) > \frac{f(x+d)+f(x-d)}{2} \geq f(x - d)$. Hence $x > x^*$. But then $f|_{(x-d,x)} \leq f(x) \leq f|_{(x,x+d)}$, where the first inequality is sometimes strict. Hence $F(x + d) - F(x) > F(x) - F(x - d)$, which contradicts the assumption that x was a steady state.

The other case is analogous. □

Proof of Proposition 5. We first construct a sequence of approximate quasi-1-equilibria

[33] If eventually $s^k(x^\epsilon) < \hat{x}$, or $s^k(x^\epsilon)$ and $s^k(y^\epsilon)$ converge to different limits, then the inequality $s^k(x^\epsilon) \lesseqgtr s^k(y^\epsilon)$ can only change a finite number of times as we vary k, and we can look at the last time it happens.

as follows. For each $i = 1, 2, \ldots$, let $\epsilon(i) = \frac{1}{i}$ and take y_1, y_2 such that $x^* < y_1 < y_2 < x^* + \epsilon(i)$ and such that, moreover, $u_{m(y_2)}(y_2) < u_{m(y_2)}(y_1)$. Define $\tilde{x}_{ik} = y_1$ for all $k > 0$ and $\tilde{x}_{i0} = y_2$. Then, for $k = -1, -2, \ldots$ define $\tilde{x}_{ik}$ such that $m(\tilde{x}_{ik})$ is indifferent between the policy $\tilde{x}_{i(k+1)}$ and the path $(\tilde{x}_{i(k+2)}, \tilde{x}_{i(k+3)}, \ldots)$. (We can show by induction that $\tilde{x}_{ik}$ is uniquely defined and strictly decreasing in k for all $k < 0$, by Corollary 2.) Let $\tilde{s}_i$ denote the associated successor function, i.e., $\tilde{s}_i(y) = \tilde{x}_{i(k+1)}$ for all $y \in [\tilde{x}_{ik}, \tilde{x}_{i(k-1)})$.

We now make some useful observations. First, $\tilde{s}_i$ satisfies all the conditions to be a quasi-1-equilibrium for $k < 0$. Indeed, $m(\tilde{x}_{ik})$ is indifferent between $S(\tilde{x}_{i(k+1)})$ and $S(\tilde{x}_{i(k+2)})$ by construction; it follows that she prefers these policy paths to any other $S(\tilde{x}_{ik'})$ from A2. Second, it can be shown by induction that $\tilde{x}_{ik}(y_1, y_2)$ is a continuous function for all $k < 0$. Third, $\tilde{x}_{ik} \leq m^{-1}(\tilde{x}_{i(k+1)})$ for all $k < 0$; in particular, $\tilde{x}_{ik} \leq m^k(\tilde{x}_{i0}) = m^k(y_2)$. Fourth, $\tilde{x}_{ik} \xrightarrow[k \to -\infty]{} x^{**}$.[34]

Next, we argue that y_1, y_2 can be chosen so that some element of the sequence $(\tilde{x}_{ik})_k$ equals $\underline{x}$. For an arbitrary initial choice of y_1, y_2 satisfying the requirements above, let k_0 be such that $\underline{x} > \tilde{x}_{ik_0}$. Now lower (y_1, y_2) continuously towards x^* while satisfying the conditions that $x^* < y_1 < y_2 < x^* + \epsilon(i)$ and $u_{m(y_2)}(y_2) < u_{m(y_2)}(y_1)$. Then $\tilde{x}_{ik_0}(y_1, y_2) \leq m^k(y_2) \xrightarrow[y_2 \to x^*]{} x^*$, so there is some intermediate value of y_1, y_2 for which $\tilde{x}_{ik_0}(y_1, y_2) = \underline{x}$. Denote $y_{i1} = y_1$, $y_{i2} = y_2$, $x_{ik} = \tilde{x}_{i(k+k_0)}(y_{i1}, y_{i2})$, $s_i = \tilde{s}_i(y_{i1}, y_{i2})$. Note that in particular $x_{i0} = \underline{x}$ for all i.

We can now construct a true quasi-1-equilibrium s by taking the limit of a subsequence of s_i. We do this by a diagonal argument: $x_{i0} \to x_0 = \underline{x}$ by construction. Next, for all i, x_{i1} must be contained in $[x^*, \underline{x}]$, so we can take a convergent subsequence such that $x_{i_j 1} \to x_1$. Next, we take a subsequence such that the $x_{i_{j_l} 2}$ also converge, and so on. By an abuse of notation, let x_{jk} denote the result of this argument, so that $x_{jk} \to x_k$ for all k.

The indifference conditions that made the s_i quasi-1-equilibria under m_i make s a quasi-1-equilibrium under m by continuity. To guarantee that s is a proper quasi-1-equilibrium, we must also show that $x_k > x_{k+1}$ for all k; $x_k \xrightarrow[k \to +\infty]{} x^*$; and $x_k \xrightarrow[k \to -\infty]{} x^{**}$.

For all these claims it is enough to show that there cannot be two sequences $x_{ik(i)}$, $x_{ik'(i)}$ such that $k(i) < k'(i)$ for all i but $\lim_{i \to \infty} x_{ik(i)} = \lim_{i \to \infty} x_{ik'(i)} \in (x^*, x^{**})$. In

[34] This is not obvious, but if $\tilde{x}_{ik} \xrightarrow[k \to -\infty]{} y$ for some $y < x^{**}$, we can obtain a contradiction by an argument analogous to the proof of Proposition 2.

turn, it is enough to show that this cannot happen for $k'(i) = k(i)+1$. Suppose it does, and relabel the sequences as follows: $y_{il} = x_{i(l+k(i))}$. (If necessary, take a subsequence such that $y_{il} \to y_l$ for all l.) Then we just have to show that $y_0 > y_1$. Clearly $y_0 \geq y_1$ as $y_{i0} > y_{i1}$ for all i, so suppose $y_0 = y_1$. If $y_2 > y_1$, then $m(y_0) = m(y_1)$ must be indifferent between y_1, $S(y_2)$ and y_2, which implies $y_2 < m(y_1) < y_1$. But then $m(y_{1i})$ would strictly prefer y_{i2} to $S(y_{i2})$ for high enough i, a contradiction. Hence $y_1 = y_2$, and by the same argument $y_2 = y_3 = y_4 = \ldots$.

This will lead to a contradiction by a similar argument as in Proposition 2. Let $V(y) = U_{m(y)}(S(s(y))) - u_{m(y)}(y)$ as in that proof. The fact that $y_1 = y_0$ implies that $V_i(y_{i0}) \to 0$. Now take an arbitrary sequence $(g(i))_i \subseteq \mathbb{N}$, and denote $y_{ig(i)} = y_{i0} - \epsilon_i$. Then, by the argument in Proposition 2,

$$V_i(y_{i0}) \geq (1-\delta)\epsilon_i \left(-\frac{\partial}{\partial x} u_{m(y_{i0})}(\tilde{y})\right) + \delta \left(V(y_{i0}) - M\epsilon_i(E(S(y_{i1})) - E(S(y_{i(g(i)+1)}))\right)$$

$$V_i(y_{i0}) \geq \epsilon_i \left(-\frac{\partial}{\partial x} u_{m(y_{i0})}(\tilde{y}_i) - \frac{\delta}{1-\delta} M(E(S(y_{i1})) - E(S(y_{i(g(i)+1)}))\right)$$

for some $\tilde{y}_i \in (y_{ig(i)}, y_{i0})$, where $E(S(y)) = (1-\delta)\sum_{t=0}^{\infty} \delta^t s^t(y)$. Given some $0 < \epsilon' < \epsilon$ and $i \in \mathbb{N}$, we say that $g(i) \in \mathbb{N}$ is ϵ',ϵ-*valid* if $\epsilon_i \in (\epsilon', \epsilon)$ and $E(S(y_{i1})) - E(S(y_{i(g(i)+1)})) \leq \frac{1-\delta}{2\delta} \frac{-\frac{\partial}{\partial x} u_{m(y_{i0})}(y_{i0})}{M}$. Clearly, if there are $0 < \epsilon' < \epsilon$ with ϵ small enough, for which we can find valid $g(i)$ for arbitrarily high i, we obtain a contradiction, as $\liminf V_i(y_{i0}) > 0$. If not, then there must be a fixed $\epsilon > 0$ and a sequence $\epsilon'_i \to 0$ for which there are no ϵ'_i,ϵ-*valid* values of $g(i)$ for any $i \geq i_0$. If there are arbitrarily high values of i for which $(y_{ik})_i \cap (y_{i0} - \epsilon, y_{i0} - \epsilon'_i)$ is empty, then let $y_{ih(i)}$ be the last element to the right of this gap, i.e., $y_{i(h(i)+1)} < y_{i0} - \epsilon, y_{i0} - \epsilon'_i < y_{ih(i)}$ and relabel the sequence so that $z_{i0} = y_{ih(i)}$. Then $z_1 < z_0 = y_0 = z_{-1}$, which leads to a contradiction by our previous argument. If there are arbitrarily high values of i for which there is $g(i)$ such that $\epsilon_i \in (\epsilon'_i, \epsilon)$, but $E(S(y_{i1})) - E(S(y_{i(g(i)+1)})) > C$ for a fixed C, this implies that there are fixed C' and k_0 such that $y_{i0} - y_{i(g(i)+k_0)} > C'$, and hence $V_i(y_{i(g(i)+k)}) \geq C''$ for some $0 < k \leq k_0$. Note that k_0, C, C' and C'' are fixed even as we take ϵ to 0, which implies that $\frac{\partial V_i(y)}{\partial y}$ must become arbitrarily large and negative as $i \to \infty$, a contradiction.

We now argue that, for any $y \in [x_n, x_{n-1})$ and for any $k \neq n+1$, a majority in $I(y)$ prefers $S(x_{n+1})$ over $S(x_k)$. Suppose $k > n+1$. By Lemma 2, there is some $\alpha_0(n+1, k)$ such that all agents in $(\alpha_0, 1]$ strictly prefer $S(x_{n+1})$ over $S(x_k)$ and

all agents in $[-1, \alpha_0)$ strictly prefer $S(x_k)$ over $S(x_{n+1})$. It is enough to show that $\alpha_0(n+1,k) \leq m(x_n)$. By construction, $m(x_n) = \alpha_0(n+1,n+2) > \alpha_0(n+2,n+3) > \ldots > \alpha_0(k-1,k)$, so $m(x_n)$ prefers $S(x_{n+1})$ to $S(x_{n+2})$ to $\ldots$ to $S(x_k)$.

Next, we show that s is a 1-equilibrium in $[x^*, x^* + d_{x^*}]$ iff $m(x_{n-1}) < x_{n+1}$ for all n.

Note first that $m(x_{n-1}) < x_n$ always holds (otherwise $m(x_{n-1})$ could not be indifferent between x_n and $S(x_{n+1})$). If $m(x_{n-1}) > x_{n+1}$, $m(x_{n-1})$ prefers $m(x_{n-1})$ to x_{n+1}; hence he prefers $S(m(x_{n-1}))$ to $S(x_{n+1})$, and also to $S(x_n)$. This implies that $S(x_n)$ cannot be a Condorcet winner in $I(x_{n-1})$, as the Median Voter Theorem must hold in this interval by Proposition 3, and thus s is not a 1-equilibrium.

Conversely, suppose that, for some $x \in [x_n, x_{n-1})$, $I(x)$ prefers $S(y)$ to $S(x_{n+1})$ for some $y \in [x_k, x_{k-1})$. If $k \leq n+1$, this is impossible as all agents in $[x - d_x, m(x) + \epsilon]$ would strictly prefer $S(x_{n+1})$ to $S(y)$. Suppose then that $k \geq n+2$. By the Median Voter Theorem, $m(x)$ prefers $S(y)$ to $S(x_{n+1})$. Suppose $m(x) \in [x_b, x_{b-1})$; we will argue that $b = k$. If $b < k$, $m(x)$ prefers $S(x_{n+1})$ to $S(x_b)$ to $S(y)$, a contradiction. If $b > k$, $m(x)$ prefers $S(x_{n+1})$ to $S(x_k)$ to $S(y)$, a contradiction.

Next, note that, if indeed $m(x)$ prefers $S(y)$ to $S(x_{n+1})$, she then prefers $S(y)$ to $S(x_{k-1})$, and so do all agents z such that $y - d_y < z < m(x)$ by A2. Hence a majority in $I(x_{k-2})$ should prefer $S(y)$ to $S(x_{k-1})$. By the above argument, since $y \in [x_k, x_{k-1})$ it must be that $m(x_{k-2}) \in [x_k, x_{k-1})$, a contradiction.

Finally, we show that, if Condition (*) holds, there is $\overline{\delta}$ such that, if $\delta > \overline{\delta}$, s is a 1-equilibrium in $[x^*, x^{**}]$.

Suppose not. Then there is a sequence $(\delta_n)_n$ with $\delta_n \to 1$ and a sequence of quasi-1-equilibria s_n for each δ_n, such that s_n is not a 1-equilibrium for all n. By a diagonal argument, assume that $s_n \to s$ in the sense of s^t. Suppose that, for each n, there is x_k^n for which $S_n(x_{k+1}^n)$ is not a Condorcet winner in $I(x_k^n)$ because a strict majority strictly prefers $S_n(y_n)$, and assume $x_k^n \to x$ and $y_n \to y$. Note that $y_n \leq x_{k+1}^n$, as otherwise all agents to the left of $m(x_k^n)$ and some to the right would strictly prefer $S_n(x_k^n)$ over $S_n(y_n)$; and thus $y \leq x$. If $x \in (x^*, x^{**})$ and $y < x$, this leads to a contradiction as $U_\alpha(S_n(y_n)) \to U_\alpha(S(y))$ and $U_\alpha(S_n(x_k^n)) \to U_\alpha(S(x))$ for all α, and $U_\alpha(S(x)) > U_\alpha(S(y))$ for all $\alpha \in (m(x) - \epsilon, x + e_x)$ for some $\epsilon > 0$. If $y = x$, suppose $y_n \in (x_{k+l_n}^n, x_{k+l_n-1}^n)$, where $l_n \geq 2$. It is clear that $U_\alpha(S_n(x_{k+1}^n)) > U_\alpha(S_n(x_{k+l_n-1}^n)) > U_\alpha(S_n(y_n))$ for all $\alpha \in (x_{k+l_n-1}^n, x_k^n + e_{x_k^n})$ and $U_\alpha(S_n(x_{k+1}^n)) > U_\alpha(S_n(x_{k+l_n}^n)) > U_\alpha(S_n(y_n))$ for all $\alpha \in (m(x), x_{k+l_n}^n)$, so it must be

that some $\alpha_n \in (x^n_{k+l_n-1}, x^n_{k+l_n})$ prefers $S_n(y_n)$ to $S_n(x^n_{k+1})$. But then

$$
\begin{aligned}
0 \geq & \frac{1-\delta}{1-\delta^{l_n-1}}\left[U_{\alpha_n}(S_n(x^n_{k+1})) - U_{\alpha_n}(S_n(y_n))\right] = \\
&= \frac{1-\delta}{1-\delta^{l_n-1}}\left[\sum_{t=0}^{l_n-1}\delta^t u_{\alpha_n}(x^n_{k+1+t}) - u_{\alpha_n}(y_n) - (1-\delta^{l_n-1})\sum_{t=1}^{\infty}\delta^t u_{\alpha_n}(x^n_{k+l_n+t})\right] = \\
&= \frac{1-\delta}{1-\delta^{l_n-1}}\left[u_{\alpha_n}(x^n_{k+1}) - u_{\alpha_n}(y_n) + \sum_{t=1}^{l_n-1}\delta^t u_{\alpha_n}(x^n_{k+1+t}) - (1-\delta^{l_n-1})\delta U_{\alpha_n}(S(x^n_{k+l_n+1}))\right] \\
&\xrightarrow[n\to\infty]{} 0 + \delta\left[u_x(x) - (1-\delta)U_x(S(x))\right] > 0,
\end{aligned}
$$

a contradiction.

An analogous proof can be written if $x = x^*$ after a normalization argument. Briefly, if $x = x^*$, assume WLOG that $x^* = 0$ to simplify notation, and denote $T_n(y) = x^n_k y$ and $U^n_\alpha(y) = U_{\alpha x^n_k}(\alpha x^n_k) - \frac{1}{(x^n_k)^2}\left(U_{\alpha x^n_k}(\alpha x^n_k) - U_{\alpha x^n_k}(y x^n_k)\right)$. In the normalized version of the problem, x^n_k maps to $y^n_k = 1 > 0$ and we can apply the above arguments. The case $x = x^{**}$ is similar. □

Additional proofs and robustness checks are found in the online Appendices B-E.

References

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin, "Coalition Formation in Non-Democracies," *The Review of Economic Studies*, 2008, *75* (4), 987–1009.

_ , _ , and _ , "Dynamics and Stability of Constitutions, Coalitions, and Clubs," *American Economic Review*, 2012, *102* (4), 1446–1476.

_ , _ , and _ , "Political Economy in a Changing World," *Journal of Political Economy*, 2015, *123* (5), 1038–1086.

Bai, Jinhui H and Roger Lagunoff, "On the Faustian Dynamics of Policy and Political Power," *The Review of Economic Studies*, 2011, *78* (1), 17–48.

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

Bewley, Truman F, "A Critique of Tiebout's Theory of Local Public Expenditures," *Econometrica: Journal of the Econometric Society*, 1981, pp. 713–740.

Bouton, Laurent and Gabriele Gratton, "Majority Runoff Elections: Strategic Voting and Duverger's Hypothesis," *Theoretical Economics*, 2015, *10* (2), 283–314.

Cremer, Helmuth and Pierre Pestieau, "Factor Mobility and Redistribution," in "Handbook of Regional and Urban Economics," Vol. 4, Elsevier, 2004, pp. 2529–2560.

Elaydi, Saber, *An Introduction to Difference Equations*, Springer, 2005.

Epple, Dennis and Thomas Romer, "Mobility and Redistribution," *Journal of Political Economy*, 1991, pp. 828–858.

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

Grossman, Gene M, "Union Wages, Temporary Layoffs, and Seniority," *The American Economic Review*, 1983, *73* (3), 277–290.

_ , "International Competition and the Unionized Sector," *Canadian Journal of Economics*, 1984, *17* (3), 541–556.

Hunt, Brian R, Tim Sauer, and James A Yorke, "Prevalence: a translation-invariant "almost every" on infinite-dimensional spaces," *Bulletin of the American mathematical society*, 1992, *27* (2), 217–238.

Jack, William and Roger Lagunoff, "Dynamic Enfranchisement," *Journal of Public Economics*, 2006, *90* (4-5), 551–572.

Ortega, Francesc, "Immigration Quotas and Skill Upgrading," *Journal of Public Economics*, 2005, *89* (9-10), 1841–1863.

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

Suwankiri, Benjarong, Assaf Razin, and Efraim Sadka, "The Welfare State and Migration: a Dynamic Analysis of Political Coalitions," *Research in Economics*, 2016, *70* (1), 122–142.

Tiebout, Charles M, "A Pure Theory of Local Expenditures," *Journal of Political Economy*, 1956, pp. 416–424.

Wooders, Myrna Holtz, "A Tiebout Theorem," *Mathematical Social Sciences*, 1989, *18* (1), 33–55.

B Proofs (Continuous Time Limit) (For Online Publication)

Lemma 5. *If $s^t(x)$ is continuously differentiable and strictly decreasing in t, then there are functions $d(x,y) : [x^*, x^{**}]^2 \to \mathbb{R}$ and $e(z) : [x^*, x^{**}] \to \mathbb{R}_+$ such that $s^{d(x,y)}(x) = y$ and $d(x,y) = \int_y^x e(z)dz$.*

$d(x,y)$ measures the time it takes the policy path to get from x to y, if $x > y$ (if $x < y$ then the time is negative). This time can be expressed as an integral of the instantaneous delay $e(z)$ at each policy z. Note that $d(x,y)$ and $e(z)$ are still well-defined even if $s^t(x)$ is only continuous in t a.e. ($s^t(x)$ may have instantaneous jumps, which correspond to $e(z) = 0$ for the policies that are jumped over).

Proof of Lemma 5. Since $s^t(x)$ is continuous and decreasing as a function of t, for each $y < x$ there is a unique $d(x,y) > 0$ such that $s^{d(x,y)}(x) = y$. Conversely, $d(y,x) < 0$ if $x > y$ (in fact, by additivity, $d(y,x) = -d(x,y)$).

Moreover, $d(x,y)$ is decreasing in y. Since $s^t(x)$ is C^1 in t by assumption, $d(x,y)$ is C^1 in y, so we can define $e(x,y) = -\frac{\partial d(x,y)}{\partial y}$, so that $d(x,y) = \int_y^x e(x,z)dz$. From the additivity of s with respect to t it follows that $\frac{\partial d(x,y)}{\partial y}$ depends only on y, so $e(x,z) = e(z)$ as desired. □

The following Proposition extends Proposition 6:

Proposition 10. *Let*

$$\tilde{e}(x) = \frac{1}{r}\left(-\frac{\frac{\partial^2 u}{\partial x^2}}{\frac{\partial u}{\partial x}} - 2\frac{\frac{\partial^2 u}{\partial \alpha \partial x}}{\frac{\partial u}{\partial x}}m'(x) - \frac{\frac{\partial^2 u}{\partial \alpha^2}}{\frac{\partial u}{\partial x}}m'(x)^2 + \frac{\frac{\partial^2 U_{m(x)}(S(x))}{\partial \alpha^2}}{\frac{\partial u}{\partial x}}m'(x)^2 + \frac{m''(x)}{m'(x)}\right). \quad (*)$$

Then, if $\tilde{e}(x) \geq 0$ for all $x \in [x^*, x^{**}]$, there is an MPE s_* given by $e \equiv \tilde{e}$. Moreover, this is the only continuous MPE.[[[adjust language]]]

Otherwise, assume that $f \in C^2$ and $A = \{x \in [x^*, x^{**}] : \tilde{e}(x) < 0\}$ is a finite union of open intervals. Let $\tilde{x} = \inf A$. Then, for generic m, s_* is given by $e(x) = \tilde{e}(x)$ for $x \leq \tilde{x}$, and by two sequences $(y_l)_{l \in \mathbb{N}_{\geq 0}}$, $(e_l)_{l \in \mathbb{N}_{\geq 1}}$ such that: $(y_l)_l$ is increasing,

$y_0 = \tilde{x}$ *and* $y_l \xrightarrow[l\to\infty]{} x^{**}$*;* $d(y_l^-, y_{l+1}^+) = 0$ *and* $d(y_l^+, y_l^-) = e_l$ *for all* $l \geq 1$*; and*[35]

$$U_{m(y_{l+1})}(S(y_l^+)) = u_{m(y_{l+1})}(y_{l+1})$$

$$e^{\frac{re_l}{2}} = 1 + m'(x)\frac{\frac{\partial V_{m(y_l)}(S(y_l^-))}{\partial \alpha}}{\frac{\partial u_{m(x)}(x)}{\partial x}}$$

*All sequences of quasi-*1*-equilibria of the* j*-refined games* $(s_j)_j$ *converge a.e. to* s_**, i.e.,* $s_j^t(x) \xrightarrow[j\to\infty]{} s_*^t(x)$ $\forall x, t$ *where* $s^t(x)$ *is continuous.*

In addition, if $m(y_l) < y_{l-1}$ *for all* l *or* $\tilde{e}(x) \geq 0 \forall x$*, all quasi-*1*-equilibria are equilibria for* δ *close enough to* 1*. (Condition (*))*

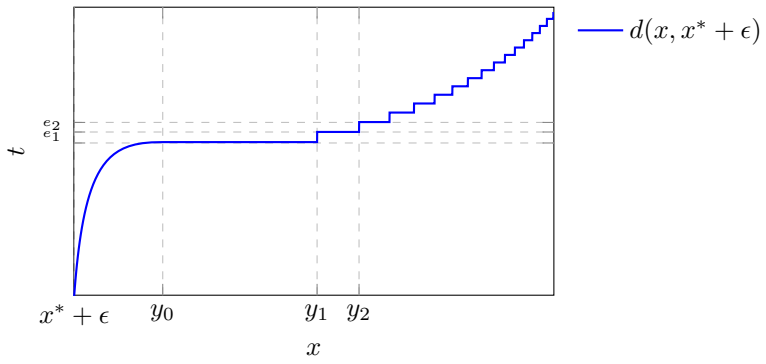


Figure 4: Equilibrium with jumps and stops in continuous time

Proof of Proposition 10. Given a policy mapping $s^t(x)$, denote $V_\alpha(S) = U_\alpha(S) - u_\alpha(m^{-1}(\alpha))$, $V(\alpha) = U_\alpha(S(m^{-1}(\alpha))) - u_\alpha(m^{-1}(\alpha))$ and $W(x) = U_{m(x)}(S(x)) - u_{m(x)}(x)$.

We proceed as follows.

First, define a continuous time quasi-equilibrium (CTQE) as a policy mapping $s^t(x)$ for the continuous time game such that:

(i) $x = \arg\max_{y\in[-1,1]} U_{m(x)}(S(y))$ $\forall x \in [x^*, x^{**}]$.

(ii) If there is $c > 0$ s.t. $W(x_0 - \epsilon) = 0$ $\forall \epsilon \in [0, c]$, then $d(x_0^+, x_0) = 0$.

[35] We denote $f(x^-) = \lim_{t\nearrow x} f(x)$, $f(x^+) = \lim_{t\searrow x} f(x)$ and $E_{y_l}(S) = E((f(s_i))_i)$, where $f(s_i) = s_i$ if $|s_i - m(y_l)| < d$ and $f(s_i) = m(y_l)$ otherwise.

(iii) If $W(x_0) = 0$ and $W(x') > 0$ for all x' in a left-neighborhood of x_0, then $d(x_0^+, x_0^-)$ satisfies

$$e^{\frac{rd(x_0^+, x_0^-)}{2}} = 1 + m'(x_0)\frac{\frac{\partial V_{m(x_0)}(S(x_0^-))}{\partial \alpha}}{\frac{\partial u_{m(x_0)}(x_0)}{\partial x}}.$$

We say a CTQE is smooth in (x^*, x_0) if $d(x, y) = \int_y^x e(z)dz$ for all $x^* < y < x \leq x_0$. We first prove some properties of a CTQE:

Lemma 6. *Let $x_0 \in (x^*, x^{**})$. If a CTQE $s^t(x)$ is such that $W(x) = 0$ for all x in a neighborhood of x_0, then $d(x_0, x)$ is differentiable with respect to its second argument at (x_0, x_0), and $e(x) = -\frac{\partial d(x_0,x)}{\partial x}$ is given by Equation (*).*

Proof. First, assume a smooth CTQE $s^t(x)$. Denote $n(\alpha) = m^{-1}(\alpha)$ and $\alpha_0 = m(x_0)$. By the envelope theorem,

$$\begin{aligned}
V'(\alpha_0) &= \frac{\partial U_\alpha(S(x))}{\partial \alpha}|_{\alpha_0,x_0} - \frac{du_\alpha(n(\alpha))}{d\alpha}|_{\alpha_0} \\
V''(\alpha_0) &= \frac{\partial^2 U_\alpha(S(x))}{\partial \alpha^2}|_{\alpha_0,x_0} + n'(\alpha_0)re(x_0)\left[\frac{\partial u_\alpha(x)}{\partial \alpha}|_{\alpha_0,x_0} - \frac{\partial U_\alpha(S(x))}{\partial \alpha}|_{\alpha_0,x_0}\right] - \frac{d^2u_\alpha(n(\alpha))}{d\alpha^2}|_{\alpha_0} \\
V''(\alpha_0) &= \frac{\partial^2 U_\alpha(S(x))}{\partial \alpha^2}|_{\alpha_0,x_0} + n'(\alpha_0)re(x_0)\left[\frac{\partial u_\alpha(x)}{\partial \alpha}|_{\alpha_0,x_0} - \frac{\partial U_\alpha(S(x))}{\partial \alpha}|_{\alpha_0,x_0}\right] \\
&\quad - \frac{\partial^2 u}{\partial \alpha^2} - 2n'\frac{\partial^2 u}{\partial \alpha \partial x} - n'^2\frac{\partial^2 u}{\partial^2 x} - n''\frac{\partial u}{\partial x}.
\end{aligned}$$

We can use the fact that $V(\alpha) \equiv 0$ in a neighborhood of α_0, and hence $V' \equiv V'' \equiv 0$, to determine $\tilde{e}(x)$:

$$\begin{aligned}
0 = V'(\alpha_0) &= \frac{\partial U_\alpha(S(x))}{\partial \alpha}|_{\alpha_0,x_0} - \frac{\partial u_\alpha(x)}{\partial \alpha}|_{\alpha_0,x_0} - n'(\alpha)\frac{\partial u_\alpha(x)}{\partial x}|_{\alpha_0,x_0} \\
\implies r\tilde{e}(x)\left[-\frac{1}{m'(x)}\frac{\partial u_{m(x)}(x)}{\partial x}\right] &= r\tilde{e}(x)\left[\frac{\partial u_{m(x)}(x)}{\partial \alpha} - \frac{\partial U_{m(x)}(S(x))}{\partial \alpha}\right] = \\
&= -m'(x)\frac{\partial^2 U_{m(x)}(S(x))}{\partial \alpha^2} + m'(x)\frac{d^2u_\alpha(n(\alpha))}{d\alpha^2}|_{\alpha_0} = \\
&= -m'(x)\frac{\partial^2 U_{m(x)}(S(x))}{\partial \alpha^2} + \frac{1}{m'(x)}\frac{\partial^2}{\partial x^2}u + 2\frac{\partial^2}{\partial \alpha \partial x}u + m'(x)\frac{\partial^2}{\partial \alpha^2}u - \frac{m''(x)}{m'(x)^2}\frac{\partial}{\partial x}u.
\end{aligned}$$

Remark. In the quadratic case, the condition imposed by Equation (*) simplifies to

$$e(x) = \frac{1}{r}\left(\frac{2m'(x)-1}{x-m(x)} + \frac{m''(x)}{m'(x)}\right) - \frac{(m'(x))^2 e^{-rt^*(x)}}{x-m(x)}\left(e(m(x)-d)d + \frac{1}{r}\right),$$

where $t^*(x) = d(x, m(x)-d)$; or

$$e(x) = \frac{1}{r}\left(\frac{2m'(x)-1}{x-m(x)} + \frac{m''(x)}{m'(x)}\right)$$

for $x \in [x^*, x^*+d)$.

Now we show that d must be differentiable at (x_0, x_0). Let $(z_n)_n$ be a sequence such that $z_n \to x_0$. WLOG assume $z_n < x_0$ for all n. Note that

$$\begin{aligned}
&\frac{\partial U_{m(x_0)}(S(x_0))}{\partial\alpha}|_{m(x_0)} - \frac{\partial U_{m(x_0)}(S(z_n))}{\partial\alpha} = \int_0^\infty re^{-rt}\left[\frac{\partial u_{m(x_0)}(s^t(x_0))}{\partial\alpha} - \frac{\partial u_{m(x_0)}(s^t(z_n))}{\partial\alpha}\right]dt \\
&= \int_0^{d(x_0,z_n)} re^{-rt}\frac{\partial}{\partial\alpha}u_{m(x_0)}(s^t(x_0))dt - \left(1-re^{-rd(x_0,z_n)}\right)\frac{\partial U_{m(x_0)}(S(z_n))}{\partial\alpha} \\
&= \left(1-re^{-rd(x_0,z_n)}\right)\left(\frac{\partial u_{m(x_0)}(\tilde{x})}{\partial\alpha} - \frac{\partial U_{m(x_0)}(S(z_n))}{\partial\alpha}\right)
\end{aligned}$$

for some $\tilde{x} \in (z_n, x_0)$.

Then

$$\begin{aligned}
0 &= V'(m(x_0)) - V'(m(z_n)) = \\
&= \frac{\partial U_\alpha(S(x))}{\partial\alpha}|_{m(x_0),x_0} - \frac{du_\alpha(n(\alpha))}{d\alpha}|_{m(x_0)} - \left(\frac{\partial U_\alpha(S(x))}{\partial\alpha}|_{m(z_n),z_n} - \frac{du_\alpha(n(\alpha))}{d\alpha}|_{m(z_n)}\right) \\
&= \frac{\partial U_\alpha(S(x))}{\partial\alpha}|_{m(x_0),x_0} - \frac{\partial U_\alpha(S(x))}{\partial\alpha}|_{m(x_0),z_n} + \frac{\partial U_\alpha(S(x))}{\partial\alpha}|_{m(x_0),z_n} - \frac{\partial U_\alpha(S(x))}{\partial\alpha}|_{m(z_n),z_n} - \\
&- (m(x_0)-m(z_n))\frac{d^2u_\alpha(n(\alpha))}{d\alpha^2}|_{\tilde{\alpha}} \\
&= \left(1-e^{-rd(x_0,z_n)}\right)\left(\frac{\partial u_{m(x_0)}(\tilde{x})}{\partial\alpha} - \frac{\partial U_{m(x_0)}(S(z_n))}{\partial\alpha}\right) + (m(x_0)-m(z_n))\frac{\partial^2 U_\alpha(S(x))}{\partial\alpha^2}|_{\hat{\alpha},z_n} - \\
&- (m(x_0)-m(z_n))\frac{d^2u_\alpha(n(\alpha))}{d\alpha^2}|_{\tilde{\alpha}} \\
0 &= \frac{\partial^2 U_\alpha(S(x))}{\partial\alpha^2}|_{\hat{\alpha},z_n} + \frac{1-e^{-rd(x_0,z_n)}}{m(x_0)-m(z_n)}\left(\frac{\partial u_{m(x_0)}(\tilde{x})}{\partial\alpha} - \frac{\partial U_{m(x_0)}(S(z_n))}{\partial\alpha}\right) - \frac{d^2u_\alpha(n(\alpha))}{d\alpha^2}|_{\tilde{\alpha}}.
\end{aligned}$$

This implies that $\lim_{n\to\infty}\frac{1-e^{-rd(x_0,z_n)}}{m(x_0)-m(z_n)} = n'(m(x_0))re(x_0)$, which in turn implies that

$\lim_{n\to\infty} \frac{d(x_0,z_n)}{x_0-z_n} = e(x_0)$, as we wanted. □

Lemma 7. *Equation (*) has a unique solution, in the following sense: for any $x_1 > x_0 \geq x^*$ and given a candidate path $S(x_0)$, there is at most one way to choose $e : (x_0, x_1) \to \mathbb{R}_{\geq 0}$ so that Equation (*) holds for all $x \in (x_0, x_1)$.*

Proof. Let $\tilde{g}(x) = r\tilde{e}(x)\left[-\frac{1}{m'(x)}\frac{\partial u_{m(x)}(x)}{\partial x}\right]$ and $g(x) = \max(\tilde{g}(x), 0)$. The issue is that Equation (*) is an integral equation, since $\frac{\partial^2 U_{m(x)}(S(x))}{\partial\alpha\partial x}$ is an integral that depends on $S(x)$, which depends on $g(x')$ for $x' < x$. We prove the result for the case $x_0 = x^* < x_1$ but other cases are analogous.

Given $x_1 \in (x^*, x^{**})$, let $\mathcal{C}_{x_1} = \{h : [x^*, x_1] \to \mathbb{R}_{\geq 0} \text{ continuous}\}$ with the norm $||h||_\infty$, and define $T_{x_1} : \mathcal{C}_{x_1} \to \mathcal{C}_{x_1}$ as follows:

$$T_{x_1}(g)(x) = \max\left(-m'(x)\frac{\partial^2 U_{m(x)}(S(x))}{\partial\alpha^2} + m'(x)\frac{d^2 u_{m(x)}(x)}{dm(x)^2}, 0\right)$$

for $x \in [x^*, x_1]$. Let g_1, $g_2 \in \mathcal{C}_{x_1}$ with $||g_1 - g_2|| \leq K$, and, for each $x_2 \in (x^*, x_1)$, define g_{x_2} by: $g_{x_2}(x) = g_1(x)$ if $x \leq x_2$ and $g_{x_2}(x) = g_2(x)$ otherwise. Then

$$\begin{aligned}
&|T_{x_1}(g_1)(x) - T_{x_1}(g_2)(x)| \leq m'(x)\left|-\frac{\partial^2 U_{m(x)}(S_{g_1}(x))}{\partial\alpha^2} + \frac{\partial^2 U_{m(x)}(S_{g_2}(x))}{\partial\alpha^2}\right| = \\
&= m'(x)\left|\int_{x^*}^{x} \frac{\partial}{\partial x_2}\left(\frac{\partial^2 U_{m(x)}(S_{g_{x_2}}(x))}{\partial\alpha^2}\right) dx_2\right| = \\
&= m'(x)\left|\int_{x^*}^{x} re^{-rd(x,x_2)}\frac{m'(x_2)}{r\frac{\partial u_{m(x)}(x)}{\partial x}}(g_1(x_2) - g_2(x_2))\left[\frac{\partial^2 U_{m(x)}(x_2)}{\partial\alpha^2} - \frac{\partial^2 U_{m(x)}(S_{g_{x_2}}(x))}{\partial\alpha^2}\right] dx_2\right| \leq \\
&\leq \int_{x^*}^{x} \frac{\overline{m}'^2 KL(x_2 - x^*)}{M'(x_2 - m(x_2))} dx_2 \leq \int_{x^*}^{x} \frac{\overline{m}'^2 KL(x_2 - x^*)}{M'(x_2 - x^*)(1 - m'(\tilde{x}(x_2)))} dx_2 \leq KC(x - x^*)
\end{aligned}$$

for some constant $C > 0$.[36] If x_1 is close enough to x^*, $C(x - x^*) < 1$ and hence T_{x_1} is a contraction. Thus g (and hence $\tilde{e}$) is uniquely determined in a neighborhood of x^*. By repeating the same argument we can extend the solution uniquely on any interval (x^*, x) where $\tilde{e}(x') > 0$ for all $x \in (x^*, x)$. □

Lemma 8. *If any CTQE $s^t(x)$ satisfying Conditions B2.2 and B2.3 is such that $W(x) > 0$ for some $x \in (x^*, x^{**})$, then there are sequences $(y_l)_{l\in\mathbb{N}_{\geq 0}}$, $(e_l)_{l\in\mathbb{N}_{\geq 1}}$ such*

[36] L is a Lipschitz constant for $\frac{\partial^2 u}{\partial\alpha^2}$, and $\overline{m}' = \sup_x m'(x)$. The argument still goes through if we only require $\frac{\partial^2 u}{\partial\alpha^2}$ to be Hölder continuous for some positive exponent.

that $(y_l)_l$ is strictly increasing in l; $W(y_l) = 0$ for all l; $W(y) > 0$ for all $y \in (y_l, y_{l+1})$ for any l; $d(y_{l+1}^-, y_l^+) = 0$; $d(y_l^+, y_l^-) = e_l$; $W'(y_l^+) > 0 > W'(y_l^-)$ for all $l \geq 1$;

$$e^{\frac{re_l}{2}} = 1 + m'(x)\frac{\frac{\partial V_{m(y_l)}(S(y_l^-))}{\partial \alpha}}{\frac{\partial u_{m(x)}(x)}{\partial x}}$$

for all l; $y_l \to x^{**}$ as $l \to \infty$; and $W(x) = 0$ for $x < y_0$.

Proof. Let (a, b) be the largest interval containing x such that $W(y) > 0$ for all $y \in (a, b)$, and denote $a = y_0$, $b = y_1$. That $d(x, x') = 0$ for all $x' < x \in (a, b)$ follows from the following argument. Take $x \in (a, b)$. Since $W(x) > 0$, there is $\tilde{x} \in (m(x), x)$ such that $u_{m(x)}(\tilde{x}) = U_{m(x)}(S(x)) > u_{m(x)}(x)$. Then $d(x, \tilde{x}) = 0$, as otherwise we would have $U_{m(x)}(S(\tilde{x})) > U_{m(x)}(S(x))$, contradicting the definition of a CTQE. Now suppose $d(x, x') > 0$ for some $x' < x \in (a, b)$. Construct a decreasing sequence $x = \tilde{x}_0 > \tilde{x}_1 > \tilde{x}_2 > \dots$ such that, for all n, $u_{m(\tilde{x}_n)}(\tilde{x}_{n+1}) = U_{m(\tilde{x}_n)}(S(\tilde{x}_n))$ and $d(\tilde{x}_n, \tilde{x}_{n+1}) = 0$ per the above argument. Let $\tilde{x}_\infty = \lim \tilde{x}_n$. If $\tilde{x}_\infty < x'$ we have a contradiction and the proof is done. If not, it follows by continuity that $u_{m(\tilde{x}_\infty)}(\tilde{x}_\infty) = U_{m(\tilde{x}_\infty)}(S(\tilde{x}_\infty))$, i.e., $W(\tilde{x}_\infty) = 0$, a contradiction.

Let $d(b^+, b^-) = e_1$. That e_1 is as required follows from the definition of CTQE. Note that Condition B2.3 implies that $e_1 > 0$. In addition, $W'(b^+) > 0$. To see this, in general let $\epsilon_l = \frac{\partial}{\partial \alpha} V_{m_l}(S(y_l^-))$ and $\epsilon_l' = \frac{\partial}{\partial \alpha} V_{m_l}(S(y_l^+))$, and suppose $\epsilon_l < 0$ as per Condition B2.3. Then

$$\begin{aligned}
\epsilon_l' &= \frac{\partial V_{m_l}(S(y_l^+))}{\partial \alpha} = e^{-re_l}\left(\frac{\partial V_{m_l}(S(y_l^-))}{\partial \alpha}\right) + \left(1 - e^{-re_l}\right)\left(-\frac{1}{m'(x)}\frac{\partial u_{m(x)}(x)}{\partial x}\right) \\
&= \frac{\frac{\partial V_{m_l}(S(y_l^-))}{\partial \alpha}\frac{1}{m'(x)}\frac{\partial u_{m(x)}(x)}{\partial x}}{-\frac{\partial V_{m_l}(S(y_l^-))}{\partial \alpha} - \frac{1}{m'(x)}\frac{\partial u_{m(x)}(x)}{\partial x}} = -\frac{\partial V_{m_l}(S(y_l^-))}{\partial \alpha}\frac{1}{1 + \frac{\frac{\partial V_{m_l}(S(y_l^-))}{\partial \alpha}}{\frac{1}{m'(x)}\frac{\partial u_{m(x)}(x)}{\partial x}}} = \\
&= -\epsilon_l \frac{1}{1 + \frac{\epsilon_l}{\frac{1}{m'(x)}\frac{\partial u_{m(x)}(x)}{\partial x}}} > 0.
\end{aligned}$$

This implies that there is $y_2 > y_1$ such that $W(y) > 0$ for $y \in (y_1, y_2)$, with $W(y_2) = 0$ and $W'(y_2^-) < 0$, and so on.

Next we argue that $y_l \to x^{**}$ as $l \to +\infty$. Suppose instead that $y_l \to y^* < x^{**}$, and let $m_l = m(y_l)$, $m^* = m(y^*)$. Since V is continuous, $V_{m(y^*)}(S(y^*)) = 0$. In addition,

$\frac{\partial}{\partial\alpha}V_{m^*}(S(y^{*-}))$ must equal zero.[37]

Suppose, then, that $\frac{\partial^2}{\partial\alpha^2}V_{m^*}(S(y^{*-}))\neq 0$. If this is positive, we have $V_{m(x)}(S(x))>0$ for all $x<y^*$ in a neighborhood of y^*, a contradiction.

If it is negative, we will obtain a contradiction by showing that $(\epsilon_l)_l$ cannot go fast enough to 0 for $(y_l)_l$ to converge. Note that $\epsilon_l<0<\epsilon_l'$ and $\epsilon_l+\epsilon_l'\in\mathcal{O}(\epsilon_l^2)$ since $\epsilon_l'=-\epsilon_l\frac{1}{1+\frac{-\epsilon_l}{-\frac{1}{m'(x)}\frac{\partial u_{m(x)}(x)}{\partial x}}}$, as shown above, and $-\frac{1}{m'(x)}\frac{\partial u_{m(x)}(x)}{\partial x}$ is bounded away from 0 in a neighborhood of y^*. Next, we argue that $\epsilon_{l+1}=\epsilon_l+\mathcal{O}(\epsilon_l^2)$.

Let $N(\alpha)=\frac{\partial V_\alpha(S(m^{-1}(\alpha)))}{\partial\alpha}$ and $M(\alpha)=\frac{\partial^2 V_\alpha(S(m^{-1}(\alpha)))}{\partial\alpha^2}$. We claim that M is left-continuous at m^*—indeed, for this to not be the case we would require $\sum_l e_l=+\infty$, which implies M is not bounded in a neighborhood of y^*, a contradiction. Thus, since $M(m^*)<0$, $M(\alpha)<0$ for all $\alpha<m^*$ in a neighborhood of m^*.

Let $\overline{M}_l=\max_{\alpha\in(\alpha_l,\alpha_{l+1})}-M(\alpha)$, $\underline{M}_l=\min_{\alpha\in(\alpha_l,\alpha_{l+1})}-M(\alpha)$. Note that $\overline{M}_l-\underline{M}_l\leq L(m_{l+1}-m_l)$ for some fixed constant L, i.e., $\overline{M}_l-\underline{M}_l\in\mathcal{O}(m_{l+1}-m_l)$.[38] Since $V_{m_l}=V_{m_{l+1}}=0$,

$$0=\int_{m_l}^{m_{l+1}}N(\alpha)=N(m_l^+)\,(m_{l+1}-m_l)+\int_{m_l}^{m_{l+1}}M(\alpha)\,(m_{l+1}-\alpha)$$

, where $N(m_l^+)=\epsilon_l'$. This implies

$$\underline{M}_l\frac{(m_{l+1}-m_l)^2}{2}\leq\epsilon_l'\,(m_{l+1}-m_l)\leq\overline{M}_l\frac{(m_{l+1}-m_l)^2}{2}$$
$$\underline{M}_l\frac{m_{l+1}-m_l}{2}\leq\epsilon_l'\leq\overline{M}_l\frac{m_{l+1}-m_l}{2}$$

Now $\epsilon_{l+1}=\epsilon_l'+\int_{m_l}^{m_{l+1}}M(\alpha)=\epsilon_l'-(m_{l+1}-m_l)\tilde{M}$, for some $\tilde{M}\in(\underline{M}_l,\overline{M}_l)$. From the above, $\epsilon_l'=\tilde{M}_l\frac{m_{l+1}-m_l}{2}+\mathcal{O}((m_{l+1}-m_l)^2)$. Then $\epsilon_{l+1}=-\epsilon_l'+\mathcal{O}((m_{l+1}-m_l)^2)$. In addition, it follows that $\mathcal{O}(\epsilon_l')=\mathcal{O}(m_{l+1}-m_l)$. Since $\epsilon_l'=-\epsilon_l+\mathcal{O}(\epsilon_l^2)$, we have that $\epsilon_{l+1}=\epsilon_l+\mathcal{O}(\epsilon_l^2)$, i.e., $(\epsilon_l)_l$ at most decays (or grows) at the rate of a harmonic series, whence $\sum_l\epsilon_l=\infty$. Since $\epsilon_l\in\mathcal{O}(m_{l+1}-m_l)$, we have $\sum_l(m_{l+1}-m_l)=\infty$ as well, which contradicts $y_l\to y^*$.

[37]Indeed, if this derivative is negative, it follows that $V_{m(x)}(S(x))\geq V_{m(x)}(S(y^{*-}))>0$ for all $x<y^*$ in a neighborhood of y^*, contradicting that $y_l\to y^*$. If it is positive, then $V_{m(x)}S(y^{*-})\leq -c(y^*-x)$ for x in such a neighborhood and some $c>0$. From the fact that $V_{m(x)}S(x)\geq 0$ and $0=V_{m^*}S(y^*)\geq V_{m(x)}S(y^*)$ it then follows that $E(S(y^*))-E(S(x))\geq c'>0$ for all $x<y^*$, which is impossible.

[38]This follows from the assumption that u is C^3.

Finally, suppose that $P = \frac{\partial^3}{\partial \alpha^3} V_{m^*}(S(y^{*-})) \neq 0$. If it is negative, we again have $V(\alpha) > 0$ for α in a left-neighborhood of m^*, a contradiction, so it must be positive; and, as before, $P(\alpha) = \frac{\partial^3}{\partial \alpha^3} V_\alpha(S(m^{-1}(\alpha)))$ must be left-continuous at m^*, i.e., it must be close to P for α close to m^*. Note that

$$\begin{aligned} 0 = \int_{m_l}^{m_{l+1}} N(\alpha) &= \epsilon_l' \left(m_{l+1} - m_l\right) + \int_{m_l}^{m_{l+1}} M(\alpha) \left(m_{l+1} - \alpha\right) \\ &= \epsilon_l' \left(m_{l+1} - m_l\right) + \frac{\left(m_{l+1} - m_l\right)^2}{2} M(\tilde{\alpha}_l) \end{aligned}$$

for some $\tilde{\alpha}_l \in (m_l, m_{l+1})$. This implies

$$\begin{aligned} \epsilon_l' = -M(\tilde{\alpha}_l) \frac{m_{l+1} - m_l}{2}, \quad \epsilon_{l+1}' &= -M(\tilde{\alpha}_{l+1}) \frac{m_{l+2} - m_{l+1}}{2} \\ \epsilon_{l+1} = \epsilon_l' + \int_{m_l}^{m_{l+1}} M(\alpha) &= \epsilon_l' + (m_{l+1} - m_l) M(\tilde{\tilde{\alpha}}_l) \\ \implies \epsilon_{l+1} = (m_{l+1} - m_l) \frac{2M(\tilde{\tilde{\alpha}}_l) - M(\tilde{\alpha}_l)}{2}, & \end{aligned}$$

where $\tilde{\alpha}_l$, $\tilde{\tilde{\alpha}}_l \in (m_l, m_{l+1})$. To finish the proof we will need to be more specific about the positions of these values in the interval (m_l, m_{l+1}). Due to the left-continuity of $P(\alpha)$, $M(\alpha)$ is roughly linear in each interval (m_l, m_{l+1}). This, coupled with the above, implies that $\tilde{\alpha}_l = \frac{2m_l + m_{l+1}}{3} + o(m_{l+1} - m_l)$ and $\tilde{\tilde{\alpha}}_l = \frac{m_l + m_{l+1}}{2} + o(m_{l+1} - m_l)$.

In addition, $M(m_l^+) - M(m_l^-) \in \mathcal{O}(\epsilon_l) \in o(m_l - m_{l-1})$. Then

$$\begin{aligned}
M(\tilde{\alpha}_{l+1}) - 2M(\tilde{\tilde{\alpha}}_l) + M(\tilde{\alpha}_l) &= M(\tilde{\alpha}_{l+1}) - M(\tilde{\tilde{\alpha}}_l) + M(\tilde{\alpha}_l) - M(\tilde{\tilde{\alpha}}_l) = \\
&= \mathcal{O}(\epsilon_{l+1}) + (P + o(l)) \left(\frac{2m_{l+1} + m_{l+2}}{3} - \frac{m_l + m_{l+1}}{2} + o(m_{l+2} - m_l) \right) + \\
&\quad + (P + o(l)) \left(\frac{2m_l + m_{l+1}}{3} - \frac{m_l + m_{l+1}}{2} + o(m_{l+2} - m_l) \right) \\
&= P \left(\frac{m_{l+2}}{3} - \frac{m_l}{3} \right) + o(m_{l+2} - m_l) > 0, \text{ so that} \\
\frac{-\epsilon_{l+1}}{-M(\tilde{\alpha}_{l+1})} &= \frac{m_{l+1} - m_l}{2} \frac{-2M(\tilde{\tilde{\alpha}}_l) + M(\tilde{\alpha}_l)}{-M(\tilde{\alpha}_{l+1})} = \\
&= \frac{m_{l+1} - m_l}{2} \left(1 + \frac{M(\tilde{\alpha}_{l+1}) - 2M(\tilde{\tilde{\alpha}}_l) + M(\tilde{\alpha}_l)}{-M(\tilde{\alpha}_{l+1})} \right) \geq \frac{m_{l+1} - m_l}{2} \\
\implies \frac{m_{l+2} - m_{l+1}}{2} &= \frac{\epsilon'_{l+1}}{-M(\tilde{\alpha}_{l+1})} = \frac{-\epsilon_{l+1} + \mathcal{O}(\epsilon_{l+1}^2)}{-M(\tilde{\alpha}_{l+1})} = \frac{-\epsilon_{l+1}}{-M(\tilde{\alpha}_{l+1})} (1 + \mathcal{O}(m_{l+1} - m_l)) \geq \\
&\geq \frac{m_{l+1} - m_l}{2} + \mathcal{O}((m_{l+1} - m_l)^2),
\end{aligned}$$

which as before implies $m_l \to \infty$, a contradiction. As a result, we can conclude that if $y_l \to y^* < x^{**}$ then m violates Condition B2.2 at y^*, which is what we wanted.

As for what happens to the left of a, if $W'(a^+) = 0$, we are done. If $W'(a^+) > 0$ instead, then by the same arguments as before $W'(a^-) < 0$ and there must be $y_{-1} < y_0$ s.t. $W(y) > 0$ for $y \in (y_{-1}, y_0)$ and $W(y_{-1}) = 0$, etc. If this sequence of intervals to the left of a is finite, re-index the sequence $(y_l)_l$ appropriately and we are done. If it is infinite, we would have an infinite decreasing sequence $y_0 > y_{-1} > \dots$ such that $y_l \to y_*$ as $l \to -\infty$. If $y_* > x^*$, we obtain a contradiction analogously to our previous arguments. If $y_* = x^*$ we still obtain a contradiction by a slightly different argument—near x^* and x^{**} it is not true that $\epsilon'_l = -\epsilon_l + \mathcal{O}(\epsilon_l^2)$, as $\frac{\partial u_{m(x)}(x)}{\partial x}$ approaches zero, but it is still true that $\epsilon'_l \leq -\epsilon_l$, so it is possible that ϵ_l is shrinking fast enough for $(y_l)_l$ to converge as $l \to +\infty$, but not as $l \to -\infty$.

Finally, the fact that $W(x) = 0$ for $x < y_0$ follows from the fact that, if this were false, there would be a sequence $(\tilde{y}_l)_l$ with $\tilde{y}_0 < x < y_0$ and $\tilde{y}_l \to x^{**}$ as $l \to +\infty$, which contradicts $W'(y_0^+) = 0$. □

Next, we construct a *canonical CTQE*, s_*, as follows. Given Condition B2.1, construct a smooth CTQE based on Lemma 6 for a maximal interval (x^*, x_0) where this is possible—either (x^*, x^{**}) if $e(x) > 0$ everywhere, or else up to a point x_0 where

$e(x_0) = 0$. In the latter case, to the right of x_0, Condition B2.2 guarantees that $W'''(x_0^+) > 0$, so $W(x) > 0$ in a right-neighborhood of x_0. We can then construct the solution based on sequences $(y_l)_l$, $(e_l)_l$ as described above, with Condition B2.3 guaranteeing that $e_l > 0$ and $y_{l+1} > y_l$ for all l.

Our last result on CTQEs is the following:

Lemma 9. *If s_* satisfies Conditions B2.1, B2.2 and B2.3, then it is the unique CTQE.*

Proof. Let $\hat{s}$ be another another CTQE. Suppose that $W(\alpha) > 0$ for some x (the case where $e(x) > 0$ for all $x \in (x^*, x^{**})$ is analogous), so s_* features a sequence $(y_l)_l$ as in Lemma 8. Let $\hat{y}_0 = \inf\{y \in (x^*, x^{**}) : W(y) > 0\}$. Note that s_* and $\hat{s}$ must be identical for x between x^* and $\min(y_0, \hat{y}_0)$ by Lemma 7.

If $y_0 < \hat{y}_0$, it follows that $\hat{V}_{m(y)}(\hat{S}(y)) = 0$ for y in a right-neighborhood of y_0, but at the same time $\hat{V}_{m(y)}(\hat{S}(y)) \geq \hat{V}_{m(y)}(\hat{S}(y_0)) = V_{m(y)}(S(y_0)) > 0$, a contradiction.

If $\hat{y}_0 < y_0$, there are two cases. First, suppose that $\hat{W}(y) > 0$ for all y in a right-neighborhood of $\hat{y}_0$. Then we can apply the previous argument at $\hat{y}_0$. Second, suppose $W(y) > 0$ and $W(y) = 0$ are both obtained for $y > \hat{y}_0$ arbitrarily close to $\hat{y}_0$. Then there must be an infinite collection of intervals $(a_n, b_n)_{n \in \mathbb{Z}_{\leq 0}}$ such that $b_n > a_n \geq b_{n-1}$ for all n; $a_n \xrightarrow[n \to -\infty]{} \hat{y}_0$; $W(y) > 0$ for all $y \in (a_n, b_n)$ and $W(y) = 0$ for all $y = a_n$ or $= b_n$. This case leads to a contradiction by arguments developed in Lemmas 6 and 8. Briefly, for $y > \hat{y}_0$ close enough to $\hat{y}_0$, $S(y)$ and $\hat{S}(y)$ are similar; $V''(m(y)) = 0$; and $e(y) \geq C > 0$, so $\hat{V}''(\alpha) \leq \tilde{C} < 0$ for any α such that $m(\alpha) \in (a_n, b_n)$. This implies that $\hat{s}$ in fact satisfies Conditions B2.2 and B2.3, which contradicts $a_n \xrightarrow[n \to -\infty]{} \hat{y}_0$ by Lemma 8.

Hence $y_0 = \hat{y}_0$. Then s_* and $\tilde{s}$ must be identical for $x > y_0$ because their behavior is uniquely pinned down by Lemma 8. □

Condition B requires m and s to be such that:

B2.1 $m'(x^*) > \frac{1}{2}$. (This implies that $e(x) > 0$ for x in a neighborhood of x^*.)[39]

B2.1' For some $C > 0$, $e(x) \geq C > 0$ for all $x \in (x^*, x^{**})$.

[39] As seen in Lemma 6, $\frac{1}{m'(x^*)} \frac{\partial^2}{\partial x^2} u|_{x^*,x^*} + 2\frac{\partial^2}{\partial \alpha \partial x} u|_{x^*,x^*} > 0$ is enough to guarantee that $e(x) > 0$ for x close to x^*—in fact, the condition guarantees $e(x) \geq \frac{C}{x - x^*}$ for some $C > 0$. In addition, Assumptions A1, A4 imply that $\frac{\partial^2}{\partial x^2} u|_{x^*,x^*} = -\frac{\partial^2}{\partial \alpha \partial x} u|_{x^*,x^*}$.

B2.2 There is no point $x \in [x^*, x^{**})$ for which $V_{m(x)}(S(x)) = \frac{\partial V_{m(x)}(S(x))}{\partial \alpha} = \frac{\partial^2 V_{m(x)}(S(x))}{\partial \alpha^2} = \frac{\partial^3 V_{m(x)}(S(x))}{\partial \alpha^3} = 0$.

B2.3 for all $x \in (x^*, x^{**})$ such that $V(x) = 0$ and $V(x') > 0$ for $x' < x$ arbitrarily close to x, $V'(x^-) < 0$. We refer to such points x as *vertex points*.

B2.1 and B2.1' are not generic, but hold in an open set; intuitively, B2.1 is true so long as the density f is not too high at x^*. B2.2 and B2.3 are generic conditions. For further details, see Appendix C.

Armed with all this, we will now show that all sequences of quasi-1-equilibria of the j-refined games $(s_j)_j$ converge to s_* a.e., i.e., $s_j^t(x) \xrightarrow[j\to\infty]{} s_*^t(x)\ \forall x, t$ where $s^t(x)$ is continuous. Take a fixed x_0 and let $p(x) = d(x, x_0)$. Then we have to show $p_j(x) \xrightarrow[j\to\infty]{} p(x)\ \forall x$ where p is continuous.

Suppose not, so there is a sequence $(s_j)_j$ and an x_1 for which p is continuous at x_1 but $p_j(x_1) \nrightarrow p(x_1)$. Take a subsequence $(s_l)_l$ such that p_l converges pointwise to some $\hat{p}$, and label the associated policy mapping $\hat{s}$.[40] $\hat{p} \neq p$ as, in particular, $\hat{p}(x_1) \neq p(x_1)$.

We will now prove the result simply by proving that $\hat{s}$ is a CTQE.

(i): Let $x \in (x^*, x^{**})$. For each j and x', $U_{m(x)}(S_j(s_j(x))) \geq U_{m(x)}(S_j(x'))$. If $s_j(x) \xrightarrow[j\to\infty]{} x$ we obtain $U_{m(x)}(\hat{S}(x))) \geq U_{m(x)}(\hat{S}(x'))$ by taking the limit. If not, and $s_{j'}(x) \xrightarrow[j'\to\infty]{} \tilde{x} < x$ for some subsequence, then $U_{m(x)}(\hat{S}(\tilde{x}))) \geq U_{m(x)}(\hat{S}(x'))$. But this also implies $\hat{p}(x) - \hat{p}(\tilde{x}) = 0$, so $U_{m(x)}(\hat{S}(x)) = U_{m(x)}(\hat{S}(\tilde{x}))) \geq U_{m(x)}(\hat{S}(x'))$.

In turn, the fact that $\hat{s}$ satisfies (i) means that Lemmas 6 and 7 apply to it.

(ii): Suppose $\hat{s}$ violates this condition at some $a \in (x^*, x^{**})$, i.e., $\hat{p}(a^+) - \hat{p}(a^-) = e^* > 0$. By an argument similar to Lemma 6, we have $\hat{W}'(a^+) > 0$ and hence $\hat{p}$ is constant on some interval (a, b).

Take $\epsilon > 0$ small, and let $(x_{jn})_n$ be the defining sequence of s_j for each j. By construction $(x_{jn})_n$, must have $je^* + j\hat{d}(a^-, a-\epsilon) + o(j)$ elements in $(a-\epsilon, a+\epsilon)$, and $j\hat{d}(a-\epsilon, a-2\epsilon) + o(j)$ elements in $(a-2\epsilon, a-\epsilon)$. In particular, given $\eta > 0$, for high enough j there must be an element $x_{jt} \in (a-2\epsilon, a-\epsilon)$ such that $x_{jt} - x_{j(t+1)} \geq \frac{1}{j\bar{e}(1+\eta)}$

[40] Use a diagonal argument to find a subsequence $(s_{l'})_{l'}$ such that $(p_{l'})_{l'}$ converges at all rational points. This guarantees convergence at all points except points of discontinuity of $\limsup_{l'\to\infty} p_{l'}$, which are countable because the function in question is increasing. Use another diagonal argument to get $(s_l)_l$ such that p_l also converges at all discontinuities of $\limsup_{l'\to\infty} p_{l'}$.

for $\overline{e} = \max_{x\in(a-2\epsilon,a-\epsilon)} e(x)$. Let $x_{jt'}$ be the right-most element of $(x_{jn})_n$ contained in $(a-\epsilon, a+\epsilon)$. The above implies $t' - t \geq je^* + j\hat{d}(a^-, a-\epsilon) + o(j)$.

Now, denoting $x_{jn} = x_n$, $m(x_{jn}) = m_n$ and $S_j(x_{jn}) = S(x_n)$, and exploiting the indifference conditions $U_{m_{n-1}}(S(x_n)) = U_{m_{n-1}}(S(x_{n+1})) = u_{m_{n-1}}(x_n)$ and $U_{m_n}(S(x_{n+1})) = U_{m_n}(S(x_{n+2})) = u_{m_n}(x_{n+1})$,

$$\begin{aligned}
V_{m_{n-1}}(S(x_{n+1})) - V_{m_n}(S(x_{n+1})) &= U_{m_{n-1}}(S(x_{n+1})) - U_{m_n}(S(x_{n+1})) - u_{m_{n-1}}(x_{n-1}) + u_{m_n}(m_n) \\
&= -(u_{m_{n-1}}(x_{n-1}) - u_{m_{n-1}}(x_n)) + (u_{m_n}(m_n) - u_{m_n}(x_{n+1})) \\
m'(\hat{x}_n)(x_{n-1} - x_n)\frac{\partial}{\partial\alpha}V_{\tilde{\alpha}_n}(S(x_{n+1})) &= -(x_{n-1} - x_n)\frac{\partial}{\partial x}u_{m_{n-1}}(\tilde{x}_n) + (x_n - x_{n+1})\frac{\partial}{\partial x}u_{m_n}(\tilde{x}_{n+1}) \\
(x_{n-1} - x_n) &= (x_n - x_{n+1})\frac{-\frac{\partial}{\partial x}u_{m_n}(\tilde{x}_{n+1})}{-m'(\hat{x}_n)\frac{\partial}{\partial\alpha}V_{\tilde{\alpha}_n}(S(x_{n+1})) - \frac{\partial}{\partial x}u_{m_{n-1}}(\tilde{x}_n)}
\end{aligned}$$

for some $\tilde{\alpha}_n \in (m_n, m_{n-1})$, $\tilde{x}_n, \hat{x}_n \in (x_n, x_{n-1})$, $\tilde{x}_{n+1} \in (x_{n+1}, x_n)$. In addition

$$\begin{aligned}
\frac{\partial}{\partial\alpha}V_\alpha(S(x_n)) &= e^{-\frac{r}{j}}\frac{\partial}{\partial\alpha}V_\alpha(S(x_{n+1})) + (1 - e^{-\frac{r}{j}})\left(\frac{\partial u_\alpha(x_n)}{\partial\alpha} - \frac{\partial u_\alpha(m^{-1}(\alpha))}{\partial\alpha} - \frac{\frac{\partial u_\alpha(m^{-1}(\alpha))}{\partial x}}{m'(m^{-1}(\alpha))}\right) \\
\frac{\partial}{\partial\alpha}V_\alpha(S(x_n)) &= e^{-\frac{rk}{j}}\frac{\partial}{\partial\alpha}V_\alpha(S(x_{n+k})) + (1 - e^{-\frac{rk}{j}})\left(\mathcal{O}(\epsilon) - \frac{\frac{\partial}{\partial x}u_\alpha(m^{-1}(\alpha))}{m'(m^{-1}(\alpha))}\right)
\end{aligned}$$

for $\alpha \in (m(a-2\epsilon), m(a+\epsilon))$, $x_n, x_{n+k} \in (a-2\epsilon, a+\epsilon)$.

Then, for $n \in \{t, \ldots, t'\}$,

$$\begin{aligned}
x_{j(n-1)} - x_{jn} &\geq (x_{jn} - x_{j(n+1)})\frac{-\frac{\partial}{\partial x}u_{m(a)}(a) - K\epsilon}{-\tilde{m}'\frac{\partial}{\partial\alpha}V_{\tilde{\alpha}_n}(S_j(x_{j(n+1)})) - \frac{\partial}{\partial x}u_{m(a)}(a) + K'\epsilon} \geq \\
&\geq (x_{jn} - x_{j(n+1)})\frac{-\frac{\partial}{\partial x}u_{m(a)}(a) - K\epsilon}{-e^{-\frac{r(t-n)}{j}}\tilde{m}'\frac{\partial}{\partial\alpha}V_{\tilde{\alpha}_n}(S_j(x_{j(t+1)})) - e^{-\frac{r(t-n)}{j}}\frac{\partial}{\partial x}u_{m(a)}(a) + K''\epsilon} \\
&\geq (x_{jn} - x_{j(n+1)})\frac{-\frac{\partial}{\partial x}u_{m(a)}(a) - K\epsilon}{G(j) - e^{-\frac{r(t-n)}{j}}\frac{\partial}{\partial x}u_{m(a)}(a) + K'''\epsilon} = \\
&= (x_{jn} - x_{j(n+1)})\frac{1 - K''''\epsilon}{e^{-\frac{r(t-n)}{j}} + \tilde{G}(j) + K'''''\epsilon}
\end{aligned}$$

for some function $G(j)$ such that $G(j) \xrightarrow[j\to\infty]{} 0$, as

$$\left|\frac{\partial}{\partial\alpha}V_{\tilde{\alpha}_n}(S_j(x_{j(t+1)}))\right| \xrightarrow[j\to\infty]{} \left|\frac{\partial}{\partial\alpha}V_{\alpha^*}(\hat{S}(x^*))\right| \leq \left|\frac{\partial}{\partial\alpha}V_{m(x^*)}(\hat{S}(x^*))\right| + C\epsilon = C\epsilon$$

for some $\alpha^* \in [m(a-2\epsilon), m(a+\epsilon)]$, $x^* \in [a-2\epsilon, a-\epsilon]$ and $C > 0$.[41] Then

$$x_{j(n-1)} - x_{jn} \geq (x_{jt} - x_{j(t+1)}) \prod_{k=0}^{t-n} \frac{1-K'\epsilon}{e^{-\frac{rk}{j}} + \tilde{G}(j) + K''''\epsilon}$$

$$\implies x_{jt'} - x_{j(t'+1)} \geq \frac{1}{j\overline{e^*}(1+\eta)} \prod_{k=0}^{je^*-1} \frac{1-K'\epsilon}{e^{-\frac{rk}{j}} + \tilde{G}(j) + K''''\epsilon}.$$

If we take ϵ small enough that $1 - K'\epsilon > 0$, the right-hand side grows to infinity as $j \to \infty$. In particular, for j high enough, $x_{jt'} - x_{j(t'+1)} > 3\epsilon$, a contradiction.

(iii): This follows from an argument analogous to the one used for (ii). Briefly, if (iii) is violated at x_0 and $\hat{p}(x_0^+) - \hat{p}(x_0^-)$ is higher than the value required by (iii), then $x_{jt'} - x_{j(t'+1)} \xrightarrow[j\to\infty]{} \infty$, a contradiction. If $\hat{p}(x_0^+) - \hat{p}(x_0^-)$ is lower than the value required by (iii), then it can be shown that $j(x_{jn} - x_{j(n+1)}) \xrightarrow[j\to\infty]{} 0$ for all n such that $x_{j(n+1)} \geq a$, which implies that the number of elements of $(x_{jn})_j$ in $(a-\epsilon, a+\epsilon)$ grows faster than j, a contradiction.

□

C Genericity of Conditions on m (For Online Publication)

In this Section we show that the conditions B1, B2.2 and B2.3 imposed on the function m are 'generic'. We employ two different notions of genericity. On the one hand, we show that these conditions hold on an open and dense set (or, at least, a residual set) within the function space with a natural metric. In addition, we are able to show that some of these conditions hold on a *prevalent* set, a notion introduced in Hunt, Sauer and Yorke (1992) that generalizes the measure-theoretic notion of 'almost everywhere' to infinite-dimensional spaces where a natural analog of the Lebesgue measure is not

[41] If necessary, take a convergent subsequence so that $(\tilde{\alpha}_{n_j})_j$ and $x_{j(t_j+1)}$ converge for this argument.

available.

Claim 1. *Consider the set of functions*

$$X_1 = \{m : [-1,1] \longrightarrow [-1,1] : m \in C^1, m \text{ weakly increasing } \}$$

with the norm $||m|| = \max(||m||_\infty, ||m'||_\infty)$. *The subset* $Y_1 \subseteq X_1$ *satisfying B1 is open and dense,*[42] *and also prevalent in the sense of Hunt et al. (1992).*

Proof. We first show Y_1 is open. Let $m_0 \in Y_1$; $x_1^* < \ldots < x_N^*$ be the fixed points of m_0; $\alpha_i = m_0'(x_i^*)$ for $i = 1, \ldots, N$; $\epsilon > 0$ and $\nu > 0$ such that $|m_0'(y) - 1| \geq \nu$ for $y \in I_i = (x_i^* - \epsilon, x_i^* + \epsilon)$ for any i; $\upsilon > 0$ such that $|m(y) - y| \geq \upsilon$ for $y \notin I_i$ for any i; $\eta = \min(\epsilon, \upsilon, \nu)$; and $m_1 \in B(m_0, \eta)$. Then $m_1(y) = y$ implies $|m_0(y) - y| < \eta \leq \upsilon$, so $y \in I_i$ for some i, so $|m_1'(y) - 1| \geq |m_0'(y) - 1| - |m_1'(y) - m_0'(y)| > \nu - \nu = 0$. This shows that $m'(y) \neq 1$ at any fixed point y of m_1. Moreover, by construction either $m_0'(y) > 1$ for all $y \in I_i$ and $m_1'(y) > 1$ for all $y \in I_i$ as well, or the reverse inequalities hold, whence m_1 can have at most one fixed point in I_i for each i, and the set of fixed points is finite.

Next, we show Y_1 is dense. Let $m_0 \in X_1$ and $\epsilon > 0$. We want to show that there is $m_1 \in B(m_0, \epsilon) \cap Y_1$. Since m_0' is continuous in $[-1,1]$, it is uniformly continuous, so we can take $\nu > 0$ such that if $|y - y'| < \nu$ then $|m_0'(y) - m_0'(y')| < \frac{\epsilon}{4}$. Partition $[-1,1]$ into intervals $I_1, I_2, \ldots, I_J$ as follows: $I_j = [y_{j-1}, y_j)$, where $y_j = -1 + j\nu$, for $j < J$, and $I_J = [y_{J-1}, 1]$. For each j, if $m_0'(y_{j-1}) \geq 1$, let $m_2'(y) = m_0'(y) + \frac{\epsilon}{4}$ for all $y \in I_j$ (which implies $m_2'(y) > 1$ for $y \in I_j$); otherwise let $m_2'(y) = m_0'(y) - \frac{\epsilon}{4}$ for all $y \in I_j$ (so $m_2'(y) < 1$ for $y \in I_j$), and then define m_2 by integrating m_2', with $m_2(-1) = m_0(-1)$. By construction m_2 has at most one fixed point in each interval I_j and $m_2' \neq 1$ at such points. Moreover, $||m_2' - m_0'|| \leq \frac{\epsilon}{4}$ and $||m_2 - m_0|| \leq 2\frac{\epsilon}{4} = \frac{\epsilon}{2}$. If $m_2(y_j) \neq y_j$ for all y_j, we can construct a 'smoothed out' version of m_2', which we'll call m_1', that is in $B(m_0, \epsilon) \cap Y_1$. If $m_2(y_j) = y_j$ for some j, and $m_2'(y) > 1$ for $y \in I_j \cup I_{j+1}$ or $m_2'(y) < 1$ for $y \in I_j \cup I_{j+1}$, this is not a problem. If $m_2(y_j) = y_j$ for some j and $m_2'(y) > 1$ for $y \in I_j$, $m_2'(y) < 1$ for $y \in I_{j+1}$, we can construct a smooth m_1' such that $m_1'(y_j) = 1$, $m_1'(y) > 1$ to the left of y_j and < 1 to the right, and $m_1(y_j) < m_2(y_j) = y_j$. The remaining case is analogous.

For the last claim, note that, if a C^1 function m defined on a compact interval has $m' \neq 1$ at all its fixed points, it automatically has a finite number of them. Consider

[42] The statement is also true within the space of C^3 functions, taken with the appropriate norm.

the translation $X_1 - v$, where v is the identity function. Then $m \in Y_1$ iff $m - v$ has no points where $(m-v)(y) = (m-v)'(y) = 0$. Finally, the fact that $Y_1 - v$ is prevalent in $X_1 - v$ follows from Proposition 3 in Hunt et al. (1992). □

Claim 2. *B2.1 holds in an open set Y_2 within*

$$X_2 = \big\{ m : [x^*, x^{**}] \longrightarrow [x^*, x^{**}] : m \in C^2, m \text{ weakly increasing}, \\ m(x^*) = x^*,\, m(x^{**}) = x^{**},\, m(x) < x\, \forall x \in (x^*, x^{**}) \big\}$$

taken with the norm $||m|| = \max(||m||_\infty, ||m'||_\infty, ||m''||_\infty)$.[43]

Proof. Trivial. □

Claim 3. *B2.1' holds in an open set Y_4 within X_2.*

Proof. This amounts to showing that $\tilde{e}$ is continuous in m'', and it follows from an argument similar to the proof of the uniqueness of $\tilde{e}$ from Lemma 7. □

Claim 4. *Let*

$$X_3 = \big\{ m \in Y_2 : m \in C^3,\, m(x^*) = x^*,\, m(x^{**}) = x^{**},\, m(x) < x\, \forall x \in (x^*, x^{**}), \\ m \text{ strictly increasing}, m \text{ satisfies } B2.1 \big\}$$

taken with the norm $||m|| = \max(||m||_\infty, ||m'||_\infty, ||m''||_\infty, ||m'''||_\infty)$.

For each $y \in (x^*, x^{**})$, the set $Y_3(y) \subseteq X_3$ of functions m for which B2.2 and B2.3 hold in $[x^*, y]$ is open and dense.

Proof. We proceed in two steps. First, we show that the set $Y_5(y) \subseteq X_3$ for which B2.2 holds in $[x^*, y]$ is open and dense. Second, we show that the set $Y_3(y)$ is open and dense within $Y_5(y)$.

To show that $Y_5(y)$ is open, take $m \in Y_5(y)$ and suppose there is a sequence $(m_n)_n$ such that $m_n \notin Y_5(y)$ for all n but $m_n \to m$. For each m_n we can construct a CTQE s_n (possibly not unique) by finding a convergent sequence of discrete-time equilibria $(s_{nj})_j$ for $\delta = e^{-rj}$ with $j \to \infty$, as in Proposition 10. Using a diagonal argument, we can find a convergent subsequence of $(s_n)_n$, which by continuity must converge to a CTQE for m, $\hat{s}$. WLOG assume $(s_n)_n \to \hat{s}$. We will need the following Lemma:

[43] Again, this is also true within the space of C^3 functions.

Lemma 10. *If s_* satisfies Conditions B.1 and B2.2, then it is the unique CTQE. Moreover, s_* has a finite number of vertex points in $[x^*, y]$ for any $y < x^{**}$.*

Proof. Briefly, if s_* has an infinite number of vertex points in $[x^*, y]$, they must accummulate at some $y^* \in (x^*, y]$, which must satisfy $V_{m(y^*)}(S(y^*)) = \frac{\partial V_{m(y^*)}(S(y^*))}{\partial \alpha} = 0$. If $\frac{\partial^2 V_{m(y^*)}(S(y^*))}{\partial \alpha^2} > 0$ we obtain $V > 0$ in a neighborhood of y^*, a contradiction. If $\frac{\partial^2 V_{m(y^*)}(S(y^*))}{\partial \alpha^2} < 0$ this guarantees Condition B2.3 in a neighborhood of y^*, which means the vertex points near y^* must be part of a single sequence, contradicting Lemma 8.

Suppose that there are infinitely many vertex points on a left-neighborhood of y^* (the other case is analogous). Similar arguments apply if $\frac{\partial^3 V_{m(y^*)}(S(y^*))}{\partial \alpha^3} < 0$ or $\frac{\partial^3 V_{m(y^*)}(S(y^*))}{\partial \alpha^3} > 0$, respectively.

As for the uniqueness of s_*, the proof in Lemma 9 can be extended to this case. □

From this we conclude that $\hat{s} = s_*$. Letting W_n be the value function for s_n, we then have $W_n \to W$. It can be shown in addition that, at every y that is not a vertex point of s_*, $W_n'(y) \to W'(y)$, $W_n''(y) \to W''(y)$ and $W_n'''(y) \to W'''(y)$, by using Lemmas 7 and 10.

Next, we show that $Y_5(y)$ is dense. Take $m \in X_3$ and $\epsilon > 0$. Consider $\hat{m}$ given by: $\hat{m(x^*)} = x^*$, $\hat{m}'(x^*) = m'(x^*)$, $\hat{m}''(x^*) = m''(x^*)$ and $\hat{m}'''(x) = m'''(x) + \eta(x)$, where $|\eta(x)| \leq \epsilon$ will be defined as 0 except where we specify otherwise. We will argue that, by picking η correctly, we can find a $\hat{m} \in Y_5(y)$ that is close to m.

Apply the following algorithm. Take $\nu > 0$ small and $N > 0$ large. Let $\eta_0 \equiv 0$ and $m_0 \equiv m$. Let

$$x_0 = \inf \left\{ x \in (x^*, y] : \max \left(\left| V_{m_0(x)}(S_0(x)) \right|, \left| \frac{\partial}{\partial \alpha} V_{m_0(x)}(S_0(x)) \right|, \right. \right.$$
$$\left. \left. \left| \frac{\partial^2}{\partial \alpha^2} V_{m_0(x)}(S_0(x)) \right|, \left| \frac{\partial^3}{\partial \alpha^3} V_{m_0(x)}(S_0(x)) \right| \right) \leq \frac{\epsilon}{N} \right\},$$

where $S_0(x)$ is a the policy path starting at x for a CTQE given median voter function m_0.[44] Let $\alpha_0 = m(x_0)$. Define $\eta_1(x) = -\epsilon$ for $x \in [x_0, x_0')$ and $\eta_1(x) = 0$ for all other x, with x_0' taken so that $m_1(x_0') = m_1(x_0) + \nu$. Next, let x_1 be the infimum of $x \in (x_0 + \nu, y]$ for which $|V_{m_1(x)}(S_1(x))|$, $|\frac{\partial}{\partial \alpha} V_{m_1(x)}(S_1(x))|$, $|\frac{\partial^2}{\partial \alpha^2} V_{m_1(x)}(S_1(x))|$ and

[44] Note that, for ϵ small enough, $x_0 > x^*$ since $e(x) \geq C > 0$ in a neighborhood of x^*, which implies $\frac{\partial^2}{\partial \alpha^2} V_{m_0(x)}(S_0(x)) \geq C' > 0$.

$|\frac{\partial^3}{\partial\alpha^3}V_{m_1(x)}(S_1(x))| \leq \frac{\epsilon}{N}$, and define $\alpha_1 = m(x_1)$ and $\eta_2(x) = -\epsilon$ for $x \in [x_1, x_1')$ and $\eta_2(x) = \eta_1(x)$ for all other x, with x_1' taken so that $m_2(x_1') = m_2(x_1) + \nu$. Define x_k, α_k, η_{k+1}, m_{k+1} for $k = 2, 3, \ldots$ in the same fashion until $x_K = +\infty$ for some K.[45] Let $\tilde{m} = m_K$.

We argue that, if ν and N are taken to be small and large enough, respectively, $\tilde{m}$ satisfies Condition B2.2. To explain why, we will need the following

Remark. A function $f : [a, b] \to \mathbb{R}$ is uniformly continuous iff there is an increasing function $h : [0, +\infty) \to [0, +\infty)$ such that $h(0) = 0$, h is continuous at 0, and $|f(x) - f(y)| \leq h(|x - y|)$ for all $x, y \in [a, b]$. We say a function h satisfying these properties is a *bounding function*.

Now note that, for any k and any $x < x'$ such that $m(x) = a$, $m(x') = a'$ satisfy $a, a' \in [\alpha_k, \alpha_k + \nu]$, we have

$$\begin{aligned}
&\left|\frac{\partial^3}{\partial\alpha^3}V_a(\tilde{S}(x)) - \frac{\partial^3}{\partial\alpha^3}V_{a'}(\tilde{S}(x'))\right| \leq \left|\frac{\partial^3}{\partial\alpha^3}U_a(\tilde{S}(x)) - \frac{\partial^3}{\partial\alpha^3}U_a(\tilde{S}(x'))\right| \\
&+ \left|\frac{\partial^3}{\partial\alpha^3}U_a(\tilde{S}(x')) - \frac{\partial^3}{\partial\alpha^3}U_{a'}(\tilde{S}(x'))\right| + \left|-\frac{d^3}{d\alpha^3}u_a(\tilde{m}^{-1}(a)) + \frac{d^3}{d\alpha^3}u_{a'}(\tilde{m}^{-1}(a'))\right| \\
&\leq \left|\frac{\partial^3}{\partial\alpha^3}U_a(\tilde{S}(x)) - \frac{\partial^3}{\partial\alpha^3}U_a(\tilde{S}(x'))\right| + h_1(a' - a) + h_2(a' - a) \\
&\leq K\frac{\epsilon}{N} + K'(a' - a) + h_1(a' - a) + h_2(a' - a) \leq K\frac{\epsilon}{N} + h_3(a' - a)
\end{aligned}$$

where h_1, h_2, h_3 are bounding functions, $K, K' > 0$, and K, h_1, h_2, h_3 are independent of ν and N.

The bound $|\frac{d^3}{d\alpha^3}u_a(\tilde{m}^{-1}(a)) + \frac{d^3}{d\alpha^3}u_{a'}(\tilde{m}^{-1}(a'))| \leq h_2(|a - a'|)$ uses the uniform continuity of $\frac{d^3}{d\alpha^3}u_{\tilde{a}}(\tilde{m}^{-1}(\tilde{a}))$, which follows from the fact that u is C^3 and $\tilde{m}$ is C^3 on $[a, a'] \subseteq [\alpha_k, \alpha_k + \nu]$. The bound $|\frac{\partial^3}{\partial\alpha^3}U_a(\tilde{S}(x')) - \frac{\partial^3}{\partial\alpha^3}U_{a'}(\tilde{S}(x'))| \leq h_1(|a - a'|)$ uses the uniform continuity of $\frac{\partial^3}{\partial\alpha^3}U$, and the fact that the mapping $x \mapsto \max(x, 0)$ is Lipschitz. The first bound is the trickiest, and is based on the idea that, if $\frac{\partial V}{\partial\alpha}$ and $\frac{\partial^2 V}{\partial\alpha^2}$ are low, then $\tilde{s}^t(x)$ changes relatively quickly as a function of t, so $\tilde{S}(x)$ and $\tilde{S}(x')$

[45] This must happen for a finite K, as $\alpha_k - \alpha_{k-1} \geq \nu > 0$ for all k.

are similar. Formally:

$$\left|\frac{\partial^3}{\partial\alpha^3}U_a(\tilde{S}(x)) - \frac{\partial^3}{\partial\alpha^3}U_a(\tilde{S}(x'))\right| \leq \left(1 - e^{-r\tilde{d}(x',x)}\right) 2\max_{a,y}\frac{\partial^3}{\partial\alpha^3}u_\alpha(y)$$

$$1 - e^{-r\tilde{d}(x',x)} \leq r\tilde{d}(x',x) = r\left(\sum_{\substack{\tilde{y}_l\in(x,x')\\ \text{vertex point}}}\tilde{e}_l + \int_x^{x'}\tilde{e}(y)dy\right) \leq r\frac{2}{r}\sum_{\tilde{y}_l\in(x,x')\text{v.p.}}(e^{\frac{r\tilde{e}_l}{2}} - 1) + r\overline{e}(x'-x) \leq$$

$$\leq 2\sum_{\tilde{y}_l\in(x,x')\text{v.p.}}\frac{|\max\tilde{m}'|}{\min_{x\in[x_0,y]}\left|\frac{\partial u_{m(x)}(x)}{\partial x}\right|}\left|\frac{\partial}{\partial\alpha}V_{a_l}(\tilde{S}(y_l^-))\right| + r\overline{e}(x'-x) \leq$$

$$\leq K''\left(\left|\frac{\partial}{\partial\alpha}V_a(\tilde{S}(x))\right| + \int_a^{a'}\left|\frac{\partial^2}{\partial\alpha^2}V_{\tilde{a}}(\tilde{S}(n(\tilde{a})))\right|d\tilde{a}\right) + r\overline{e}(x'-x) \leq K''\frac{\epsilon}{N} + K'''(a'-a).$$

Pick ν and N so that $K\frac{\epsilon}{N} + h_3(\nu) \leq \frac{\epsilon}{2}$ and $N \geq 4$. Now $\tilde{m}$ satisfies Condition B2.2 because, for $a \in [\alpha_k, \alpha_k + \nu)$,

$$\left|\frac{\partial^3}{\partial\alpha^3}V_a(\tilde{S}(x))\right| \geq \left|\frac{\partial^3}{\partial\alpha^3}V_{\alpha_k}(\tilde{S}(x_k))\right| - \left|\frac{\partial^3}{\partial\alpha^3}V_{\alpha_k}(\tilde{S}(x_k)) - \frac{\partial^3}{\partial\alpha^3}V_a(\tilde{S}(x))\right| \geq \epsilon - \frac{\epsilon}{N} - \frac{\epsilon}{2} \geq \frac{\epsilon}{4} > 0.$$

On the other hand, if $a \notin [\alpha_k, \alpha_k + \nu)$ for any k, then $|\frac{\partial^i}{\partial\alpha^i}V| > \frac{\epsilon}{N}$ for some $i = 0, 1, 2, 3$ by construction.

The only remaining issue is that $\tilde{m}$ is not C^3 because η_K is not continuous at α_k and $\alpha_k + \nu$ for $k = 0, 1, \ldots, K-1$. However, it is easy to construct a continuous η close to η_K that fixes this problem.[46]

Next, we argue that $Y_3(y)$ is open. As shown in Proposition 10, Conditions B2.1, B2.2 and B2.3 taken together imply that the equilibrium path $s^t(x)$ will be given by either a smooth path with $e(x) > 0$ for all $x \in (x^*, y]$ or a smooth path up to some y_0 followed by a finite sequence of jumps and stops with stops at $y_1, y_2, \ldots, y_l$. It is enough to show that e' is continuous in m''', which follows from the arguments in Proposition 10, and that y_i is continuous in m''' for $i = 1, 2, \ldots, l$, which is elementary (in fact, y_i is continuous in m').

Finally we argue that $Y_3(y)$ is dense. Take $m \in X_3$ and $\epsilon > 0$. Because $Y_5(y)$

[46]WLOG, take $a = \alpha_k$. If $V_a(S(x^-)) > 0$, it is easy to perturb η_K to make it continuous at a without violating Condition B2.2. If not, but $V_{\tilde{a}}(S(\tilde{x}^-)) > 0$ for $\tilde{a} < a$ arbitrarily close to a, we can perturb η_K at one such $\tilde{a}$ instead. If $V_{\tilde{a}}(S(\tilde{x}^-)) = 0$ for all $\tilde{a} < a$ close to a, but $\frac{\partial^2}{\partial\alpha^2}V$ is nonzero close to a, we can do the same argument. If $\frac{\partial^2}{\partial\alpha^2}V$ is also zero in an interval to the left of a, then Condition B2.2 would be violated for $\tilde{a} < a$, a contradiction.

is dense, there is $\hat{m} \in B(m, \epsilon) \cap Y_5(y)$. Because $Y_5(y)$ is open, there is $\epsilon' > 0$ such that $B(\hat{m}, \epsilon') \subseteq B(m, \epsilon) \cap Y_5(y)$. Next, we claim that there is $\hat{\hat{m}} \in B(\hat{m}, \epsilon') \cap Y_3(y)$, which completes the proof. This can be shown by construction. If $\hat{e}(x) > 0$ for all $x \in [x^*, y]$, we are done. If not, $\hat{s}$ induces a policy path that is continuous up to some y_0 and then features a sequence of jumps and stops with stops at $y_1, y_2, \ldots, y_l$. (By the arguments in Proposition 10, this sequence cannot be infinite.) If Condition B2.3 holds at $y_1, \ldots, y_l$, we are done. If not, suppose WLOG that it first fails at y_l. $\hat{m}$ can be perturbed near y_l to obtain $\hat{m}_2 \in B(\hat{m}, \epsilon')$ that satisfies Condition B2.3 at $y_1, \ldots, y_l$. Similarly, if $\hat{m}_2$ first fails Condition B2.3 at some $y_{l'} > y_l$, we can construct $\hat{m}_3 \in B(\hat{m}, \epsilon')$, a perturbation of $\hat{m}_2$ near $y_{l'}$, that satisfies Condition B2.3 up to $y_{l'}$. If this process stops in a finite number of steps, we are done. If not, let $\hat{m}_\infty$ be the pointwise limit of $(\hat{m}_k)_k$. $\hat{m}_\infty$ must feature an infinite sequence of vertex points $y_1 < y_2 < \ldots$ with $y_l \xrightarrow[l \to +\infty]{} y^* \leq y$, but, as $\hat{m}_\infty \in Y_5(y)$, $\hat{m}_\infty$ satisfies Condition B2.2, leading to a contradiction. □

Corollary 3. *The set $Y_3 \subseteq X_3$ of functions for which B2.2 and B2.3 hold in $[x^*, x^{**}]$ is a residual set.*

Claim 5. *In the case of quadratic utility, the set of functions m for which B2.2 holds for $x \in (x^*, x^* + d)$ is prevalent.*

Proof. The result follows from Theorem 3 in Hunt et al. (1992). Following their notation, take $M = \{y \in \mathbb{R}^4 : \frac{2y_3 - 1}{y_1 - y_2} + \frac{y_4}{y_3} = 0\}$ and $Z = \{y \in \mathbb{R}^5 : (y_1, y_2, y_3, y_4) \in M \text{ and } \frac{2y_4(y_1 - y_2) - 2y_3(1 - y_3)}{(y_1 - y_2)^2} + \frac{y_5 y_3 - y_4^2}{y_3^2} = 0\}$. We need to check that M is a manifold of codimension 1, and that the projection $\pi : M \longrightarrow \mathbb{R}$ given by $y \longmapsto y_1$ is a submersion; both follow from the Implicit Function Theorem. Finally, we need to check that Z is a zero set in $M \times \mathbb{R}$, which can also be shown using the Implicit Function Theorem. Theorem 3 from Hunt et al. (1992) then implies that the set of functions m for which there is an x such that

$$e(x) = \frac{2m'(x) - 1}{x - m(x)} + \frac{m''(x)}{m'(x)} = 0 =$$
$$= \frac{2m''(x)(x - m(x)) - 2m'(x)(1 - m'(x))}{(x - m(x))^2} + \frac{m'''(x)m'(x) - m''(x)^2}{m'(x)^2} = e'(x)$$

is shy, i.e., its complement is prevalent. □

I conjecture that B2.2 and B2.3 hold in a prevalent set even for general utility functions, but this is hard to prove.

D Other Equilibria (For Online Publication)

The discrete time model in Section 2 may admit solutions other than 1-equilibria. We discuss two possible types here: k-equilibria, which are given by k interleaved sequences (as opposed to one sequence for 1-equilibria), and continuous equilibria. Formally

Definition 6. Let s be a MPE on $[x^*, x^{**}]$. s is a k-equilibrium if there is a sequence $(x_n)_{n\in\mathbb{Z}}$ such that $x_{n+1} < x_n$ for all n, $x_n \to x^{**}$ as $n \to -\infty$, $x_n \to x^*$ as $n \to \infty$, and $s(x) = x_{n+k}$ if $x \in [x_n, x_{n-1})$.[47]

A continuous equilibrium is one where s is continuous.[48]

Figure 5 shows a 2-equilibrium (5b) compared to a 1-equilibrium (5a).

Although these cases don't exhaust the set of possible solutions, studying them sheds light on the general behavior of non-1-equilibria. The general thrust of the results is that these solutions are not as well-behaved as 1-equilibria, and their existence is not robust in any sense analogous to Proposition 5, which is why the paper does not focus on them.

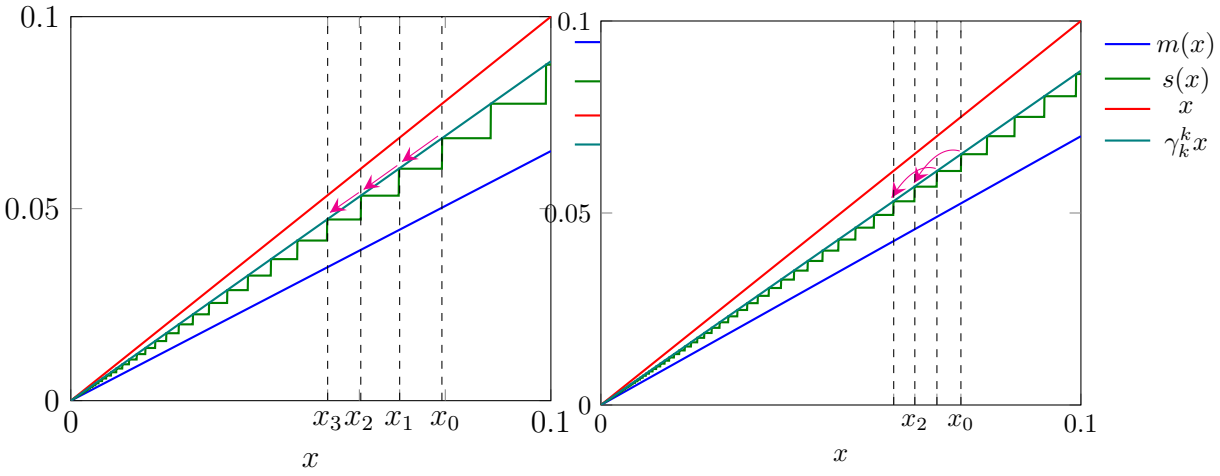
When $m(x)$ is linear[49] and u is quadratic, we can find k-equilibria for all k and a continuous equilibrium, as illustrated in Figures 5a and 5b:

Proposition 11. *Let $u_a(x) = C - (a - x)^2$, and f be such that $m(x) = \alpha x$ for $x \in [-e, e]$, where $\alpha < 1$ and $e \leq d$. Assume $\delta \geq \frac{2}{3}$ and $\alpha \geq \frac{1}{2}$. Then, for each k and $\underline{x} < e$, there is a k-equilibrium s_k^* such that $x_0 = \underline{x}$, given by $x_n = \gamma_k^n \underline{x}$, where $0 < \gamma_k < 1$. There is also a continuous equilibrium s_∞^* given by $s_\infty^*(x) = \gamma_\infty x$. Moreover, γ_k^k is decreasing in k and $\gamma_k^k \to \gamma_\infty$.*

[47] If the basin of attraction is of the form $[x^*, 1]$ then the sequence would be of the form $(x_n)_{n\in\mathbb{N}}$.

[48] Note that this is a different concept from the smooth continuous-time equilibria considered in Proposition 10—there, discrete-time 1-equilibria were given by discontinuous policy functions, but they converged to a continuous limit as $\delta \to 1$.

[49] We can construct densities f such that $m(x) = \alpha x$ for $x \in [-d, d]$. For example, for a continuous f symmetric around the steady state $x = 0$, take $f(y) = 1 - \frac{1-\alpha}{d} y$ for $y \in [0, d]$ and $f(y) = \alpha + (1 - \alpha)(2\alpha^2 + 1) - \frac{(1-\alpha)(2\alpha^2+1)}{d} y$ thereafter.



(a) 1-equilibrium

(b) 2-equilibrium

Figure 5: Equilibria for $m(x) = 0.7x$, $\delta = 0.7$

Proof of Proposition 11. Given $k \geq 1$, assume a k-equilibrium of the form $s(x_n) = \gamma_k^k x_n$. Since $s(x_n) = x_{n+k}$ but $s(x_n - \epsilon) = x_{n+k+1}$, $m(x_n)$ must be indifferent between choosing x_{n+k} and x_{n+k+1}. This implies

$$
\begin{aligned}
-\sum \delta^t (\alpha x_n - x_{n+(t+1)k})^2 &= -\sum \delta^t (\alpha x_n - x_{n+(t+1)k+1})^2 \\
\sum \delta^t (\alpha - \gamma^{(t+1)k})^2 &= \sum \delta^t (\alpha - \gamma^{(t+1)k+1})^2 \\
\frac{\alpha^2}{1-\delta} - 2\frac{\alpha\gamma^k}{1-\delta\gamma^k} + \frac{\gamma^{2k}}{1-\delta\gamma^{2k}} &= \frac{\alpha^2}{1-\delta} - 2\frac{\alpha\gamma^{k+1}}{1-\delta\gamma^k} + \frac{\gamma^{2k+2}}{1-\delta\gamma^{2k}} \\
\frac{\gamma^k(1+\gamma)}{1-\delta\gamma^{2k}} &= \frac{2\alpha}{1-\delta\gamma^k}
\end{aligned}
$$

We now argue that there is a unique solution $0 < \gamma_k < 1$. Let $V(\gamma) = \frac{\gamma^k(1+\gamma)}{1-\delta\gamma^{2k}} - \frac{2\alpha}{1-\delta\gamma^k}$. Since $V(0) = -2\alpha < 0$ and $V(1) = \frac{2(1-\alpha)}{1-\delta} > 0$, by continuity, there is at least one solution between 0 and 1. Besides

$$
\begin{aligned}
V(\gamma) \propto W(\gamma) &= (\gamma^k + \gamma^{k+1})(1-\delta\gamma^k) - 2\alpha(1-\delta\gamma^{2k}) \\
&= -2\alpha + \gamma^k + \gamma^{k+1} + (2\alpha - 1)\delta\gamma^{2k} - \delta\gamma^{2k+1}.
\end{aligned}
$$

Since the highest order term has a negative coefficient, $W(M) < 0$ for large M; hence there is also a solution larger than 1. But, by Descartes' rule of signs, W has at most two positive roots. Hence there is a unique solution $0 < \gamma_k < 1$.

We can also see that γ_k^k is decreasing in k. Let $\tilde{W}(\gamma) = W(\gamma^{\frac{1}{k}})$. Then $\tilde{W}(\gamma, k) = \gamma(1+\gamma^{\frac{1}{k}})(1-\delta\gamma) - 2\alpha(1-\delta\gamma^2)$ is increasing in k for fixed $0 < \gamma < 1$. Since W is increasing around the solution, this means that the $\tilde{\gamma}_k$ that sets $\tilde{W}(\tilde{\gamma}_k, k) = 0$ must be decreasing in k, i.e., $W(\tilde{\gamma}_k^{\frac{1}{k}}, k) = 0$ where $\tilde{\gamma}_k$ is decreasing. Setting $\gamma_k = \tilde{\gamma}^{\frac{1}{k}}$, we conclude that γ_k^k is decreasing.

We now show that the constructed s_k supports an MPE. By increasing differences, if $m(x_n)$ is indifferent between $S(x_{n+k})$ and $S(x_{n+k+1})$, all $m(x) > m(x_n)$ strictly prefer $S(x_{n+k})$ between the two, and $m(x) < m(x_n)$ strictly prefer $S(x_{n+k+1})$. Hence, $m(x_n)$ prefers x_{n+k} to all x_r with $r > n+k+1$ or $r < n+k$.

Next, we show that $m(x_n)$ prefers x_{n+k} to other policies x not belonging to the sequence. We do this in two steps. First, we argue that $\gamma^{k+1} > \alpha$, which implies $x_{n+k+1} > m(x_n)$. Second, we note that this yields our result by a similar argument as in Proposition 5. For the first part, note that

$$\begin{aligned} & \gamma^{k+1} > \alpha \\ \iff & (\gamma^k + \gamma^{k+1})(1-\delta\gamma^k) < 2\gamma^{k+1}(1-\delta\gamma^{2k}) \\ \iff & (1-\gamma) < \delta\left(\gamma^k(1-\gamma^{k+1}) + \gamma^{k+1}(1-\gamma^k)\right) \\ \iff & 1 < \delta\left(\gamma^k + 2\gamma^{k+1} + \ldots + 2\gamma^{2k}\right) \end{aligned}$$

Consider two cases. If $k = 1$, then the required inequality is $1 < \delta(\gamma + 2\gamma^2)$. Since $\delta \geq \frac{2}{3}$, this holds as long as $\gamma \geq \frac{2}{3}$, since $1 < \frac{28}{27}$. Next, we check that $W(\frac{2}{3}) < 0$, which guarantees that $\gamma > \frac{2}{3}$. Clearly the worst case is when δ is minimal, so take $\delta = \frac{2}{3}$. Then $W(\frac{2}{3}) = \frac{10}{9}\frac{5}{9} - 2\alpha\frac{19}{27} < 0$ whenever $\alpha > \frac{25}{57}$. If $k \geq 2$, then it is enough to satisfy $1 < \frac{2}{3}(\gamma^k + 4\gamma^{2k})$, which is true whenever $\gamma^k \geq \frac{1}{2}$. We then check that $W(\frac{1}{2}) < 0$. Again, the worst case is when δ is minimal, and we can bound $\gamma^{k+1} \leq \gamma^k$, so $W(\frac{1}{2}) \leq \frac{2}{3} - 2\alpha\frac{5}{6} < 0$ whenever $\alpha > \frac{2}{5}$.

Finally, we construct a continuous equilibrium. In general, s must solve

$$\begin{aligned} s(x) &= \arg\max_y \sum_{t=0}^{\infty} \delta^t \left(C - (m(x) - s^t(y))^2\right) \\ \implies 0 &= \sum_{t=0}^{\infty} \delta^t \left(-2(m(x) - s^t(y)) \prod_{i=0}^{t-1} s'(s^i(y))\right) \end{aligned}$$

if s is smooth. Since $m(x) = \alpha x$, we look for a solution of the form $s_\infty(x) = \gamma x$:

$$\sum_{t=0}^{\infty} \delta^t \left((\alpha - \gamma^{t+1}) \prod_{i=0}^{t-1} \gamma \right) = \sum_{t=0}^{\infty} \delta^t \left((\alpha - \gamma^{t+1}) \gamma^t \right) = 0,$$

whence $\frac{\alpha}{1-\delta\gamma} = \frac{\gamma}{1-\delta\gamma^2}$. By similar arguments as before, there is a unique solution $0 < \gamma_\infty < 1$ to this equation, and $\gamma_k^k \to \gamma_\infty$ because the equations pinning down γ_k^k converge to this one. Finally, $\frac{\partial U_{m(x)}(S(y))}{\partial y}|_{y=y_0} > 0$ for $y_0 < s(x)$ by increasing differences, since $\frac{\partial U_{m(s^{-1}(y_0))}(S(y))}{\partial y}|_{y=y_0} = 0$; similarly, $\frac{\partial U_{m(x)}(S(y))}{\partial y}|_{y=y_0} < 0$ for $y_0 > s(x)$. Hence $y = s(x)$ maximizes $U_{m(x)}(S(y))$. □

In the general case, however, k-equilibria for $k > 1$ and continuous equilibria are not well-behaved, in the following sense. Suppose that a k-equilibrium s_k exists in a right-neighborhood of a stable steady state, $[x^*, x^* + \epsilon)$ (note that even existence in a neighborhood of x^* is not guaranteed in general). s_k can then be extended to $[x^*, x^{**}]$ but it may lose its interesting properties beyond $x^* + \epsilon$, i.e., it may turn into a 1-equilibrium to the right of a certain $y \in (x^* + \epsilon, x^{**})$. Similarly, when a continuous equilibrium is extended away from x^*, discontinuities may appear; and whether this happens depends on arbitrarily small details of m.

We can see this in an example. Suppose that $u_a(x) = C - (a - x)^2$, $x^* = 0$, $\delta > \frac{2}{3}$, $\alpha > 0.5$, and $\tilde{m}(x) = \alpha x + \frac{\alpha}{4} \max(c - |x - x'|, 0)$, where c is small. We are in the linear case, except $\tilde{m}$ has a small "bump" around x'. Let s be a 2-equilibrium for $m(x) = \alpha x$ such that $x_0 = x'$, and let $\tilde{s}$ be a 2-equilibrium for $\tilde{m}$ such that $\tilde{x}_n = x_n$ for $n > 0$. As $m(x_0)$ and $\tilde{m}(\tilde{x}_0)$ must both be indifferent between $S(x_2)$ and $S(x_3)$, $m(x_0) = \tilde{m}(\tilde{x}_0)$, but $\tilde{m}(x_0) > m(x_0)$, so we must have $\tilde{x}_0 < x_0$. Meanwhile $\tilde{x}_1 = x_1$. But $m(\tilde{x}_{-2})$, being indifferent between $\tilde{S}(\tilde{x}_0)$ and $S(x_1)$, must be *lower* than $m(x_{-2})$ because $\tilde{x}_0$ being lower makes the former path more attractive than $S(x_0)$, so $\tilde{x}_{-2}$ is lower. On the other hand $\tilde{x}_{-3} > x_{-3}$ because it is defined by indifference between S_1 and $\tilde{S}_0$ (more attractive than S_0). Continuing in this fashion, the subsequence $(\tilde{x}_0, \tilde{x}_{-2}, \tilde{x}_{-4}, \ldots)$ is lower than $(x_0, x_{-2}, \ldots)$, and the opposite is true for the odd elements. It can be shown that eventually $\tilde{x}_{2l} < \tilde{x}_{2l+1}$ for some l, i.e., the even subsequence becomes so attractive that a voter $m(\tilde{x}_{2l+1})$, supposed to be indifferent between $\tilde{S}_{2l+3}$ and $\tilde{S}_{2l+4}$, instead prefers $\tilde{S}_{2l+2}$ to both, so no one votes for x_{2l+1} and s becomes a 1-equilibrium beyond that point.

This same dynamic makes k-equilibria for $k > 1$ unstable in general. To see this,

let $W((s_0, s_1, \ldots)) = (1-\delta)\sum_{t=0}^{\infty} \delta^t s_t^2$, and characterize a k-equilibrium recursively as follows, using the definition of $m(x_n)$ being indifferent between S_{n+k} and S_{n+k+1}:

$$W_n = W(S(x_n)) = (1-\delta)\left[m^{-1}\left(\frac{1}{2}\frac{W_{n+k}-W_{n+k+1}}{E_{n+k}-E_{n+k+1}}\right)\right]^2 + \delta W_{n+k}$$
$$E_n = E(S(x_n)) = (1-\delta)m^{-1}\left(\frac{1}{2}\frac{W_{n+k}-W_{n+k+1}}{E_{n+k}-E_{n+k+1}}\right) + \delta E_{n+k}$$

Taking $Y_n = (E_n, \ldots, E_{n+k+1}, W_{n+1}, \ldots, W_{n+k+1})$ as the state variable of the recursion, its linearization around an equilibrium is given by $Y_n = M_n Y_{n+1}$, where

$$M_n = \begin{pmatrix} 0 & \ldots & 0 & A & B & 0 & \ldots & 0 & C & D \\ 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 2x & 0 & \ldots & 0 & -2\delta x & 0 & 0 & \ldots & 0 & \delta \\ 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{matrix} \} 1 \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} k \\ \} 1 \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} k-1 \end{matrix}$$

where $x = x_{n+1}$; $B = \frac{\partial E_n}{\partial E_{n+k+1}}$, $D = \frac{\partial E_n}{\partial W_{n+k+1}}$ and so on. Now note that

$$\det(M_n) = -\delta B - 2\delta x_{n+1} D = \delta(1-\delta)\frac{x_{n+1}-m(x_n)}{m'(x_n)(E_{n+k}-E_{n+k+1})}$$
$$\begin{aligned} \det(M_n M_{n-1}\ldots M_{n-k+1}) &\geq \delta^k(1-\delta)^k\left[\min_{0\leq l\leq k-1}\left(\frac{x_{n-l+1}-m(x_{n-l})}{m'(x_{n-l})}\right)\right]^k \frac{1}{\prod_{l=0}^{k-1}(E_{n-l+k}-E_{n-l+k+1})} \\ &\geq \delta^k(1-\delta)^k\left[\min_{0\leq l\leq k-1}\left(\frac{x_{n-l+1}-m(x_{n-l})}{m'(x_{n-l})}\right)\right]^k \frac{k^k}{(E_{n+1}-E_{n+k+1})^k} \\ &\geq \delta^k\left[\min_{0\leq l\leq k-1}\left(\frac{x_{n-l+1}-m(x_{n-l})}{m'(x_{n-l})}\right)\right]^k \frac{k^k}{(x_{n+1}-E_{n+k+1})^k} \end{aligned}$$

Now, if δ is close to 1 and the equilibrium is approximately smooth in the sense of Proposition 10, then $\frac{x-m(x)}{m'(x)(x-E(S(s(x))))} \approx 1$ (see Appendix B) and $\det(M_n \ldots M_{n-k+1}) \approx k^k$. In particular, there must be an eigenvalue of absolute value at least $k^{\frac{k}{2k+1}} > 1$.

Hence any deviation from an equilibrium resulting from a local perturbation of m which adds a nonzero component to a generalized eigenvector of this eigenvalue (in the Jordan form decomposition of the matrix) will grow exponentially.

In similar fashion, if we consider a continuous equilibrium in the example given above, the bump would generate a discontinuity around $s^{-1}(x_0)$. More generally

Proposition 12. *Assume $u_a(x) = C - (a-x)^2$. Let $s : [x^*, x^{**}] \to [x^*, x^{**}]$ be a continuous equilibrium for a given m and parameters δ, C. Let $x_0 \in (x^*, x^{**})$. A perturbation $\tilde{m}$ of m is an increasing function $\tilde{m} = m + \rho\kappa$ where $\kappa : [x^*, x^{**}] \to [x^*, x^{**}]$ has support $(x_0 - \epsilon, x_0 + \epsilon)$. For each $\tilde{m}$, let $\tilde{s}$ be an equilibrium under $\tilde{m}$ such that $\tilde{s}|_{[x^*, x_0-\epsilon)} = s|_{[x^*, x_0-\epsilon)}$.*

Suppose m is C^∞ in $(s^l(x_0 - \epsilon), s^l(x_0 + \epsilon))$ for all $l \in \mathbb{Z}$. Then, if κ is C^k but its $(k+1)$th derivative has a discontinuity somewhere in $(x_0 - \epsilon, x_0 + \epsilon)$, $\tilde{s}$ has a discontinuity in $[x^*, s^{-k-1}(x_0 + \epsilon)]$ for arbitrarily small ρ.

Proof. Let $\frac{1}{2}\frac{\partial W(y)}{\partial E(y)} = L^{-1}(y)$. Then

$$\begin{aligned} s(x) &= \arg\max_y -m(x)^2 + 2m(x)E(y) - W(y) \Rightarrow m(x) = \frac{1}{2}\frac{\partial W(y)}{\partial E(y)}|_{s(x)} \\ s(x) &= \left(\frac{1}{2}\frac{\partial W(y)}{\partial E(y)}\right)^{-1}(m(x)) = L(m(x)) \\ (W, E)(x) &= \big((1-\delta)x^2 + \delta W(L(m(x))), (1-\delta)x + \delta E(L(m(x)))\big) \end{aligned}$$

In particular, $W(E)$ must be a strictly convex function so that s is surjective, and it must have no kinks, i.e., s must be strictly increasing (if s is locally constant at x, it will be discontinuous at $s^{-1}(x)$ as long as $s(y) > m(y)$ in this area), so L^{-1} and L are strictly increasing and well-defined. Moreover, it can be shown that s must be C^∞ where m is C^∞.[50] If κ is discontinuous at x, so is $\tilde{m}$, and so is $\tilde{s}$, for any ρ.

Now suppose E is C^{l+1} around $s(s(x))$ but s has a $(l+1)$-kink at $s(x)$, i.e., it is C^{l+1} in $(s(x) - \eta, s(x)) \cup (s(x), s(x) + \eta)$ but only C^l in $(s(x) - \eta, s(x) + \eta)$. Then s'

[50] The argument is similar to the rest of the proof: if (W, E) is not C^∞, take an y where it has a l-kink with l minimal, and find a kink of lower degree at $s^{-1}(y)$.

has a l-kink at $s(x)$. Since

$$\frac{\partial W(y)}{\partial E(y)} = \frac{\frac{\partial W(y)}{\partial y}}{\frac{\partial E(y)}{\partial y}} = \frac{2(1-\delta)y + \delta W'(s(y))s'(y)}{(1-\delta) + \delta E'(s(y))s'(y)}$$
$$= \frac{W'(s(y))}{E'(s(y))} + \frac{(1-\delta)(2y - \frac{W'(s(y))}{E'(s(y))})}{1-\delta+\delta E'(s(y))s'(y)} = 2m(y) + \frac{(1-\delta)(2y-2m(y))}{1-\delta+\delta E'(s(y))s'(y)},$$

L^{-1} has a l-kink at $s(x)$; L has a l-kink at $m(x)$; and s has a l-kink at x.

Now we argue that, if κ has a $(k+1)$-kink at x, then $\tilde{m}$ has a $(k+1)$-kink at x; $\tilde{s}$ has a $(k+1)$-kink at x; $\tilde{s}$ has a k-kink at $\tilde{s}^{-1}(x)$; ...and $\tilde{s}$ has a discontinuity at $\tilde{s}^{-k-1}(x)$. This follows from the same logic as above. In particular, since $E'(s(y)) = (1-\delta)(1+\delta s'(s(y)) + \ldots)$, if s has a $(l+1)$-kink at y, it is C^{l+1} at $s(y)$, so E' is C^l at $s(y)$, and our previous argument applies. □

I conjecture that, even if κ is C^∞, perturbations generically lead to discontinuities devolving into 1-equilibria for arbitrarily small ρ, but in that case the number of steps until the discontinuity will depend on ρ.

E Additional Results (For Online Publication)

E.1 Non-Markov Equilibria

The restriction to Markov equilibria may appear restrictive—after all, allowing strategies to condition only on the current policy prevents agents from doling out history-dependent rewards and punishments in ways that might be plausible in some applications. To address this concern, this Section discusses the set of non-Markov equilibria of the game.

We make two important points. First, if non-Markov equilibria are allowed, many other outcomes are possible; in fact, an anything goes-type result can be obtained, so not much can be said about expected behavior if we take SPE as our solution concept. Second, there are substantively interesting perturbations of the game which happen to rule out all non-Markov equilibria. This suggests that Markov equilibria may indeed be the most sensible in this setting.

For simplicity, we will restrict our analysis in the following ways. First, as elsewhere, we restrict our analysis to the right side of the basin of attraction of a sta-

ble steady state; that is, we consider an interval $[x^*, x^{**}]$ such that $m(x^*) = x^*$, $m(x^{**}) = x^{**}$ and $m(x) < x$ for all $x \in (x^*, x^{**})$. Second, we assume the Median Voter Theorem as a primitive, i.e., we assume that, given a current policy x, $m(x)$ gets to choose her most preferred policy.[51]

For each $x \in [x^*, x^{**}]$, let $z(x)$ be the unique policy such that $u_{m(x)}(z(x)) = u_{m(x)}(x)$ and $z(x) \neq x$. Then we have the following

Proposition 13. *Suppose that u and δ are such that $(1-\delta)u_{m(x)}(x') + \delta u_{m(x)}(z(x')) \leq u_{m(x)}(x)$ for all x and $x' \in (z(x), x)$.*

Then, for every weakly decreasing path $(y_t)_{t\in\mathbb{N}_0} \subseteq [x^*, x^{**})$ such that $U_{m(y_t)}((y_{t+1}, y_{t+2}, \ldots)) \geq u_{m(y_t)}(y_t)$ for all t, there is an SPE with policy path $(y_t)_t$.

Proof. We will construct a suitable successor function $s(x, T)$, where T is a payoff-irrelevant variable that will summarize the effect of the history on current behavior. T can take on the values 0, 1 or 2. s is defined as follows:

- $s(y_t, 0) = y_{t+1}$ for all t, and $s(x, 0) = x$ for all $x \notin (y_t)_t$;
- $s(x, 1) = x$;
- $s(x, 2) = z(x)$.

As for T, assume that $T_0 = 0$ and $(T_\tau)_\tau$ behaves according to the following mapping H:

- If $T = 0$ and $x = y_t$, $x' = y_{t+1}$ for some t, then $H(x, T, x') = 0$;
- else, if $x' \notin (z(x), x)$, $H(x, T, x') = 1$;
- else $H(x, T, x') = 2$.

In other words, in state 0, the policy follows the intended equilibrium path, $(y_t)_t$, and T_τ remains equal to zero. In state 1, the policy path is constant and T_τ remains equal to 1. In state 2, the current decisionmaker, $m(x)$, chooses the lowest policy that she weakly prefers to x, and the state then changes to 1. In all cases, deviations to myopically attractive policies (that is, policies that $m(x)$ strictly prefers to x) are

[51] In the model of political power from Section 5, the MVT is always true, so this assumption is just for brevity. In the main model, the MVT is only guaranteed to hold within $I(x^*)$, but well-behaved equilibria—in particular, 1-equilibria—satisfy the MVT everywhere, so the case of equilibria satisfying the MVT is still the most enlightening one to study.

punished by switching to state 2, while deviations to myopically unattractive policies are punished by switching to state 1.

Let us verify that this is an SPE. If $(x_\tau, T_\tau) = (y_t, 0)$ for some t, then $m(x_\tau)$'s utility from following the equilibrium path is $U_{m(y_t)}(S(y_{t+1}))$. On the other hand, if she chooses a policy $x' \notin (z(x_\tau), x_\tau)$, then $T_{\tau+1} = 1$ and the continuation is given by $x_s = x'$ for all $s > \tau$, yielding utility $u_{m(y_t)}(x') \leq u_{m(y_t)}(y_t)$. If she chooses a policy $x' \in (z(x_\tau), x_\tau)$ different from y_{t+1}, then $T_{\tau+1} = 2$ and the continuation is given by $x_{\tau+1} = x'$, $x_s = z(x')$ for all $s > \tau + 1$, yielding utility $(1-\delta)u_{m(x)}(x') + \delta u_{m(x)}(z(x'))$.

If $(x_\tau, T_\tau) = (x, T)$ with $T = 1$ or $T = 2$, then $m(x)$'s utility from following the equilibrium path is $u_{m(x)}(x)$. If she chooses a policy $x' \notin (z(x), x)$, she gets utility $u_{m(x)}(x') < u_{m(x)}(x)$. If she chooses a policy $x' \in (z(x), x)$, she gets utility $(1-\delta)u_{m(x)}(x') + \delta u_{m(x)}(z(x'))$.

To make this strategy profile an SPE, it is sufficient that $(1-\delta)u_{m(x)}(x') + \delta u_{m(x)}(z(x')) \leq u_{m(x)}(x)$ for all x and all $x' \in (z(x), x)$, as we have assumed. □

Note that the condition $U_{m(y_t)}((y_{t+1}, y_{t+2}, \ldots)) \geq u_{m(y_t)}(y_t)$ is clearly necessary: otherwise, $m(y_t)$ could unilaterally deviate to staying at policy y_t forever. What this result shows is that, aside from this common-sense restriction, anything goes.[52] In particular, for each $x \in [x^*, x^{**}]$ there is an SPE with policy path constantly equal to x, so any policy can become an (intrinsic) steady state in the right SPE.

There are several reasons to think this is not too concerning:

- Non-Markovian behavior must be supported by non-Markovian behavior happening arbitrarily close to x^*. Equivalently, if we think it is reasonable to assume Markov behavior in a very small interval around x^*, then the set of equilibria collapses to, in fact, a single equilibrium that is Markov. Formally:

  Lemma 11. *Let $\epsilon > 0$ and $\tilde{s} : [x^*, x^* + \epsilon] \to [x^*, x^* + \epsilon]$ such that $\tilde{s}(x) < x$ for all x. Let s, s' be two SPEs on $[x^*, x^{**}]$ such that $s(x, h) = s'(x, h) = \tilde{s}(x)$ for all $x \in [x^*, x^* + \epsilon]$, and assume that s and s' obey the following tie-breaking rule: if the set $\arg\max_{y \leq x} U_{m(x)} S(y, h)$ has multiple elements, then $s(y)$ is the highest element of the set. Then $s \equiv s'$ and s is an MPE, i.e., $s(x, h)$ is independent of h.*

[52] The joint condition imposed on u and δ, namely, that $(1-\delta)u_{m(x)}(x') + \delta u_{m(x)}(z(x')) \leq u_{m(x)}(x)$ for all x and $x' \in (z(x), x)$, is in general satisfied if u is well-behaved and δ is high enough. For example, in the quadratic-linear case given by $u_\alpha(x) = C - (\alpha - x)^2$, $x^* = 0$ and $m(x) = \alpha x$, the condition holds if $2\alpha\delta \geq 1$, i.e., if $\alpha > \frac{1}{2}$ and δ is high enough.

Proof of Lemma 11. The intuition behind this result is a simple unraveling argument: suppose two equilibria coincide up to some point $x^* + \epsilon$. Then, for y slightly above $x^* + \epsilon$, $I(y)$ will be choosing between successors in $[x^*, x^* + \epsilon]$, which have the same continuation in both equilibria, so the same choice will be made.

Formally, let

$$A = \{x \in [x^*, x^{**}] : \exists \hat{s} \text{ s.t. } \forall h, \forall y \in [x^*, x],\ s(y,h) = s'(y,h) = \hat{s}(y)\},$$

and $x_0 = \sup(A)$. By assumption, $x_0 \geq x^* + \epsilon$. Suppose $x_0 < x^{**}$.

There are two cases. First, suppose $x_0 \notin A$. Then the same proof as in Proposition 2 can be used to show that $u_{m(x_0)}(x_0) < \max_{y \in [x^*, x_0)} U_{m(x_0)}\hat{S}(y)$, whence $s(x_0, h),\ s'(x_0, h) < x_0$ for all h. Moreover, the tiebreaking rule means that $s(x_0, h) = s'(x_0, h) = \max(\arg\max_{y \in [x^*, x_0)} U_{m(x_0)}\hat{S}(y))$ for all h, whence $\hat{s}$ can be extended to x_0, a contradiction.

Second, suppose $x_0 \in A$. Then there is a sequence $(x_n)_n$ such that $x_n \to x_0$ and $x_n > x_0\ \forall n$ such that, for each n, $s(x_n, h_n) > x_0$ for some history h_n, as otherwise the tiebreaking rule would guarantee that $s(x_n, h) \equiv s'(x_n, h)$ are independent of h.

Note that, for each n, $m(x_n)$ always has the option of jumping to any policy $z \in [x^*, x_0]$, and that the continuation would be the history-independent path $\hat{S}(z)$; hence, the optimality of s requires that $U_{m(x_n)}(S(s(x_n, h_n), h_n)) \geq U_{m(x_n)}(\hat{S}(z))$.

For each n, label the continuation path starting at $s(x_n, h_n)$ as $S_n = (s_0^n, s_1^n, \ldots)$, where $s_0^n = s(x_n, h_n)$. Let $s_{k_n}^n$ be the first policy in this path that is in $[x^*, x_0]$. Note that $(s_{k_n}^n, s_{k_n+1}^n, \ldots) = \hat{S}(s_{k_n}^n)$ is history-independent, and $m(x_n)$ always has the option of jumping directly to policy $s_{k_n}^n$, whence

$$U_{m(x_n)}(\hat{S}(s_{k_n}^n)) \leq U_{m(x_n)}(S_n) \leq \frac{1-\delta}{1-\delta^{k_n+1}} \sum_{t=0}^{k_n-1} u_{m(x_n)}(s_t^n) \leq u_{m(x_n)}(x_0).$$

In turn, as $m(x_n)$ can also choose any $z \in [x^*, x_0]$ and get a history-independent path $\hat{S}(z)$, it must be that

$$\max_{z \in [x^*, x_0]} U_{m(x_n)}(\hat{S}(z)) \leq U_{m(x_n)}(S_n) \leq u_{m(x_n)}(x_0).$$

By continuity, it follows that $U_{m(x_0)}(S(x_0)) \leq u_{m(x_0)}(x_0)$, which again leads to a contradiciton as in Proposition 2. (If $k_n = \infty$ for arbitrarily high n, then we obtain $U_{m(x_0)}(S(x_0)) \leq u_{m(x_0)}(x_0)$ directly.) □

- If we consider an approximation of the problem with a discrete policy space, so that agents are restricted to a finite set of policies X, then for any choice of $X = \{x_1, \ldots, x_N\} \subseteq [-1, 1]$ there is generically a unique subgame perfect equilibrium, which is Markov.[53] Hence, if we are interested in equilibria that can be obtained as limits of discrete-policy space equilibria,[54] we need only to study Markov equilibria.

- If we consider a variant of the game with a finite number of periods, so that $t = 0, 1, \ldots, T$, then for each T the game has a unique equilibrium s_T, which is Markov in (x, t). A limit of such equilibria as we take $T \to \infty$ may not be Markov in x exclusively, but Propositions 2 and 3 can still be extended to this case, so intrinsic steady states are also ruled out under this refinement.

- The MPEs we construct in the main text, which involve a slow drift towards a myopically stable state, are strictly preferred by all pivotal decisionmakers to an SPE where the policy never changes. In other words, the fall down the slippery slope is *desired* by agents. More precisely, assume an initial policy x_0, and consider an MPE s where the equilibrium path $(x_t)_t$ is such that $x_t \xrightarrow[t\to\infty]{} x^*$, and an SPE $\tilde{s}$ where the policy remains constant at x_0 on the equilibrium path. By Lemma 2, there is some α_0 such that all agents to the left of α_0 strictly prefer $(x_t)_t$ to x_0; since $m(x_0)$ has this preference, $\alpha_0 \geq m(x_0)$. Consequently, all agents in $[x^*, m(x_0)]$—that is, all agents who might be pivotal decisionmakers on the equilibrium path—have the same preference. In other words, it is not the case that, by ruling out non-Markov equilibria, we are removing from consideration desirable equilibria that just require mutually desirable coordination on the part of the players to arise.

[53] This can be shown by proving Proposition 1 in the discrete version of the problem, and then applying backward induction, as done in Acemoglu et al. (2015); the equilibrium is unique so long as there are no indifferences.

[54] Formally, denoting s_X to be the unique equilibrium for policy space s_X, an equilibrium s is a limit of discrete-policy space equilibria if there is a sequence $(X_n)_n$ such that $\max_{y\in[-1,1]} d(X_n, y) \xrightarrow[n\to\infty]{} 0$, i.e., the sets X_n are progressively finer, and $s_{X_n}(x_n) \to s(x)$ for any sequence $(x_n)_n$ s.t. $x_n \in X_n \forall n$ and $x_n \to x$.

E.2 Positive Entry and Exit Costs

The main model features perfectly free entry and exit. This assumption adds a lot of tractability, but if it were crucial for the results the reader might worry about its affecting the relevance of the model, as it is rarely going to be exactly true in a descriptive sense.

This subsection considers a variant of the game with positive entry costs (the case with exit costs is similar; see the end of the subsection). In other words, every time an outsider chooses to join the organization, she must pay a cost $c > 0$.

A full extension of the results in the paper to this case is difficult because the introduction of entry costs adds intertemporal concerns to entry and exit decisions—that is, agents considering to enter the organization now need to think about what the policy will be several periods from now, whether they will want to leave later, etc., in deciding whether paying the entry cost will be worth it. Relatedly, agents with identical preferences might end up behaving differently depending on their current status: if the club's policy, x, is stable over time, and an agent α's flow payoff from membership, $u_\alpha(x)$, is positive but very small, then α would choose to remain in the club if she is already a member but not bother entering otherwise. As a result, we can no longer take the club's current policy to be the only payoff-relevant state variable; in general, I_t is now an (infinite dimensional) state variable as well.

In spite of this, the main thrust of the paper's results—namely, that the club should converge to a myopically stable policy—still carries over in this model if we impose some reasonable simplifications. Concretely, we will assume the following:

- As in Section 4, we restrict our attention to an interval $[x^*, x^{**}]$ where x^* is a stable steady state, x^{**} is an unstable steady state, and $m(x) < x$ for all $x \in (x^*, x^{**})$.
- We assume that $x_{t+1} \leq x_t$ for all $x \in [x^*, x^{**}]$, i.e., the policy cannot move to the right.
- We assume an initial $x_0 \in [x^*, x^{**}]$ such that $u_{x_0 - d_{x_0}}(x^*) \geq (1-\delta)c$. In other words, x_0 is close enough to x^* that all agents who might consider joining the club as the policy moves left from x_0 will strictly prefer not to quit later.[55]
- Assume an initial set of members $I_0 = I(x_0)$.

[55] Effectively, this means we find equilibria restricted to $[x^*, x_0]$.

- We require that voting behavior at time t can condition on I_t but only up to a set of measure zero, i.e., if $I_1 - I_2$ and $I_2 - I_1$ have Lebesgue measure zero then $s(I_1) = s(I_2)$.

An MPE of this game is given by mappings $s(I)$ and $I(x, I')$ satisfying the above conditions such that I reflects optimal entry and exit decisions given current policy x, an existing set of members I' and the expected continuation; and $S(s(I))$ is a Condorcet winner among all $S(y)$ for the set of voters I.

It turns out that the set of equilibria of this game corresponds exactly with the set of equilibria of a game with free entry and exit but modified utility functions. Let $G(u, c)$ denote the game just described above, with cost of entry c and utility functions $u_\alpha(x)$. Similarly, let $G(v, 0)$ denote the game with free entry and exit and utility functions $v_\alpha(x)$ given by $v_\alpha = u_\alpha$ for $\alpha \geq x_0 - d_{x_0}$ and $v_\alpha = u_\alpha - (1 - \delta)c$ for $\alpha < x_0 - d_{x_0}$.

Then we have the following

Proposition 14. *In any equilibrium (s, I) of $G(u, c)$, $I(x, I') = I_v(x)$ for all (x, I') on the equilibrium path.*

As a result, for any equilibrium (s, I) of $G(u, c)$, there is an equilibrium $\tilde{s}$ of $G(v, 0)$ given by $\tilde{s}(x) = s(I_v(x))$.

Conversely, for any equilibrium $\tilde{s}$ of $G(v, 0)$, there is an equilibrium (s, I) of $G(u, c)$ given by $s(I) = \tilde{s}(m_v(I))$ and $\tilde{I}(x, I') = I_v(x)$.

Proof. For the first claim, suppose $(x_t, I_t)_t$ is an equilibrium path, and assume that $I_t = I_v(x_t)$. We aim to show that $I_{t+1} = I_v(x_{t+1})$.

There are four types of agents to consider. First, suppose $\alpha \notin I_t$ and $\alpha > x_t$, i.e., α is an outsider with a policy preference to the right of x_t. Then $\alpha \notin I_v(x_t)$, i.e., $u_\alpha(x_t) < 0$, so $u_\alpha(x_{t+1}) \leq u_\alpha(x_t) < 0$, whence $\alpha \notin I_v(x_{t+1})$. Moreover, since $u_\alpha(x_s) < 0$ for all $s \geq t$, it follows that α should not join the club at time $t + 1$, i.e., $\alpha \notin I_{t+1}$.

Second, suppose $\alpha \in I_t$ and $\alpha \geq x_0 - d_{x_0}$. Then α is an incumbent member at time $t + 1$, and will choose to remain a member iff $u_\alpha(x_{t+1}) \geq 0$, which is also the condition that determines whether $\alpha \in I_v(x_{t+1})$ as $u_\alpha = v_\alpha$ for this agent.

Third, suppose $\alpha \in I_t$ and $\alpha < x_0 - d_{x_0}$. Since $\alpha \in I_t$, α is an incumbent member at time $t + 1$. Since $\alpha < x_0 - d_{x_0}$, we have $u_\alpha(x_s) \geq (1 - \delta)c > 0$ for all $s \geq t$. This

means both that α will choose to remain a member forever (in particular, at time $t+1$) and that $\alpha \in I_v(x_{t+1})$.

Fourth, suppose $\alpha \notin I_t$ and $\alpha < x_t$, i.e., α is an outsider with a policy preference to the left of x_t. Then α should join at time $t+1$ iff $u_\alpha(x_{t+1}) \geq (1-\delta)c$, i.e., iff $\alpha \in I_v(x_{t+1})$.

(Note that, in all of these arguments, α expects the equilibrium path not to change as a function of her behavior, because her joining or leaving the club amounts to a measure zero change to I_{t+1}.)

Finally, since $G(u,c)$ has the same membership behavior as $G(v,0)$, the two games are effectively equivalent, whence the set of successor functions compatible with equilibrium are the same. □

As for the case of positive exit costs, it can be shown that, with a positive exit cost $c' > 0$, the game $G(u,c,c')$ is still equivalent to a game with free entry and exit, except that now the relevant utility functions are $v_\alpha = u_\alpha + (1-\delta)c'$ for $\alpha \geq x_0 - d_{x_0}$ and $v_\alpha = u_\alpha - (1-\delta)c$ for $\alpha < x_0 - d_{x_0}$.

With this result, we can apply Propositions 2 and 3 to $G(u,c,c')$ to determine the organization's long-run behavior in this setting. Let $m_v(y) = m(I_v(y))$, and note that $m_v(y) \geq m(y)$ for all $y \in [x^*, x_0]$. Let $y^*(c,c',\delta)$ be the highest $y \in [x^*, x_0]$ for which $m_v(y^*) = y^*$. Then it follows that $x_t \to y^*$ for any equilibrium path $(x_t, I_t)_t$.

A few interesting observations can be made. First, in general $y^* > x^*$, that is, the existence of entry and exit costs does affect the long-run policy. Second, y^* is also a function of δ, unlike x^*: this is because the entry and exit decisions of marginal agents now compare one-time costs against lifetime payoff streams, and the optimal trade-off depends on their patients. Third, it can be shown that $y^*(c,c',\delta) \to x^*$ as $c, c \to 0$, or as $\delta \to 1$ if we take c, c' as fixed. In both cases, this follows from the fact that $v \to u$ as $(1-\delta)c$, $(1-\delta)c' \to 0$.

Effectively, once the policy is near x^*, the only effect of entry costs is to keep out agents near $x^* - d_{x^*}$ who, in the absence of entry costs, would have entered the club to reap a positive but small payoff. Similarly, the effect of exit costs is to keep in agents near $x^* + e_{x^*}$ who would otherwise quit to avoid small negative payoffs. As a result, small entry and exit costs only have a small effect on the organization's long-run policy, and even sizable costs matter less and less as agents become more patient, as they only have to be paid once.

E.3 Supermajority Requirements and Other Decision Rules

Although the paper focuses on a case where chosen policies must be Condorcet winners—often leading to a Median Voter Theorem—other decision rules can be used without greatly affecting the analysis. We briefly discuss two.

First, suppose that, given a policy x and a set of members $I(x)$, an unmodeled political process results in some agent $n(x)$ being given the right to choose the policy for the next period. (For example, the function $n(x)$ might embed the notion that the policy x also affects the relative power of different agents within the organization.) The results of the paper can be extended to this case, substituting $n(x)$ for $m(x)$ everywhere, regardless of the fact that $n(x)$ is not a median voter function.

Second, consider an organization with a bias towards inaction, i.e., requiring a supermajority $\rho > \frac{1}{2}$ to make policy changes. Define $m_p(x)$ as the pth percentile member of $I(x)$, and assume that a policy $y > x$ can only be chosen over staying at the current policy x if $m_{1-\rho}(x)$ votes for it, while a change to a policy $y < x$ is only possible if m_ρ votes for it. It follows that, in intervals where $x > m_\rho(x)$, the game is equivalent to the main model with $n(x) = m_\rho(x)$; in intervals where $x < m_{1-\rho}(x)$, it is equivalent to setting $n(x) = m_{1-\rho}(x)$; and in intervals where $m_{1-\rho}(x) < x < m_\rho(x)$, no policy changes are possible. In other words, steady states are now intervals rather than points, and we will observe somewhat lower policy drift, but the gist of the results is unchanged.

E.4 Non-Monotonic Equilibria

We have established that equilibria must be monotonic in a neigborhood of a stable steady state, as shown in Proposition 3, and 1-equilibria are monotonic everywhere. However, one may wonder if nonmonotonic equilibria exist at all. Here we show by sketching an example that the answer is yes.

Assume that $u_\alpha(x) = C-(\alpha-x)^2$ and let $d = \sqrt{C} = d_x = e_x$ for all x. In addition, suppose that $m(x) = x - \rho d$ for all $x \in \mathbb{R}$, where $\rho \in \left[\frac{1}{2}, 1\right)$ is a parameter.[56] Finally, for simplicity we will take the MVT as a primitive, i.e., we assume that $S(s(x))$ is $m(x)$'s most-preferred path.

[56] m can be obtained as a median voter function if $f(x) = ke^{-\hat{\rho}x}$ for an appropriately chosen $\hat{\rho} > 0$. This is a degenerate example in the sense that there is an infinite mass of voters distributed on the real line, as opposed to a unit mass with support $[-1, 1]$, but it allows for a simpler construction.

Recall that, if $\delta = 0$, then $s(x) = m(x)$. Here this means $S(s(x)) = (x - \rho d, x - 2\rho d, x - 3\rho d, \ldots)$. Crucially $x - \rho d, x - 2\rho d \in (m(x) - d, m(x) + d)$ but $x - 3\rho d \notin (m(x) - d, m(x) + d)$ by the assumption that $\rho \in \left[\frac{1}{2}, 1\right)$.

Now, suppose that δ is small but positive. What are some possible equilibria?

First, assume a successor function s_1 of the form $s_1(x) = x - \rho d + \eta_1$ with η_1 small, such that $s_1^2(x) > m(x) - d$ but $s_1^3(x) < m(x) - d$.

For s_1 to be an equilibrium, it must hold that

$$\eta_1 = \arg\max_{\eta'} u_{m(x)}(x - \rho d + \eta') + \delta u_{m(x)}(s_1(x - \rho d + \eta')) = C(1 + \delta) - \eta'^2 - \delta(\rho d - \eta_1 - \eta')^2$$
$$-2\eta' - 2\delta(\eta' + \eta_1 - \rho d) = 0$$
$$\eta_1 = \frac{\delta}{1 + 2\delta}\rho d.$$

Note that, since $\eta_1 > 0$, this calculation will be invalid for ρ close enough to $\frac{1}{2}$, as in fact we will then have $s_1^3(x) > m(x) - d$.

Next, assume a successor function s_1 of the form $s_1(x) = x - \rho d + \eta_2$ with η_2 small, such that $s_1^3(x) > m(x) - d$ but $s_1^4(x) < m(x) - d$.

For s_2 to be an equilibrium, it must hold that

$$\begin{aligned}\eta_2 &= \arg\max_{\eta'} u_{m(x)}(x - \rho d + \eta') + \delta u_{m(x)}(s_2(x - \rho d + \eta')) + \delta^2 u_{m(x)}(s_2^2(x - \rho d + \eta')) \\ &= C(1 + \delta + \delta^2) - \eta'^2 - \delta(\rho d - \eta_2 - \eta')^2 - \delta^2(2\rho d - 2\eta_2 - \eta')^2 \\ &\quad - 2\eta' - 2\delta(\eta' + \eta_2 - \rho d) - 2\delta^2(\eta' + 2\eta_2 - 2\rho d) = 0 \\ \eta_2 &= \frac{\delta + 2\delta^2}{1 + 2\delta + 3\delta^2}\rho d.\end{aligned}$$

Now consider the related problem of an agent who expects everyone to play strategy s_2, but who considers deviating to some η_3 such that $s_2^2(x - \rho d + \eta_3) < m(x) - d$. Then it must be that

$$\begin{aligned}-2\eta_3 - 2\delta(\eta_3 + \eta_2 - \rho d) &= 0 \\ \eta_3 = \frac{\delta}{1 + \delta}(\rho d - \eta_2) &= \frac{\delta}{1 + \delta}\frac{1 + \delta + \delta^2}{1 + 2\delta + 3\delta^2}\rho d.\end{aligned}$$

Note that $\eta_2 > \eta_1 > \eta_3 > 0$. In particular, since $\eta_2 > \eta_3$, we can choose ρ, δ so that $x - 3\rho d + 3\eta_2 > m(x) - d > x - 3\rho d + \eta_1 + 2\eta_2$. Furthermore, we can choose

them so that $U_{m(x)}(S_2(x-\rho d+\eta_3)) = U_{m(x)}(S_2(x-\rho d+\eta_2))$.[57] If we choose our parameters this way then $m(x)$ is indifferent (of course, this is in fact true for all x). Now construct a successor function s as follows: $s(x) = s_2(x)$ for all $x < x_0$; $s(x_0) = x - \rho d + \eta_3$; and $s(x)$ for $x > x_0$ is defined by backward induction. This is a non-monotonic equilibrium by construction. Finally note that, while this example relies on indifference, we can adjust it by tweaking m to make it strict.

E.5 Explicit Voting Protocols

Throughout the paper we assume that, for a policy y to be chosen in equilibrium by a set of voters $I(x)$, $S(y)$ must be a Condorcet winner, i.e., for any other available path $S(z)$, a weak majority of $I(x)$ prefers $S(y)$ to $S(z)$. This assumption stays silent on the actual voting process taking place. Here, we discuss two natural microfoundations.

The first is Downsian competition. Suppose that there are two politicians A_t, B_t who simultaneously propose policies x_{At}, x_{Bt}; voters observe the two proposals and then vote for a winner. Assume either that the politicians are short-lived (they are replaced every period) or that they play Markov strategies, and they are office-motivated, i.e., they obtain a payoff $R > 0$ from winning and zero from losing. An equilibrium is given by policy proposal strategies $x_A(I)$, $x_B(I)$ and voting strategies $v_\alpha(x_A, x_B)$ such that: for each candidate i, offering $x_i(I)$ maximizes i's probability of winning given a set of voters I, and $v_\alpha(x_A, x_B) = i$ if $U_\alpha(S(x_i)) > U_\alpha(S(x_{-i}))$, where $S(x_i)$ is the equilibrium policy path starting at x_i. It is then clear that:

Remark. Given an equilibrium successor function s from the main model, we can construct an equilibrium of the Downsian model as follows: $x_A(I(x)) = x_B(I(x)) = s(x)$ and $v_\alpha(x_A, x_B) = \mathbb{1}_{\{U_\alpha(S(x_A))>U_\alpha(S(x_B))\}}$ for all x, x_A, x_B, where $S(x) = (x, s(x), \ldots)$.

Conversely, if an equilibrium of the Downsian model features pure proposal strategies $x_i(I)$ for every i, I, then $x_i(I)$ must be a Condorcet winner for every i, I. Moreover, if $x_A(I) = x_B(I)$ for all I, then $s(x) = x_i(I(x))$ constitutes an equilibrium of the main model.

In other words, requiring $S(s(x))$ to be a Condorcet winner among $I(x)$ is equivalent to assuming Downsian competition at the voting stage, except that we implicitly

[57] This follows from a continuity argument: when $s_2^2(x-\rho d+\eta_3) = m(x) - d$, $U_{m(x)}(S_2(x-\rho d+\eta_3)) < U_{m(x)}(S_2(x-\rho d+\eta_2))$, whereas when $s_2^2(x-\rho d+\eta_2) = m(x)-d$ then $U_{m(x)}(S_2(x-\rho d+\eta_3)) > U_{m(x)}(S_2(x-\rho d+\eta_2))$, so we can choose intermediate values of ρ, δ for which $U_{m(x)}(S_2(x-\rho d+\eta_3)) = U_{m(x)}(S_2(x-\rho d+\eta_2))$.

rule out situations where there is no Condorcet winner (in which case the Downsian model might still have equilibria with mixed proposal strategies), and we rule out mixed policy choices by voters ($s(x)$ is assumed to be deterministic).

Another possible microfoundation is a sequential proposal protocol similar to that used in Acemoglu et al. (2008). Suppose that, at each voting stage, there is a continuous time period $[-1, 1]$ (this is measured on a different scale from the time that passes between periods, and no discounting accrues during the voting stage). At each instant $y \in [-1, 1]$, policy y is proposed to the organization and each voter α casts a vote $v_\alpha(y, I) \in \{0, 1\}$. If a majority votes 1 when y is under consideration, the voting stage ends and y is chosen; if no policy receives a majority then the policy next period is equal to the current policy. Voters are strategic and that voting strategies are pure and weakly undominated.

Given an equilibrium successor function $s(x)$ from the main model, we can construct an equilibrium of this model such that, if $I(x)$ is the set of members, then $s(x)$ is chosen.