Policy Persistence and Drift in Organizations*

Germán Gieczewski†

June 2016

Abstract

We analyze the evolution of organizations that allow free entry and exit of members, such as cities, trade unions, religious organizations, cooperatives, and so on. The organization chooses a policy, which influences the set of agents who want to become members, but in turn current members get to decide policy in the next period. This generates feedback effects: an organization with a policy x may attract a population where the median-preferred policy is higher than x, which will make the chosen policy higher in the next period; but the new policy will attract members wanting an even higher policy, and so on. We characterize a set of steady states uniquely determined by the distribution of preferences; equilibrium paths converge to these steady states depending on the starting position. Moreover, organizations may drift to the mainstream or become extremist. Unlike in models with a fixed population, a small change in the distribution of preferences can lead to dramatic changes in the long-run equilibrium policy. We also extend the model to allow for competition between multiple organizations and endogenous entry, and find that there are inefficiencies both in how many organizations are created, and what policies they choose in steady state.

1 Introduction

Many organizations provide certain services to their members, and must decide exactly what those services are, or who they are geared towards–this is what we call a *policy*. In many cases, free entry and exit of members is allowed, or the cost of changing status is low. Members often have a say in what the policy should be. But, in turn, the chosen policy affects the population of future members: some current members may become disillusioned after the latest changes, or outsiders may be enticed to join. Thus, the organization's policy and set of members constantly influence each other, potentially leading to large changes over time. The main question of this paper is: what outcomes can we expect from an organization ruled by these dynamics, especially in the long term?

Our framework applies to many important organizations in the real world. For instance:

*I am deeply indebted to my advisors, Daron Acemoglu, Juuso Toikka and Muhamet Yildiz for their support, guidance and suggestions that substantially improved the paper. I would also like to thank Alessandro Bonatti, Glenn Ellison, Robert Gibbons, Mihai Manea, Robert Townsend, Alex Wolitzky, and seminar participants at MIT for helpful comments and suggestions. All remaining errors are my own.

†Department of Economics, Massachusetts Institute of Technology.

- Workers in a given industry may have the option to join a trade union. The union makes collective demands on their members' behalf, as well as providing additional services; different policies cater to different subsets of the population. In particular, demand for high wages is beneficial for workers who keep their jobs, but may reduce employment overall. If senior workers have higher job stability, they will push for more aggressive demands, as they do not suffer the cost. Hence a union with more senior members will be more aggressive, which in turn means junior workers will not want to unionize, and vice versa.[1]

- Sports clubs and other neighborhood clubs allow people to become members for a fee, and provide access to facilities, priority tickets to games, etc., in exchange. Many clubs hold elections where members vote for the club leadership, and hence indirectly for policies; on the other hand, the club attracts different demographics depending on what it offers. For example, if the club adds a new golf course, it may attract more upscale families in the neighborhood, leading to further "gentrification" of the club.

- Other non-profit organizations, like churches and universities, exhibit similar dynamics, albeit through somewhat different channels. In both cases, people can join the community (by converting or applying, respectively) and choose partly based on cultural fit. Unlike the examples above, members often have no direct channel to choose the leaders, although they may still influence decisions informally. However, current members directly influence the community's culture by belonging to it, as peer effects are important, so the same feedback mechanics are in place.

- Even entities not usually considered organizations can exhibit these dynamics. For instance, a city can be construed as a club with a set of members (i.e., people living within its limits), drawn from a larger population of people who can move in and out. The city chooses policies such as its local taxes, school quality, housing regulations, and so on, which affect who wants to move there. In turn, citizens can vote for mayor as well as in referendums, and they can belong to homeowners' associations, write petitions, etc.[2]

These examples differ in the details of exactly how current members influence future policies, and whether there is some cost of entry, but they share the same essential features. For tractability, we study a stylized version of this problem, in which members vote for policies directly,[3] and entry and exit are completely free at any time. Members choose whether to belong to the club based on the current policy, and they are small and numerous enough to behave like "policy-takers". Finally, and most importantly, we assume agents to be forward-looking. This means that, when they are

[1] This effect is highlighted in Grossman (1984), which studies the change in wage demands in response to a drop in labor demand. By the same logic we describe, if the shock leads to junior people losing their jobs–and hence their voting rights as well–the remaining members will be more aggressive, resulting in rigid wages.

[2] Glaeser and Shleifer (2005) study the case of Mayor Curley in Boston, who used wasteful policies to induce rich citizens to move out, since he was mainly popular among the poor Irish population.

[3] It is equivalent to assume that members vote for one of two policy-motivated candidates, so this is not too restrictive.

voting, they take into account the fact that feedback effects may cause the current policy to drift away, according to equilibrium behavior.

We can answer several interesting questions in this setting. For instance, does the feedback effect in this model lead to unstable or knife-edge outcomes, or multiple equilibria? Does policy tend to drift toward the center? Or, on the contrary, can a moderate club be captured by extremists? How do different clubs interact?

Our main result for the basic model is a characterization of the equilibrium paths that the club's policy can follow. We find that, although there are generally multiple equilibria, they all yield similar outcomes. Namely, given an initial policy x, future policies drift away from x in the same direction, and towards the same limit, in all equilibria, although the speed of convergence varies somewhat across equilibria. (In particular, policy paths are always monotonic, i.e., they do not double back on themselves). However, there may be multiple steady states; if so, the one the club goes to is a function of the initial policy.

The steady states, as well as their basins of attraction, are uniquely pinned down by the distribution of preferences in the population in a simple way. Intuitively, if a policy x attracts a group of members with median higher than x, then policy will drift upward from x in equilibrium, and vice versa. In particular, the long-run outcome is independent of the agents' discount factor: it is the same that would obtain if they were completely myopic, although the speed of convergence will be proportionally lower when agents are patient. In other words, agents understand future drift and react to it by doing what they can to slow it down, but they do not stop it completely.

Armed with this result, we can show when clubs will drift to the mainstream or become extremist in concrete examples. Generally speaking, drift leads clubs towards high-density areas of the preference distribution, which favors centrism. However, a pocket of agents concentrated at an extreme can also support a steady state. 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 club than vice versa).

A related feature of the model is that steady state policies are more sensitive to the exact shape of the distribution than in models with a fixed population of voters. Rather than being close to the global median voter, steady states tend to be close to modes of the distribution. Hence, when the distribution is close to uniform, small changes to its density can result in dramatic swings in the steady states. In practice, this means that a slow, continuous demographic change may at some point trigger a change in the club's dynamics, resulting in policy changes which would appear sudden in comparison.

Finally, we also extend the model to allow for competition between multiple clubs and endogenous creation of clubs. It turns out that both cases add room for new inefficiencies: in the former, clubs tend to cluster too close together, effectively competing for members instead of spreading their benefits to as many people as possible. In the latter, the willingness of agents to create clubs is lowered insofar as they expect their chosen policy to drift, resulting in underprovision of clubs.

The paper contributes to the growing literature on dynamic policy selection and "elite clubs".

Compared to Roberts (1999), Barbera, Maschler and Shalev (2001) and Acemoglu, Egorov and Sonin (2008, 2012, 2015), our framework is in the same tradition, but with three important differences. First, in these papers, it is assumed–with some variants–that current members get to directly restrict entry of newcomers and do so strategically, while in our model the dynamics are driven by free entry and exit of members. Second, our model features a continuum of voters and policies, whereas the others are discrete. Although this seems like an innocuous difference, it turns out to have important implications: with a discrete policy space, policies stop drifting short of the steady state when agents are patient, because they anticipate that further shifts will happen too quickly for their taste, as a result of the discreteness of the options. On the other hand, with continuous policies, this sort of stalling does not occur, so policy converges to a steady state regardless of the discount factor. (In turn, which conclusion is right in a concrete case may depend on how fine the space of available policies is). Third, our assumption that members can stop receiving payoffs from the club by quitting technically breaks the assumption of increasing differences used in previous papers, but it turns out that the results still follow through.

On the other hand, the story behind our model is closely related to more applied papers, which focus on concrete examples from different literatures. For instance, Grossman (1984) highlights the interaction between wage demands and the membership of trade unions. And Glaeser and Shleifer (2005), as well as the literature on Tiebout competition–starting with Tiebout (1956) and continuing with Epple, Filimon and Romer (1984) and Epple and Romer (1991)–study the interaction between policies chosen by a city and the citizens' decision to relocate. These papers share the premise that policies and membership decisions must be in mutual equilibrium, but they simply assume that this must be so statically. In other words, they study the steady states of the model. In contrast, we allow our organizations to start with non-steady state policies and characterize the equilibria of the dynamic game where the policy and set of members adjust endogenously over time.

We will proceed as follows. In Section 2, we set up the basic model. In Section 3, we first show some common properties of all equilibria, then characterize a particular family with useful properties. In addition, we show a continuous time limit of the game, where some of the results are simplified, yielding a closed-form expression for the equilibrium. In Section 4, we expand on the intuition behind the results in Section 3, as well as their implications in practice. In Section 5, we discuss an extension with multiple clubs and entry of new clubs. Section 6 concludes.

2 The Model

There is a club existing in discrete time $t = 0, 1, \ldots$. The population of *potential members* of the club is given by a continuous density f with support $[-1, 1]$. The timing is as follows: at $t = 0$, the club starts with an initial policy x_0. At each integer $t \geq 1$, existing members vote on the policy $x_t \in [-1, 1]$ to be implemented during the period $(t, t + 1]$. In addition, at each $t + \epsilon$ ($t \geq 0$), agents can choose to enter the club (if they are outsiders) or leave (if they are currently members) at no cost, where $\epsilon > 0$ is small. We denote by $I_t \subseteq [-1, 1]$ the set of members for the period

$(t+\epsilon, t+1+\epsilon]$. The essential feature of this setup is that membership affects both an agent's utility and his right to vote; agents will 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 ultimately affect future policy.

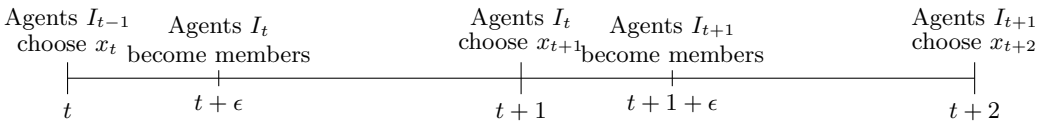


Figure 1: Order of moves in each period of the game

The assumption that entry/exit decisions happen shortly after voting serves two purposes. First, since $\epsilon > 0$, current members want to vote for policies they like, even if they plan to quit the club at the next possible opportunity. (Otherwise, members who intend to quit immediately after the vote would be indifferent and thus willing to vote for any policy). Second, since ϵ is small, potential entrants and quitters at $t+\epsilon$ mostly base their decisions on the policy chosen at time t, since they are locking themselves in for the period $(t+\epsilon, t+1+\epsilon]$ which is mostly contained in $(t, t+1]$. Otherwise, if ϵ were large, voters would be concerned instead about the policy they expect will be chosen at $t+1$, which could lead to multiple equilibria. For example, given a left-wing club, a completely different set of right-wing agents could enter (and the current cohort would abandon the club) based on a self-fulfilling expectation that a right-wing policy will be chosen at $t+1$. We think it is reasonable to rule out these outcomes.[4]

Agent preferences

Potential member α has utility

$$U_\alpha\left((x_t, I_{\alpha t})\right) = \sum_{t=0}^{\infty} \beta^t \left(\epsilon I_{\alpha(t-1)} + (1-\epsilon) I_{\alpha t}\right)\left(C - (x_t - \alpha)^2\right).$$

Here x_t is the club's policy at time t, α is the agent's bliss point, and $I_{\alpha t} = \mathbb{1}_{\alpha \in I_t}$ denotes whether α is a member during $(t+\epsilon, t+1+\epsilon]$. $C > 0$ is the maximum utility the agent can get from being a member, if the club has optimal policy α. Intuitively, the agent wants x_t to be as close as possible to α. But, if the distance is large enough, he will instead quit the club. Whenever α is not a member, his flow payoff is 0. Note also that agents are forward-looking with discount factor $\beta > 0$. This greatly affects our equilibrium analysis, since agents, when voting for a current policy, take into account how it will drift in the future.

Finally, for the rest of the paper we will make the simplifying assumption that $\epsilon \approx 0$. In other words, we will study the limit case where $\epsilon = 0$, so that an agent's utility is effectively:

[4] Although this explanation might make the behavior of the model seem sensitive to the choice of ϵ, there are other natural assumptions that would give us the same result: for example, if agents had to be members for a period of time before becoming "full members" (and thus being allowed to vote), this would also rule out unstable outcomes, independent of ϵ.

$$U_\alpha\left((x_t, I_{\alpha t})\right) = \sum_{t=0}^{\infty} \beta^t I_{\alpha t}\left(C - (x_t - \alpha)^2\right),$$

with the caveat that, when a member is voting between two policies which would both induce him to quit, he will vote for the one closest to α, as he would if ϵ were positive but arbitrarily small.

Equilibrium Concept

Although we do not model the voting process explicitly we assume that, if there is a Condorcet winner, then this will be the chosen policy.[5]

We will focus on Markov Perfect Equilibria, meaning that

- at time t, when votes are cast, the only relevant state variable which voters condition on is the set of current members I_{t-1};
- at time $t + \epsilon$, when entry and exit decisions are made, the only relevant state variable which agents condition on is the current policy x_t.

We can spell out the structure of a MPE for this game as follows:

Definition 1. A MPE of the game is given by a policy function $\tilde{s} : 2^{[-1,1]} \to [-1, 1]$ and a membership correspondence $I : [-1, 1] \rightrightarrows [-1, 1]$, such that:

- Given the current policy x, it is optimal for agents in $I(x)$ to be in the club, and no others.
- Given a set of voters J, choosing $\tilde{s}(J)$ as the policy next period is the Condorcet winner.

We denote by $s = \tilde{s} \circ I$ the *successor* function. For any current policy x, the induced set of members will be $I(x)$, and they will vote for policy $\tilde{s}(I(x)) = s(x)$. Hence, given an initial policy x_0, the equilibrium path will be given by $x_{t+1} = s(x_t)$.

For the rest of the paper, we will describe equilibria by the functions I and s rather than I and $\tilde{s}$. This is without loss of detail: we only lose the description of voting that happens when sets of voters are not of the form $I(x)$, which never occurs on the equilibrium path.

Finally, it will be useful in describing equilibria to have a notion of steady states:

Definition 2. Given a successor function s, we say that $x \in [-1, 1]$ is a *steady state* if $s(x) = x$. Moreover, x is *stable* if there is a neighborhood $(a, b) \ni x$ such that $s^k(y) \to x$ for all $y \in (a, b)$.

3 Equilibrium Characterization

In this section we show some common properties of all Markov Perfect Equilibria, which in particular pin down the long-run behavior of any equilibrium. First, the following Lemma allows us to simplify our description of MPE:

[5] In particular, this would be the outcome if in each period there are two representatives, who choose a policy to win the election – or if there is some rotating proposal mechanism where voters get a turn to propose a policy to replace the current one.

Lemma 1. *In the game with $\epsilon \approx 0$, in any MPE, we must have $I(x) = (x - d, x + d)$, where $d = \sqrt{C}$.*

Intuitively, since members can enter or leave at any time $t+\epsilon$, the optimal membership decision is simply to join myopically whenever the flow payoff of the current policy is positive. When $\epsilon \approx 0$, this happens when $C - (x - \alpha)^2 \geq 0$, i.e., when $\alpha \in (x - \sqrt{C}, x + \sqrt{C})$. In other words, d is the maximum distance between the club's policy and a voter's bliss point, such that the voter would still rather belong to the club. Since the chosen policy x pins down $I(x) = (x - d, x + d)$, we can describe an MPE solely as a successor function $s(x)$.

It will be useful to define the *median voter function* m. For $x \in [-1, 1]$, let $m(x)$ as the median voter among the set of agents who would choose to be club members if the club's policy is x, that is, the median voter in $I(x)$. Formally $m(x)$ is defined by the condition

$$F(m(x)) - F(x - d) = \frac{F(x + d) - F(x - d)}{2}.$$

For any path $S = (s_1, s_2, \ldots)$, let $E(S) = \sum \beta^t s_t$ be the discounted average policy in S. Define $S(y) = (y, s(y), s(s(y)), \ldots)$ as the equilibrium path following from a policy choice y. We first establish two auxiliary lemmas that allow us to compare paths based on their average policies, in the vein of increasing differences. Note that $u_\alpha(x) = C - (x - \alpha)^2$ has increasing differences in α and x, and $U_\alpha(S)$ would have increasing differences in α and $E(S)$ if α intended to always stay in the club under path S, but this property is broken when α may choose to leave at different times given different policy paths. However, we can still show that:

Lemma 2. *Let $S = (s_1, s_2, \ldots)$ and $T = (t_1, t_2, \ldots)$ be policy paths, and $\alpha_0 < \alpha_1$ two voters such that $s_j - d < \alpha_i < s_j + d$ and $t_j - d < \alpha_i < t_j + d$ for all i, j (in other words, both voters are always in the club under both paths). If $E(T) > E(S)$ and α_0 prefers T to S, so does α_1.*

Lemma 3. *Let S be a path such that $\sup(S) \leq x$ and $S \neq (x, x, \ldots)$. Then there is $\alpha_0 \leq x$ such that voters in $[-1, \alpha_0)$ strictly prefer S to a constant path with policy x, and voters in $(\alpha_0, 1]$ strictly prefer x to S.*

In other words, the expected increasing differences result holds when the two voters being compared never want to exit the club under either path. In addition, if one of the paths is constant and the paths do not overlap, the result holds for all voters.

Next, we characterize the possible successor functions s. First, we show that equilibrium paths must be monotonic:

Lemma 4. *In any MPE, for any y, $S(y)$ is monotonic: i.e., if $s(y) \geq y$ then $s^k(y) \geq s^{k-1}(y)$ for all k and vice versa, where $s^k(y) = s^{k-1}(s(y))$.*

This rules out paths that increase up to some point and then double back, or vice versa. Intuitively, such paths are incompatible with equilibrium: imagine a path $(s_1, s_2, \ldots)$ which increases up to s_k and decreases afterwards. Then voters in $I(s_{k-1})$ prefer the path $(s_k, s_{k+1}, \ldots)$, while voters

in $I(s_k)$ prefer $(s_{k+1}, s_{k+2}, \ldots)$. Note that the latter path has a lower average policy, 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 $s_k > s_{k-1}$. The main difficulty is to prove this contradiction occurs even in cases where the path doubles back on itself infinitely many times.

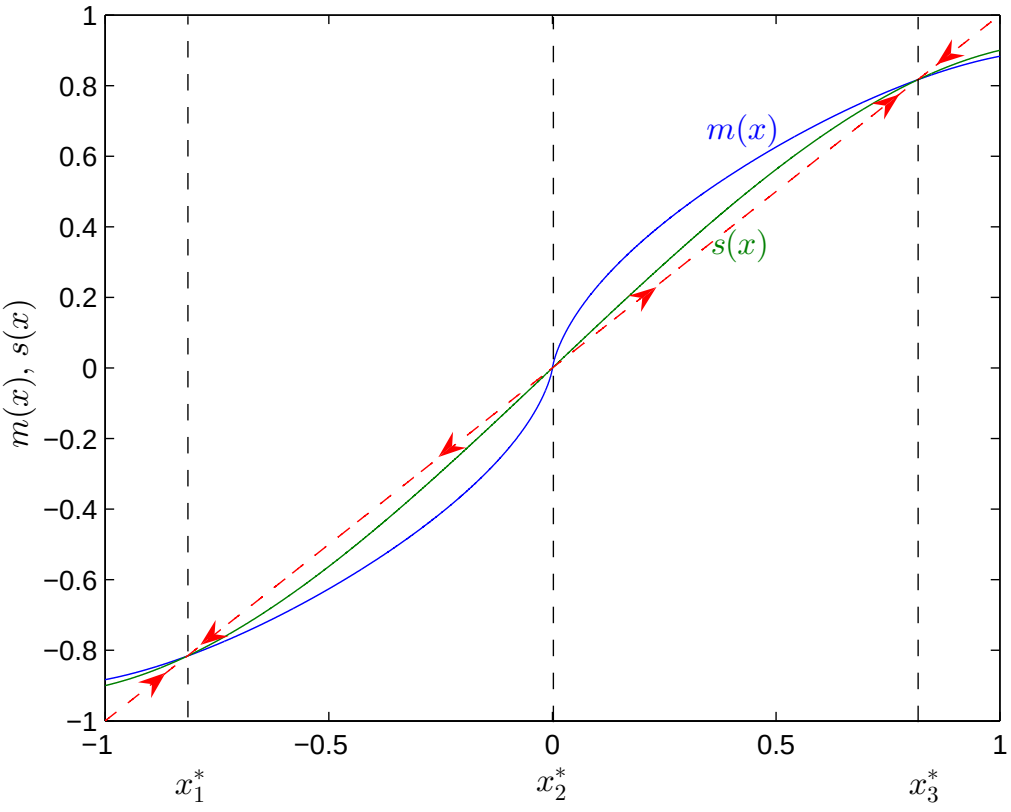
Armed with this result, we can pin down the general shape and long-run behavior of any MPE:

Proposition 1. *Let $m^*(y) = \lim_{n\to\infty} m^n(y)$. Then, in any MPE and for any y,*

- *If $m(y) = y$, then $s(y) = y$;*
- *If $m(y) > y$, then $m^*(y) > s(y) > y$;*
- *If $m(y) < y$, then $m^*(y) < s(y) < y$.*

In particular, $s^k(y) \to m^*(y)$ as $k \to \infty$.

Figure 2: Convergence to steady states in MPE



Proposition 1 provides a natural interpretation for the steady states of s: they 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 always coincide.

Intuitively, it is easy to see why policies should converge to a stable fixed point of m. Let x^* be such a point, and suppose the club is initially at policy $x > x^*$. To the right of x^*, we have

$m(y) < y$, meaning that any given policy attracts a set of members whose median's bliss point is to the left of the current policy. If voters were myopic, they would choose $s(x) = m(x) < x$. As they are forward-looking and wary of future changes, they will usually choose a smaller shift, but in the same direction, i.e., $m(x) \leq s(x) < x$: the tendency to move left is still present.[6] In turn, the new policy attracts more left-wing voters, leading to a lower $m(s(x)) < m(x)$, and so on. Conversely, to the left of x^* we have $m(y) > y$, so the members attracted by y would rather move to the right.

Figure 2 illustrates this result in an example with three steady states: x_1^* and x_3^* are stable, while x_2^* is unstable. Although the successor function s differs from m (given that voters are forward-looking), the two mappings lead to the same limit policies when iterated.

The structure of alternating stable and unstable steady states shown in the example is actually quite general:

Corollary 1. *Let f be such that $m(y) = y$ has finitely many solutions, and call these values $x_1^* < x_2^* < \ldots < x_n^*$. Suppose that $m'(x_i^*) \neq 1$ for all i.*[7] *Then n must be odd; x_i^* must be stable for odd i and unstable for even i; and, for any MPE, all equilibrium paths starting at any $y \in (x_{2j}^*, x_{2j+2}^*)$ must converge to x_{2j+1}^*.*

In particular, Proposition 1 shows that the long-run behavior of the club does not depend on the players' discount factor, or on the fact that they are forward-looking. Indeed, in any MPE starting at some y, the club's policy converges in the long run to $m^*(y)$. This is the same result that would obtain if $\beta = 0$, in which case $s(y) = m(y)$.

In other words, the equilibria of the model do not feature any unnatural stopping points. That is, there are no policies y such that $m(y) \neq y$, but $s(y) = y$ (in some MPE) because the voters in $I(y)$ are afraid of further changes that will occur if they move closer to $m(y)$.

This contrasts with other papers in the literature, such as Roberts (1999), where both *extrinsic* and *intrinsic* steady states are possible (the former correspond to fixed points of m in our model, while the latter are created by dynamic considerations); and Acemoglu et al. (2012, 2015), where some dynamically stable states may be Pareto inefficient. In both cases, policies considered sub-optimal by the current voters can be sustained indefinitely in equilibrium due to the fear that, if a line is crossed, future agents will move towards a different policy too quickly: the so-called *slippery slope* argument.[8] The models leading to this result share two important assumptions: a discrete policy space and patient agents.

On the other hand, in our model, a continuous policy space always affords the option to move slowly enough towards the steady state, so that there is a better alternative to moving too fast or not moving at all. Hence, whether slippery slope concerns would stall policy change in a real-life setting may depend on institutional details, namely, on whether the exact speed of change can be regulated by using incremental changes, or whether only certain large changes are possible. For

[6] The statement that $m(x) \leq s(x) < x$, as opposed to the weaker $x^* < s(x) < x$, is necessarily true when $x \in (x^* - d, x^* + d)$ by Proposition 2, but may be false otherwise.

[7] m is differentiable since f is continuous.

[8] See Schauer (1985) for an explanation of slippery slope arguments in judicial reasoning. Volokh (2003) provides examples in other areas, including shifts in political power.

example, take a polity with limited franchise considering whether to extend the franchise.[9] Suppose that voters are ordered by their income and high-income voters generally prefer a limited franchise, but a bit laxer than the smallest one they would be in (i.e., say a voter in the top 10% of income would want the top 15% of voters to be enfranchised, a voter in the top 20% would want the top 25% to be enfranchised and so on). Then, if at each time current voters can choose to grant voting rights to the top x% of voters for any x, slippery slope concerns would not prevent full democracy from obtaining in the long run, through a series of small changes. However, if voting rights can only be extended based on a coarse set of categories (e.g., only to men who can read; only to property owners; only to taxpayers, and so on), stalling is much more likely to occur.

Of course, the agents' patience and forward-looking behavior still have an impact on the equilibrium, since they affect the speed of convergence. This can also be seen in Figure 2: generally speaking, $s(x)$ traces a similar shape to $m(x)$, but it will be closer to the identity when β is high (in other words, each jump $|s(x) - x|$ will be smaller when β is high).[10]

Finally, we address the question of whether s must be monotonic (a stronger property than path monotonicity). It turns out that s must be monotonic, and the Median Voter Theorem must hold, around each stable steady state:

Proposition 2. *Let x^* be a stable steady state. Let $I = (x^{**}, x^{***}) \ni x^*$ be the basin of attraction of x^*, and let $J = [x^* - d, x^* + d]$. Then, in any MPE, for all $y \in I \cap J$:*

- *s is weakly increasing;*
- *$s(y)$ is $m(y)$'s most-preferred policy; in other words, the Median Voter Theorem holds.*

However, there may be non-monotonicities in equilibrium, away from the steady state. These are driven by the interaction between different voters' bliss points and their optimal times to quit the club. Intuitively, voters with a higher bliss point prefer paths with a higher average policy, yielding monotonic choices. But whenever a policy is far enough from the bliss point to become unacceptable, it essentially drops out of the average. Hence different voters may disagree about how two paths compare in terms of their *effective* average policy. In the same vein, the Median Voter Theorem may fail away from a steady state because the set of people preferring one policy over another may not be a single interval (which would need to contain the median to be a majority), but rather a collection of disjoint intervals, each reflecting a different drop-out time.

On the other hand, as we will see next, monotonic equilibria always exist when voting is frequent enough.

k-Equilibria and Continuous Equilibria

While we have shown that all possible MPEs share some important properties, there are typically multiple equilibria. In this section, we illustrate what drives this multiplicity, and how equilib-

[9] This example does not fit exactly into our model but the same techniques we use can be readily applied to it.

[10] This is not an exact result because there are many equilibria for each β, which may not be directly comparable, but becomes clear in our next two sections, where we show explicit equilibria as well as a continuous time limit.

ria may differ, by characterizing two particular classes of equilibria: k-equilibria and continuous equilibria.

Without loss of generality, we will study the game restricted to the right side of the basin of attraction of a stable steady state. In other words, let $x^* < x^{**}$ such that $m(x^*) = x^*$, $m(x^{**}) = x^{**}$ and $m(y) < y$ for all $y \in (x^*, x^{**})$. Then we will study s restricted to $[x^*, x^{**}]$. In any MPE s, $s(y) \in (x^*, y)$ for all $y \in (x^*, x^{**})$, and $I(y)$ will never want to choose a policy outside of (x^*, x^{**}), so $s|_{[-1,1]-(x^*,x^{**})}$ is irrelevant for determining whether $s|_{(x^*,x^{**})}$ is compatible with MPE. (The case where $m(y) > y$ is analogous).

First, the following Lemma shows that multiplicity is pinned down by behavior that occurs arbitrarily close to x^*:

Lemma 5. *Let s, s' be two MPEs on $[x^*, x^{**}]$ such that $s(y) = s'(y)$ for all $y \in [x^*, x^* + \epsilon]$. Suppose s and s' obey the following tie-breaking rule: if the set of Condorcet winners for $I(y)$ has multiple elements, then the highest policy in the set is chosen. Then $s = s'$ on $[x^*, x^{**}]$.*

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 unless there is indifference. Conversely, if there are multiple equilibria generically, their differences must start from the beginning.

Next, we define a k-equilibrium:

Definition 3. 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 $x \in [x_n, x_{n-1})$.[11]

On the other hand, a continuous equilibrium is one where s is continuous.

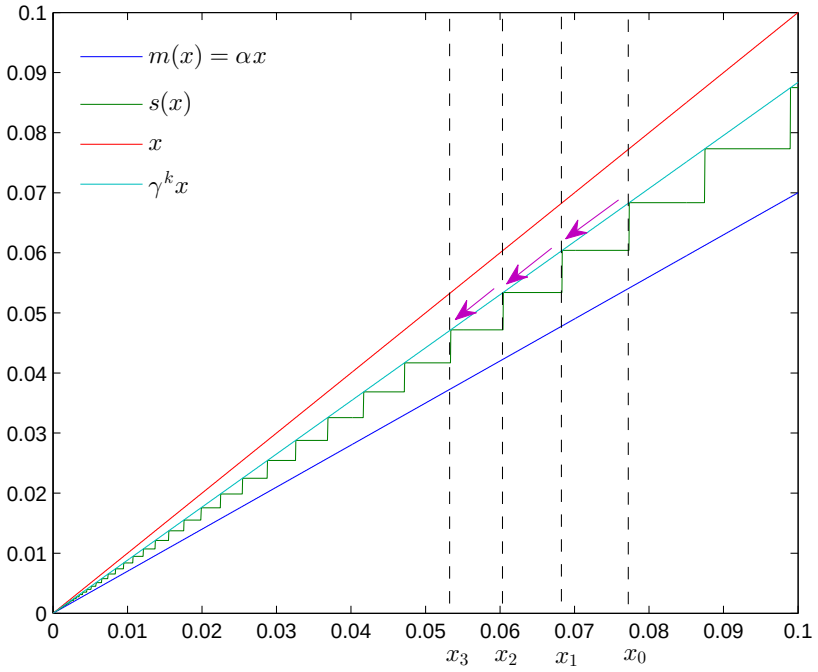
Intuitively, in a k-equilibrium, there are k staggered sequences, and each policy in a sequence leads to the next policy of that sequence being chosen in the next period. Policies that are not in any sequence are never chosen. It is easy to guess why: by construction, any x between x_{n+1} and x_n leads to the same continuation as x_{n+1} would, so the only difference between $S(s(x))$ and $S(s(x_{n+1}))$ is the first flow payoff. If β is relatively high, so that convergence to x^* takes several periods, then we expect that $x_{n+1} > m(x_{n-k})$, so there is no benefit to choosing $x \in (x_{n+1}, x_n)$ instead of x_{n+1}. In addition, since the path is discontinuous, there must be indifference at every discontinuity, i.e., voters in $I(x_{n-k})$ must be exactly split in their preference for x_n or x_{n+1}.

For our next result, we assume that $m(x)$ is linear around a stable steady state. Although this is not true in general, we can construct densities f such that m is indeed linear,[12] and it serves

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

[12] We need $F(\alpha x) = \frac{F(x+d)+F(x-d)}{2}$. Assuming a symmetric f around the steady state $x = 0$, this boils down to $f(x+d) = 2\alpha f(\alpha x) - f(d-x)$. Hence we can choose f freely on $[0, d]$ and it becomes uniquely determined from d on. For example, we can 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, which makes f continuous.

Figure 3: 1-equilibrium for $m(x) = 0.7x$, $\beta = 0.7$



as a useful approximation of the general case, since a differentiable m is approximately linear in a neighborhood of the steady state. It turns out that, in the linear case, we can find k-equilibria for all k, as well as a continuous equilibrium:

Proposition 3. *Let $x = 0$ be a stable steady state and let f be such that $m(x) = \alpha x$ for $x \in [-e, e]$, where $\alpha < 1$ and $e \leq d$. Furthermore, suppose that $\beta \geq \frac{2}{3}$ and $\alpha \geq 0.44$. 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 $0 < \gamma_k < 1$. In addition, there is a continuous equilibrium s_∞^* given by $s_\infty^*(x) = \gamma_\infty x$. Moreover, γ_k^k is decreasing in k, and $\gamma_k^k \to \gamma_\infty$.*

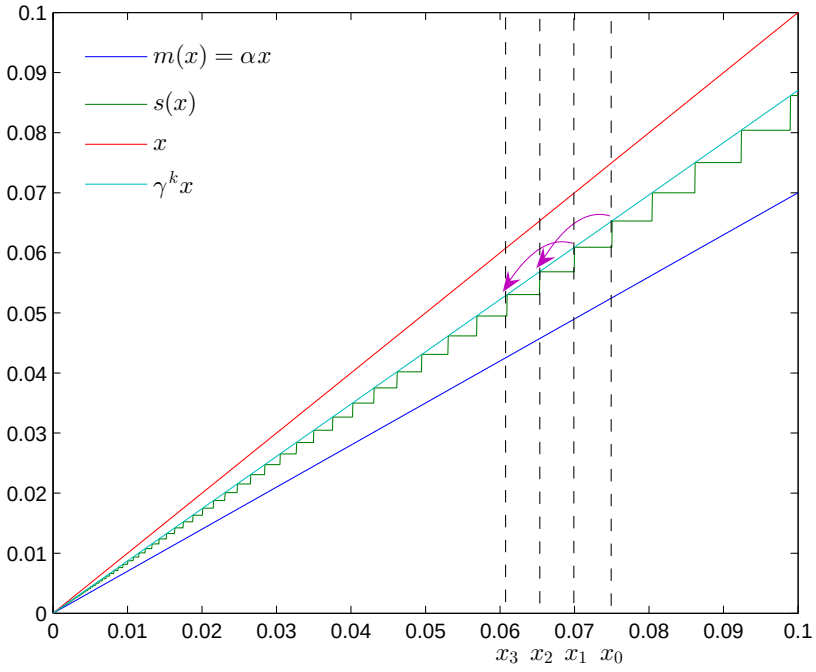
Figures 3, 4 and 5 illustrate 1 and 2-equilibria, as well as a continuous equilibrium, for the linear case.

Although these equilibria are highly structured, there is one degree of freedom in picking the sequence $(x_n)_n$, namely, the choice of x_0. Intuitively, since both m and our constructed s_k are linear, a rescaling of one linear k-equilibrium would be another k-equilibrium, so the exact placement of the points $(x_n)_n$ is arbitrary; what is pinned down by our construction is the ratio between the elements of each sequence.

While the case of linear m is useful to fix ideas, we want to know what equilibria exist for more general densities. It turns out that the existence of 1-equilibria is robust, and they have some additional desirable properties.

For the following result, we use the following definition. Let s be a 1-profile given by sequence

Figure 4: 2-equilibrium for $m(x) = 0.7x$, $\beta = 0.7$



$(x_n)_n$. We say that s is a *1-quasi-equilibrium* if, for all n, $m(x_n)$ is indifferent between x_{n+1} and x_{n+2} and prefers them to all other x_k (but not necessarily to other points outside of the sequence). Then

Proposition 4. *Consider the game restricted to $[x^*, x^{**}]$, and let $\underline{x} \in (x^*, x^{**})$. Then there is a 1-quasi-equilibrium $s_{\beta,\underline{x}}$ defined on $[x^*, x^{**}]$, such that $x_0 = \underline{x}$. In addition, s is weakly increasing, depends only on m (rather than on f), and the Median Voter Theorem holds for all $x \in [x^*, x^{**}]$.*

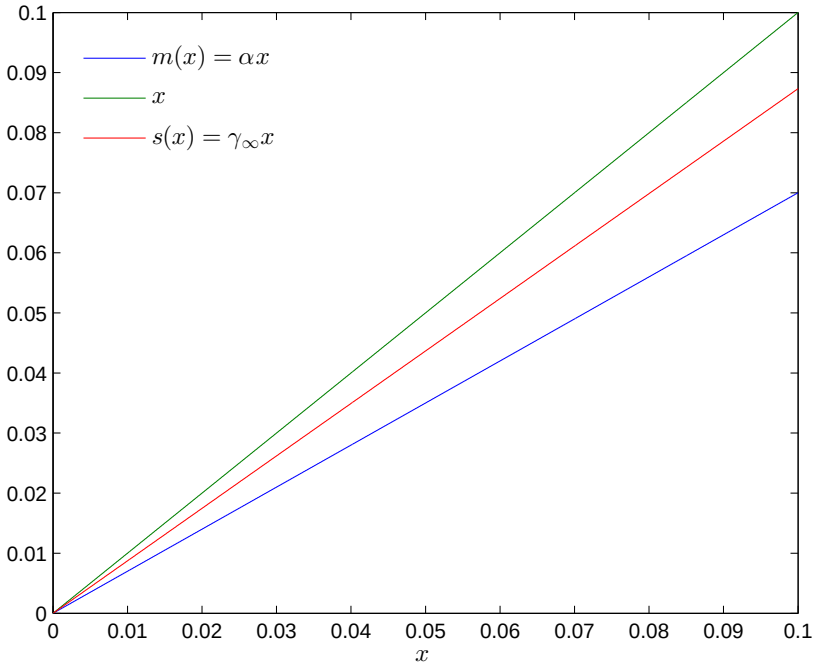
Conjecture: if the equilibrium in Proposition 7 has no jumps, then there exists $\overline{\beta} < 1$ such that, if $\beta < \overline{\beta}$, then $s_{\beta,\underline{x}}$ is a 1-equilibrium.

Conjecture: s is unique given β, $\underline{x}$.

In a 1-equilibrium, there is just one sequence $(x_n)_n$ such that $s(x_n) = x_{n+1}$. Besides always existing for any m, a 1-equilibrium is always guaranteed to extend nicely to the whole basin of attraction of x^*. This result is driven by the special structure of 1-equilibria, wherein a comparison between $S(x_n)$ and $S(x_{n+1})$ is essentially a comparison between x_n and $S(x_{n+1})$, which allows us to apply Lemma 3 and recover the Median Voter Theorem, even away from the steady state.

k-equilibria for $k > 1$ and continuous equilibria do not share these properties. While they–like any other equilibrium–can be extended to $[x^*, x^{**}]$ if defined on a neighborhood of x^* by Lemma 5, even when m is linear they may lose their properties beyond $x^* + d$, meaning that in a k-equilibrium some of its sequences may disappear, or in a continuous equilibrium discontinuities may appear, and so on. Since comparisons between $S(x_n)$ and $S(x_{n+1})$ are comparisons between interleaved

Figure 5: Continuous equilibrium for $m(x) = 0.7x$, $\beta = 0.7$



sequences, Lemma 3 does not apply and voters may have nonmonotonic preferences depending on their exit times.

Moreover, for general m, k-equilibria with $k > 1$ and continuous equilibria may or may not exist. To see this, consider the following example. Suppose that $x^* = 0$, β is high and $m(x) = \alpha x + \frac{\alpha}{4}\max(c - |x - x'|, 0)$, where c is small. This is similar to the linear case, but m has a small "bump" around x'.[13] Suppose further that s is a k-equilibrium ($k > 1$) such that $x_0 = x'$, and x_n is as in Proposition 3 for $n > 0$. Because $m(x_0)$ must be indifferent between x_k and x_{k+1}, $m(x_0)$ must also be the same as in the linear case, but as m is higher around x' than in the linear case, x_0 must be *lower* as a result of the bump. Then the break-even point between x_0 and x_1 must be *lower* than in the linear case because x_0 being lower makes it closer to $m(x_{-k})$, hence more attractive, so x_{-k}–who is indifferent between x_0 and x_1–is lower, hence more attractive himself to x_{-2k}. Continuing in this fashion, the subsequence $(x_0, x_{-k}, x_{-2k}, \ldots)$ is more attractive than it should be due to the bump; eventually, it may become so attractive that a voter $m(x_{-kr+1})$, supposed to be indifferent between $x_{-k(r-1)+1}$ and $x_{-k(r-1)+2}$, instead likes $x_{-k(r-1)}$ better than both, so no one votes for $x_{-k(r-1)+1}$ and s becomes a $(k-1)$-equilibrium beyond that point. This "domino effect" may continue until there is just one sequence left; for high β, this sort of dynamic makes k-equilibria for $k > 1$ unstable. In similar fashion, if we consider a continuous equilibrium in this example, the bump would generate a discontinuity around $s^{-1}(x_0)$.

Although there is no guarantee that continuous equilibria exist, even in a neighborhood of x^*,

[13] We can also construct smooth versions of this example.

we can use two arguments to find some. First, if m is analytic, we can construct an analytic continuous s locally around x^*:

Proposition 5. *If m is analytic at x^*, then there is at most one analytic function s defined in a neighborhood of x^* such that s supports a MPE on $[x^*, x^* + e)$ for some $e > 0$.*

Essentially, our argument is to show that the first-order conditions around the steady state uniquely pin down all the derivatives of s at the steady state. If the resulting Taylor series has positive radius of convergence, this is guaranteed to be a local solution to the problem.

Second, given a continuous successor function $s(x)$ with certain properties, we can reverse-engineer a median voter function $m(x)$ such that $s(x)$ supports an MPE for $m(x)$.

Given a path $S = (s_0, s_1, \ldots)$, let $W(S) = -(1-\beta)\sum_{t\geq 0}\beta^t s_t^2$. (As before, $E(S) = (1-\beta)\sum_{t\geq 0}\beta^t s_t$ is the average policy). Take a successor function s and resulting paths $S(y)$ for each y, and graph $(E(S(y)), W(S(y)))$ in $\mathbb{R}^2$. We say s is *well-behaved* if this graph is concave, i.e., if W is concave when taken as a function of E. Then

Proposition 6. *Let $\hat{x} = \min(x^{**}, x^* + d)$. Let $s : [x^*, \hat{x}] \to [x^*, \hat{x}]$ be continuously differentiable, increasing, such that $s(x^*) = s^*$; $s(x^{**}) = x^{**}$ if applicable; and $s(x) < x$ for $x \in (x^*, \hat{x})$. If s is well-behaved, then there is $m : [x^*, \hat{x}] \to [x^*, \hat{x}]$ continuous, increasing, satisfying $m(x^*) = x^*$; $m(x^{**}) = x^{**}$ if applicable; and $m(x) < x$ for $x \in (x^*, \hat{x})$, such that s supports a MPE on $[x^*, \hat{x}]$ given median voter function m.*

The condition that s be well-behaved is not as strange as it looks: it simply guarantees that, given any voter, the utility offered by different paths is concave in the average policy of said paths (note that this is a necessary condition for a continuous solution: if concavity did not hold, there would be some paths which are dominated for all voters, so any solution would have to jump over them, creating discontinuities). In addition, since linear functions of the form $s(x) = x^* + \gamma(x - x^*)$ are strictly well-behaved, functions that are approximately linear will also be at least locally well-behaved; we can construct more examples in this fashion. However, it is also easy to construct functions s with many oscillations that are not well-behaved, e.g., $s(x) = \gamma x + \rho \sin(x)\frac{x}{1+x}$, where $\rho < \gamma, 1-\gamma$ and $x^* = 0$.

In the next Section, we show how many of these results simplify when we consider the limit case of continuous time.

Continuous Time Limit

So far we have analyzed a model in discrete time. This seems more natural and intuitive, given that in most organizations where a policy is chosen through voting, decisions are made periodically (e.g., at weekly meetings, annual elections, etc.). However, as we have seen, it creates many technical problems, which ultimately arise from the need to jump through a finite set of policies while avoiding the rest. Especially in equilibria other than 1-equilibria, where many policy paths can run parallel to each other, this makes comparisons between different paths difficult and often non-monotonic.

In this Section, we show how these problems are solved in the continuous time limit. Formally, we suppose that decisions are now made increasingly often, so that each period $[t, t+1]$ is broken into j periods of length $\frac{1}{j}$, with discount factor between sub-periods $\beta^{\frac{1}{j}}$. Then we take the limit as $j \to \infty$, denoting $e^{-r} = \beta$.

Define $s^t(x)$ as the successor policy if we start at policy x and let an amount of time t pass. (A discrete successor policy $s(x)$ is no longer meaningful). Note that s must be additive in t, i.e., $s^t(s^{t'}(x)) = s^{t+t'}(x)$. Suppose s is smooth as a function of t. Then, in the limit, $I(x)$ chooses $s(x) = x$. A solution $s^t(x)$ must then solve

$$x = \arg\max_y \int_0^\infty re^{-rt} \max\left(C - (m(x) - s^t(y))^2, 0\right) dt.$$

(This condition makes $s(x) = x$ optimal for $m(x)$, rather than for $I(x)$. However, this is enough because x is higher than $S(x)$, so the Median Voter Theorem applies to this problem, by a continuous version of Lemma 3).

Continue working within an interval $[x^*, x^{**}]$ as before, where x^* is a stable steady state, so that $s^t(x)$ is decreasing in x and $s^t(x) \to x^*$ as $t \to \infty$. A second simplifying observation is that, in the continuous case, all paths must run through the same set of values at the same speed, just with a different starting point. Hence, the first order condition boils down to choosing a starting point that gives the same payoff as the average for the rest of the path. In other words,

$$u_{m(x)}(x) = C - (m(x) - x)^2 = \int_0^\infty re^{-rt} \max\left(C - (m(x) - s^t(x))^2, 0\right) dt = U_{m(x)}(x).$$

The following Lemma provides a useful characterization of s:

Lemma 6. *If $s^t(x)$ is continuously differentiable and decreasing in t; $s^0(x) = x$; $\lim_{t\to\infty} s^t(x) = x^*$; and $s^t(s^{t'}(x)) = s^{t+t'}(x)$ for all $t, t' \geq 0$, 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 $d(x, y) = \int_y^x e(z)dz$.*

Intuitively, $d(x, y)$ 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). Since the path is memory-free (i.e., the speed at which $s^t(x)$ decreases does not depend on previous policy values), this time can be expressed as an integral of the instantaneous delay $e(z)$ at each policy z.

In addition, note that differentiability of s^t is sufficient but not necessary: $d(x, y)$ and $e(z)$ may be well-defined, with $d(x, y)$ differentiable, even if $s^t(x)$ is not continuous in t. (It may be that $s^t(x)$ has some instantaneous jumps, which would correspond to $e(z) = 0$ for the policies that are jumped over). In what follows, we will call an equilibrium *smooth* if it can be expressed as per Lemma 6.

It turns out that there is a unique smooth equilibrium, and we can characterize it explicitly:

Proposition 7. *Let*

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

for $x \in [x^*, m^{-1}(x^* + d))$ *and*

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

otherwise, where $t^*(x) = d(x, m(x) - d)$. *Then there is a unique smooth MPE* $s^t(x)$, *given by* $e(x) = \tilde{e}(x)$ *whenever* $\tilde{e}(x) \geq 0$ *and* $u_{m(x)}(x) = U_{m(x)}(x)$, *and* $e(x) = 0$ *otherwise.*

Figure 6: Smooth equilibrium in continuous time

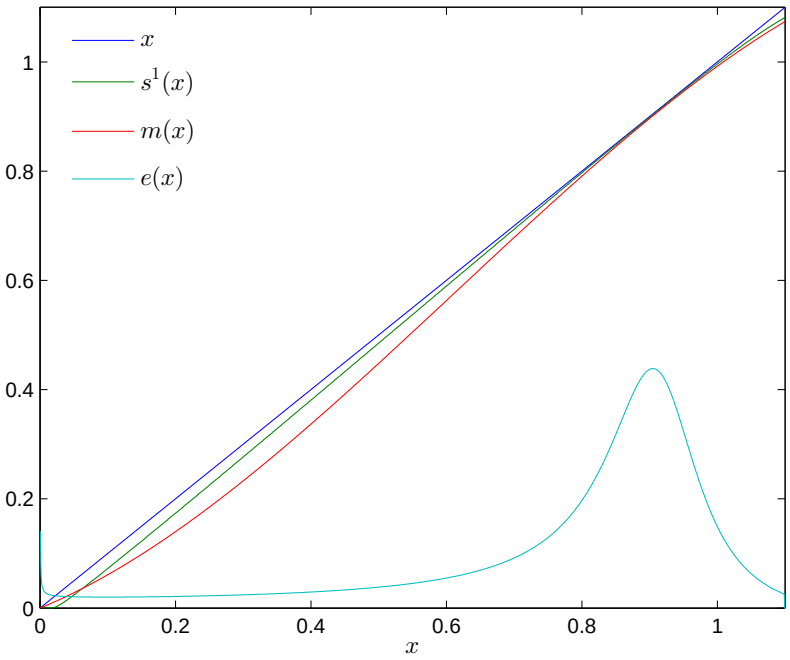


Figure 6 illustrates such an equilibrium. (Here $s^1(x)$ is $s^t(x)$ for $t = 1$; s^1 does not provide a full description of the equilibrium, but unlike $e(x)$ it is comparable with previous graphs). Intuitively, in the continuous time limit, $m(x)$ is indifferent about whether to include x at the beginning of the policy path $S(x)$; but the choice of $e(x)$ is crucial in making that indifference possible for slightly higher x's. The required delay $e(x)$ tends to be high when $m(x)$, $m'(x)$ and $m''(x)$ are high: if $m(x)$ is close to x, this means that the median voter is relatively happy to stay at x, and so $S(x)$ must move slowly away from x to be weakly preferable. In addition, when $x > m^{-1}(x^* + d)$ (i.e., when the current policy is high enough that the median voter expects to quit in finite time), the required delay tends to be lower because the current median voter expects not to suffer the full cost from the policy drifting too far away. As expected, the discount rate r plays a part, essentially as a rescaling factor: when r is low (agents are patient), the delay at all points becomes proportionally higher, so that the effective delay (measured against patience) remains constant. This result mirrors the discrete time case, where an increase in β would slow down convergence to x^* at a proportional

rate, but there a change in β had real effects since it also affected the effective frequency of voting (which has been made infinite here).

Finally, note that the formulas only make sense when their results are non-negative; this condition is violated in degenerate cases, where the equilibrium needs to jump through some interval of policies at infinite speed (i.e., with zero delay). Generally, this case arises when $m(x)$ is flat. For example, in the linear case where $m(x) = \alpha(x - x^*) + x^*$, it arises iff $\alpha \leq \frac{1}{2}$. To see why, consider the decision to be made by $m(x)$. When $\alpha > \frac{1}{2}$, $m(x)$ likes x better than the far end of the tail of $S(x)$, which approaches x^*; hence, he is happy to start at x given some expected delay in the path towards x^*. Analogously, in the discrete version of the model, $s(x)$ would be arbitrarily close to x as $\beta \to 1$. However, if $\alpha < \frac{1}{2}$, then $m(x)$ actually prefers x^* (and hence all points in $[x^*, x)$) to x, so there is no way that he will include x in the path–in fact, no points to the right of $x^* + 2\alpha(x - x^*)$ can be included. In the discrete version, this means $s(x) \leq x^* + 2\alpha(x - x^*)$ regardless of β. Hence, in the continuous time limit, the path travels to x^* instantly.

4 Discussion

In this Section, we further discuss the implications of our results. First, the model provides sharp predictions about the long-term behavior of policy paths: every policy (save for unstable steady states) is in the basin of attraction of some stable steady state, and will eventually drift to it, independently of the MPE in consideration. However, at the same time, the limit policy depends on the initial position of the club, so the model exhibits path dependence.[14] Strikingly, and unlike previous models in the literature, long-run policies are also independent of the discount factor (although β does affect the shape of the equilibrium path, it has no impact on the set of steady states, which is pinned down only by m).

Second, our characterization of steady states reflects a natural intuition: stable steady states should correspond to maxima of the density function f, while unstable ones correspond to minima. Formally:

Lemma 7. *If x is a stable (unstable) steady state, then $(x - d, x + d)$ contains a local maximum (minimum) of f.*

In particular, if f is increasing (decreasing) everywhere, there will be a unique steady state close to 1 (-1); and, if f is symmetric and single-peaked–say, a truncated normal distribution–then 0 will be the unique steady state. In general, then, clubs drift over time to policies around which the density of voters is higher.

In terms of welfare analysis, this is relatively good news, yet not optimal, for two reasons. First, while a stable steady state x^* must be close to a local maximum of f, it does not have to maximize the number of members $F(x^* + d) - F(x^* - d)$, nor the sum of their utilities or other such metrics,

[14] Note that the multiplicity of MPEs and the multiplicity of long-run outcomes are different and unrelated phenomena; a single successor function s can support the entire set of long-run limit policies if we vary the initial policy, and conversely, even with multiple MPEs we may have a unique limit policy, if there is only one fixed point of m.

even locally. Second, even if it were a local optimum in some sense, the choice of steady state is still based on the (arbitrary) starting policy; it is possible that some steady states yield more social welfare than others (i.e., they serve more members) but there is no guarantee that the club will converge to the former.

We may also be interested in the impact of policy drift on extremism: does convergence to steady states favor moderates? Or can it create situations where extremist factions "capture" the club?

At face value, the above result suggests that policy drift encourages centrist policies: if moderate voters are abundant, there will be a stable steady state close to 0 with a large basin of attraction. Yet there are several cases where the club can exhibit extremism, even the kind we would consider dysfunctional (i.e., when the club is extremist but most of the population is moderate).

First, if most of the voters are moderates but f has local maxima near the extremes, there might be multiple steady states, including some near the extremes. This is especially easy to support if d is low, i.e., if the nature of the club is to attract a relatively small niche of members, so that a club with an extreme policy would attract the nearby extremists (who are locally strong) and not be disrupted by a large mass of moderates. Of course, if f increases towards the endpoints and decreases towards the center, then the only stable steady states would be at the extremes.

Second, even when the distribution has a single steady state, its location may be "unstable" when f is close to being 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 similar, and close to uniform, but f_1 has a unique steady state close to -1, while f_2 has one close to 1, and f_3's is equal to 0. Hence, small demographic changes can have a dramatic impact on the equilibrium policy of the club. This example contrasts with models of voting with a fixed population, where a small change in the density function would produce a small change in the median voter and chosen policies.

Third, and most importantly, the tendency towards moderate policies depends on the assumed symmetry of preferences. Namely, in our basic model, a policy x always induces the interval $(x - d, x + d)$ to become members: there is no distinction based on whether x is a right-wing or left-wing policy to begin with, whether it is extreme or moderate, etc. In particular, it is equally bad to be in a club that is too moderate as it is to be in a club that is too extremist.

For an example where this assumption may be unreasonable, suppose that the club in question is a local nationalist organization. Moderate agents want to join to engage in benign activities, such as getting together for traditional meals, publishing a local newspaper for their community, etc. Hard-line nationalists want the club to organize harassment or attacks against foreigners living in the area, but they would still rather join the club than not, even if it is too moderate for their tastes. On the other hand, moderate agents would want to leave the club if it turns xenophobic.

Formally, if preferences are no longer symmetric, $I(x)$ would no longer be of the form $(x-d, x+d)$, but would become some other interval $(x - d_-(x), x + d_+(x))$; this, in turn, would induce a different median voter function $m(x)$. Reflecting our example above, suppose that preferences are such that $I(x) = [x - d, 1]$ if $x > 0$, $I(x) = [-d, d]$ if $x = 0$ and $I(x) = [-1, x + d]$ if $x > 0$, or some

continuous approximation thereof. Then, even if f is symmetric and single-peaked at 0, $x = 0$ is an *unstable* steady state. If d is low and f is relatively flat, there are two stable steady states, close to -1 and 1 respectively, so the resulting policy will always be extreme. The moral of this example is that, if extremist members are more willing to join the club, they may end up capturing it even if they are a minority, and vice versa.

In particular, this raises questions about the merits of social discouragement as a tool to prevent extremism. A society that becomes less tolerant towards undesirable behavior (e.g., by punishing it with ostracism, boycotts, or by making it illegal) may dissuade people from engaging in said behavior individually, as well from joining the "wrong" clubs. But, if the punishments are only strong enough to make moderate members quit, they will lead existing clubs to become more extremist.

A related point is that, if current members can somehow manipulate the pool of potential members in addition to choosing the current policy, they may want to do so–and this may substantially affect the club's trajectory. For example, suppose that a city is divided into two districts, A and B, and a club can vote to only admit members from district A rather than the whole city.[15] If A's demographics differ from B's, replacing $f_A + f_B$ with f_A may lead to a completely different set of steady states.

5 Multiple Clubs

In this section, we analyze an extension of the model where multiple clubs compete for potential members. This raises important questions: how are clubs affected by nearby clubs? Are their sets of members or possible long-term policies changed? Will clubs distribute themselves efficiently along the preference distribution? Can the presence of one club force another one off of a certain demographic, or make it disappear entirely?

For tractability, we first consider the case where a fixed number $k > 1$ of clubs is given and focus on analyzing the steady states. Afterwards we briefly discuss the dynamics leading to the steady state, as well as the possibility of agents creating new clubs.

Multi-Club Steady States

In this case, there is a set of k clubs with initial policy positions $x_{1,0} < x_{2,0} < \ldots < x_{k,0}$. For simplicity, assume that this order is always maintained. First, we will characterize the sets of members for each club, given current policies: $I(x_1, x_2, \ldots, x_k) = (I_1(x_1, \ldots, x_k), \ldots, I_k(x_1, \ldots, x_k))$. This is the multi-club equivalent of $I(x)$. We assume that agents can only belong to one club, so $I_1, \ldots, I_k$ are always pairwise disjoint and $I_l \subseteq [-1, 1]$. It turns out that

$$I_l(x_1, \ldots, x_l) = \left(\max\left(\frac{x_{l-1} + x_l}{2}, x_l - d \right), \min\left(\frac{x_l + x_{l+1}}{2}, x_l + d \right) \right).$$

[15] In our model, the population is exogenously given and we always allow free entry, but it can be extended to this case.

In other words, if the next club to the left has policy $x_{l-1} \leq x_l - 2d$, the two clubs do not interfere, and the leftmost member of I_l is $x_l - d$. Otherwise, each agent goes to the closest club and the one with bliss point $\frac{x_{l-1}+x_l}{2}$ is indifferent. The other side of the interval is analogous.

We formalize this notion that clubs' bases of support may overlap with the next definition. A *cluster* of clubs is a subset $\{i, i+1, \ldots, j\}$ of consecutive clubs such that $x_l - x_{l-1} < 2d$ for $l = i+1, \ldots, j$ but $x_i - x_{i-1} \geq 2d$ and $x_{j+1} - x_j \geq 2d$. In other words, all voters with bliss points between x_i and x_j belong to one of the clubs, but voters at $x_i - d$, $x_j + d$ are indifferent about being in any club.

Next, we extend our definition of steady states to the multi-club case. We say that $x_1 < \ldots < x_k$ form a *steady state* if $m(I_l(x_1, \ldots, x_k))) = x_l$ for all l. In other words, given the intervals I_l defined above, the median voter in each I_l has no interest in moving.[16] Note that this condition is different from $m(x_l) = x_l$, as $I_l(x_1, \ldots, x_k)$ is different (generally smaller) than $I(x_l)$, except in the case when $\{l\}$ is a cluster.

Our main result in this section characterizes the possible steady state distributions, cluster by cluster:

Proposition 8. *Let $(x_1, \ldots, x_k)$ be a steady state. If $\{i\}$ is a cluster, then x_i is compatible with steady state iff $m(x_i) = x_i$. If clubs $\{i, i+1, \ldots, j\}$ form a cluster and $j > i$, then $m(x_i) > x_i$ and $m(x_j) < x_j$; in particular, the interval $[x_i, x_j]$ must contain a stable steady state x_0 of the single-club game.*

Moreover, if f is given by a non-constant polynomial, then given a cluster size $j - i + 1$ and a stable steady state x_0, there is only a finite number of policy $(j-i+1)$-tuples $(x_i, x_{i+1}, \ldots, x_j)$ compatible with steady state and containing x_0 (i.e., such that $x_i \leq x_0 \leq x_j$). Alternatively, if f is strictly log-concave in $[x_0 - 2(j-i+1)d, x_0 + 2(j-i+1)d]$, then there is a unique valid $j-i+1$-tuple $(x_i, x_{i+1}, \ldots, x_j)$ contained in $[x_0 - 2(j-i+1)d, x_0 + 2(j-i+1)d]$.

Note that the Proposition says nothing about the distribution of clubs into clusters: this is generally arbitrary, and results in multiple steady states.[17] For example, suppose there are three single-club steady states $x_1 < x_2 < x_3$, where x_1, x_3 are stable and x_2 is unstable. Assume that there are two clubs and the steady states are far away, and f is log-concave in a large interval around x_1, as well as x_3. Then there is a two-club steady state where both clubs form a cluster around x_1; one where they form a cluster around x_3; and three cases where the clubs are separate and occupy two separate single-club steady states. Hence, there are multiple possibilities even if there is a unique r-cluster for each r around each single-club steady state.

However, if f is *globally* strictly log-concave, then there is a single multi-club steady state: there must be a unique single-club steady state, so all k clubs must be clustered around it, and there is a unique position for the cluster.

[16] Note that, in the single-club case, the definition of steady state was that $s(x) = x$ was compatible with MPE; the fact that steady states were fixed points of m was a result. Here, we take it as a primitive for tractability.

[17] The caveat is that, if two steady states are close to each other, then two clusters centered at each would bump into each other and become a single cluster if they are too large. This puts a joint constraint on the size of adjacent clusters.

These results have important implications for welfare analysis. First, as in the single club case, clubs will be centered around stable steady states, which are higher density areas; but there is no guarantee that, given several steady state positions giving different welfare, they will center around the "best" one. Second, there is now an additional inefficiency that comes from clubs bunching together: a club standing next to another creates a welfare loss, because it gives multiple options to certain agents in the population (who can only take advantage of one) while leaving other agents without any club, but clubs do not internalize this.

Dynamics

Here we briefly discuss some issues that arise in the dynamics of the multi-club case. The general gist is that, when clubs cluster together, it makes each club's median voter less responsive to changes in policy, which in turn induces faster convergence.

Remember that, in the single-club case, the median voter $m(x)$ is given by the condition $F(x+d) - F(m(x)) = F(m(x)) - F(x-d)$. In particular, then,

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

Suppose now that i is the first element of a cluster. Then $I_i = (x_i - d, \frac{x_i + x_{i+1}}{2})$. Taking club $i+1$'s policy as fixed, we instead have

$$m_i'(x_i) = \frac{f(\frac{x_i+x_{i+1}}{2})}{4} + \frac{f(x_i - d)}{2}.$$

If l is in the middle of a cluster, then

$$m_l'(x_l) = \frac{f(\frac{x_l+x_{l+1}}{2}) + f(\frac{x_l+x_{l-1}}{2})}{4}.$$

In particular, there are many possible densities f for which $\frac{1}{2} < m'(x) < 1$, but $m_i'(x) < \frac{1}{2}$, and even more for which $m_l'(x) < \frac{1}{2}$ (taking values of x_{i+1}, x_{l-1}, x_{l+1} consistent with the cluster). In other words, $m_i'(x)$ is effectively lower than in the single-club case because, if i moves to the right, it will not pick up as many right-wing members as it would if alone, as it has to compete for them with club $i+1$; the effect is doubled for clubs in the interior of a cluster.

As we saw in Proposition 7, once $m'(x) < \frac{1}{2}$, convergence to the steady state is much faster–instant in the continuous time limit–because the median voter prefers the steady state policy to anything close to x. Hence equilibria with instant convergence in the multi-club case will be much more prevalent. (Although our argument is built assuming that other clubs stay constant, this is consistent with an instant convegence equilibrium, as other clubs always expect the rest to converge instantly, so it makes sense for them to converge instantly as well, and stay fixed thereafter).

Asymmetric Clubs

Here, we suggest how the model's results might change when asymmetric clubs interact. In other words, if two adjacent clubs offer different payoff profiles to prospective members (in particular, with different functional forms), which one will succeed in attracting the marginal members between them? And how will this affect their policies in the steady state?

Consider the following benchmark case. Suppose that there are two clubs 1, 2, and f is symmetric and log-concave, so there is a unique stable steady state at $x = 0$ and both clubs are clustered around it.

Now assume that, if club 1 implements policy x_1, it generates utility $a_1 > 0$ for members with bliss points in $(x_1 - d_1, x_1 + d_1)$ and -1 to the rest. On the other hand, given a policy x_2, club 2 provides utility $a_2 > 0$ to $(x_2 - d_2, x_2 + d_2)$ and -1 to the rest. Suppose that $x_1 < x_2$. What would "competition" between these clubs look like?

In this stark example, it turns out that the deciding factor is which club provides more utility to its members, regardless of whether they have wide appeal or not. Concretely, if $a_1 > a_2$, then in the steady state $x_1 = 0$ and $x_2 > d_1$. The reason is simple: since $a_1 > a_2$, all voters who get positive utility from both clubs choose 1; hence $I_1(x_1) = (x_1 - d_1, x_1 + d_1)$ always. Thus, for 1 to be in steady state, we need $m_1(x_1, x_2) = m(x_1) = x_1$, so $x_1 = 0$. Then 1 captures all the members in $[-d_1, d_1]$ and x_2 must be to the right of this point.[18]

If $a_1 > a_2$ and $d_1 > d_2$ (i.e., 1 is superior to 2 in every way) this may seem natural, but the result still holds even if $d_1 < d_2$. In other words, a niche club can displace a mainstream club.

Of course, the conclusion in this example hinges on the assumption that utility is either a_i or 0. Other functional forms lead to more complex interactions. For example, suppose now that club 1 gives utility A_1 to voters within distance d_1 and utility a_1 to voters within D_1, where $D_1 > d_1$ and $A_1 > a_2 > a_1$; club 2 is as before. Then, if the overlap between the clubs is relatively small (e.g., if f is close to uniform, so the pressure to move towards the steady state is weak) 2 would now displace 1 at the margin: although 1 still has a strong niche, its appeal to its marginal members is weaker than club 2's. However, if the overlap is large (e.g., if f is steeply concave, making the clubs aggressively cluster together) so that 1's core overlaps with 2, then 1 would again displace 2.

Free Entry

In this section, we discuss an extension of the model where agents (or groups of agents) can agree to create new clubs.

We will sketch a very simple example, as the general case is very complicated. Suppose that there are no clubs initially. Out of the entire population, there is a single agent $\overline{x}$ capable of creating a club, at cost K. (For instance, $\overline{x}$ is rich, or has the right connections). If the club is created, $\overline{x}$ gets to decide the initial policy of the club. The question is: when will $\overline{x}$ decide to create the club, and when would it be socially optimal to do so?

[18] The result presumes that $d_1 < 2d_2$ and $d_2 < 2d_1$, to avoid the pathological case where one club's members could "wrap around" the other's.

To fix ideas, suppose $\overline{x} \in [x^*, x^{**}]$, where x^*, x^{**} are consecutive steady states and x^* is stable. If $\overline{x}$ indeed creates the club, he will choose initial policy $x_0 = s(m^{-1}(\overline{x}))$ (note that this is generally higher than $\overline{x}$: he will make the club more extreme than his own preference, taking into account that it will later drift towards the steady state). Given this, it is optimal for him to create the club iff $U_{\overline{x}}(S(s(m^{-1}(\overline{x})))) \geq K$.

In the continuous time limit, this condition boils down to $\frac{u_{\overline{x}}(m^{-1}(\overline{x}))}{1-\beta} \geq K$. Hence, $\overline{x}$'s willingness to create the club depends on how lopsided the density is around him. If f is close to flat (or, alternatively, if $\overline{x}$ is close to the steady state), so that $\overline{x}$ is close to $m^{-1}(\overline{x})$, the club will move away slowly so he is happy to create it (in the limit, if $\overline{x}$ *is* the steady state, his condition is $\frac{C}{1-\beta} \geq K$). On the other hand, if f is very lopsided, the drift will occur relatively quickly, lowering $\overline{x}$'s interest (potentially to zero if $m(x)$ approaches $x - d$).

On the other hand, the social planner would not care about the speed at which the club drifts, since by drifting it brings benefits to other members (which $\overline{x}$ cannot internalize). Thus, there is a discrepancy between private and social value, insofar as the prospective founder cannot guarantee that the club will stay where it was intended to.

This will generally lead to under-provision of clubs. The problem persists if some small interval (e.g., $(\overline{x} - \eta, \overline{x} + \eta)$) of voters can cooperate to create the club, since they all face similar incentives. Even if all agents have resources they could contribute, it may not be enough to solve the problem: the theoretical solution–where all benefits are internalized–would involve future members contributing now to the club's creation, anticipating that it will later move towards them, but this would be unimplementable in practice.

Even if several agents have the power to create clubs, the same qualitative issue arises. However, it will be mitigated if founders are abundant and well-distributed enough that clubs can always be founded close to their intended long-term positions. On the other hand, if there are many founders, this can also reduce other inefficiencies found previously: e.g., stable steady states with more voters around them will be more likely to have clubs created nearby.

6 Conclusions

In this paper we have studied a model of policy choice in clubs with endogenous membership, which is relevant to a variety of organizations in the real world. Our analysis yields interesting technical results as well as important empirical implications.

On the technical side, the model predicts that the club's policy will drift in the long run towards a steady state of the preference distribution, that is, a policy attracting a set of members that have no inclination to change the policy further, even myopically. This result is driven by the feedback effect at the heart of the model: if a policy attracts a majority of voters wanting to increase it, the policy will drift up, attracting voters with even higher bliss points, and vice versa. Crucially, this drift is slowed down by the concern that further drift will occur, but it never stops completely. This result, enabled by a continuous policy space, is true for any discount factor and contrasts with

other models in the literature, where dynamic concerns often make stable states out of policies that are myopically undesirable.

Despite this strong prediction about long-run behavior, two factors can still generate a variety of outcomes, even if we keep the density of potential members fixed. First, depending on the preference distribution, there may be one or several steady states. In the latter case, each has a basin of attraction and the club's initial policy determines where it will drift towards in the long run; in other words, there is path dependence. Second, in the discrete time version of the model, there can be many equilibrium paths, all leading to the same steady state but exhibiting discontinuities of different sizes (or none at all) on the path.

On the applied side, the model tells us when we can expect organizations to become mainstream or drift towards extremism. With symmetric preferences, steady states are near maxima of the density function; in particular, if moderates are the majority, there will be a moderate steady state–but there may be others depending on the shape of the density at the tails. On the other hand, extreme policies are much easier to support if preferences are asymmetric, with extremists being more willing to belong to the club than moderates.

There are three extensions of the model which we discuss briefly, but which deserve a complete analysis in future work. First, several interesting issues arise when we allow multiple clubs to interact. We only provide a characterization of the natural steady states in this case. Although it is likely that an analog of Proposition 1 holds–that is, the clubs collectively converge to a natural steady state in the long run–a full characterization of the dynamics is needed to prove this. As we note in our discussion, convergence is likely to be much faster than in the single-club case once clubs become clustered, and we would like to have a general condition on when this is the case.

In addition, if clubs offer different payoff profiles, very complex interactions can arise between them. We would like to understand: under what general conditions does one club displace another? This issue arises very often in practice, as competing organizations are rarely copies of each other; they have some built-in structure, culture, and institutions which affect how wide their base of support can be, and how committed their members are. For example, two political activism groups may differ on their policy prescriptions on a left-right spectrum, but one has a culture of tolerance while the other caters to fanatics: for a given policy, the latter group attracts fewer members, but those are more committed.

Secondly, although our analysis explains the evolution of already-existing organizations, these must be created in the first place, and we know that the policy where a club begins matters a great deal. Hence the explanatory power of the model would be improved by a general model of entry decisions when clubs are created endogenously. In particular, such a model would have to account for free-riding issues (if several nearby agents can create similar clubs, who will pay the cost of doing it?) as well as the cost of maintaining the club (would it charge a membership fee? If so, how would it be distributed among the members, and how would it affect membership?).

Third, the assumptions that members can freely enter and exit, and choose the club's policy by majority vote, constitute a useful benchmark but are rarely exactly true in practice. On the

one hand, there is usually some fixed cost of entering, and sometimes a cost of leaving (e.g., imagine a situation where migrants choose among several cities in a new country, but after settling in cannot easily move again). On the other hand, most organizations have leaders who choose policies, and members either vote for the leaders or influence them through non-electoral channels, rather than deciding policy directly. Even when voting, members may be weighted by seniority or other categories (e.g., consider a university where administrators, faculty and students have different standing). It would be useful to extend the paper's main results to a more general setting encompassing all these cases. In turn, allowing for different distributions of power within the club would allow for new comparative statics. For instance, if more senior members have more votes, does that slow down convergence to the steady state? And if the leader has agency, can he–by choosing the right policies–reshape the electorate to fit him instead of the other way around, as in the case of Mayor Curley?

A Appendix

Proofs

Proof of Lemma 2. We can write

$$U_\alpha(S) - U_\alpha(T) = \sum_t \beta^t \left(u_\alpha(s_t) - u_\alpha(t_t)\right).$$

Then

$$\frac{\partial(U_\alpha(S) - U_\alpha(T))}{\partial\alpha} = \sum_t \beta^t \frac{\partial(u_\alpha(s_t) - u_\alpha(t_t))}{\partial\alpha},$$

where the terms are positive by increasing differences since $s_t \geq t_t$. □

Proof of Lemma 3. If $\alpha \geq x$, clearly α prefers x to S since the comparison holds point-wise. When $\alpha \in (x, x+d)$ the pointwise inequality is strict for at least some terms. Now consider $\alpha \in (x-d, x)$. We can write

$$U_\alpha(x) - U_\alpha(S) = \sum_{t\in T} \beta^t \left(u_\alpha(x) - u_\alpha(s_t)\right) + \sum_{t\notin T} \beta^t u_\alpha(x),$$

where T are the times for which $u_\alpha(s_t) > 0$. Then

$$\frac{\partial(U_\alpha(x) - U_\alpha(S))}{\partial\alpha} = \sum_{t\in T} \beta^t \frac{\partial(u_\alpha(x) - u_\alpha(s_t))}{\partial\alpha} + \sum_{t\notin T} \beta^t \frac{\partial u_\alpha(x)}{\partial\alpha},$$

where the first set of terms is positive by increasing differences since $x \geq s_t$, and the second term is positive because $\alpha < x$. (We can also check that the derivative must be *strictly* positive so long as $S \neq (x, x, \ldots)$). Hence, if there is $\alpha_0 \in (x-d, x)$ that is indifferent between x and S, then voters in (α_0, x) must strictly prefer x, while voters in $(x-d, \alpha_0)$ must strictly prefer S. On the other hand, voters in $[-1, x-d)$ weakly prefer S since they get utility 0 from x.

If there is no such α_0, since $U_x(x) > U_x(S)$, by continuity all voters in $(x-d, x)$ must strictly prefer x. □

Proof of Lemma 4. Suppose that some $S(y)$ is not monotonic. Let $\underline{y} = \inf(S(y))$ and $\overline{y} = \sup(S(y))$. We consider two cases:

Case 1: $S(y)$ attains $\underline{y}$ or $\overline{y}$. In other words, $\exists k \in \mathbb{N}$ such that $s^k(y) = \overline{y}$ or $s^k(y) = \underline{y}$. Suppose WLOG that the former is true. Then there is a $k \in \mathbb{N}$ such that $s^{k-1}(y) < \overline{y}$, $s^k(y) = \overline{y}$ and $s^{k+1}(y) < \overline{y}$.[19]

We then consider the decision made by voters in $I(s^{k-1}(y))$ and in $I(s^k(y))$. Since $s^k(y)$ is the Condorcet-winning policy in $I(s^{k-1}(y))$, in particular at least half of the voters must prefer it to $s^{k+1}(y)$. At the same time, $s^{k+1}(y)$ is Condorcet-winning in $I(s^k(y))$, so in particular at least half of the voters prefer it to $s^k(y)$.

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

Consider now the intervals $(s^{k-1}(y)-d, s^{k-1}(y)+d)$ and $(s^k(y)-d, s^k(y)+d)$. We can divide them into $A=(s^{k-1}(y), s^k(y)-d)$, $B=(s^k(y)-d, s^{k-1}(y)+d)$, $C=(s^{k-1}(y)+d, s^k(y)+d)$. [20] Voters in B are present in both cases so they contribute the same votes; voters in A are present only in the first vote; C is only present in the second vote. (Note: Lemma 3 guarantees that there is at most one indifferent voter in B, so we don't have to worry about a set of positive measure being indifferent and voting differently in each case).

Note also that a voter will prefer $S(s^k(y))$ to $S(s^{k+1}(y))$ iff he prefers the constant policy $s^k(y)$ to the path $S(s^{k+1}(y))$. Now, voters in A can never prefer $s^k(y)$ to $S(s^{k+1}(y))$ because by construction $s^k(y)$ gives them zero utility.[21] On the other hand, voters in C with bliss points $\alpha > s^k(y)$ will always prefer $s^k(y)$ to $S(s^{k+1}(y))$. If $s^{k-1}(y)+d > s^k(y)$, we have a contradiction: all voters in C prefer $s^k(y)$ and all voters in A prefer $s^{k+1}(y)$, so $B\cup C$ has more votes for $s^k(y)$ and fewer for $s^{k+1}(y)$ than $A\cup B$. If not, consider voters in $(s^{k-1}(y)+d, s^k(y))$. By Lemma 3, there is α_0 in $(s^k(y)-d, s^k(y))$ that is indifferent. If $\alpha_0 < s^{k-1}(y)+d$, then all voters in C prefer $s^k(y)$, and we have the same contradiction. If $\alpha_0 > s^{k-1}(y)+d$, then all voters in $A\cup B$ prefer $S(s^{k+1}(y))$, a contradiction, since a majority in $A\cup B$ must prefer $S(s^k(y))$.

Case 2: $S(y)$ never attains its infimum nor its supremum. Then there must be a subsequence $s^{k_i}(y)$ (with increasing k_i) such that $s^{k_i}(y) \underset{i\to\infty}{\longrightarrow} \overline{y}$. Given this subsequence, construct a sub-subsequence $s^{k_{i_j}}(y)$, such that $s^{k_{i_j}}(y) \underset{j\to\infty}{\longrightarrow} \overline{y}$ and $s^{k_{i_j}-1}(y) \underset{j\to\infty}{\longrightarrow} s_*^{-1}$ for some limit s_*^{-1}. Essentially, we take a subsequence such that the elements of the original sequence immediately preceding the k_{i_j} are also converging to some limit, not necessarily $\overline{y}$ (we can always do this because all the $s^k(y)$ are in $[-1,1]$, which is compact). Iterating this, we can construct a nested list of subsequences $s^{k_{im}}(y)$ such that k_{im} is increasing in i for each m; $K_m=\{k_{im}: i\geq 0\}\supseteq K_{m'}$ for $m'\geq m$; and, for each m, $s^{k_{im}+r}(y) \underset{i\to\infty}{\longrightarrow} s_*^r$ for any $r\in\{-m,\ldots,m\}$, where s_*^r is independent of m and in particular $s_*^0=\overline{y}$.

Now we consider four sub-cases. First, 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. Then consider the decision made by $I(s^{k_{im}+\underline{r}}(y))$ vs. the decision made by $I(s^{k_{im}+\overline{r}-1}(y))$, for m high enough. In the limit, these decisions imply that a weak majority in $I(s_*^{\underline{r}})$ prefers $\overline{y}$ to $S(s_*^{\overline{r}})$, while a weak majority in $I(\overline{y})$ prefers $S(s_*^{\overline{r}})$ to $\overline{y}$. (Here $S(s_*^{\overline{r}})$ is the limit of the paths $S(s^{k_{im}+\overline{r}}(y))$ as $i,m\to\infty$). This leads to a contradiction by the same arguments as in Case 1, since the path $S(s_*^{\overline{r}})$ is strictly to the left of $\overline{y}$.

In the fourth sub-case, $s_*^r=\overline{y}$ for all r. In other words, the sequence spends arbitrarily long times near $\underline{y}$ and $\overline{y}$ (it must be true for both boundaries, as otherwise one would fall under the first case and we would have a contradiction). We first prove the following sub-lemma: it must be that

[20] It must be that $s^{k-1}(y)+d > s^k(y)-d$. If not, it would mean that all the voters in $I(s^{k-1}(y))$ will get utility 0 during the immediate next period when $s^k(y)$ is implemented, so they would always switch to $s^{k+1}(y)$, since the total payoff of the continuation must be positive for a majority.

[21] Voters in A could be almost indifferent if the rest of the path stayed close to $s^k(y)$, so that they never joined the club again and got zero utility either way. However, in that case, the tie-breaker is that they still get the payoff of their immediate choice for a period of length ϵ, so they would prefer $s^{k+1}(y)<s^k(y)$.

$m(y) = y$ for all $y \in [\underline{y}, \overline{y}]$.

To do this, take any $y_0 \in (\underline{y}, \overline{y})$ and construct a subsequence $s^{k_n}(y)$ such that: $s^{k_n}(y) > y_0$ but $s^{k_n+i}(y) \leq y_0$ for $i = 1, \ldots, n$ and there are n consecutive elements $s^{k_n+r+i}(y) < \underline{y} + \frac{1}{n}$ for some $r \geq 0$ and $i = 1, \ldots, n$ before s reaches above y_0 again. In other words, $s^{k_n}(y)$ are the last elements of the sequence above y_0 before the sequence goes near $\underline{y}$ for a long time. Now take iterated subsequences so that $s^{k_n+i}(y)$ has a limit s^i_* for all i. Clearly $s^0_* \geq y_0$ and $s^i_* \leq y_0$ for all $i > 0$.

Consider the decision made by $I(s^{k_n}(y))$. Almost all voters $\alpha \geq s^{k_n}(y)$ strictly prefer staying at $s^{k_n}(y)$ over going to $s^{k_n+1}(y)$,[22] so if $s^{k_n+1}(y)$ is the Condorcet winner, $m(s^{k_n}(y)) \leq s^{k_n}(y) + \epsilon_n$, where ϵ_n goes to 0 as n goes to ∞. Now, if $s^0_* = y_0$, then $m(y_0) \leq y_0$. If $s^0_* > y_0$, then a weak majority in $I(s^0_*)$ prefers the continuation (which is below y_0) to s^0_*, hence $m(s^0_*) \leq \frac{s^0_* + y_0}{2}$. Moreover, we can do the same argument with subsequences that go near $\overline{y}$.

Now suppose that $m(y_0) \neq y_0$ for some $y_0 \in [\underline{y}, \overline{y}]$. Let (y', y'') be a connected component of $m|^{-1}_{[\underline{y},\overline{y}]}(\mathbb{R} - \{0\})$ of maximal size, and suppose WLOG that $m(y) > y$ for all $y \in (y', y'')$. Apply the above argument to $y = y' + \nu$ for small ν. Since $m(y) > y$, it must be that the associated subsequence constructed above has $s^0_* > y$, and $m(s^0_*) \leq \frac{s^0_* + y}{2}$. Hence $\frac{s^0_* + y}{2} \geq y''$, so $s^0_* - y'' \geq y'' - y$. Take a subsequence with $\nu \to 0$ and $s^0_*(\nu)$ converging to a limit s^0_{**}. This satisfies the above, and moreover, by the strict monotonicity of m, $y'' \leq m(m(s^0_{**})) < m(s^0_{**})$, so $s^0_{**} - y'' > y'' - y'$. Since m is strictly monotonic, we must have $m(y) > y$ for all $y \in (y'', s^0_{**})$, which contradicts the maximality of (y', y''). Hence the only case left is if $m(y) = y$ for all $y \in [\underline{y}, \overline{y}]$.

For this last case, we employ the following

Lemma 8. *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 $x' - x > d$.*

Proof. First, note that it is enough to check the case when T is contained in $[x, x']$: if not, then 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. Clearly T' is weakly better for both x and x' than T.

Now, if $x' - x \leq d$, then both x and x' derive non-negative utility from all elements of T'. But then, if $E(T) < y$, then x' must strictly prefer S, since it has higher mean and no variance; if $E(T) > y$, then x must strictly prefer S; and if $E(T) = y$, then both x and x' must strictly prefer S since T cannot be constant, hence it has positive variance. □

Intuitively, the Lemma says that non-constant paths cannot appeal to too many voters on both sides of a constant path. Now, take a subsequence $s^{k_n}(y)$ that is above y_0 as before, and such that after $s^{k_n}(y)$ the sequence stays below y_0 for at least n periods and also stays near $\underline{y}$ for n consecutive periods before returning above y_0. Take $y_0 = \overline{y} - \nu$ with ν small. Consider the decision made by

[22] Voters $\alpha \geq s^{k_n}(y)$ prefer $s^{k_n}(y)$ to any path contained in $[-1, s^{k_n}(y))$. The path $S(s^{k_n+1}(y))$ is not strictly contained in there, but it only has elements higher than $s^{k_n}(y)$ after an arbitrarily high number of periods spent close to $\underline{y}$. Hence, for any $\delta > 0$ fixed, voters in $[s^{k_n}(y), s^{k_n}(y) + d - \delta]$ prefer $s^{k_n}(y)$ for n high enough. The exception is that voters very close to $s^{k_n}(y) + d$ are almost indifferent between $s^{k_n}(y)$ and lower policies, so they may prefer a path with lower policies just because it eventually travels close to $s^{k_n}(y) + d$.

$I(s^{k_n}(y))$. Clearly there is $\epsilon > 0$ such that voters in $(s^{k_n}(y) - \epsilon, s^{k_n}(y) + \epsilon)$ strictly prefer $s^{k_n}(y)$, so $s^{k_n+1}(y)$ can only be preferred by a majority if there are voters both above and below $s^{k_n}(y)$ who prefer it. Let $y' < s^{k_n}(y) < y''$ be the closest voters to $s^{k_n}(y)$ who prefer the continuation.

By the Lemma, $y'' > y' + d$. Moreover, as n goes to infinity, y' must converge to y_0 and y'' to $y_0 + d$.[23] For each $x \in [y', y'']$ consider the utility given by x to the two agents y', y'': $\tilde{U}(x) = (U_{y'}(x), U_{y''}(x)) \in \mathbb{R}^2$ (in particular both coordinates are non-negative since agents can always quit).

Given the path $T = S(s^{k_n+1}(y))$, construct T' as follows. First, fix $\nu > 0$. Replace elements below $y' - \nu$ with $y' - \nu$. Replace elements between $y' - \nu$ and $y'' - d$ with y'. Replace all elements $s^{k_n}(y) > s_t > y'' - d$ by their average, i.e., $y_1 = \frac{\sum_{s^{k_n}(y) > s_t > y'' - d} \beta^t s_t}{\sum_{s^{k_n}(y) > s_t > y'' - d} \beta^t}$. Replace all elements $s_t > s^{k_n}(y)$ by their average y_2. Finally, if y_1, y_2 have discounted weights w_3, w_4 respectively, replace y_1, y_2 with $s^{k_n}(y)$, $y_3 = y_2 + \frac{w_3}{w_4} y_1 - s^{k_n}(y)$.[24] These changes make T' weakly better for both y' and y'' than T, because both prefer elements below y' being shifted up; y' prefers elements in $[y', y'' - d]$ being shifted to y', and y'' is indifferent; and both get positive utility from all elements in $[y'' - d, \overline{y}]$, so they prefer the decrease in variance that results from averaging or partially averaging subsets of them. Moreover, T' is a linear combination of at most 4 policies:

$$\tilde{U}(T') = w_1 \tilde{U}(y') + w_2 \tilde{U}(y' - \nu) + w_3 \tilde{U}(s^{k_n}(y)) + w_4 \tilde{U}(y_3)$$

where $w_1 + w_2 + w_3 + w_4 = 1$. In addition, because the sequence spends a long time near $\underline{y}$ (hence under $y' - \nu$) before going back up, we have that $\frac{w_4^n}{w_2^n}$ goes to zero as n goes to infinity.

Since y', y'' prefer T' to $s^{k_n}(y)$, we have that:

$$\begin{aligned}
-(y' - s^{k_n}(y))^2 &\leq -w_2\nu^2 - w_3(y' - s^{k_n}(y))^2 - w_4(y' - y_3)^2 \\
-(w_1 + w_2 + w_4)(y' - s^{k_n}(y))^2 &\leq -w_2\nu^2 - w_4(y' - y_3)^2 \\
(w_1 + w_2 + w_4)(y' - s^{k_n}(y))^2 &\geq w_2\nu^2 \\
s^{k_n}(y) - y' &\geq \sqrt{\frac{w_2}{w_1 + w_2 + w_4}}\nu \geq \frac{w_2}{w_1 + w_2 + w_4}\nu
\end{aligned}$$

[23] This happens because the continuation stays under y_0 for a long time, and only goes back over y_0 much later. First, $s^{k_n}(y)$ must be converging to y_0, else a majority would prefer to stay at $s^{k_n}(y)$ for large n. Second, voters above $s^{k_n}(y)$ can't prefer the continuation unless they are very close to $s^{k_n}(y) + d$ and get utility almost zero from $s^{k_n}(y)$. Hence y'' is close to $y_0 + d$. Then y' must be close to y_0, since otherwise a majority would prefer to stay at $s^{k_n}(y)$ for large n.

[24] Note that y_3 must be higher than $s^{k_n}(y)$; if not, then y'' would prefer $s^{k_n}(y)$ to T', a contradiction.

$$C - (y'' - s^{k_n}(y))^2 \leq w_3(C - (y'' - s^{k_n}(y))^2) + w_4(C - (y'' - y_3)^2)$$
$$-(w_1 + w_2 + w_4)(y'' - s^{k_n}(y))^2 \leq -(w_1 + w_2)C - w_4(y'' - y_3)^2$$
$$(w_1 + w_2 + w_4)(y'' - s^{k_n}(y))^2 \geq (w_1 + w_2)C$$
$$y'' - s^{k_n}(y) \geq \sqrt{\frac{w_1 + w_2}{w_1 + w_2 + w_4}}d$$
$$s^{k_n}(y) + d - y'' \leq d\left(1 - \sqrt{\frac{w_1 + w_2}{w_1 + w_2 + w_4}}\right) \leq d\left(1 - \frac{w_1 + w_2}{w_1 + w_2 + w_4}\right) = d\frac{w_4}{w_1 + w_2 + w_4}$$

At the same time, a weak majority in $I(s^{k_n}(y))$ prefers $S(s^{k_n+1}(y))$ to $s^{k_n}(y)$. Since voters in $[y', y'']$ prefer $s^{k_n}(y)$, we must have $F(s^{k_n}(y)+d) - F(y'') + F(y') - F(s^{k_n}(y)-d) \geq F(y'') - F(y')$. Since m equals the identity, we have $F(s^{k_n}(y)+d)) - F(s^{k_n}(y)) = F(s^{k_n}(y)) - F(s^{k_n}(y)-d))$, hence $F(s^{k_n}(y)+d) - F(s^{k_n}(y)) \geq F(y'') - F(y')$, or $F(s^{k_n}(y)+d) - F(y'') \geq F(s^{k_n}(y)) - F(y')$. But by the above, this implies that there are x, x' such that $\frac{f(x)}{f(x')} \leq \frac{w_4}{w_2}\frac{d}{\nu}$; for large n, this implies that $\frac{f(x)}{f(x')}$ can be arbitrarily small, a contradiction.

□

Proof of Proposition 1. Suppose $m(y) = y$, and compare the path $T = (y', s(y'), \ldots)$ with $S = (y, y', s(y'), \ldots)$. WLOG $y' > y$, so by the above Lemma every element of T is higher than y. Then all voters below y, and some above y, prefer S to T. This argument holds for any T, so $s(y) = y$ is the Condorcet winner.

If $m(y) \neq y$, suppose WLOG that $m(y) < y$. By a similar argument, all voters below $m(y)$ and some above $m(y)$ would prefer $s(y) = y$ to any increasing path. Hence $s(y) \leq y$. On the other hand, all voters above $m(y)$ and some below $m(y)$ would prefer $s(y) = m^*(y)$ (followed by $m^*(y)$ forever, since $m^*(y)$ is a stable steady state) to any decreasing path that starts below $m^*(y)$. Hence $s(y) \geq m^*(y)$.

Now consider $y' = \frac{m(y)+m^*(y)}{2}$. Since $y > m(y)$, we must have $m(y) > m^*(y)$ so $y' > m^*(y)$. The path starting at y' would be bounded within $[m^*(y), y']$ by the previous results, and all voters in $[y', y+d]$ would strictly prefer it. Since $y' < m(y)$, this contains a strict majority of $I(y)$. Hence $s(y) > m^*(y)$.

Finally, we show that $s(y) < y$. Suppose that $s(y) = y$. First, note that there must be ϵ_0 such that $s(y-\epsilon) < y - \epsilon$ for all $\epsilon < \epsilon_0$ (otherwise, a strict majority in $I(y)$ 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 $s_-(y) = y$, this implies that $s^k(y-\epsilon) \to y$ as $\epsilon \to 0$ for all k. Note that, since a majority in $I(y-\epsilon)$ prefers $S(s(y-\epsilon))$ to $y-\epsilon$, $m(y-\epsilon)$ in particular must have this preference, so $U_{m(y-\epsilon)}(S(s(y-\epsilon))) \geq U_{m(y-\epsilon)}(y-\epsilon)$ for all $\epsilon > 0$ small enough, i.e.

$$(1-\beta)\sum_{t=0}^{k} \beta^t \left(C - (s^{t+1}(y-\epsilon) - m(x))^2\right) - C + (x - m(x))^2 \geq 0,$$

where $x = y - \epsilon$. The derivative of the above expression with respect to x is

$$
\begin{aligned}
& 2(1-\beta)m'(x)\sum_{t=0}^{k}\beta^t\left(s^{t+1}(y-\epsilon)-m(x)\right)+2(x-m(x))(1-m'(x)) \\
\propto & -(1-\beta)m'(x)\sum_{t=0}^{k}\beta^t\left(s^{t+1}(y-\epsilon)-m(x)\right)+(x-m(x))(1-m'(x)) \\
= & (1-\beta)m'(x)\sum_{t=0}^{k}\beta^t\left(s^{t+1}(y-\epsilon)-m(x)\right)+(x-m(x)) \\
& \qquad -(x-m(x))m'(x)(1-\beta^{k+1})-(x-m(x))m'(x)\beta^{k+1} \\
= & (1-\beta)m'(x)\sum_{t=0}^{k}\beta^t\left(s^{t+1}(y-\epsilon)-x\right)+(x-m(x))-(x-m(x))m'(x)\beta^{k+1}
\end{aligned}
$$

As ϵ goes to 0, k goes to infinity (because the policy stays close to y, thus above $y-d$, for a long time) and $s^{t+1}(y-\epsilon)-x$ goes to 0 for an arbitrarily high number of terms. Thus the above converges to $y - m(y) > 0$. Hence, for $\epsilon > 0$ small enough, $U_{m(y)}(S(s(y-\epsilon))) > U_{m(y)}(y)$, 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 $s_1 = s_-(y)$.

On the one hand, $m(y)$ must prefer y over $S(s(y_n))$. On the other hand, $m(y_n)$ must prefer $S(s(y_n))$ over y. Hence $m(y)$ must be indifferent between y and (s_k). Moreover, $m(y_n)$ prefers $S(s(y_n))$ to all other $S(s(y_{n'}))$, hence to (s_k). All this implies

$$
\begin{aligned}
& 0 \geq U_{m(y)}(S(s(y_n))) - U_{m(y)}(y) = (1-\beta)\sum_{t=0}^{k}\beta^t\left(C-(s^{t+1}(y_n)-m(y))^2\right)-C+(y-m(y))^2 = \\
& (1-\beta)\sum_{t=0}^{k}\beta^t\left(C-(s^{t+1}(y_n)-m(y))^2\right)-(1-\beta)\sum_{t=0}^{k'}\beta^t\left(C-(s_{t+1}-m(y))^2\right) \geq \\
& (1-\beta)\sum_{t=0}^{k}\beta^t\left(C-(s^{t+1}(y_n)-m(y))^2\right)-(1-\beta)\sum_{t=0}^{k'}\beta^t\left(C-(s_{t+1}-m(y))^2\right) \\
& +(1-\beta)\sum_{t=0}^{k''}\beta^t\left(C-(s_{t+1}-m(y_n))^2\right)-(1-\beta)\sum_{t=0}^{k'''}\beta^t\left(C-(s^{t+1}(y_n)-m(y_n))^2\right)
\end{aligned}
$$

If $k = k' = k'' = k'''$ for n high enough,[25] then this is equal to

$$(1-\beta)\sum_{t=0}^{k'}\beta^t\left((s_{t+1}-m(y))^2-(s^{t+1}(y_n)-m(y))^2+(s^{t+1}(y_n)-m(y_n))^2-(s_{t+1}-m(y_n))^2\right)=$$
$$(1-\beta)\sum_{t=0}^{k'}\beta^t 2\left(s_{t+1}-s^{t+1}(y_n)\right)\left(m(y_n)-m(y)\right)$$

Now, crucially,

$$\frac{2(1-\beta)\sum_{t=0}^{k'}\beta^t\left(s_{t+1}-s^{t+1}(y_n)\right)\left(m(y_n)-m(y)\right)}{y-y_n}\xrightarrow[n\to\infty]{}0.$$

If $C-(s_{t+1}-m(y))^2=0$ for some t's, the sums may have different numbers of terms, but the same result holds.[26]

Consider now the possibility of $m(y)$ choosing $S(y_n)$ instead (i.e., the path starting at y_n instead of at $s(y_n)$. We can see that

$$\begin{aligned}U_{m(y)}(S(y_n))-U_{m(y)}(y)&=(1-\beta)\left(u_{m(y)}(y_n)-u_{m(y)}(y)\right)+\beta\left(U_{m(y)}(S(s(y_n)))-U_{m(y)}(y)\right)=\\&=(1-\beta)\left((y-m(y))^2-(y_n-m(y))^2\right)+\beta\left(U_{m(y)}(S(s(y_n)))-U_{m(y)}(y)\right)=\\&=(1-\beta)\left(y+y_n-2m(y)\right)\left(y-y_n\right)+\beta\left(U_{m(y)}(S(s(y_n)))-U_{m(y)}(y)\right)>0\end{aligned}$$

for high n, since $U_{m(y)}(S(s(y_n)))-U_{m(y)}(y)$ is small relative to $y-y_n$, and $y+y_n-2m(y)\to 2(y-m(y))>0$, a contradiction.

Finally, we will show that $s^k(y)$ must converge to $m^*(y)$. Suppose WLOG that $m(y)<y$, so $m^*(y)<y$. 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, we know that $m(s^*)<s^*$, so there is k_0 such that $m(s^k(y))<s^*$ for all $k\geq k_0$. Then a strict majority of voters in $I(s^k(y))$ (all voters to the left of $m(s^k(y))$ and some to the right) would prefer $S(s^{k+2}(y))$ over $S(s^{k+1}(y))$, a contradiction. □

Proof of Corollary 1. Let $x_i^*<x_{i+1}^*$ be two consecutive fixed points of m. Since m is continuous, either $m(y)>y$ for all $y\in(x_i^*,x_{i+1}^*)$ or $m(y)<y$ for all such i. Moreover, if $m(y)>y$ for all $y\in(x_i^*,x_{i+1}^*)$, we must have $m'(x_i^*)\geq 1$ and $m'(x_{i+1}^*)\leq 1$; these inequalities become strict by our assumption that $m'(x_j^*)\neq 1$, which in turn implies that the intervals must alternate (i.e., if $m(y)>y$ for $y\in(x_i^*,x_{i+1}^*)$, then $m'(x_{i+1}^*)<1$, so $m(y)<y$ for $y\in(x_{i+1}^*,x_{i+2}^*)$ and so on).

Note that a fixed point of m is stable if $m'(x^*)<1$ and unstable if $m'(x^*)>1$. Since $m(-1)>-1$ and $m(1)<1$, x_1^* and x_n^* must both be stable, and stable and unstable fixed points must

[25]Note that k' is independent of n. If $C-(s_{t+1}-m(y))^2\neq 0$ for all t, then for high enough n, k, k'' and k''' all equal k' because $y_n\to y$ and $s^{t+1}(y_n)\to s_{t+1}$.

[26]The fact that terms are replaced by 0 when negative can only reduce the difference between terms, i.e., $f(x)=\max\{x,0\}$ is Lipschitz with constant 1.

alternate in between. Hence $n = 2k+1$ must be odd; x_i^* must be stable for odd i and unstable for even i; and all equilibrium paths starting at any $y \in (x_{2k}^*, x_{2k+2}^*)$ must converge to x_{2k+1}^*. □

Proof of Proposition 2. First, let $x < x' \in [x^*, x^{***}]$, where $m(x^*) = x^*$, $m(x^{***}) = x^{***}$ and $m(y) < y$ for all $y \in (x^*, x^{***})$. This is without loss of generality. Suppose $x' \leq x^* + d$ and $s(x) > s(x')$. Then $s(x)$ must be preferred to $s(x')$ by a weak majority in $I(x)$, and the opposite must happen in $I(x')$.

We consider three cases depending on how $E(S(s(x)))$ compares to $E(S(s(x')))$.

Suppose first that $E(S(s(x))) > E(S(s(x')))$. Let $I(x) = A \cup B \cup C$ where $A = [x-d, x'-d,)$, $B = [x'-d, x^*+d)$, $C = [x^*+d, x+d]$, and $I(x') = B \cup C \cup D$ where $D = (x+d, x'+d]$. Voters in $A \cup B$ would never leave the club under either path,[27] so by Lemma 2 there is some $\alpha_0 \in [x-d, x^*+d]$ such that voters to the left of α_0 prefer $s(x')$ and voters to the right prefer $s(x)$.[28] On the other hand, voters in D must prefer $s(x)$ because they will quit immediately under both paths, so the tie-breaker is that they myopically like $s(x)$ better since $x + d > s(x) > s(x')$.

Hence, if $\alpha_0 > x' - d$, then all voters in A prefer $s(x')$ and all voters in D prefer $s(x)$, a contradiction, since $I(x)$ prefers $s(x)$ but $I(x')$ prefers $s(x')$. Thus it must be that $\alpha_0 < x' - d$, so all voters in B prefer $s(x)$. But, since $m(x') < x' \leq x^* + d$,[29] B is a strict majority of $I(x')$, so $I(x')$ would prefer $s(x)$, a contradiction.

Now, suppose that $E(S(s(x))) \leq E(S(s(x')))$. Since $s(x) > s(x')$, this implies that $s^k(x) < s^k(x')$ for some $k > 1$; let k_0 be the smallest such k. Then $s^{k_0-1}(x) > s^{k_0-1}(x')$ but $s^{k_0}(x) < s^{k_0}(x')$. In addition, $E(S(s^{k_0}(x))) < E(S(s^{k_0}(x')))$ because otherwise $E(S(s(x)))$ would be higher than $E(S(s(x')))$. Hence $s^{k_0-1}(x)$ and $s^{k_0-1}(x')$ satisfy all the assumptions of our first case, which we already know leads to a contradiction.

Finally, we prove that the Median Voter Theorem must hold. Let $y \in [x^*, \min(x^{***}, x^* + d)]$ as above and suppose $m(y)$ strictly prefers $y' < s(y)$ to $s(y)$. Then, since s is increasing, $E(S(y')) < E(S(s(y)))$, so by increasing differences and Lemma 2 all voters to the $x < m(y)$ prefer y' to $s(y)$. Some voters $x > m(y)$ close enough to $m(y)$ will also prefer y' by continuity. Hence $s(y)$ is not the Condorcet winner in $I(y)$, a contradiction. On the other hand, suppose $m(y)$ strictly prefers $s(y) < y' < y$ to $s(y)$. Then all voters in $[m(y), x^* + d]$ prefer $s(y)$ by increasing differences, and some to the left of $m(y)$ prefer y' by continuity. On the other hand, voters $x \in (x^* + d, y + d]$ prefer y' to $s(y)$ because $x > x^* + d \geq y$ (x's bliss point is higher than all the policies in both paths) and $s^k(y') \geq s^{k+1}(y)$ for all k (since s is increasing), which is strict for $k = 0$. Hence $s(y)$ is not the Condorcet winner in $I(y)$, a contradiction. □

Proof of Lemma 5. Suppose that $s \neq s'$; in other words, the set $A = \{y \in [x^*, x^{**}] : s(y) \neq s'(y)\}$ is nonempty. Let $\underline{y} = \inf A \in [x^* + \epsilon, x^{**})$ ($\underline{y}$ cannot be x^{**} because $s(x^{**}) = s'(x^{**}) = x^{**}$).

[27] voters in A would not be members under current policy x', but $s(x') < s(x) < x$ and both paths are decreasing, so they would be members in both continuations.

[28] There are also the degenerate cases where all voters in the interval prefer the same policy.

[29] $m(x') < x'$ must hold unless $x' = x^{**}$, but in that case we would have $s(x') = x' > x > s(x)$.

Also, note that the rule to always pick the highest Condorcet winner is well-defined because, by continuity, a limit of Condorcet winners must also be a Condorcet winner.

There are two cases. First, suppose $s(\underline{y}) = s'(\underline{y})$. Then there is a sequence $y_n \to \underline{y}$ of policies for which $s(y_n) \neq s'(y_n)$. If $s(y_{n_k}) \to \underline{y}$ or $s'(y_{n_k}) \to \underline{y}$ for some subsequence y_{n_k}, then by continuity $\underline{y}$ is an optimal policy for $I(\underline{y})$, contradicting Proposition 1. Hence $s(y_n)$, $s'(y_n) \leq \underline{y} - \delta$ for $n \geq n_0$ and some $\delta > 0$. Since the continuations to these policies are contained in $[x^*, \underline{y} - \delta]$, they are the same under s and s'. Hence $S(s(y_n))$ and $S(s'(y_n))$ must both be Condorcet winners in $I(y)$ under s. Hence, by assumption, $s(y_n) = s'(y_n)$, a contradiction.

Second, suppose $s(\underline{y}) \neq s'(\underline{y})$. Since both values are below $\underline{y}$, $S(s(y_n))$ and $S(s'(y_n))$ must both be Condorcet winners in $I(y)$ under s, leading to the same contradiction. □

Proof of Proposition 3. First, 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

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

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

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

Since the highest order term has a negative coefficient, we know that $W(M) < 0$ for large M; hence there is also a solution larger than 1. On the other hand, by Descartes' rule of signs, W has at most two positive roots. Hence there is a unique solution $0 < \gamma_k < 1$. However, given the right γ_k, any x_0 can be used to start the sequence.

From here we can also show that γ_k^k is decreasing in k: let $\tilde{W}(\gamma) = W(\gamma^{\frac{1}{k}})$. Then $\tilde{W}(\gamma, k) =$

$\gamma(1+\gamma^{\frac{1}{k}})(1-\beta\gamma)-2\alpha(1-\beta\gamma^2)$. Clearly this 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.

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

Second, we want to 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 argue that this yields our result.

For the first part, note that

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

Consider two cases. If $k=1$, then the required inequality is $1<\beta(\gamma+2\gamma^2)$. Since $\beta\geq\frac{2}{3}$, 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 $\gamma>\frac{2}{3}$. It is easy to see that the worst case is when β is minimal, so take $\beta=\frac{2}{3}$. Then $W(\frac{2}{3})=\frac{10}{9}\frac{5}{9}-2\alpha\frac{19}{27}<0$ whenever $\alpha>\frac{25}{57}<0.44$. 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 β 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 $\alpha>\frac{2}{5}=0.4$.

Now, let's see that $m(x_n)$ prefers x_{n+k} to any x not in the sequence. If $x\in(x_{n+k+1},x_{n+k})$, then $s(x)=s(x_{n+k+1})$. Since $m(x_n)<x_{n+k+1}<x$, $m(x_n)$ prefers x_{n+k+1} to x, and the continuations are identical. Similarly, if $x\in(x_{n+k+1-r},x_{n+k-r})$ for $r\geq 1$, then $m(x_n)$ prefers $x_{n+k+1-r}$ to x, and in turn prefers x_{n+k} to $x_{n+k+1-r}$. On the other hand, if $x\in(x_{n+k+1+r},x_{n+k+r})$ for $r\geq 1$, then we know by the previous argument that $m(x_{n+r})$ prefers x_{n+k+r} to x. Then $m(x_n)$ must also prefer x_{n+k+r} to x by increasing differences, and in turn he prefers x_{n+k} to x_{n+k+r}.

Next, we check that, if $x\in(x_n,x_{n-1})$, $m(x)$ prefers x_{n+k} to any other x. That he prefers x_{n+k} any $x<x_{n+k}$ follows from the fact that x_{n+k} is optimal for $m(x_n)$, plus increasing differences. On the other hand, he prefers x_{n+k} any $x>x_{n+k}$ because x_{n+k} is optimal for $m(x_{n-1})$.

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

$$s(x)=\arg\max_y\sum_{t=0}^{\infty}\beta^t\left(C-(m(x)-s^t(y))^2\right).$$

If s is smooth, then $y = s(x)$ must satisfy the first order condition

$$\sum_{t=0}^{\infty} \beta^t \left(-2(m(x) - s^t(y)) \prod_{i=0}^{t-1} s'(s^i(y)) \right) = 0.$$

Since $m(x) = \alpha x$, we look for a solution of the form $s_\infty(x) = \gamma x$. We obtain

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

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

Proof of Proposition 4. First, we construct a sequence of approximate 1-equilibria as follows. Note that, since f is continuous, m must be C^1, so $m'(x^*) = \alpha$ is well defined. For each i, construct an increasing $m_i \in C^1[x^*, x^{**}]$ such that $m_i(x) = \alpha(x - x^*) + x^*$ for $x < x^* + \frac{1}{i}$ and $m_i \to m$ in the C^1 norm, i.e., $||m_i - m||_\infty \to 0$ and $||m_i' - m'||_\infty \to 0$.

Now, for each m_i, we construct a 1-equilibrium s_i as follows. Let $x' \in [x^*, x^* + \frac{1}{i})$ and define $s_i(x) = s_1^*(x)$ for $x < x^* + \frac{1}{i}$, where $s_1^*(x)$ is the 1-equilibrium constructed in Proposition 3 with $x_0 = x'$. Then extend $s_i(x)$ beyond $x^* + \frac{1}{i}$ in the usual way: WLOG let x_0' be the highest element of the sequence below $x^* + \frac{1}{i}$. Then, by Lemma 3, there is a unique y that is indifferent between x_0' and $S(x_1')$ (hence indifferent between $S(x_0')$ and $S(x_1')$), so define $x_{-1}' = m^{-1}(y)$. Proceed likewise to define all x_n.

Next, we check that the constructed sequence indeed yields a 1-equilibrium. By the same arguments as in Proposition 3, $m_i(x_n)$ is indifferent between x_{n+1} and x_{n+2}, and prefers these to all other elements of the sequence; and $x \in [x_n, x_{n-1})$ strictly prefers x_{n+1} to all other elements of the sequence.

Finally, the so far unproven part of the result is that, if β is high enough and m_i is well-behaved, then policies outside of the sequence are never optimal. This is equivalent to the condition $m_i(x_n) < x_{n+2}$, which guarantees that x_n prefers x_{n+2} to any point in (x_{n+2}, x_{n+1}).

Next, we argue that there is a way to choose x' so that $\underline{x}$ is an element of the sequence. By construction, for each n, x_n is a continuous function of x'. If $\underline{x}$ is never an element of the sequence, by reducing x' and making it arbitrarily close to x^*, we can obtain equilibria where the interval $[x^* + \frac{1}{n}, \underline{x}]$ has no elements of the sequence in it, which easily leads to a contradiction. (The idea is that elements of the sequence initially above $\underline{x}$ cannot leapfrog it by continuity, and there are only finitely many elements initially in $[x^* + \frac{1}{n}, \underline{x}]$, which can all be reduced below $x^* + \frac{1}{n}$ by lowering x' enough). Let s_i^* be a 1-equilibrium setting $x_0 = \underline{x}$.

Now, we construct a 1-equilibrium s for the true median voter function m, such that the s_i^*

converge to s in terms of their defining sequences, i.e., $x_{in} \to x_n$ for all n. We do this by a diagonal argument: we already know that $x_{i0} = \underline{x}$ for all i, so $x_0 = \underline{x}$ works. Next, for all i, x_{i1} must be contained in $[x^*, \underline{x}]$, so we can take a subsequence of the s_i such that $x_{i1} \to x_1$. (Again, x_1 cannot equal $\underline{x}$ by a similar argument as in Proposition 1). Next, we take a subsequence such that the x_{i2} also converge, and so on. Finally, all the optimality conditions that made the s_i 1-equilibria under m_i make s a 1-equilibrium under m by continuity.

Next, we check that $x_n \to x^{**}$ as $n \to -\infty$, as intended. This follows from an argument similar to Proposition 1: if x_n instead converged to some $x < x^{**}$, then it would be optimal for x to choose $s(x) = x$, which in fact is always worse than some $x' < x$.

Finally, note that s is weakly increasing by design, and our consruction does not rely on being within the interval $[x^*, x^* + d]$. The Median Voter Theorem also holds even outside of $[x^*, x^* + d]$, which is clear from our arguments above (all voters above $m(x_n)$ prefer x_{n+1} to any other policy, and all voters below $m(x_n)$ prefer x_{n+2} to any other policy). □

Proof of Proposition 5. We will assume that s is analytic and characterize it in a small interval around x^*, exploiting the condition that m is analytic.

Consider a candidate equilibrium given by $s(x)$. For x close to x^*, $s(x)$ solves

$$s(x) = \arg\max_y \sum_{t=0}^{\infty} \beta^t \left(C - (s^t(y) - m(x))^2\right).$$

Thus, if s is smooth, then $y = s(x)$ satisfies the first order condition

$$\sum_{t=0}^{\infty} \beta^t \left((m(x) - s^t(y)) \prod_{i=0}^{t-1} s'(s^i(y)) \right) = 0.$$

WLOG let $x^* = 0$. Remember that, by Proposition 3, if $m(x) = \alpha x$ is linear around the steady state, there is a unique linear solution $s(x) = \gamma x$. We can in fact extract a stronger conclusion: suppose that $m(x) = \sum_1^\infty \alpha_n x^n$ and we are trying to choose a $\tilde{s}(x) = \sum_1^\infty \gamma_n x^n$ such that

$$\sum_{t=0}^{\infty} \beta^t \left((m(x) - \tilde{s}^{t+1}(x)) \prod_{i=0}^{t-1} \tilde{s}'(\tilde{s}^{i+1}(x)) \right) = O(x^2),$$

i.e., such that both the LHS and its first derivative vanish at $x = 0$. Then, it is necessary and sufficient to choose $\gamma_1 = \gamma$ as above. Intuitively, the γ_n for $n \geq 2$ will be multiplied by higher order terms, which have no effect on the first derivative. Indeed, the LHS can be rewritten as

$$\sum_{t=0}^{\infty} \beta^t \left((\alpha x + O(x^2) - \gamma_1^{t+1} x - O(x^2)) \prod_{i=0}^{t-1} (\gamma_1 + O(x)) \right) = \sum_{t=0}^{\infty} \beta^t \left((\alpha x - \gamma_1^{t+1} x)\gamma_1^t + O(x^2) \right),$$

so the condition pinning down γ_1 is the same as in Proposition 3.

We can then use a similar method to determine the rest of the coefficients inductively. In

general, we argue that given $m(x)$ as above, there are unique values $\gamma_1^*, \ldots, \gamma_n^*$ such that

$$\sum_{t=0}^{\infty} \beta^t \left((m(x) - \tilde{s}^{t+1}(x)) \prod_{i=0}^{t-1} \tilde{s}'(\tilde{s}^{i+1}(x)) \right) = O(x^{n+1})$$

iff $\gamma_1 = \gamma_1^*, \ldots, \gamma_n = \gamma_n^*$. We have already argued the case $n = 1$. For $n > 1$, we know by assumption that $\gamma_1, \ldots, \gamma_{n-1}$ are already determined. In addition, terms containing a γ_n factor are always of order at least n; and for terms of order exactly n, there is never a γ_m factor for $m > n$, or more than one γ_n factor. Finally, if we choose $\gamma_1, \ldots, \gamma_{n-1}$ as before, terms of order $n-1$ and below will vanish. Thus, the above expression is of the form

$$(\rho_n + \eta_n \gamma_n) x^n + O(x^{n+1}),$$

where ρ_n and η_n are polynomials in $\gamma_1, \ldots, \gamma_{n-1}$, so that we can simply choose $\gamma_n = -\frac{\rho_n}{\eta_n}$ and this is the only option. The main complication here is showing that η_n is not zero. To do this, we characterize it explicitly. Note that our expression can be written as

$$\sum_{t=0}^{\infty} \beta^t \left((-P_{t+1} x^n) \gamma_1^t \right) + \sum_{t=0}^{\infty} \beta^t \left((\alpha x - \gamma_1^{t+1} x) Q_t x^{n-1} \right) + \sum_{t=0}^{\infty} \beta^t R_t x^n + O(x^{n+1}),$$

where P_{t+1} are factors in $s^{t+1}(x)$ with order n and coefficients divisible by γ_n; Q_t are factors in $(s^t(y))'|_{y=s(x)}$ with order $n-1$ and coefficients divisible by γ_n; and R_t is a polynomial in $\gamma_1, \ldots, \gamma_{n-1}, \alpha_1, \ldots, \alpha_n$, so that $\sum_t \beta^t R_t = \rho_n$.

Next, by inspection, we can show that

$$P_{t+1} = \gamma_1 P_t + \gamma_n \gamma_1^{tn} = \gamma_n \gamma_1^t (1 + \gamma_1^{n-1} + \ldots + \gamma_1^{t(n-1)}) = \gamma_n \gamma_1^t \frac{1 - \gamma_1^{(n-1)(t+1)}}{1 - \gamma_1^{n-1}}$$

$$Q_t = n P_t \gamma_1^{n-1} x^{n-1}.$$

Hence, focusing on the coefficients containing γ_n, our expression becomes

$$
\begin{aligned}
&\sum_{t=0}^{\infty}\beta^t\left(-P_{t+1}\gamma_1^t\right)+\sum_{t=0}^{\infty}\beta^t\left((\alpha-\gamma_1^{t+1})Q_t\right)\\
=&\sum_{t=0}^{\infty}\beta^t\left(-P_{t+1}\gamma_1^t\right)+\sum_{t=0}^{\infty}\beta^t\left((\alpha-\gamma_1^{t+1})nP_t\gamma_1^{n-1}\right)\\
=&\sum_{t=0}^{\infty}\beta^t\left(-P_{t+1}\gamma_1^t+(\alpha-\gamma_1^{t+1})nP_t\gamma_1^{n-1}\right)\\
=&\sum_{t=0}^{\infty}\frac{\beta^t\gamma_n}{1-\gamma_1^{n-1}}\left(-\gamma_1^t(1-\gamma_1^{(n-1)(t+1)})\gamma_1^t+(\alpha-\gamma_1^{t+1})n\gamma_1^{t-1}(1-\gamma_1^{(n-1)t})\gamma_1^{n-1}\right)\\
=&\sum_{t=0}^{\infty}\frac{\beta^t\gamma_n}{1-\gamma_1^{n-1}}\left(-\gamma_1^{2t}+\gamma_1^{(n+1)t}\gamma_1^{n-1}+\alpha n\gamma_1^{n-2}\gamma_1^t-n\gamma_1^{n-1}\gamma_1^{2t}-\alpha n\gamma_1^{n-2}\gamma_1^{nt}+n\gamma_1^{n-1}\gamma_1^{(n+1)t}\right)\\
\propto&\sum_{t=0}^{\infty}\beta^t\left(-\gamma_1^{2t}+\gamma_1^{(n+1)t}\gamma_1^{n-1}+\alpha n\gamma_1^{n-2}\gamma_1^t-n\gamma_1^{n-1}\gamma_1^{2t}-\alpha n\gamma_1^{n-2}\gamma_1^{nt}+n\gamma_1^{n-1}\gamma_1^{(n+1)t}\right)
\end{aligned}
$$

Adding up all the infinite series, we obtain

$$
\begin{aligned}
=&-\frac{1}{1-\beta\gamma_1^2}+\frac{\gamma_1^{n-1}}{1-\beta\gamma_1^{n+1}}+\frac{\alpha n\gamma_1^{n-2}}{1-\beta\gamma_1}-\frac{n\gamma_1^{n-1}}{1-\beta\gamma_1^2}-\frac{\alpha n\gamma_1^{n-2}}{1-\beta\gamma_1^n}+\frac{n\gamma_1^{n-1}}{1-\beta\gamma_1^{n+1}}\\
=&-\frac{1}{1-\beta\gamma_1^2}+\frac{\gamma_1^{n-1}}{1-\beta\gamma_1^{n+1}}+\frac{n\gamma_1^{n-1}}{1-\beta\gamma_1^2}-\frac{n\gamma_1^{n-1}}{1-\beta\gamma_1^2}-\frac{\alpha n\gamma_1^{n-2}}{1-\beta\gamma_1^n}+\frac{n\gamma_1^{n-1}}{1-\beta\gamma_1^{n+1}}\\
=&-\frac{1}{1-\beta\gamma_1^2}-\frac{\alpha n\gamma_1^{n-2}}{1-\beta\gamma_1^n}+\frac{(n+1)\gamma_1^{n-1}}{1-\beta\gamma_1^{n+1}}\\
\propto&-\frac{\gamma^2}{1-\beta\gamma_1^2}-\frac{\alpha n\gamma_1^n}{1-\beta\gamma_1^n}+\frac{(n+1)\gamma_1^{n+1}}{1-\beta\gamma_1^{n+1}}\\
=&-\frac{\alpha\gamma_1}{1-\beta\gamma_1}-\frac{\alpha n\gamma_1^n}{1-\beta\gamma_1^n}+\frac{(n+1)\gamma_1^{n+1}}{1-\beta\gamma_1^{n+1}}
\end{aligned}
$$

We can then check manually that this expression equals zero for $n=1$ and is decreasing in n; hence, for $n\geq 2$, $\eta_n\neq 0$.

By proceeding this way we can define a s such that the Taylor series of the expression above is identically 0. Since the expression is analytic by construction, it must in fact be 0, which means s is a solution. If the constructed series has a positive radius of convergence, this defines an analytic solution within some interval $[x^*, x^*+e)$, which can then be extended as per Lemma 5, although the extension may no longer be analytic (or even continuous). □

Proof of Proposition 6. A continuous solution $s(x)$ must satisfy the first-order condition

$$
\sum_{t=0}^{\infty}\beta^t\left(-2(m(x)-s^{t+1}(x))\prod_{i=1}^{t}s'(s^i(x))\right)=0.
$$

Since s is given here, this implies

$$m(x) = \frac{\sum_{t=0}^{\infty} \beta^t s^{t+1}(x) \prod_{i=1}^{t} \tilde{s}'(\tilde{s}^i(x))}{\sum_{t=0}^{\infty} \beta^t \prod_{i=1}^{t} \tilde{s}'(\tilde{s}^i(x))} = \frac{s(x) + \beta s^2(x)s'(s(x)) + \beta^2 s^3(x)s'(s^2(x))s'(s(x)) + \dots}{1 + \beta s'(s(x)) + \beta^2 s'(s^2(x))s'(s(x)) + \dots},$$

pinning down $m(x)$. It is easy to check that m is continuous, $m(x^*) = x^*$ and $m(x^{**}) = x^{**}$; in addition $m(x)$ maps into $[x^*, x)$ for interior x because it is a weighted average of values of the form $s^t(x)$, which are in $[x^*, x)$ by assumption.

Next, we show that s supports an MPE given m. Since s is well-behaved, $W(E)$ is concave. We can check that $U_{m(x)}(S) = -(m(x) - E(S))^2 + W(S) + E(S)^2 = -m(x)^2 + 2m(x)E + W$. Written in these terms, the first-order condition is $2m(x) = -W'(E)$. Since W is concave in E, we know that this guarantees optimality. Moreover, $s(x)$ is increasing in $x \Rightarrow E$ is increasing in $x \Rightarrow -W'$ is increasing in x by the concavity of $W \Rightarrow m(x)$ is increasing, as intended. □

Proof of Lemma 6. 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. □

Proof of Proposition 7. Consider first $x \in [x^*, m^{-1}(x^* + d))$, so that the median voter $m(x)$ never leaves the club. Then, the condition

$$C - (m(x) - x)^2 = \int_0^{\infty} re^{-rt} \max\left(C - (m(x) - s^t(x))^2, 0\right) dt$$

boils down to

$$\begin{aligned}
(m(x) - x)^2 &= \int_0^{\infty} re^{-rt} \left(m(x) - s^t(x)\right)^2 dt \\
m(x)^2 - 2m(x)x + x^2 &= m(x)^2 - 2m(x) \int_0^{\infty} re^{-rt} s^t(x) + \int_0^{\infty} re^{-rt} (s^t(x))^2 \\
2m(x)(x - E(x)) &= x^2 - W(x),
\end{aligned}$$

where $E(x) = \int_0^{\infty} re^{-rt} s^t(x)$ and $W(x) = \int_0^{\infty} re^{-rt} (s^t(x))^2$. Now, note that $E'(x) = re(x)(x - E(x))$ and $W'(x) = re(x)(x^2 - W(x))$. (This can be seen easily if we rewrite the integrals as $E(x) = \int_{x^*}^x re^{-r\int_y^x e(z)} e(y)ydy$, $W(x) = \int_{x^*}^x re^{-r\int_y^x e(z)} e(y)y^2dy$ by a change of variables). In particular, this

together with the above implies $2m(x)E'(x) = W'(x)$. Now derive the above equation to get

$$\begin{aligned}
2m'(x)(x - E(x)) + 2m(x)(1 - E'(x)) &= 2x - W'(x) \\
2m'(x)(x - E(x)) + 2m(x) &= 2x \\
E(x) &= \frac{m'(x)x + m(x) - x}{m'(x)} = x + \frac{m(x) - x}{m'(x)} \\
E'(x) &= 1 + \frac{m'(x) - 1}{m'(x)} - \frac{(m(x) - x)m''(x)}{(m'(x))^2} \\
re(x)\frac{x - m(x)}{m'(x)} &= re(x)(x - E(x)) = \frac{2m'(x) - 1}{m'(x)} - \frac{(m(x) - x)m''(x)}{(m'(x))^2} \\
e(x) &= \frac{1}{r}\left(\frac{2m'(x) - 1}{x - m(x)} + \frac{m''(x)}{m'(x)}\right),
\end{aligned}$$

as desired.

Now suppose $x > m^{-1}(x^* + d)$. Let $\tilde{E}(x) = \int_0^\infty re^{-rt} \max(s^t(x), m(x) - d)$, $\tilde{W}(x) = \int_0^\infty re^{-rt} \max((s^t(x))^2, (m(x) - d)^2)$ and $\hat{E}_{x_0}(x) = \int_0^\infty re^{-rt} \max(s^t(x), m(x_0) - d)$, $\hat{W}_{x_0}(x) = \int_0^\infty re^{-rt} \max((s^t(x))^2, (m(x_0) - d)^2)$. Essentially, $\tilde{E}$ and $\tilde{W}$ are statistics for a truncated path where policies below $m(x) - d$ are instead set equal to $m(x) - d$; $\tilde{E}_{x_0}$, $\tilde{W}_{x_0}$ are based on the same concept but make the lower bound for truncating policies, $m(x_0) - d$, independent of x. Note that, in particular, $\hat{E}_x(x) = \tilde{E}(x)$ and $\hat{W}_x(x) = \tilde{W}(x)$.

By the same arguments as before, $2m(x)(x - \tilde{E}(x)) = x^2 - \tilde{W}(x)$. In addition

$$\begin{aligned}
\frac{\partial \hat{E}_y(x)}{x} &= re(x)(x - \hat{E}_y(x)) \\
\frac{\partial \hat{W}_y(x)}{x} &= re(x)(x^2 - \hat{W}_y(x)) \\
\tilde{E}'(x) &= re(x)(x - \tilde{E}(x)) + e^{-rt^*(x)}m'(x) \\
\tilde{W}'(x) &= re(x)(x^2 - \tilde{W}(x)) + 2e^{-rt^*(x)}m'(x)(m(x) - d)
\end{aligned}$$

The former two equations follow as before; the latter include a second term which results from the lower bound, $m(x) - d$, being shifted up as x increases. Here $t^*(x)$ is the time it takes to reach the lower bound, i.e., $t^*(x) = d(x, m(x) - d)$.

Deriving our first order condition, we now get

$$
\begin{aligned}
2m'(x)(x-\tilde{E}(x))+2m(x)(1-\tilde{E}'(x)) &= 2x-\tilde{W}'(x)\\
2m'(x)(x-\tilde{E}(x))+2m(x)-2m(x)re(x)(x-\tilde{E}(x))-2e^{-rt^*(x)}m(x)m'(x) &=\\
=2x-re(x)(x^2-\tilde{W}(x))-2e^{-rt^*(x)}m'(x)(m(x)-d)&\\
2m'(x)(x-\tilde{E}(x))+2m(x) &= 2x+2e^{-rt^*(x)}m'(x)d\\
\tilde{E}(x) &= \frac{m'(x)x+m(x)-x-e^{-rt^*(x)}m'(x)d}{m'(x)}\\
\tilde{E}(x) &= x+\frac{m(x)-x}{m'(x)}-de^{-rt^*(x)}
\end{aligned}
$$

We can now derive this expression to obtain

$$
\begin{aligned}
\tilde{E}'(x) &= A+de^{-rt^*(x)}r(e(x)-e(m(x)-d)m'(x))\\
re(x)(x-\tilde{E}(x))+e^{-rt^*(x)}m'(x) &= A+de^{-rt^*(x)}r(e(x)-e(m(x)-d)m'(x))\\
re(x)\left(\frac{x-m(x)}{m'(x)}+de^{-rt^*(x)}\right)+e^{-rt^*(x)}m'(x) &= A+de^{-rt^*(x)}re(x)-de^{-rt^*(x)}re(m(x)-d)m'(x)\\
re(x)\frac{x-m(x)}{m'(x)}+e^{-rt^*(x)}m'(x) &= A-de^{-rt^*(x)}re(m(x)-d)m'(x),
\end{aligned}
$$

where $A=\frac{2m'(x)-1}{m'(x)}-\frac{(m(x)-x)m''(x)}{(m'(x))^2}$. Hence

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

as desired.

Finally, note that all these calculations are made assuming that the first-order condition $u_{m(x')}(x')=U_{m(x')}(x')$ holds in a neighborhood of x. In particular, this implies $e(x)\geq 0$. If at any point the $e(x)$ calculated above becomes negative, it means that no non-negative delay around x can sustain the FOC, so $e(x)=0$ and $u_{m(x')}(x')<U_{m(x')}(x')$ for $x'>x$ close to x. Then, so long as $u_{m(x')}(x')<U_{m(x')}(x')$, we must have $e(x')=0$ (since $m(x')$ would strictly prefer to move away from x'). □

Proof of Lemma 7. 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 8. The case where the cluster is composed of a single club is trivial.

If $j > i$, we first argue that $m(x_i) > x_i$. Suppose otherwise that $m(x_i) \leq x_i$; this would imply that a club with policy x_i would drift downward in the single-club game, or at best stay put. In this case, the actual median voter of club i is lower than in the single-club case, because some voters to the right of x_i who would belong to club i will instead be in club $i+1$. In other words, $m(I_i) < m(x_i) \leq x_i$. Thus club i would drift away from the cluster in this case as well. Similarly, $m(x_j) < x_j$.

Next, we will characterize the tuple $(x_i, x_{i+1}, \ldots, x_j)$ as a function of x_i. Let e_l be the rightmost member of I_l for $l = i, \ldots, j$. First, since there is no club immediately to the left of club i, all the voters in $(x_i - d, x_i)$ belong to i. For i to be in steady state, x_i must be the median member, so $F(e_i) - F(x_i) = F(x_i) - F(x_i - d)$. This condition pins down a unique value for e_i. Since e_i must be indifferent between belonging to clubs i and $i+1$, they must give the same utility, i.e. $e_i = \frac{x_i + x_{i+1}}{2}$. This pins down $x_{i+1} = 2e_i - x_i$. Then again, for $i+1$ to be in equilibrium, there must be equal numbers of voters to the left and right of x_{i+1} in the club; this imposes the condition $F(e_{i+1}) - F(x_{i+1}) = F(x_{i+1}) - F(e_i)$, which pins down e_{i+1}, and so on. Thus, given x_i, there is always at most one tuple $(x_{i+1}, \ldots, x_j)$ compatible with steady state. Finally, for club j, there is the extra condition that its rightmost voter e_j must happen to be $x_j + d$, which serves to pin down x_i. If f is a non-constant polynomial, then we can check that the entire system has a finite number of solutions by Bézout's theorem.

Finally, suppose that f is log-concave. We will show that $e_j - x_j$ is an increasing function of x_i, so there is a unique valid choice of x_i that sets $e_j - x_j = d$. The argument has three parts:

- Let $f(x) = e^{g(x)}$ with $g(x)$ concave. Then, for any $x' > x$,

$$g'(x') \leq \frac{f(x') - f(x)}{F(x') - F(x)} \leq g'(x).$$

  To see this, first note that if $h(z) = ke^{gz}$ and $H(z) = \int_{z_0}^{z} h(w)dw$ then for any $z' > z$

$$\frac{h(z') - h(z)}{H(z') - H(z)} = g.$$

  Now take k, g such that $h(x) = f(x)$ and $h(x') = f(x')$. Since f is log-concave and h is log-linear, $h(x'') \leq f(x'')$ for all $x'' \in (x, x')$. Then $H(x') - H(x) \leq F(x') - F(x)$. In addition,

$g'(x) \geq g$, as otherwise $h(x) = f(x)$ would imply $h(x') > f(x')$. Thus

$$g'(x) \geq g = \frac{h(x') - h(x)}{H(x') - H(x)} \geq \frac{f(x') - f(x)}{F(x') - F(x)}.$$

On the other hand, take k, g such that $h(x') = f(x')$ and $g = g'(x')$. This implies $h(x'') \geq f(x'')$ for all $x'' < x'$, so $h(x') - h(x) \leq f(x') - f(x)$ and $H(x') - H(x) \geq F(x') - F(x)$, hence

$$g'(x') = \frac{h(x') - h(x)}{H(x') - H(x)} \leq \frac{f(x') - f(x)}{F(x') - F(x)}.$$

- Take an interval $(x, x+r)$ with r fixed and let $q(x)$ be such that $F(x+q(x)) - F(x+r) = F(x+r) - F(x)$. We want to show that $q(x)$ is increasing. To do this, note that $F(x+q) - F(x+r)$ is increasing in q, so $q(x)$ is increasing around x_0 iff $f(x_0+q) - f(x_0+r) \leq f(x_0+r) - f(x_0)$ (i.e., if we leave $q = q(x_0)$ fixed and increase $x + q$ at the same rate as x, then the left hand side does not increase fast enough).

  Note that, by assumption, $F(x_0 + q) - F(x_0 + r) = F(x_0 + r) - F(x_0)$ so it is equivalent to show

  $$\frac{f(x_0 + q) - f(x_0 + r)}{F(x_0 + q) - F(x_0 + r)} \leq \frac{f(x_0 + r) - f(x_0)}{F(x_0 + r) - F(x_0)}.$$

  This is true because, by the previous item,

  $$\frac{f(x_0 + q) - f(x_0 + r)}{F(x_0 + q) - F(x_0 + r)} \leq g'(x_0 + r) \leq \frac{f(x_0 + r) - f(x_0)}{F(x_0 + r) - F(x_0)}.$$

- The above argument implies that $e_i(x_i) - x_i$ is increasing in x_i, if we take $x = x_i - d$, $x + r = x_i$, $e_i = x + q(x)$. Then, in particular, $e_i(x_i)$ is also increasing, as is $x_{i+1} - e_i(x_i)$ (because it is equal to $e_i(x_i) - x_i$). If we denote $x - r = e_i$, $x = x_{i+1}$, then the previous argument tells us that $e_{i+1} - x_{i+1}$ is increasing if we increase $x - r$, x by the same amount, and it is also clearly increasing in r (since lowering $x - r$ increases $F(x) - F(x - r)$, it requires increasing $F(x + q) - F(x)$). Thus $e_{i+1} - x_{i+1}$ is increasing in x_i. The same argument can be carried through by induction to $e_j - x_j$. Moreover, $e_j - x_j$ is strictly increasing in x_i unless f is exponential, i.e., log-linear (in which case all our inequalities above are equalities).

□

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.

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

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

_ , Radu Filimon, and Thomas Romer, "Equilibrium Among Local Jurisdictions: Toward an Integrated Treatment of Voting and Residential Choice," *Journal of Public Economics*, 1984, *24* (3), 281–308.

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

Roberts, Kevin, "Dynamic Voting in Clubs," *LSE STICERD Research Paper No. TE367*, 1999.

Schauer, Frederick, "Slippery Slopes," *Harvard Law Review*, 1985, *99* (2), 361–383.

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

Volokh, Eugene, "The Mechanisms of the Slippery Slope," *Harvard Law Review*, 2003, *116* (4), 1026–1137.