Policy Persistence and Drift in Organizations[*]

Germán Gieczewski[†]

October 2017

Abstract

I analyze the evolution of organizations that allow free entry and exit of members, such as cities, trade unions, sports clubs and cooperatives. Current members choose a policy for the organization, but this, in turn, may lead to new agents joining or dissatisfied members leaving, yielding a new set of policymakers tomorrow. The resulting feedback effects may take the organization down a "slippery slope", which agents may allow in equilibrium despite being forward-looking and patient, a result that contrasts with existing models of elite clubs. The model explains how quickly the organization approaches a steady state; how this limit depends on the distribution of agents' preferences and the initial policy; and, in particular, when a population of mostly moderate agents might support extremist organizations. The model can also be extended to allow for competition between multiple organizations.

Keywords: dynamics, median voter, slippery slope, endogenous population, extremism

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

[†] Department of Politics, Princeton University.

1 Introduction

Many organizations provide certain services to their members, and must decide exactly what those services are, or who they are geared towards. For example, consider a professional sports club: it may build more swimming pools or tennis courts for members; it may expand its academy for junior players, or spend more on signing famous players for its professional teams. This is what we call a *policy*, and each policy will please some members more than others: some may become disillusioned after the latest changes and leave, or outsiders may be enticed to join. Since there is free entry and exit, the chosen policy affects the population of future members. But, in turn, members of many clubs have a say in what the policy should be, as their membership entitles them to vote for the president or board of directors. Thus, the organization's policy and set of members influence each other, potentially leading to large changes over time. For instance, if the club's neighborhood undergoes gentrification for exogenous reasons, more well-off members will join; this may lead the new leadership to add a new golf course, which will attract even more upscale families, and so on and so forth. The main question of this paper is: what outcomes can we expect from an organization ruled by these dynamics, especially in the long term?

Before we continue, note that the basic mechanics embedded in this story apply to many other organizations in the real world. For instance, trade unions make demands on behalf of their members, but unionized workers elect their leaders; in turn, a firm's workers may choose whether to join a union or not based on its performance. A union known for being aggressive will attract workers from firms with inelastic labor demand, who will push to maintain these policies.[1] Other non-profit organizations, like churches and universities, exhibit similar dynamics: people can join the community (by converting or applying, respectively) and choose partly based on cultural fit. Existing members then influence policy choices as, for instance, alumni usually vote for a university's trustees or directors. 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 or out. The city chooses policies such as its local taxes,

[1] A related idea appears 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 senior members, who have higher job security, will be more aggressive, resulting in rigid wages.

school quality, housing regulations, and so on, which affect who wants to move there. In turn, citizens elect the mayor and local legislators.[2]

These examples differ in the details of exactly how current members influence future policies, as well as whether there is some cost of entry or exit, but they share the same essential features. For tractability, I study a stylized version of this problem, in which policy is one-dimensional, 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 its policy, and they are small and numerous enough to behave like "policy-takers". Finally, and most importantly, agents are forward-looking: when they vote, they take into account the fact that feedback effects may cause the current policy to drift away from their original choice, according to equilibrium behavior.

The benchmark model yields a characterization of the equilibrium paths that the club's policy can follow. Although there are 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, at roughly the same speed in all equilibria.

The characterization of long-run behavior is straightforward. The steady states, as well as their basins of attraction, are uniquely pinned down by the distribution of agents' preferences, as follows: if a policy x attracts a group of members with its median preferring a policy higher than x, then policy will drift upward from x in equilibrium, and vice versa; x is a steady state if the resulting median's bliss point is exactly x. In particular, the set of steady states is independent of the agents' discount factor: it is the same that would obtain if agents were completely myopic, although the speed of convergence will be proportionally lower when agents are more 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.

The implications are both substantively and technically important. On the one hand, the model yields concrete predictions of when organizations will drift to the mainstream or become extremist, given information about the preference distribution. Generally speaking, drift leads organizations towards high-density areas of the preference distribution, which favors centrism if the distribution is unimodal with a

[2] Glaeser and Shleifer (2005) study the case of Mayor Curley in Boston, who used wasteful policies to induce rich citizens to move out, as 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.

centrist mode. 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). Relatedly, steady state policies are more sensitive to the shape of the preference distribution than in models with a fixed population of voters, as they tend to be close to modes of the distribution rather than to the global median voter. In particular, when the distribution is close to uniform, small changes to its density can result in dramatic swings in the set of steady states. In practice, this means that a slow, continuous demographic change may at some point trigger a change in the organization's dynamics, resulting in policy changes which would appear sudden in comparison.

These results are relevant to existing applied theory papers which have studied some of the examples mentioned above. For instance, Grossman (1984) highlights the interaction between wage demands and the membership of trade unions, while 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 assume that this must be so statically, i.e., they study the steady states of the model. In contrast, I allow organizations to start with non-steady state policies and characterize the transition dynamics. Since the transition is generally slow as measured by the players' patience (i.e., they spend most of the equilibrium path, as weighted by their discount factor, relatively far from the steady state), this turns out to be an important consideration, as organizations we study in practice may often be far from a steady state when we observe them.

On the other hand, the paper makes several contributions to the growing literature on dynamic policy selection and "elite clubs". First, it shows that, in fact, the setting of elite clubs—where current members can strategically restrict the entry of newcomers—and the one in this paper—where agents can freely enter and exit, and members choose a policy—are closely related despite the apparent differences. Indeed, they both involve a "coupling" of policy and decision-making power (i.e., agents cannot choose policy and future decision-makers independently), which introduces similar intertemporal trade-offs. Secondly, a novel result arises compared to

Roberts (1999), Barbera, Maschler and Shalev (2001), Acemoglu, Egorov and Sonin (2008, 2012, 2015) and related work: as noted above, in this paper policy always converges towards myopically stable steady states, even with patient agents. In comparison, the aforementioned papers feature "intrinsic" steady states, where agents refrain from changing the policy in desirable ways due to the fear of going down a slippery slope. This difference stems from the fact that I model policy as continuous, rather than discrete; in fact, my results can be translated into existing work, yielding novel predictions about long-run behavior of elite clubs if decision-makers are assumed to have access to a rich set of policy choices.[4] Thirdly, this paper provides a more comprehensive characterization of equilibrium paths: in particular, in the continuous time limit with frequent elections, a closed form solution is reached under fairly general conditions. Finally, within this literature, this paper is the first to obtain tractable results in a setting that violates the single-crossing assumption on preferences—a necessary complication stemming from the fact that agents too far from the chosen policy can cut their losses by leaving the organization.

The paper is structured as follows. The basic model is presented in Section 2. Section 3 shows some common properties of all equilibria and discusses the practical implications of the results. In Section 4, the transition dynamics are characterized in discrete and continuous time. Section 5 discusses an extension of the model to allow for competition between multiple clubs. Section 6 concludes. All proofs can be found in the Appendices.

2 The Model

There is a club existing in discrete time $t = 0, 1, \dots$. 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

[4] Bai and Lagunoff (2011) also consider continuous policies, so their conclusions regarding slippery slopes are similar to mine, but their general results are restricted to smooth equilibria, which at least in my model do not exist generically.

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

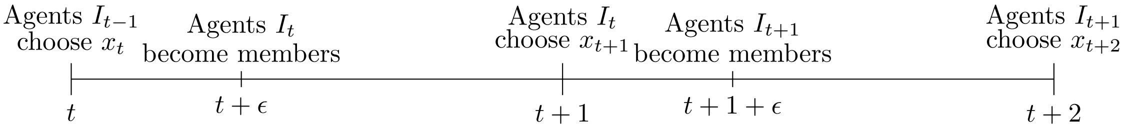

Agents I_{t-1} choose x_t
Agents I_t become members
Agents I_t choose x_{t+1}
Agents I_{t+1} become members
Agents I_{t+1} choose x_{t+2}
t
$t+\epsilon$
$t+1$
$t+1+\epsilon$
$t+2$

Figure 1: Order of play

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 $t+1$. It seems reasonable to rule out these outcomes.[5] Hereafter, I will focus for simplicity on the limit case where $\epsilon = 0$, but preserving the refinements derived when ϵ is small and positive.

Preferences

A potential member α has utility

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

[5] There are other natural assumptions that would give us the same result: for example, if agents have 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.

where 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 $(t, t+1]$. $C > 0$ is the maximum utility the agent can get from being a member. 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, obtaining a flow payoff of 0. Although the assumption of quadratic utility simplifies the analysis, it is not essential to most of the results.[6]

$\delta > 0$ is the agents' common discount factor. That players are forward-looking greatly affects the equilibrium analysis: when choosing a policy, they must take into account how it will drift in the future according to the equilibrum path.

Equilibrium Concept

Without modeling the voting process explicitly we assume that, if there is a Condorcet winner, this will be the chosen policy. In addition, we will focus on Markov Perfect Equilibria (MPE): that is, when votes are cast at time t, the only relevant state variable that voters can condition on will be the set of current members, I_{t-1}; similarly, when entry and exit decisions are made, the only state variable will be the chosen policy x_t. Formally

Definition 1. An MPE is given by a policy function $\tilde{s} : \mathcal{L}([-1,1]) \to [-1,1]$ and a membership function $I : [-1,1] \to \mathcal{L}([-1,1])$ such that:[7]

1. Given a policy x, it is optimal for agents in $I(x)$ to be in the club, and no others.

2. Given a set of voters J, the policy $\tilde{s}(J)$ is a Condorcet winner.[8]

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 $x_{t+1} = s(x_t)$.

[6] In a more general version, preferences would be given by $U_\alpha((x_t, I_{\alpha t})) = \sum_{t=0}^{\infty} \delta^t I_{\alpha t} u_\alpha(x)$ where $u_\alpha(x)$ is strictly concave and C^2 in x for each α; it is maximized at $x = \alpha$, with $u_\alpha(\alpha) \geq C$ for all α; and it has increasing differences in α and x. The results in Section 3 would go through, but a closed-form solution as in Section 4 would be much harder to obtain.

[7] $\mathcal{L}([-1,1])$ is the set of measurable sets contained in $[-1,1]$. We need the set of voters to be measurable for Condorcet winners to be well-defined; in equilibrium, $I(x)$ will always be an interval so this will not be an issue.

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

From here on we will describe equilibria by the functions I and s rather than I and $\tilde{s}$. This is without loss of detail: the set of voters is always of the form $I(x)$, so this fully describes the equilibrium path. Finally, it will be useful 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 $s(x) = x$. A steady state x is *stable* if there is a neighborhood $(a, b) \ni x$ such that $s^k(y) \xrightarrow[k \to \infty]{} x$ for all $y \in (a, b)$.

3 Equilibrium Characterization

We first show some common properties of all MPEs, which in particular pin down the long-run behavior of any equilibrium. We start by solving for the optimal membership decision, which is simple:

Lemma 1. *In any MPE, $I(x) = (x - d, x + d)$, where $d = \sqrt{C}$.*

Indeed, since members can enter or leave at any time, it is optimal to join whenever the flow payoff of the current policy is positive and leave when it isn't; α's flow payoff is positive when $C - (x - \alpha)^2 \geq 0$, i.e., when $\alpha \in (x - \sqrt{C}, x + \sqrt{C})$. We can then describe MPEs solely in terms of successor functions.

For any path $S = (s_1, s_2, \ldots)$, let $E(S) = \sum \delta^t s_t$ be the discounted average policy in S, and define $S(y) = (y, s(y), s^2(y), \ldots)$ as the equilibrium path following from a policy choice y. Two auxiliary lemmas allow us to compare paths based on their average policies, in the vein of increasing differences:[9]

Lemma 2. *Let $S = (s_1, s_2, \ldots)$ and $T = (t_1, t_2, \ldots)$ be policy paths, and let $\alpha_0 < \alpha_1$ be 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. If $E(T) > E(S)$ and α_0 prefers T to S, so does α_1.*

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

[9] $u_\alpha(x) = C - (x - \alpha)^2$ 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 chooses to leave at different times given different policy paths.

In other words, preference for a high average policy is increasing in α only when the two voters being compared never want to exit the club under either path. However, if one of the paths is constant and the paths do not overlap, the result holds across all voters.

Armed with these tools, we can show that equilibrium paths must be monotonic:

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

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 $s_k > s_{k-1}$.[10]

In describing the equilibrium it will be useful to define the *median voter function* m. For $x \in [-1, 1]$, let $m(x)$ be the median among the set of agents who would choose to join if the policy were x, i.e., the median voter in $I(x)$. Formally

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

We can now 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 s and for any y:*

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

(ii) If $m(y) > y$ then $m^*(y) > s(y) > y$.

(iii) If $m(y) < y$ then $m^*(y) < s(y) < y$.

Moreover, $s^k(y) \xrightarrow[k\to\infty]{} m^*(y)$.

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

[10] The same result is shown in Acemoglu et al. (2015), but the proof is more involved in this setting because, owing to the infinite policy space, we also have to rule out cases where the path doubles back on itself infinitely many times.

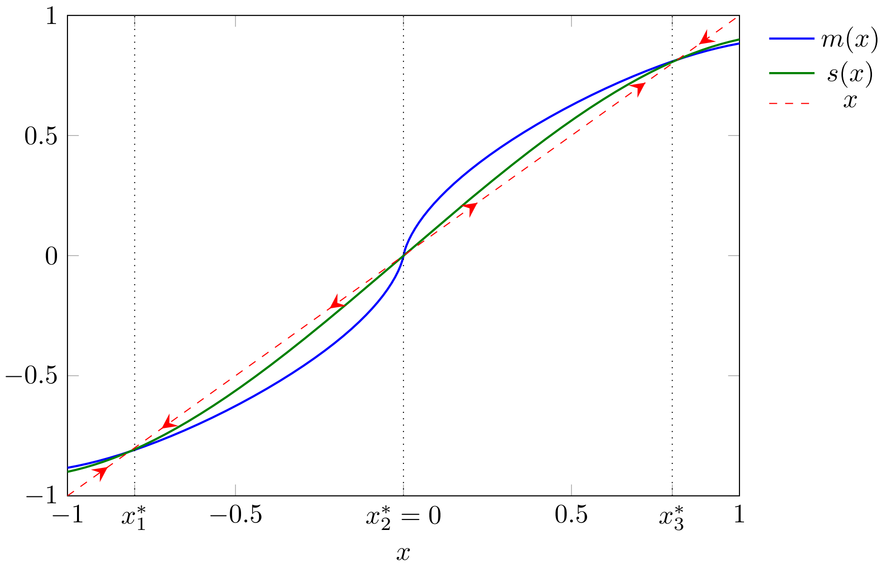


Figure 2: Convergence to steady states in MPE

states of s are also stable (unstable) fixed points of m, and their basins of attraction coincide. Since the long-run policy depends on the initial position of the club, equilibria exhibit path-dependence.[11] Figure 2 illustrates this result in an example with three steady states: x_1^* and x_3^* are stable, while x_2^* is unstable. This alternation of stable and unstable steady states is fairly 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.*[12] *Then n is odd; x_i^* is stable for odd i and unstable for even i; and, in any MPE, equilibrium paths starting at any $y \in (x_{2j}^*, x_{2j+2}^*)$ converge to x_{2j+1}^*.*

The intuition for why the policy should move towards a stable fixed point of m is straightforward. Let x^* be such a point, and suppose the club is initially at a slightly higher policy $x > x^*$. To the right of x^* we have $m(y) < y$, i.e., any policy attracts a set of members whose median's bliss point wants to lower the policy. If voters were myopic, they would choose $s(x) = m(x) < x$. As they are forward-looking, they will typically choose to move more slowly, but still in the same direction, and they will

[11] The multiplicity of MPEs and the multiplicity of steady states are different phenomena: a single successor function s can lead to different steady states given different initial policies; conversely, we may have multiple MPEs sharing a unique steady state if m has a unique fixed point.

[12] m is differentiable since f is continuous.

still converge to policies that are myopically stable.[13]

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:*

(i) s is weakly increasing in $I \cap J$.

(ii) $s(y)$ is $m(y)$'s most-preferred policy for all $y \in I \cap J$; in other words, the Median Voter Theorem holds in this region.

There may be non-monotonicities away from the steady state, driven by the interaction between different voters' bliss points and their optimal times to quit the club: voters with a higher bliss point prefer paths with a higher average policy, 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 be a collection of disjoint intervals, reflecting different drop-out times. On the other hand, as we will see later, equilibria that are monotonic in the entire policy space exist when voting is frequent.

Long-Run Behavior

As noted above, the long-run behavior of the club does not depend on the players' discount factor at all: indeed, in any MPE starting at some y, the club's policy converges to $m^*(y)$, the same result that would obtain if $\delta = 0$. In the language of Roberts (1999) all steady states are "extrinsic", i.e., pinned down by the fundamentals independently of the equilibrium transition paths. This contrasts with other papers in this literature such as Roberts (1999) and Acemoglu et al. (2012, 2015), where "intrinsic" steady states exist: that is, policies considered suboptimal by the current voters can be sustained indefinitely in equilibrium due to the fear that, if a line is

[13] As shown later, the speed of convergence is roughly inversely proportional to the discount rate.

crossed, future agents will move towards a different policy too quickly, the so-called *slippery slope* argument.[14]

The models leading to that result share two important assumptions: a discrete policy space and patient agents. In contrast, the continuous policy space in our model always affords the option to move slowly enough towards the steady state, so there is a better alternative to moving too fast or not at all. Indeed, when $m(x)$ chooses[15] between staying at x indefinitely vs. moving to a policy that is myopically slightly better, the potential cost of the latter option is that $m(x)$ will relinquish the choice of the continuation path to a slightly different voter next period; however, since their preferences are similar and utilities are flat close to the optimum, this is a second-order cost, while the benefit of getting a better policy even for one period is a first-order benefit. Of course, this argument does not carry through with a discrete policy space.[16]

Hence, whether slippery slope concerns would stall policy change in a real-life setting may depend on institutional details, namely, on whether the speed of change can be regulated by using incremental changes, or whether only certain large reforms are possible. For example, take a polity with limited franchise considering whether to extend the franchise. (This example does not exactly fit our story but the same techniques we use can be readily applied to it.) Suppose that voters are ordered by their income and high-income voters 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% to be enfranchised, a voter in the top 20% would want the top 25% to be enfranchised, etc.). Then, if it is possible 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, etc.), indefinite stalling is much more likely.

[14] 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.

[15] For simplicity, we suppose $m(x)$ is pivotal in $I(x)$, as per Proposition 2.

[16] If we simultaneously take δ to 1 and make the (discrete) policy space increasingly fine, whether we get intrinsic steady states or not depends on the order of limits.

Distribution of Steady States

A natural application of the model concerns the impact of policy drift on extremism: does convergence to steady states lead to moderate policies in the long run? Or can extremist factions "capture" the club indefinitely?

The distribution of steady states reflects the following intuition: if f is increasing around x then $m(x) > x$, so the policy should drift upward, and vice versa. Hence stable steady states correspond roughly to maxima of the density function:

Lemma 5. *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 is a unique steady state close to 1 (-1); if f is symmetric and single-peaked, 0 is the unique steady state.[17]

At face value, this suggests that policy drift encourages moderate policies: if centrist voters are abundant, there will be a stable steady state close to 0 with a large basin of attraction. Yet there are three reasons why extremism may be sustainable, in the sense that the club will converge to a policy more extreme than the bliss points of most voters.

First, if most voters are moderates but f has local maxima near the extremes, there may be multiple stable steady states, including some near the extremes. This is especially likely if d is low, i.e., if the club attracts a relatively narrow niche, 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.

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, but f_1 has a unique steady state close to -1, f_2 has one close to 1, and f_3's is at 0. Hence, small demographic changes can have a dramatic impact on the long-run policy. This example contrasts with models of voting with a fixed population, where

[17] In terms of welfare this is good news, yet still potentially inefficient, for two reasons. First, while a stable steady state x^* must be close to a local maximum of f, it does not necessarily maximize the number of members $F(x^* + d) - F(x^* - d)$ or the sum of their utilities, even locally. Second, the choice of steady state is still based on the starting policy; there is no guarantee that the club will converge to steady states that serve more members.

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 hinges 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, 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 radical.

For an example where this assumption is unreasonable, suppose that the club is a nationalist organization. Moderate agents want to engage in benign activities, such as enjoying traditional meals and music, publishing a local newspaper for their community, etc., while hard-liners want to organize attacks against immigrants. Importantly, hard-liners would likely still join the club even if it were too moderate for their tastes, but moderates would want to leave the club if turned 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 $(x-d_-(x), x+d_+(x))$, inducing a different median voter function $m(x)$. For a stark reflection of our example, suppose that $I(x) = [x-d, 1]$ if $x > 0$, $I(x) = [-d, d]$ if $x = 0$ and $I(x) = [-1, x+d]$ if $x > 0$, with $d < 1$. 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 will be two stable steady states, close to -1 and 1 respectively, so the club will always go to an extreme in the long run. In general, if extremists are more willing to join the club, they may end up capturing it even if they are a minority.

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 moderates quit, they will lead existing groups to radicalize as only the extremists stay in them.

4 Transition Dynamics

Although all MPEs share the same long-run behavior, there are typically multiple equilibria with slightly different transition dynamics. In this Section, we illustrate what drives this multiplicity and study a natural class of equilibria.

Without loss of generality, we restrict the game to the right side of the basin of attraction of a stable steady state. That is, let $x^* < x^{**}$ such that $m(x^*) = x^*$, $m(x^{**}) = x^{**}$ and $m(y) < y$ for all $y \in (x^*, x^{**})$. Then we study s restricted to $[x^*, x^{**}]$. (The analysis is similar for a basin of attraction of the form $[x^*, 1]$.) In any MPE s, $s(y) \in (x^*, x^{**})$ 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.

It turns out that multiplicity is pinned down by behavior that occurs arbitrarily close to x^*:

Lemma 6. *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. Conversely, if there are multiple equilibria, their differences must start from the beginning.

Next, we define a special set of equilibria, which we will focus on for the rest of the paper:

Definition 3. Let s be an MPE on $[x^*, x^{**}]$. s is a *1-equilibrium* if there is a sequence $(x_n)_{n\in\mathbb{Z}}$ such that $x_{n+1} < x_n$ for all n, $x_n \xrightarrow[n\to-\infty]{} x^{**}$, $x_n \xrightarrow[n\to\infty]{} x^*$, and $s(x) = x_{n+1}$ if $x \in [x_n, x_{n-1})$.[18]

In a 1-equilibrium, the equilibrium path always follows a single sequence of policies; other points are never chosen. It is easy to guess why this might happen: 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 δ is relatively high, so that convergence to x^* takes time, then we expect that $x_{n+1} > m(x_{n-1})$, so $m(x_{n-1})$ sees no benefit to choosing $x \in (x_{n+1}, x_n)$ instead of

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

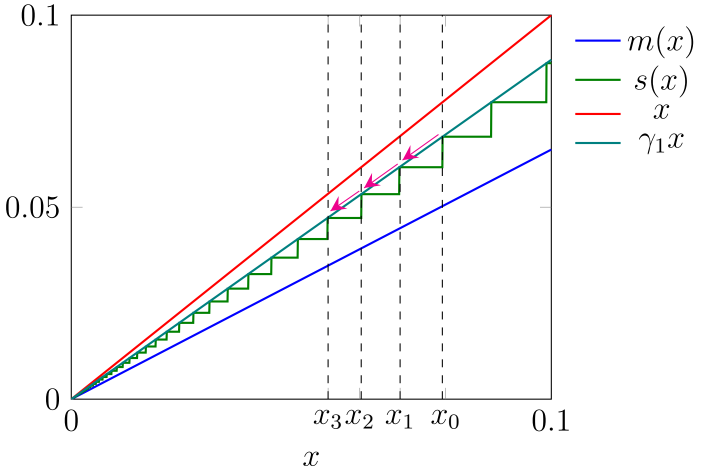


Figure 3: 1-equilibrium for $m(x) = 0.7x$, $\delta = 0.7$

x_{n+1}. In addition, there is indifference at every discontinuity, i.e., $m(x_{n-1})$ must in fact be indifferent between x_n and x_{n+1}.

As illustrated in Figure 3, we can find explicit 1-equilibria when m is linear.[19] This serves as a useful approximation of the general case, since a differentiable m is approximately linear in a neighborhood of a steady state.

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 $e \leq d$. Assume $\delta \geq \frac{2}{3}$ and $\alpha \geq \frac{1}{2}$. Then, for each $\underline{x} < e$, there is a 1-equilibrium s^* in $[-e, e]$ such that $x_0 = \underline{x}$, given by $x_n = \gamma_1^n \underline{x}$, where $0 < \gamma_1 < 1$.*

Although these equilibria are highly structured there is one degree of freedom, namely, the choice of x_0. The exact placement of the points $(x_n)_n$ is arbitrary, but the ratio between the elements of the sequence is uniquely determined.

In the general case, 1-equilibria have additional desirable properties. A 1-equilibrium is guaranteed to exist in a neighborhood of a stable steady state, and to extend to its whole basin of attraction, under general conditions. These results are driven by the special structure of 1-equilibria, whereby 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.

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

To state these results formally we need the following definition. Let s be a 1-equilibrium candidate given by sequence $(x_n)_n$. We say that s is a *quasi-1-equilibrium* if, for all n, $m(x_n)$ is indifferent between $S(x_{n+1})$ and $S(x_{n+2})$ and prefers them to all other $S(x_k)$ (but not necessarily to other $S(x)$). Then

Proposition 4. *Let $\underline{x} \in (x^*, x^{**})$. Then there is a quasi-1-equilibrium $s_{\delta,\underline{x}}$ 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 $x \in [x^*, x^{**}]$.*

$s_{\delta,\underline{x}}$ is an equilibrium in $[x^*, m^{-1}(x^* + d)]$ iff $m(x_n) < x_{n+2}$ for all n such that $m(x_n) < x^* + d$. Moreover, if Condition (*) is satisfied, there exists $\overline{\delta} < 1$ such that $s_{\delta,\underline{x}}$ is an equilibrium in $[x^*, x^{**}]$ for all $\delta < \overline{\delta}$.[20]

The game may also admit other equilibria, such as k-equilibria (featuring k interleaved sequences) and continuous equilibria (where s is continuous). However, the existence of these equilibria is not robust as they are inherently unstable. For more detail on these, see Appendix D.

Continuous Time Limit

So far we have analyzed a discrete time model. This is the natural choice, given that in most organizations voting happens periodically (e.g., at weekly meetings, annual elections, etc.), but it creates technical issues which are partially avoided, allowing for cleaner results, if we move to continuous time.

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

An MPE will now be given by a function $s^t(x)$ denoting the successor policy starting from x after a length of time t has passed. Note that s must be additive in t: $s^t(s^{t'}(x)) = s^{t+t'}(x)$ and $s^0(x) = x$. Moreover, if s is the limit of a sequence of discrete time equilibria, $s^t(x)$ must be decreasing in t (at least in a neighborhood of x^*, by Proposition 2) and $s^t(x) \xrightarrow[t\to\infty]{} x^*$ for all $x \in [x^*, x^{**})$. We say s is continuous if $s^t(x) \xrightarrow[t\to 0]{} x$ for all x.

The following Lemma provides a useful characterization of s:

[20] Condition (*) requires the equilibrium in the continuous time limit not to have large instantaneous jumps. It is fully stated in Appendix B.

Lemma 7. *If $s^t(x)$ is continuously differentiable and strictly decreasing in t, then there are functions $d(x,y) : [x^*, x^{**}]^2 \to \mathbb{R}$ and $e(z) : [x^*, x^{**}] \to \mathbb{R}_+$ such that $s^{d(x,y)}(x) = y$ and $d(x,y) = \int_y^x e(z)dz$.*

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

We can now give an equilibrium characterization in closed form:

Proposition 5. *Suppose $f \in C^1$, and 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, if $\tilde{e}(x) \geq 0$ for all $x \in [x^*, x^{**}]$, there is an MPE s_* given by $e \equiv \tilde{e}$. Moreover, this is the only continuous MPE; all sequences of quasi-1-equilibria of the j-refined games $(s_j)_j$ converge a.e. to s_*, i.e., $s_j^t(x) \xrightarrow[j\to\infty]{} s_*^t(x)$ a.e.; and all quasi-1-equilibria s_j are 1-equilibria for $j \geq j_0$.

The intuition behind the solution, illustrated in Figure 5,[21] is as follows. Suppose s is continuous in t; then it is as if $I(x)$ were choosing $s(x) = x$ in the limit, so[22]

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

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, i.e., $u_{m(x)}(x) = U_{m(x)}(S(x))$: indeed, if $u_{m(x)}(x) < U_{m(x)}(S(x))$ then $m(x)$ could increase his utility by starting his

[21] $s^1(x)$, the policy after a delay of length 1, does not provide a full description of the equilibrium, but it serves as a direct point of comparison with discrete time equilibria.

[22] This condition makes $s(x) = x$ optimal for $m(x)$. This is enough because x is higher than $S(x)$, so Lemma 3 applies.

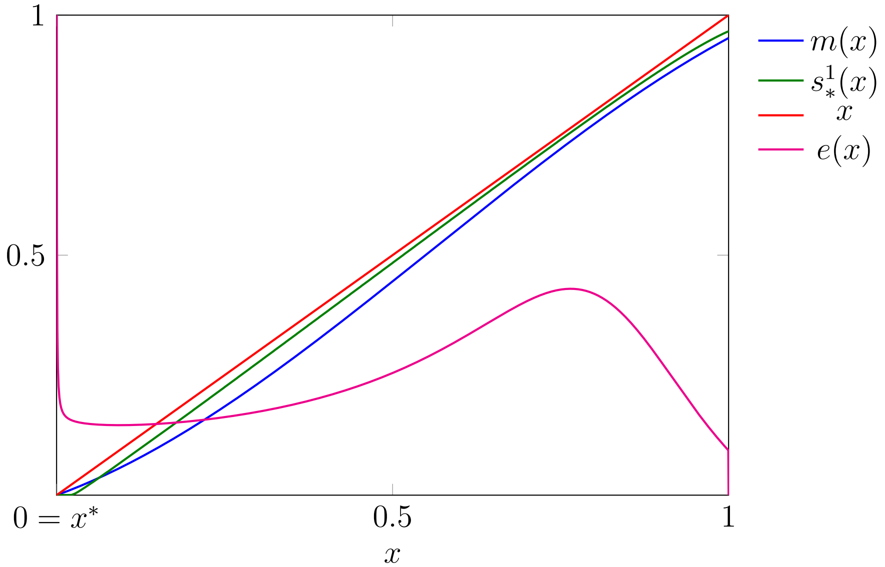


Figure 4: Equilibrium in continuous time

path at a lower policy, while if $u_{m(x)}(x) > U_{m(x)}(S(x))$ he'd rather not move at all. Hence

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

whence we can back out the required delay, $e(x)$. $e(x)$ tends to be high when $m(x)$, $m'(x)$ and $\frac{m''(x)}{m'(x)}$ are high. To see why, note that if $x - m(x)$ is small, $m(x)$ is already close to his bliss point, so $S(x)$ must move slowly away from x to be preferable to staying at x; if $m'(x)$ is high, lowering the policy requires handing power to more different decision-makers, so $m(x)$ is reluctant to do it. The intuition behind the second derivative is subtler: an *increase* in $m'(x)$ requires a faster increase in $E(S(x))$ to sustain the equilibrium, which is accomplished by means of a high $e(x)$. Finally, when $x > m^{-1}(x^* + d)$ (i.e., $m(x)$ expects to quit in the future), $e(x)$ is lower because $m(x)$ expects not to suffer the full cost from the policy drifting away. The discount rate r acts as a rescaling factor: when r is lower, delay becomes proportionally higher, so that the effective delay (measured against patience) remains constant.[23]

Note however that the continuous time model also admits spurious equilibria where, e.g., a certain $x \in (x^*, x^{**})$ decides not to move at all, and higher x's adapt accordingly. This is a sustainable deviation from s_* because, given s_*, $m(x)$ is indif-

[23] In discrete time, an increase in δ would also slow down convergence at a roughly proportional rate, but there a change in δ has real effects since it also affects the effective frequency of voting.

ferent between $S(x)$ and the path $s^t(x) \equiv x$, but we know from Proposition 1 that this is impossible in the discrete time model. This is why we are interested in limits of discrete time equilibria.

Finally, $\tilde{e}(x)$ may be negative in some interval(s), in which case there is no way to maintain the indifference between $S(x)$ and x, and the equilibrium must involve an instantaneous jump. This generally happens when $m(x)$ is low—for example, in the linear case $m(x) = \alpha x$, it happens when $\alpha \leq \frac{1}{2}$. To see why, consider $m(x)$'s decision in this case. When $\alpha < \frac{1}{2}$, $m(x)$ likes the tail of $S(x)$ strictly better than the current policy x (as, in particular, 0 is preferable to x) so $s^t(x) \underset{t \to 0}{\longrightarrow} x$ is impossible–in fact, $s^t(x) \leq 2\alpha x \; \forall t > 0$, and it follows that the path travels to 0 instantly. The general solution in this case involves a series of jumps and temporary stops and is technically complicated, so it is relegated to Appendix B.

5 Multiple Clubs

In this Section, we discuss an extension of the model where multiple clubs compete for potential members, which can tractably answer questions such as: are the long-term policies or transition dynamics of clubs affected by the presence of competing clubs? Will clubs spread themselves out efficiency along the distribution of agents?

For tractability, I consider the case where a fixed number of clubs is given and focus on analyzing the steady states. Afterwards I briefly discuss the dynamics leading to the steady state.

Steady States

Suppose there are $k > 1$ clubs with initial policy positions $x_{1,0} < x_{2,0} < \ldots < x_{k,0}$. For simplicity, this order is always maintained. We can immediately 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 $I(x)$.) Assume that agents can only belong to one club, so $I_1, \ldots, I_k$ are always pairwise disjoint and $I_l \subseteq [-1, 1]$. Then

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

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. $x_1 < \ldots < x_k$ 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.[24] Note that this condition is different from $m(x_l) = x_l$, as $I_l(x_1, \ldots, x_k)$ is different from $I(x_l)$, except when $\{l\}$ is a cluster.

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

Proposition 6. *Let $(x_1, \ldots, x_k)$ be a steady state. If $\{i\}$ is a cluster, x_i is compatible with steady state iff $m(x_i) = x_i$. If $\{i, i+1, \ldots, j\}$ is 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.*

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 $(j-i+1)$-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 $[x_0 - 2(j-i+1)d, x_0 + 2(j-i+1)d]$, there is a unique valid $(j-i+1)$-tuple $(x_i, x_{i+1}, \ldots, x_j)$ containing x_0.

Note that the Proposition says nothing about the distribution of clubs into clusters: this is generally arbitrary, and constitutes a new source of steady state multiplicity.[25] 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 $k = 2$ and f is log-concave in a

[24] 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. For tractability, here we take the definition of steady states in terms of myopic stability as a primitive.

[25] 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.

large interval around x_1, as well as around x_3. Then there is a two-club steady state where both clubs cluster around x_1; one where they cluster around x_3; and three cases where the clubs are separate and occupy different single-club steady states.

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.

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 states, 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 voters do not internalize this.

Dynamics

A general analysis of transition dynamics in the multi-club case is beyond the scope of this paper, but the main substantive difference compared to the single-club case is simple: when clubs cluster together, the identity of each club's median voter is less responsive to changes in policy due to competition for voters with the other club, which in turn induces faster convergence.

This can be illustrated in an example. Suppose that $f(x) = 1 - \frac{x^2}{4d^2}$ for $x \in [-2d, 2d]$ (in particular, it can be checked that there is slow convergence to the steady state in the single-club case as per Proposition 5). Assume further that there are two clubs with initial positions $x_{10} = -x^* - \eta$, $x_{20} = x^* + \eta$, where $(-x^*, x^*)$ is the steady state from Proposition 6 and $\eta > 0$ is small.

Then there is an MPE of the continuous time game where $s_1^t(a, b) = -x^*$ and $s_2^t(a, b) = x^*$ for all $t > 0$ and $-x^* - \eta \leq a \leq -x^*$, $x^* \leq b \leq x^* + \eta$. In particular, the equilibrium path starting at $(-x^* - \eta, x^* + \eta)$ has both clubs traveling instantly to $(-x^*, x^*)$.

To see why, consider the behavior of $m_i(x)$. Remember that, in the single-club case,$m(x)$ is given by $F(x+d) - F(m(x)) = F(m(x)) - F(x-d)$, so

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

In particular, $m'(0) = \frac{3}{4} > \frac{1}{2}$. In the two-club example, $I_1 = (x_1 - d, \frac{x_1+x_2}{2})$ and $I_2 = (\frac{x_1+x_2}{2}, x_2 + d)$. Taking the other club's policy as fixed, we have

$$\frac{\partial m_1}{\partial x_1} = \frac{f(\frac{x_1+x_2}{2})}{4f(m_1(x_1))} + \frac{f(x_1 - d)}{2f(m_1(x_1))}, \quad \frac{\partial m_2}{\partial x_2} = \frac{f(\frac{x_1+x_2}{2})}{4f(m_2(x_2))} + \frac{f(x_2 + d)}{2f(m_2(x_2))}.$$

Now look at club 2's problem. Since it expects $x_{1t} \equiv x^*$ regardless of the path of x_2, it is effectively solving a single-club problem with the modified median voter function $m_2(-x^*, x_2)$. Compared to $m(x_2)$, $m_2(-x^*, x_2)$ tends to have a lower derivative because, when x_2 moves to the left, club 2 only picks up half as many voters on its left side as it would if it were not competeing for them with club 1, so there are fewer newcomers to drag down the average preference. In particular, for η small,

$$\frac{\partial m_2}{\partial x_2} \approx \frac{f(0)}{4f(x^*)} + \frac{f(x^*+d)}{2f(x^*)} = \frac{1}{1 - \frac{x^{*2}}{4d^2}} \left(\frac{1}{4} + \frac{1 - \frac{(x^*+d)^2}{4d^2}}{2} \right) = \frac{1}{2} + \frac{1}{1 - \frac{x^{*2}}{4d^2}} \left(\frac{1}{8} - \frac{x^*}{4d} \right)$$

It can be shown directly that $x^* > \frac{d}{2}$, so this expression is smaller than $\frac{1}{2}$. Hence the club will move instantly to x^*, as in our example of instant convergence in the previous Section.

6 Conclusions

I have studied a model of policy choice in clubs with endogenous membership, which is relevant to a variety of organizations in the real world. The technical results advance our understanding of both the long-run behavior and the transition dynamics in this problem, and they extend to many existing models of policy choice that feature a linkage between the current policy and the identity of the policymaker. On the substantive side, the model tells us when we can expect organizations to become mainstream or drift towards extremism. In particular, stable steady states are near maxima of the density function when preferences are symmetric, but 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 multiple extensions of the model which deserve a complete analysis in future work; I will briefly discuss three. First, as seen in Section 5, interesting issues arise when multiple clubs interact. My general results in this case are limited to a characterization of the natural steady states. 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. In particular, this would also settle the question of exactly when the presence of competing clubs enables instant convergence.

A different set of very complex interactions arises if clubs offer different payoff profiles. In that context, we would like to understand the conditions under which one club can displace another. This is a relevant question 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 slightly 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. Which strategy would prevail? A preliminary analysis suggests that a niche club can displace a mainstream club, as it will always win the fight for the marginal members; but a club with much wider appeal may instead "wrap around" the base of support of the niche club. With general payoff profiles, which club is stronger may depend on how much pressure there is for their policies to converge close to each other, which depends on the shape of f.

Secondly, although the model studies the evolution of already-existing organizations, these must be created in the first place, and the model shows that the initial policy can matter even in the long run. This raises two questions: when would it be socially optimal to create a club? And when would it be incentive-compatible to do so for whoever is considering it? For instance, if only certain agents have the resources to create clubs, but their bliss points are far from any stable steady state of the preference distribution, they are unlikely to go through with it, as they understand that any club they create will eventually get away from them; or, if it is an option, they may restrict membership to certain demographics to reshape the population of available agents and shift the position of the steady states.

Thirdly, 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., in a situation where migrants choose among several cities in a new country, but after settling in cannot easily move again). On the other hand, many organizations have differing levels of influence among members according to their seniority, rank, etc. (e.g., consider a university where faculty, students and administrators have different standing), and they typically have leaders who choose policies with some leeway, as electoral control is imperfect. It seems likely that the basic results of the model—in particular, regarding long-run behavior—would extend to a more general setting encompassing these cases, but the complexity of the model would quickly grow, as all of these changes would require keeping track of a multi-dimensional state variable. 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 she–by choosing the right policies–reshape the electorate to fit her instead of the other way around, as in the case of Mayor Curley?

A Discrete Time (Proofs)

Proof of Lemma 2.

$$U_\alpha(S) - U_\alpha(T) = \sum_t \delta^t \left(u_\alpha(s_t) - u_\alpha(t_t)\right)$$
$$\implies \frac{\partial(U_\alpha(S) - U_\alpha(T))}{\partial\alpha} = \sum_t \delta^t \frac{\partial(u_\alpha(s_t) - u_\alpha(t_t))}{\partial\alpha}$$

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

Proof of Lemma 3. If $\alpha \geq x$, 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

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

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

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

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) strictly prefer x, while voters in $(x-d, \alpha_0)$ 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 $(x-d, x)$ 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 $s^{k+1}(y) < \overline{y}$.[26]

[26] If we relax the definition of s, there could be paths where $s^k(y) = \ldots = s^{k+m}(y) >$

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 $s^k(y)$.

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 $A=(s^{k-1}(y), s^k(y)-d)$, $B=(s^k(y)-d, s^{k-1}(y)+d)$, $C=(s^{k-1}(y)+d, s^k(y)+d)$. [27] 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.[28] 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,

$s^{k-1}(y), s^{k+m+1}(y)$, 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.

[27] 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.

[28] 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 $s^{k+1}(y) < s^k(y)$.

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 $r \in \{-m, \dots, m\}$, 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 $\overline{y}$.

Second, suppose that $s_*^r < \overline{y}$ for some $r < 0$ but never for $r > 0$. Let $\underline{r}$ be the number closest to 0 satisfying this; then $s_*^r = \overline{y}\ \forall r \geq \underline{r}$. For any high enough i and m, let $r'(i,m)$ be such that $s^{k_{im}+r'(i,m)}(y)$ is the first element of the sequence after $s^{k_{im}(y)}$ that is weakly smaller than $s_*^{\underline{r}}$. By an abuse of notation, pick a large n and construct a subsequence where $s^{k_{im}+r'(i,m)+l}(y) \to s_{**}^l$ for $l = -1, \dots, n$ (in particular $s_{**}^{-1} > s_*^{\underline{r}} \geq s_{**}^0$). Now compare the decisions made by $I(s^{k_{im}+\underline{r}}(y))$ and $I(s^{k_{im}+r'(i,m)-1}y)$. In the limit, they imply that a weak majority in $I(s_*^{\underline{r}})$ prefers $\overline{y}$ to $S(s_{**}^0)$, while a weak majority in $I(s_{**}^{-1})$ prefers the opposite (here $S(s_{**}^0)$ is the limit of the continuation paths as $n \to \infty$). (This follows because the path chosen by intervals arbitrarily close to $I(s_*^{\underline{r}})$, for high i and m, is arbitrarily close to $\overline{y}$ for an arbitrarily long time; on the other hand, the path $S(s_{**}^0)$ is strictly to the left of $\overline{y}$). This also leads to a contradiction as in Case 1.

Third, suppose that $s_*^r < \overline{y}$ for some $r > 0$ but never for $r < 0$, and let $\overline{r}$ be the smallest r satisfying this. Let $\nu > 0$ be small (in particular $\nu < \overline{y} - s_*^{\overline{r}}$) and $0 < \nu' << \nu$. For i and m high enough, let $r'(i,m)$ be such that $s^{k_{im}+r'(i,m)}(y)$ is the last element of the sequence before $s^{k_{im}(y)}$ that is weakly smaller than $\overline{y} - \nu$. By assumption, $r'(i,m) \to -\infty$ as $i, m \to \infty$.

In particular, by considering the decision of $I(s^{k_{im}+\overline{r}-1}(y))$ in the limit, we see that $I(\overline{y})$ prefers some path to the left of $\overline{y}$ over $\overline{y}$; this means that $m(\overline{y}) < \overline{y}$. By monotonicity of m, $m^k(\overline{y})$ is strictly decreasing in k and converges to a limit $\tilde{m}$; moreover, $m(y) < y$ for all $y \in (\tilde{m}, \overline{y}]$. Call $s^{k_{im}+r'(i,m)}(y) = s_{im}$. Now consider the decision made by $I(s_{im})$ for large i, m. In the limit, $s_{im} < \overline{y} - \nu$, but $I(s_{im})$ is choosing a path that stays to the right of $\overline{y} - \nu$ for arbitrarily long; moreover, since $s_*^r = \overline{y}$ for all $r < 0$, we know that the path also stays to the right of $\overline{y} - \nu'$ arbitrarily long before going anywhere else. Since $I(s_{im})$ chooses this over simply staying put at s_{im}, we must have $m(s_{im}) \geq s_{im}$, hence $s_{im} \leq \tilde{m}$. Now construct a subsequence such that s_{im} converges to a limit s_0. Then, taking ν arbitrarily small, $I(s_0)$ prefers $\overline{y}$ over any other path, including $S(s_*^0)$, while $I(\overline{y})$ prefers the opposite, a contradiction as in Case 1.

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 $m(y) = y$ 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 $s^{k_n+1}(y)$,[29] 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

[29] 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 $\upsilon > 0$ fixed, voters in $[s^{k_n}(y), s^{k_n}(y) + d - \upsilon]$ 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 $s^{k_n}(y) + d$.

$\nu \to 0$ 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 $I(s^{k_n}(y))$. 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 $y_0 + d$.[30] 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).

[30] 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.

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., $y_1 = \frac{\sum_{s^{k_n}(y)>s_t>y''-d} \delta^t s_t}{\sum_{s^{k_n}(y)>s_t>y''-d} \delta^t}$. 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 $s^{k_n}(y)$, $y_3 = y_2 + \frac{w_3}{w_4} y_1 - s^{k_n}(y)$.[31] 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:

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

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

[31] 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.

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.

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

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

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

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

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

If $k = k' = k'' = k'''$ for n high enough,[32] then this is equal to

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

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

Now, crucially,

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

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.[33]

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

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

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. □

[32] 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 $s^{t+1}(y_n) \to s_{t+1}$.

[33] 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.

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 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,[34] 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 $s(x)$.[35] 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 $m(x') < x' \leq x^* + d$,[36] 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

[34] 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.

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

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

$E(S(s(x)))$ 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. 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 Lemma 6. 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 $s(x^{**}) = s'(x^{**}) = x^{**}$). 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 $s(y_n)$, $s'(y_n) \leq \underline{y} - \upsilon$ for $n \geq n_0$ and some $\upsilon > 0$. Since the continuations to these policies are contained in $[x^*, \underline{y} - \upsilon]$, 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 4. First, we construct a sequence of approximate 1-equilibria as follows. 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 quasi-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 8 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.

This yields a quasi-1-equilibrium. By construction, $m_i(x_n)$ is indifferent between S_{n+1} and S_{n+2}, and prefers these to all other elements of the sequence by increasing differences (e.g., $m_i(x_{n+1})$ is indifferent between S_{n+2} and S_{n+3}, so $m_i(x_n)$ strictly prefers S_{n+2} over S_{n+3}), and $x \in [x_n, x_{n-1})$ strictly prefers S_{n+1} to all other S_k.

Next, we 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'. Moreover, some x_n must fall above $\underline{x}$ by a similar argument to Proposition 1. 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 $[x^* + \epsilon, \underline{x}]$ has no elements of the sequence in it, which leads to a contradiction. (Elements of the sequence initially above $\underline{x}$ cannot leapfrog it by continuity, and there are only finitely many elements in $[x^* + \epsilon, \underline{x}]$, which can be reduced below $x^* + \epsilon$ by lowering x' enough). Let s_i be a quasi-1-equilibrium for m_i with $x_0 = \underline{x}$.

Now, we construct a quasi-1-equilibrium s for m such that the s_i converge to s, i.e., $x_{in} \to x_n$ for all n. We do this by a diagonal argument: $x_{i0} = \underline{x}$ for all

i, so $x_0 = \underline{x}$. 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 $x_{i1} \to x_1$. (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. The indifference conditions that made the s_i quasi-1-equilibria under m_i make s a quasi-1-equilibrium under m by continuity.

$x_n \to x^{**}$ as $n \to -\infty$, again by the argument in Proposition 1. s is weakly increasing by design, and our construction does not rely on being within the interval $[x^*, x^* + d]$. The Median Voter Theorem also holds even outside of $[x^*, x^* + d]$ (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).

Finally we will show that s is a 1-equilibrium in $[x^*, m^{-1}(x^*+d)]$ iff $m(x_{n-1}) < x_{n+1}$ for all n. First, note that $m(x_{n-1}) < x_n$ always (otherwise he could not be indifferent between x_n and S_{n+1}). Now, ($\Rightarrow$) is easy: if $m(x_{n-1}) < x_{n+1}$, $m(x_{n-1})$ prefers $m(x_{n-1})$ to x_{n+1}; hence he prefers $S(m(x_{n-1}))$ to S_{n+1}, and also to S_n, so s is not a 1-equilibrium.

For ($\Leftarrow$), suppose that some $m(x_{n-1})$ strictly prefers $S(x)$ to x_n. Then we argue that he prefers $S(m(x_{n-1}))$ to S_n. Let $l \geq n$ be such that $m(x_{n-1}) \in [x_l, x_{l-1})$. Clearly $l \geq n + 1$ as above.

Suppose $x \in (x_b, x_{b-1})$ for $b < l$. Then $m(x_{n-1})$ strictly prefers $S(x_b)$ over $S(x)$ (as x_b is better than x and continuations are the same) and $S(x)$ over x_n, a contradiction. Next, suppose $x \in (x_b, x_{b-1})$ for $b > l$. Then $m(x_{n-1})$ strictly prefers $S(x_{b-1})$ over $S(x)$ (as his bliss point is to the right of both paths and $S(x_{b-1})$ is higher in the FOSD sense) and $S(x)$ over x_n, a contradiction. Then $x \in (x_l, x_{l-1})$. But then $x = m(x_{n-1})$ is optimal, so $S(m(x_{n-1}))$ beats x_n.

Now we argue that this implies $m(x_{j-1}) > x_{j+1}$ for some $j \geq n$. Suppose this is false for $j = n$. Then, by increasing differences, $m(x_n)$ prefers $S(m(x_{n-1}))$ to $S(x_{n+1})$. By our previous arguments, $m(x_{n-1})$ is in the same interval $[x_l, x_{l-1})$ as $m(x_n)$, and $m(x_n)$ prefers $S(m(x_n))$ to x_{n+1}. We can continue this argument to show that $m(x_{l-2}) > x_l$. The case where some $x \neq x_n$ wants to deviate is analogous. □

B Continuous Time Limit (Proofs)

Proof of Lemma 7. 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. □

Proposition 7 (Proposition 5 restated). *Let*

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

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

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

Otherwise, assume that $f \in C^2$ *and* $A = \{x \in [x^*, x^{**}] : \tilde{e}(x) < 0\}$ *is a finite union of open intervals. Let* $\tilde{x} = \inf A$. *Then, for generic* m, s_* *is given by* $e(x) = \tilde{e}(x)$ *for* $x \leq \tilde{x}$, *and by two sequences* $(y_l)_{l \in \mathbb{N}_{\geq 0}}$, $(e_l)_{l \in \mathbb{N}_{\geq 1}}$ *such that:* $(y_l)_l$ *is increasing,* $y_0 = \tilde{x}$ *and* $y_l \xrightarrow[l \to \infty]{} x^{**}$; $d(y_l^-, y_{l+1}^+) = 0$ *and* $d(y_l^+, y_l^-) = e_l$ *for all* $l \geq 1$; *and*[37]

$$\begin{gathered} U_{m(y_{l+1})}(S(y_l^+)) = u_{m(y_{l+1})}(y_{l+1}) \\ \frac{y_l - m(y_l)}{m'(y_l)(y_l - E_{y_l}(S(y_l^-)))} = e^{-\frac{re_l}{2}}. \end{gathered}$$

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

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

Proof of Proposition 5. Consider first $x \in [x^*, m^{-1}(x^* + d))$, so that the median voter $m(x)$ never leaves the club. Assume a smooth s. Then

$$\begin{aligned} C - (m(x) - x)^2 &= \int_0^\infty re^{-rt} \max\left(C - (m(x) - s^t(x))^2, 0\right) dt \\ \Longrightarrow 2m(x)(x - E(x)) &= x^2 - W(x), \end{aligned}$$

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

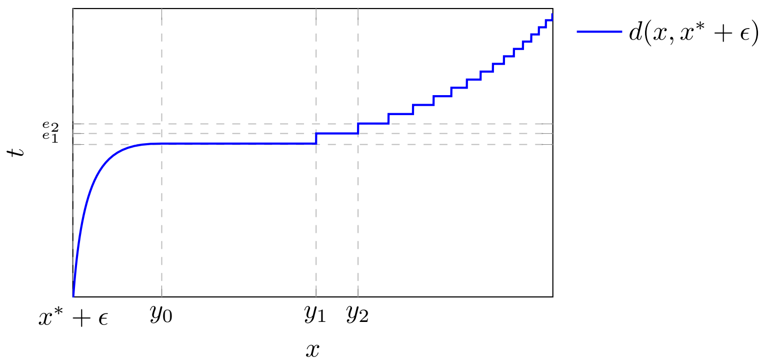


Figure 5: Equilibrium with jumps and stops in continuous time

where $E(x) = \int_0^\infty re^{-rt}s^t(x)$ and $W(x) = \int_0^\infty re^{-rt}(s^t(x))^2$. Note that $E'(x) = re(x)(x - E(x))$ and $W'(x) = re(x)(x^2 - W(x))$.[38] This implies $2m(x)E'(x) = W'(x)$. Derive the above equation to get

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

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 $\hat{E}_{x_0}(x) = \int_0^\infty re^{-rt}\max(s^t(x), m(x_0) - d)$, $\hat{W}_{x_0}(x) = \int_0^\infty re^{-rt}\max((s^t(x))^2, (m(x_0) - d)^2)$. $\tilde{E}$ and $\tilde{W}$ are statistics for a truncated path where policies below $m(x) - d$ are instead set equal to $m(x) - d$; $\tilde{E}_{x_0}$, $\tilde{W}_{x_0}$ make the lower bound for truncating policies $m(x_0) - d$, independent of x. (In particular, $\hat{E}_x(x) = \tilde{E}(x)$, $\hat{W}_x(x) = \tilde{W}(x)$ and $E_x(x) = \tilde{E}(x) + de^{-rt^*(x)}$.)

[38] 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.

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

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

The new terms in the two last equations result from the lower bound, $m(x) - d$, being shifted up as x increases. Deriving the first order condition, we now get

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

Deriving this expression,

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

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

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

We now consider the case where $\tilde{e}$ is negative in an interval starting at $y_0 = \tilde{x}$. Denote $V_{m(x)} = U_{m(x)}(S(x)) - u_{m(x)}(x)$. Here, no non-negative delay around $\tilde{x}$ can sustain the FOC, so $V_{m(x)} > 0$ for $x > \tilde{x}$ close to $\tilde{x}$, and hence $e(x) = 0$ (as $m(x)$ strictly wants to move away from x, i.e., $x - \lim_{t \to 0} s^t(x)$ is bounded away from 0).

Let y_1 be the lowest $x > y_0$ for which $V_{m(x)} = 0$. The "jump" that started at y_0 must end here. However, generically $V'_{m(y_1^-)} < 0$, in which case V must have a kink at $m(y_1)$ so that $V'_{m(y_1^+)} \geq 0$ (otherwise the condition $V \geq 0$ would be violated). This is achieved by means of a temporary stop at y_1 that increases $E(y_1^+)$, i.e., $d(y_1^+, y_1^-) =$

$e_1 > 0$. In principle e_1 could be any length of time high enough that $V'_{m(y_1^+)} \geq 0$; the lowest possible would induce $V'_{m(y_1^+)} = 0$, taking us back to a continuous equilibrium. However, this is *not* the limit of discrete equilibria. Instead, the "correct" e_1 used by s_* causes $V'_{m(y_1^+)} > 0$, which starts another jump up to some value y_2, followed by another stop of length e_2, etc.[39]

Next, we show that $y_l \to x^{**}$. Suppose instead that $y_l \to y^* < x^{**}$. Note that

$$E_{y_l}(y_l^+) = (1 - e^{-re_l})y_l + e^{-re_l}E_{y_l}(y_l^-) \Longrightarrow y_l - E_{y_l}(y_l^+) = e^{-re_l}(y_l - E_{y_l}(y_l^-)).$$

Let $m(y_l) = m_l$, $m'(y_l) = m'_l$, and ϵ_l, δ_l, ν_l, γ_l such that

$$\begin{aligned} \frac{y_l - m_l}{m'_l(y_l - E_{y_l}(y_l^-))} &= \frac{1}{1+\epsilon_l}, \qquad \frac{y_l - m_l}{m'_l(y_l - E_{y_l}(y_l^+))} = 1 + \epsilon_l \\ m'_{l+1} &= m'_l(1+\nu_l), \quad \frac{m_{l+1} - m_l}{y_{l+1} - y_l} = m'_l(1+\delta_l) \\ y_{l+1} - y_l &= \gamma_l(y_l - E_l^+) \end{aligned}$$

Note that $V'_{m_l^-} < 0$ implies $\epsilon_l > 0$. Let $\mathcal{O}((y_{l+1} - y_l)^k) = \mathcal{O}((m_{l+1} - m_l)^k) = \mathcal{O}^k$. Since $V_{m_l} = V_{m_{l+1}} = 0$,

$$\begin{aligned} -m_l^2 + 2m_l\hat{E}_{y_l}(y_l^+) - \hat{W}_{y_l}(y_l^+) &= -m_l^2 + 2m_l y_l - y_l^2 \\ -m_{l+1}^2 + 2m_{l+1}\hat{E}_{y_{l+1}}(y_l^+) - \hat{W}_{y_{l+1}}(y_l^+) &= -m_{l+1}^2 + 2m_{l+1}y_{l+1} - y_{l+1}^2 \\ 2m_{l+1}\hat{E}_{y_{l+1}}(y_l^+) - 2m_l\hat{E}_{y_l}(y_l^+) - \hat{W}_{y_{l+1}}(y_l^+) + \hat{W}_{y_l}(y_l^+) &= 2m_{l+1}y_{l+1} - 2m_l y_l - y_{l+1}^2 + y_l^2 \end{aligned}$$

Now, since we assumed f is C^2, F is C^3; m is C^3; and $\tilde{e}$ is C^1.[40] Moreover, either $e(y) = \tilde{e}(y)$ for y in a neighborhood of $m(y^*) - d$; or $m(y^*) - d > x^*$ and the equilibrium has had a finite number of jumps and stops up to then (so $e(y) = 0$ for $y \in (m(y^*) - d - \rho, m(y^*) - d)$ which applies for l high enough), so in either case e is C^1; or else $t^*(y) = \infty$ so this is irrelevant.

Let $D_1 = e^{-rd(y_l^+, m_l - d)}$, $D_2 = e^{-rd(y_l^+, \frac{m_l + m_{l+1}}{2} - d)}$, $D_3 = e^{-rd(y_l^+, m_{l+1} - d)}$, $D_4 = 1 +$

[39] In the non-generic case where $V'_{m(y_l^-)} = 0$ for some l, s_* would become continuous to the right of y_l.

[40] In fact, m'' only has to be Lipschitz for this argument.

$dre(m_l - d)$. Then

$$\begin{aligned}
&D_1 + D_3 re(m_l - d)\frac{m_{l+1} - m_l}{2} = D_3 - D_3 re(m_l - d)\frac{m_{l+1} - m_l}{2} + \mathcal{O}^2 = D_2 + \mathcal{O}^2 \\
&\hat{E}_{y_{l+1}}(y_l^+) - \hat{E}_{y_l}(y_l^+) = (m_{l+1} - m_l)e^{-rd(y_l^+, \frac{m_l}{2} - d)} + \int_{m_l - d}^{m_{l+1} - d} re^{-rd(y_l^+, y)}(m_{l+1} - d - y)e(y) \\
&\Delta_1 = \hat{E}_{y_{l+1}}(y_l^+) - \hat{E}_{y_l}(y_l^+) = (m_{l+1} - m_l)D_2 + \mathcal{O}^3 \\
&E_{y_{l+1}}(y_l^+) - E_{y_l}(y_l^+) = (m_{l+1} - m_l)D_2 + d(D_3 - D_1) + \mathcal{O}^3 = (m_{l+1} - m_l)D_1 D_4 + \mathcal{O}^2 \\
&\Delta_2 = \hat{W}_{y_{l+1}}(y_l^+) - \hat{W}_{y_l}(y_l^+) = (m_{l+1} + m_l - 2d)(m_{l+1} - m_l)D_2 + \mathcal{O}^3
\end{aligned}$$

$$\begin{aligned}
&(2m_{l+1} - 2m_l)\hat{E}_{y_{l+1}}(S(y_l^+)) + 2m_l\Delta_1 - \Delta_2 = 2m_{l+1}(y_{l+1} - y_l) + 2(m_{l+1} - m_l)y_l - y_{l+1}^2 + y_l^2 \\
&m_l'(1 + \delta_l)\left[2\hat{E}_{y_{l+1}}(S(y_l^+)) + (m_l - m_{l+1} + 2d)D_2\right] = 2m_{l+1} + 2m_l' y_l(1 + \delta_l) - y_{l+1} - y_l + \mathcal{O}^2 \\
&y_{l+1} - y_l = m_l'(1 + \delta_l)\left[2(y_l - \hat{E}_{y_{l+1}}(y_l^+)) - (m_l - m_{l+1} + 2d)D_2\right] + 2m_{l+1} - 2y_l + \mathcal{O}^2 \\
&\;= m_l'(1 + \delta_l)\left[2(y_l - E_{y_{l+1}}(y_l^+)) + (m_{l+1} - m_l)D_3(1 + dre(m_l - d))\right] + 2m_{l+1} - 2y_l + \mathcal{O}^2 \\
&\;= m_l'(1 + \delta_l)\left[2(y_l - E_{y_{l+1}}(y_l^+)) + (m_{l+1} - m_l)D_3 D_4\right] + 2m_{l+1} - 2m_l + 2m_l - 2y_l + \mathcal{O}^2
\end{aligned}$$

Dividing by $y_l - E_{y_l}(y_l^+)$,

$$\gamma_l = m_l'(1 + \delta_l)\left[2\frac{y_l - E_{y_{l+1}}(y_l^+)}{y_l - E_{y_l}(y_l^+)} + m_l'(1 + \delta_l)\gamma_l D_3 D_4\right] + 2m_l'(1 + \delta_l)\gamma_l - 2(1 + \epsilon_l)m_l' + \mathcal{O}^2$$

Clearly $\gamma_l \in \mathcal{O}^1$. Since $m_l'(1 + \delta_l) = m'(y)$ for some $y \in [y_l, y_{l+1}]$ by construction, and m' is C^1, $\delta_l \in \mathcal{O}^1$. It can also be shown that $2\delta_l = \nu_l + \mathcal{O}^2$. Hence

$$\begin{aligned}
&\gamma_l = 2m_l'(1 + \delta_l)\left(1 - \frac{E_{y_{l+1}}(y_l^+) - E_{y_l}(y_l^+)}{y_l - E_{y_l}(y_l^+)}\right) + m_l'^2\gamma_l D_3 D_4 + 2m_l'\gamma_l - 2(1 + \epsilon_l)m_l' + \mathcal{O}^2 \\
&\gamma_l = 2m_l'(\delta_l - \epsilon_l) - 2m_l'(1 + \delta_l)\frac{(m_{l+1} - m_l)D_3 D_4}{y_l - E_{y_l}(y_l^+)} + m_l'^2\gamma_l D_3 D_4 + 2m_l'\gamma_l + \mathcal{O}^2 \\
&\gamma_l = 2m_l'(\delta_l - \epsilon_l) - m_l'^2\gamma_l D_3 D_4 + 2m_l'\gamma_l + \mathcal{O}^2 \\
&\gamma_l\left(\frac{1}{m_l'} - 2 + m_l' D_3 D_4\right) = 2(\delta_l - \epsilon_l) + \mathcal{O}^2
\end{aligned}$$

It follows that $\epsilon_l \in \mathcal{O}$. Since $\epsilon \to 0$ near y^*, $\tilde{e} = \frac{1}{rm'}\left(\frac{-1 + 2m' - m'^2 D_3 D_4}{y - E(y)} + m''\right)$ at y^*.

As $\nu_l \approx \frac{m''}{m'}(y_{l+1}-y_l)$ and $\gamma_l = \frac{y_{l+1}-y_l}{y_l - E_{y_l}(y_l^+)}$, if $\tilde{e}(y^*) \neq 0$, $\gamma_l \in \mathcal{O}(\epsilon_l)$ as well.[41] Finally

$$\frac{1}{1+\epsilon_{l+1}} = \frac{y_{l+1}-m_{l+1}}{m'_{l+1}(y_{l+1}-E_{y_{l+1}}(y_l^+))}$$

$$\begin{aligned}
&= \frac{1}{1+\nu_l}\frac{y_l - m_l + y_{l+1} - y_l - (m_{l+1}-m_l)}{m'_l(y_l - E_{y_l}(y_l^+) + y_{l+1} - y_l + E_{y_l}(y_l^+) - E_{y_{l+1}}(y_l^+))} \\
&= \frac{1}{1+\nu_l}\frac{m'_l(y_l - E_{y_l}(y_l^+))(1+\epsilon_l) + \gamma_l(y_l - E_{y_l}(y_l^+))(1-m'_l)}{m'_l(y_l - E_{y_l}(y_l^+) + \gamma_l(y_l - E_{y_l}(y_l^+)) - (m_{l+1}-m_l)D_3D_4)} + \mathcal{O}^2 \\
&= \frac{1}{1+\nu_l}\frac{1+\epsilon_l+\gamma_l\left(\frac{1}{m'_l}-1\right)}{1+\gamma_l - m'_l\gamma_l D_3D_4} + \mathcal{O}^2 \\
&= 1+\epsilon_l+\gamma_l\left(\frac{1}{m'_l} - 2 + m'_l D_3 D_4\right) - \nu_l + \mathcal{O}^2 = 1 - \epsilon_l + \mathcal{O}^2.
\end{aligned}$$

This means that $\epsilon_{l+1} = \epsilon_l + \mathcal{O}(\epsilon_l^2)$, i.e., $(\epsilon_l)_l$ at most decays (or grows) at the rate of a harmonic series, whence $\sum_l \epsilon_l = \infty$. Since $\epsilon_l \in \mathcal{O}(\gamma_l)$, we have $\sum_l \gamma_l = \infty$ which contradicts $y_l \to y^*$.

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

Suppose not, so there is a sequence $(s_j)_j$ and an x_1 for which n is continuous at x_1 but $n_j(x_1) \nrightarrow n(x_1)$. Take a subsequence $(s_l)_l$ such that n_l converges pointwise to some $\tilde{n}$.[42] $\tilde{n} \neq n$ as, in particular, $\tilde{n}(x_1) \neq n(x_1)$.

Let $a = \inf\{x : n \text{ is continuous at } x \text{ and } \tilde{n}(x) \neq n(x)\}$ and $\tilde{V}$ analogous to V for the equilibrium generated by $\tilde{n}$. There are a few cases. For brevity, we assume $a < m^{-1}(x^* + d)$, but the proof readily generalizes.

(Case 1): $a > x^*$.

[41] $\tilde{e}(y^*) < 0$ is impossible, as it would imply $\tilde{e}(y) < 0$ in a neighborhood of y^*, and the first jump to start at some y_l in this neighborhood would have $y_{l+1} > y^*$. The result is still true if $\tilde{e}(y^*) = 0$, but a more careful argument is needed to rule out convergence in this case, which is omitted.

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

(1i) n is continuous in $[x^*, a+\epsilon')$, $\epsilon' > 0$. Since $\tilde{n} = n$ up to a and n is smooth, $V = 0$ up to a. If $\tilde{V} = 0$ up to $a + \epsilon''$ for $\epsilon'' > 0$, $\tilde{n} = n$ there as well, contradiction. So $\tilde{V} > 0$ arbitrarily close to a.

(1ia) A "big" jump starts at a, i.e., $\exists \rho > 0$ s.t. $\tilde{V} > 0$ on $(a, a+\rho)$. Then $\tilde{n}$ is constant on $(a, a+\rho)$, since $x - \lim_{t \to 0} s^t(x)$ is bounded away from 0. However, if $\tilde{n}$ is also continuous at a, this would yield a contradiction, as $\tilde{V}' \leq V' = 0$ on $(a, a+\rho)$ and $\tilde{V}(a) = 0$. Hence $\tilde{n}$ must be discontinuous at a, i.e., there is a temporary stop at a of some length e^*.

Take $\epsilon > 0$ small. Then s_j's defining sequence must have $je^* + jd(a+\epsilon, a-\epsilon) + o(j)$ elements in $(a-\epsilon, a+\epsilon)$, and $jd(a-\epsilon, a-2\epsilon) + o(j)$ elements in $(a-2\epsilon, a-\epsilon)$. In particular, given $\eta > 0$, for high enough j there must be an element of s_j's sequence $x_{j0} \in (a-2\epsilon, a-\epsilon)$ such that $x_{j0} - x_{j1} \geq \frac{1}{j\overline{e}(1+\eta)}$ for $\overline{e} = \max_{x \in (a-2\epsilon, a-\epsilon)} e(x)$. Exploiting m_{n-1} and m_n's indifference conditions, in general

$$-m(x_{n-1})^2 + 2m_{n-1}E_{n+1} - E_{n+1}^2 - V_{n+1} = -m_{n-1}^2 + 2m_{n-1}x_n - x_n^2$$

$$-m_n^2 + 2m_nE_{n+1} - E_{n+1}^2 - V_{n+1} = -m_n^2 + 2m_nx_{n+1} - x_{n+1}^2$$

$$2(m_{n-1} - m_n)E_{n+1} = 2(m_{n-1} - m_n)(x_{n+1}) + 2m_{n-1}(x_n - x_{n+1}) - x_n^2 + x_{n+1}^2$$

$$\frac{m_{n-1} - m_n}{x_n - x_{n+1}} = \frac{\frac{x_{n+1}+x_n}{2} - m_{n-1}}{x_{n+1} - E_{n+1}}, \quad \frac{x_{n-1} - x_n}{x_n - x_{n+1}} = \frac{\frac{x_{n+1}+x_n}{2} - m_{n-1}}{m'(y)(x_{n+1} - E_{n+1})} \text{ for } y \in [x_n, x_{n-1}]$$

So, for $n \in \{-T, \ldots, 0\}$

$$\frac{x_{j(n-1)} - x_{jn}}{x_{jn} - x_{j(n+1)}} \geq \frac{a - 2\epsilon - m(a+\epsilon)}{\overline{m}'(a + \epsilon - E_{j(n+1)})}$$

$$E_{j(n-1)} \geq e^{-\frac{r}{j}} E_{jn} + (1 - e^{-\frac{r}{j}})(a - 2\epsilon) \implies (a - 2\epsilon - E_{j(n-1)}) \leq e^{-\frac{r}{j}}(a - 2\epsilon - E_{jn})$$

$$x_{j(n-1)} - x_{jn} \geq (x_{j0} - x_{j1}) \prod_{n-1}^{0} \frac{a - 2\epsilon - m(a+\epsilon)}{\overline{m}'(a + \epsilon - E_{j(t+1)})}$$

$$\geq \frac{1}{j\overline{e^*}(1+\eta)} \left[\frac{a - 2\epsilon - m(a+\epsilon)}{\overline{m}'(a + \epsilon - E_{j1})} \right]^n \frac{1}{e^{-\frac{r}{j}\frac{n(n-1)}{2}}}.$$

Since n is smooth around a, $\frac{a - m(a)}{m'(a - E(a))} = 1$. Take ϵ small enough and j high enough that $\left[\frac{a-2\epsilon-m(a+\epsilon)}{\overline{m}'(a+\epsilon-E_{j1})} \right] > e^{-r\frac{e^*}{4}}$. Then it can be shown that $x_{j(-je)} - x_{j0} > 4\epsilon$ for high j, a contradiction.

(1ib) There is no large jump at a. Instead, there are points $x > a$ arbitrarily close to a for which $V_{m(x)} > 0$ or $= 0$. For each $\epsilon > 0$, let $x_\epsilon = \inf\{x \geq a : V_{m(y)}|_{(x,x+\epsilon)} > 0\}$ be the beginning of the first jump of size ϵ or higher. Then there are two possibilities: if $\tilde{s}$ follows a series of jumps and stops given by sequences $(y_l)_{l\in\mathbb{Z}}$, $(e_l)_{l\in\mathbb{Z}}$ as determined in the Proposition, we obtain a contradiction from the fact that $y \xrightarrow[l\to-\infty]{} a > x^*$, similarly to the proof that $y \xrightarrow[l\to\infty]{} x^{**}$. Else, for some ϵ, $\tilde{s}$'s behavior in (x_ϵ, x^{**}) implies a violation of Case (1iia) or (1iib).

(1iia) n is not continuous up to a, and n is locally constant at a ($\tilde{n}$ stops a jump prematurely). This is impossible, as it would imply $\tilde{V}_{m(a+\epsilon)} > V_{m(a+\epsilon)} \geq 0$ for small $\epsilon > 0$, whence $\tilde{n}$ must in fact feature a jump through $a + \epsilon$.

(1iib) n is not smooth up to a, and it is discontinuous at a ($\tilde{n}$'s stop at a is too long, too short or nonexistent). This leads to a contradiction by a similar argument to Case (1ia). Informally, e_l must be such that, if s_j has x_{jn} at the beginning of a stop and x_{jm} at the end, $x_{j(n-1)} - x_{jn}$ and $x_{j(m-1)} - x_{jm}$ must have the same order of magnitude.

(Case 1): $a = x^*$.

(2i) n is continuous in $[x^*, x^* + \epsilon']$.

(2ia) $\tilde{V}_{m(x)} > 0$ for $x \in (x^*, x^* + \epsilon)$. This implies that $\tilde{n}$ jumps to x^*, i.e., $\tilde{V}_{m(x)} = u_{m(x)}(x^*)$, but then $V_{m(x)} > 0$ unless n also jumps directly to x^*.

(2ib) There is a sequence of small jumps converging to x^*. This is similar to Case (1ib).

(2ii) n is not continuous in $[x^*, x^* + \epsilon]$. By our assumptions, this means $e < 0$ in a neighborhood of x^*, so $V > 0$ and a jump is forced in any solution.

Finally, if Condition (*) is met, suppose there is a sequence of quasi-1-equilibria $(s_{j'})_{j'}$ that are not equilibria, with $j' \to \infty$. For each, let $x_{j'}$ be such that $m(x_{j'})$ prefers $S(m(x_{j'}))$ over $S(s(x_{j'}))$. Take a subsequence $(s_j)_j$ such that $(x_j)_j$ has a limit x_*.

If $x_* \in (x^*, x^{**})$, there are two cases. If $V_{m(x_*)} = 0$, Condition (*) implies that the path $S(x_*)$ spends positive time on policies in $(m(x^*), x^*)$ which are strictly preferred to x^*, so $U_{m(x_*)}(S(m(x_*))) < U_{m(x_*)}(S(x_*)) = u_{m(x_*)}(x_*)$. By the convergence of s_j to s_*, it is also unprofitable for x_j to deviate to $m(x_j)$ for high j. If $V_{m(x_*)} > 0$, suppose $x_* \in (y_l, y_{l+1})$. By construction $m(y_l)$ is indifferent between y_l and $S(y_l)$; since $x_* > y_l$, he strictly prefers the former. Since $m(x_*) < m(y_{l+1}) < y_l$, again deviating to $m(x_*)$ means dropping policies with a higher than average payoff.

If $x_* = x^*$, normalize the sequence of quasi-1-equilibria like so: $z_{jn} = \frac{x_{jn}-x^*}{x_{j0}-x^*}$, with $x_j = x_{j0}$; and then take a convergent subsequence with limit $(z_n)_n$. This must converge to the continuous equilibrium for the game with $\tilde{m}(z) = m'(x^*)z$, which is smooth, and so covered by the previous argument. □

C Proofs of Section 5

Proof of Proposition 6. 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, i.e., $m(I_i) < m(x_i) \leq x_i$, which violates the conditions for a steady state. Similarly, $m(x_j) < x_j$.

Next, we 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 $l = i, \ldots, j$. 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 pins down e_i. Since e_i must be indifferent between clubs i and $i+1$, $e_i = \frac{x_i + x_{i+1}}{2}$. This pins down $x_{i+1} = 2e_i - x_i$. 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, i.e., $F(e_{i+1}) - F(x_{i+1}) = F(x_{i+1}) - F(e_i)$, which pins down e_{i+1}, and so on. Thus, given x_i, there is always at most one steady state $(x_i, x_{i+1}, \ldots, x_j)$. Finally, for club j, there is the extra condition that its rightmost voter e_j must happen to be $x_j + d$, which pins down x_i. If f is a non-constant polynomial, the entire system has a finite number of solutions by Bézout's theorem.

Suppose now 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

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

The other side is analogous.

- 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 $f(x_0+q) - f(x_0+r) \leq f(x_0+r) - f(x_0)$.

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

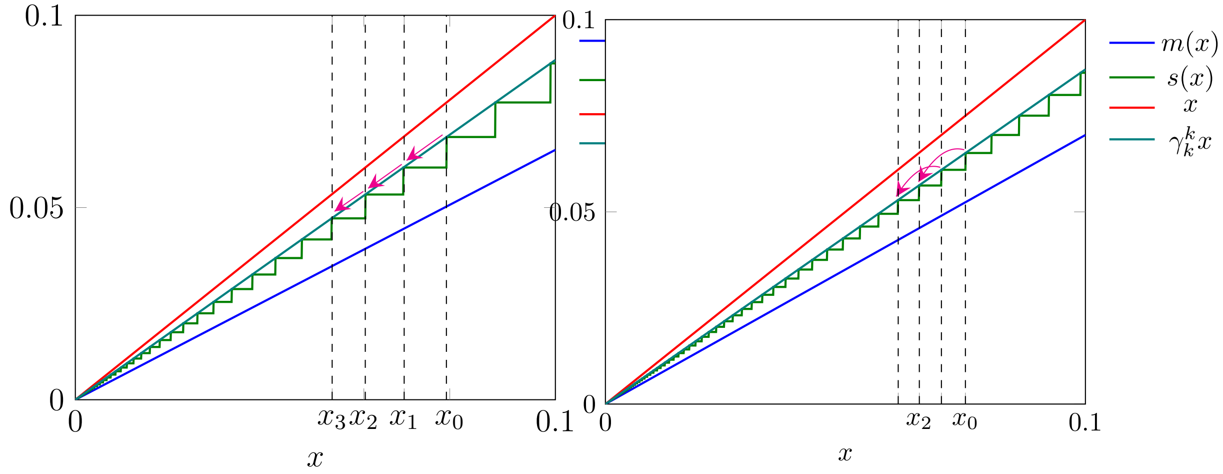
- The above argument implies that $e_i(x_i) - x_i$ is increasing in x_i, if we take $x = x_i - d$, $x + r = x_i$, $e_i = x + q(x)$. Then, in particular, $e_i(x_i)$ is also increasing, as is $x_{i+1} - e_i(x_i)$ (because it is equal to $e_i(x_i) - x_i$). If we denote $x - r = e_i$, $x = x_{i+1}$, then 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, …, and so is $e_j - x_j$. Moreover, $e_j - x_j$ is strictly increasing in x_i unless f is exponential, i.e., log-linear.

□

D Other Equilibria (For Online Publication)

In the discrete time model, we consider k-equilibria, consisting of k interleaved sequences, as well as continuous equilibria. Although they don't exhaust the set of possible solutions, studying them sheds light on the general behavior of non-1-equilibria.

Definition 4. 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 $x \in [x_n, x_{n-1})$.[43] A continuous equilibrium is one where s is continuous.



(a) 1-equilibrium

(b) 2-equilibrium

Figure 6: Equilibria for $m(x) = 0.7x$, $\delta = 0.7$

When $m(x)$ is linear, we can find k-equilibria for all k and a continuous equilibrium, as illustrated in Figures 6a and 6b:

Proposition 8. *Let f be such that $m(x) = \alpha x$ for $x \in [-e, e]$, where $\alpha < 1$ and $e \le d$. Assume $\delta \ge \frac{2}{3}$ and $\alpha \ge \frac{1}{2}$. Then, for each k and $\underline{x} < e$, there is a k-equilibrium s_k^* such that $x_0 = \underline{x}$, given by $x_n = \gamma_k^n \underline{x}$, where $0 < \gamma_k < 1$. There is also a continuous equilibrium s_∞^* given by $s_\infty^*(x) = \gamma_\infty x$. Moreover, γ_k^k is decreasing in k and $\gamma_k^k \to \gamma_\infty$.*

Proof of Proposition 8. Given $k \ge 1$, 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

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

choosing x_{n+k} and x_{n+k+1}. This implies

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

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

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

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

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

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

Next, we show that $m(x_n)$ prefers x_{n+k} to other policies x not belonging to the sequence. We do this in two steps. First, we argue that $\gamma^{k+1} > \alpha$, which implies $x_{n+k+1} > m(x_n)$. Second, we note that this yields our result by a similar argument as in Proposition 4. For the first part, note that

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

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

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

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

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

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

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

However, in the general case, k-equilibria for $k > 1$ and continuous equilibria are not well-behaved. While they can be extended to $[x^*, x^{**}]$ if defined in a neighborhood of x^*, they may lose their properties, i.e., a k-equilibrium may turn into a

1-equilibrium beyond a certain x, or in a continuous equilibrium discontinuities may appear; and whether this happens depends on arbitrarily small details of m.

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

This same dynamic makes k-equilibria for $k > 1$ unstable in general. To see this, let $W((s_0, s_1, \ldots)) = (1-\delta)\sum_{t=0}^{\infty} \delta^t s_t^2$, and characterize a k-equilibrium recursively as follows, using the definition of $m(x_n)$ being indifferent between S_{n+k} and S_{n+k+1}:

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

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

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

$$M_n = \left(\begin{array}{cccccccccc}
0 & \dots & 0 & A & B & 0 & \dots & 0 & C & D \\
1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\
2x & 0 & \dots & 0 & -2\delta x & 0 & 0 & \dots & 0 & \delta \\
0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0
\end{array}\right)
\begin{array}{l}
\left.\vphantom{0}\right\} 1 \\
\left.\begin{array}{l} \\ \\ \\ \\ \end{array}\right\} k \\
\left.\vphantom{0}\right\} 1 \\
\left.\begin{array}{l} \\ \\ \\ \\ \end{array}\right\} k-1
\end{array}$$

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

$$\det(M_n) = -\delta B - 2\delta x_{n+1} D = \delta(1-\delta)\frac{x_{n+1} - m(x_n)}{m'(x_n)(E_{n+k} - E_{n+k+1})}$$

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

Now, if δ is close to 1 and the equilibrium is approximately smooth in the sense of Proposition 5, then $\frac{x - m(x)}{m'(x)(x - E(S(s(x))))} \approx 1$ (see Appendix B) and $\det(M_n \dots M_{n-k+1}) \approx k^k$. In particular, there must be an eigenvalue of absolute value at least $k^{\frac{k}{2k+1}} > 1$. Hence any deviation from an equilibrium resulting from a local perturbation of m which adds a nonzero component to a generalized eigenvector of this eigenvalue (in the Jordan form decomposition of the matrix) will grow exponentially.

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

Proposition 9. *Let $s : [x^*, x^{**}] \to [x^*, x^{**}]$ be a continuous equilibrium for a given m and parameters δ, C. Let $x_0 \in (x^*, x^{**})$. A perturbation $\tilde{m}$ of m is an increasing function $\tilde{m} = m + \rho\kappa$ where $\kappa : [x^*, x^{**}] \to [x^*, x^{**}]$ has support $(x_0 - \epsilon, x_0 + \epsilon)$. For each $\tilde{m}$, let $\tilde{s}$ be an equilibrium under $\tilde{m}$ such that $\tilde{s}|_{[x^*, x_0 - \epsilon)} = s|_{[x^*, x_0 - \epsilon)}$.*

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

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

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

In particular, $W(E)$ must be a strictly convex function so that s is surjective, and it must have no kinks, i.e., s must be strictly increasing (if s is locally constant at x, it will be discontinuous at $s^{-1}(x)$ as long as $s(y) > m(y)$ in this area), so L^{-1} and L are strictly increasing and well-defined. Moreover, it can be shown that s must be C^∞ where m is C^∞.[44] If κ is discontinuous at x, so is $\tilde{m}$, and so is $\tilde{s}$, for any ρ.

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

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

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

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

I conjecture that, even if κ is C^∞, perturbations generically lead to discontinuities

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

devolving into 1-equilibria for arbitrarily small ρ, but in that case the number of steps until the discontinuity will depend on ρ.

References

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

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

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

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

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

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.