

Searching for Good Policies: Repeated Elections, Learning, and Policy Dynamics*

Steven Callander[†]

April 16, 2008

Abstract

I study a model of dynamic policy making in which citizens do not have complete knowledge of how policies are mapped into outcomes. Citizens are able to order policies in the single-dimensional space according to the outcomes they are expected to produce but not according to the realized outcomes they do produce. They learn about the mapping through repeated elections as policies are implemented and outcomes observed. I characterize for this environment the policy trajectory with impatient voters. Although the trajectory is path-dependent, I show that basic patterns emerge. Policy making passes sequentially through three phases: a monotonic phase, a triangulating phase, and ultimately a stable phase. Policy making generally stabilizes at outcomes close to the median voter's preference, although I show that it can at times get stuck, in which case any outcome can prove stable. I also consider how the structure of political competition affects experimentation and learning.

*Preliminary version. Comments welcome. I thank Martin Osborne and Abhinav Muthoo for helpful discussions.

[†]Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston, IL 60208; scal@kellogg.northwestern.edu.

1 Introduction

In October 1979 the United States embarked on a policy making experiment. For three years the Federal Reserve targeted growth in the money supply in its attempts to influence economic conditions, ushering in the era of monetarism. Although the experiment produced mixed results – and was abandoned in October 1982 – it revealed valuable information about how the economy works, information that was used to shape the replacement policy as well as macroeconomic policy generally. Indeed, Benjamin Friedman (1984) argues that the experiment provided such a valuable learning experience precisely because it was so radical.

The monetarism experiment is not unique within policy making. Throughout history experimentation and learning has been central to policy choice. From *laissez-faire* to the New Deal, from tradeable pollution permits to school vouchers, the search for good policies has been guided by trial-and-error.

The objective of this paper is to study the challenges posed by policy making in a dynamic and uncertain world. In particular, the paper seeks to provide answers to such questions as: When do policy makers experiment with policy and when do they settle for the status quo? When do they make radical changes to policy, and when are changes incremental? How much is learned from the policy experience and how does the history of choices affect future choices? What is the trajectory of policy through time and does the choice of policy ever settle down to a stable selection?

To answer these questions I develop a model of repeated two-candidate elections in a single dimensional policy space. The key ingredient of the model is that citizens – voters and candidates – have imperfect information about how policies are transformed into outcomes. Thus, finding the policy that delivers the desired outcome is not straightforward. Aiding the policy making process is the ability to learn from experience. In each period the winner of the election implements his campaign promise and the outcome is observed (and experienced). If the policy choice is *experimental* (not previously tried) its outcome reveals whether the policy is itself good and also provides information about the likely outcomes of other policy alternatives. Citizens update their beliefs accordingly and use the information to *predict* the outcomes of other policies, guiding their future choices. I characterize for this environment with impatient voters the optimal choice of policy in each period and describe the policy trajectory.

A novelty of the model is the specification of the policy process. I suppose policies are mapped into outcomes by the realized path of a Brownian motion, where citizens know the parameters of the motion (the drift and variance) but not the path. Although used in a non-standard manner – with policies acting as the independent variable – the Brownian motion captures many realistic properties of policy making and does so in a highly tractable form.¹ This structure endows citizens with the ability to order policies

¹The usefulness of the Brownian motion in modeling the policy process suggests other stochastic

along the standard liberal-conservative continuum according to *expected* outcomes but not according to *realized* outcomes. Thus, citizens know which policies are *more likely* to deliver liberal (or conservative) outcomes but do not know which policies *do* deliver liberal outcomes. Moreover, the non-monotonicity of a Brownian path captures a key difficulty – and risk – of policy making: that outcomes may move in the opposite direction to that intended from a change in policy. This possibility formalizes Merton’s (1936) famous notion of “unanticipated consequences of purposive social action” (known as the Law of Unintended Consequences). The Brownian motion also possesses attractive learning properties that I elaborate on in more detail momentarily.

The equilibrium trajectory of policy choices is path-dependent, varying in the outcomes realized as citizens progressively learn about the policy process. Yet, basic patterns emerge. Policy making progresses through at most three distinct phases. The first two phases – the *monotonic* and *triangulating* phases – are experimental with new, untested policies being chosen. In the final *stable* phase experimentation stops and the same policy is retained thereafter. The phases are traversed in fixed sequence, although one or both of the experimental phases may be empty, and with probability one a stable policy is ultimately obtained.

The two experimental phases are distinguished by whether both sides of the *political divide* have been established. At the beginning of play citizens know only one point in the mapping – the status quo – and this provides a foundation for one side of politics. It leaves, however, a majority of voters – centrists and those on the other side – who support experimentation and who agree as to which direction to move policy. Policy continues to move in this direction – the monotonic phase – until an experimental policy overshoots its intended target, realizes an outcome on the other side of the median voter, and establishes the other side of the political divide. Although at this point a majority of voters may prefer to not experiment further, they may not be able to agree where to stabilize and experimentation enters the triangulating phase and continues. Thereafter the policy choice oscillates between known policies as the candidates (*a la* Bill Clinton) triangulate between the two sides of the political divide in the hope of finding an appealing middle ground.

In addition to the direction of search, the experimental phases generate a pattern in the size of policy changes. This pattern speaks to a long-running debate on the optimal style of policy making. On one side, Lundblom (1959) famously argues in favor of *incrementalism* – that policy changes should be small and incremental. In contrast, a competing school of thought contends that bold changes are more valuable (as Friedman argues with regard to the monetarism experiment). My results show that neither style is optimal always and I identify the conditions when each works well. I find that bold changes are preferable early during the monotonic phase, whereas small tweaks to policy are optimal only once an issue matures and reaches the triangulating phase.

processes may also be applicable, an issue I take up in the discussion section.

Ultimately the policy choice stabilizes and a key question is whether stable policies deliver outcomes attractive to the electorate. I show that generally – but not always – policy stabilizes at outcomes sufficiently close to the median voter’s ideal. The boundary on what is deemed *good enough*, however, is not constant through time, tightening progressively throughout the triangulating phase as more is learned about the policy process. Consequently, an outcome that earlier may have been deemed good enough can later be discarded and experimentation continued.

Policy stability is not limited to outcomes that are good enough and I identify the possibility of policy making getting *stuck* at less appealing outcomes. In fact, I prove that policy making can get stuck at policies that deliver *any* outcome, including outcomes arbitrarily distant from those preferred by voters. This possibility depends on the particular sequence of realized outcomes and is triggered by different circumstances depending on the experimental phase. In the monotonic phase policy making gets stuck following very bad outcomes. Whereas in the triangulating phase only moderately bad outcomes trigger stability and very bad outcomes actually make further experimentation *more* attractive. Policy making that is stuck represents a novel informational failure of policy making, showing how informational traps can appear and lock-in bad outcomes.

A key feature of the model is that voters are impatient, choosing policy each period to maximize their immediate payoff (discounting the future entirely).² In modeling elections this assumption is not entirely inappropriate. It is consistent with the myopia and inattention mass electorates are well-known for. Moreover, it provides a reasonable approximation to a rational world in which policies take considerable time to implement and produce results. Indeed, in a related model, Piketty (1995) supposes that each generation has the opportunity to make policy only once, and impatience follows from each generations’ disregard for the welfare of successor generations. That said, concerns about the future are surely relevant to policy making. Although the results derived here are not knife-edged,³ extending the analysis to the case of patient voters is nevertheless of obvious interest (but also considerable difficulty).⁴

The results for the benchmark model are built upon a simple model of political competition. This simplicity raises a fundamental question: How does the structure of politics affect experimentation and policy outcomes? In a second set of results I consider this question by extending the model of political competition in two natural directions. The answers suggest how a social planner – who cares about all generations – may

²An indirect benefit of this assumption is that it clearly distinguishes my results from other models of repeated elections in which patience and reputation drive policy dynamics (e.g., Duggan 2000).

³It is straightforward to establish that for generic histories, the policy choice when patience is positive but sufficiently low is in a neighborhood of the choice characterized here.

⁴In this way my focus differs from the experimentation literature in that learning is *passive* rather than *active*. I adopt the approach (see, for example, Piketty (1995), Bala and Goyal (1998)) of simplifying preferences to gain tractability in a general decision environment. The literature on active experimentation, in contrast, restricts attention to much simpler environments (such as bandit models) or generates only limiting behavior results (Aghion et al 1991).

use the design of political systems to encourage (or discourage) experimentation and overcome the short-termism inherent in politics.

The first extension I pursue is the inclusion of candidate uncertainty over the preferences of voters. As is well known (Wittman 1983; Calvert 1985), this uncertainty induces divergence in candidate platforms. My main result is to show that this divergence has contrasting effects: *within* period the divergence is inefficient and strictly disliked by a majority of voters. *Across* periods, however, divergence can improve efficiency as it acts as a catalyst to experimentation by providing voters with a greater array of policy alternatives (two versus one). Surprisingly, divergence does not always increase experimentation and in some circumstances it induces policy stability where it otherwise wouldn't emerge.

The second extension is to allow for voter abstention. I show how abstention distorts political outcomes towards *safe* alternatives, reducing the willingness of society to experiment. I find that although a society may be clamoring for "change," the inability of disillusioned voters to coalesce around a particular alternative means the will to change is often less than the desire.

In what follows I discuss related literature before turning to the benchmark model and results. I then consider issues of robustness and the extensions just described. In the concluding discussion I take up briefly several further generalizations and consider how the policy structure can be applied more broadly.

1.1 Related Literature

This paper is related to several distinct literatures inside and outside of political economy. One broad connection is to the large literature in economics on experimentation and learning. Formally, modeling uncertainty as a Brownian path corresponds to a bandit problem with a continuum of correlated, deterministic arms. To the best of my knowledge, no existing model considers such a structure.⁵

A benefit of the Brownian structure is that it retains the standard notion of a continuous policy space, allowing me to draw conclusions on both the size and direction of policy changes. In politics, Gilligan and Krehbiel (1987) offer a model with a continuous policy space to study the role of expertise in policy making. Tailored to one-shot games, however, the non-expert's uncertainty in their setting is limited to a single piece of information, which in a dynamic environment induces full revelation of the policy mapping after the first period.⁶ In contrast, the richness of the Brownian motion generates uncertainty that is a continuum of random variables and learning that is always incomplete (in countable time).

⁵To be sure, I am able to manage this general structure by simplifying voter preferences. See footnote 4 and the associated discussion on this point.

⁶In a companion paper, Callander (2007), I apply the Brownian motion structure to a one-shot interplay between an expert and non-expert.

Despite its practical relevance, the notion of searching for good policies has received limited attention in political economy. Most closely related is Lundblom’s (1959) boundedly-rational theory of “Muddling Through” (see Bendor (1995) and Kollman, Miller, and Page (2000) for related models). This line of work views policy makers as naively searching around the policy space for better policies according to some predetermined algorithm. I differ in that I solve for the fully rational (albeit impatient) policy choice. In addition to providing a contrast to boundedly rational results, my findings complement them by showing when particular search algorithms are optimal, thereby suggesting when a triangulating candidate or a candidate of bold change may do well.

My model differs also in the nature of search. In the boundedly rational literature policy makers observe precisely the outcomes for a limited set of policies before they are implemented and change policy only when an available outcome is an improvement over the status quo. The search for good policies, therefore, is not experimental as the possibility (that is present here) of making outcomes worse is not present.

Other work on policy dynamics is more distant as it assumes the policy mapping to be common knowledge.⁷ An exception is Piketty (1995) who also assumes agents are impatient but differs in having agents make individual economic decisions as well as vote. He also restricts outcomes to be binary rather than continuous. Moreover, Piketty’s focus differs from mine: he establishes conditions such that heterogenous beliefs (about social mobility) persist through time, whereas here all players have symmetric (but incomplete) information.

In one-shot models of policy making, Fernandez and Rodrik (1991) were the first to observe that policy making may be inefficient if the identities of a policy’s beneficiaries are ex-ante unknown.⁸ More recently, in a binary policy, two-period model Majumdar and Mukand (2004) show that incumbents’ electoral incentives can induce inefficient experimentation due to signaling effects. Neither of these sources of inefficiency is present in my model.

In the study of elections Roemer (1994) is unique in relaxing the connection between policies and outcomes, although his model and focus differs from mine in most other respects. In addition to being one-shot, Roemer explores communication between informed candidates and uninformed voters, whereas here all agents are symmetrically informed.

⁷A literature on repeated elections originating with Barro (1973) and Ferejohn (1986) deals with adverse selection and moral hazard issues in selecting candidates, neither of which is present here (Duggan (2000) extends these models to the choice of a policy rather than effort). A more recent literature studies repeated legislative bargaining in various environments (Baron 1996 is an early reference).

⁸In an interesting recent contribution to this stream, Strulovici (2008) allows for an infinite horizon but retains the assumption that policy is binary (one riskless and one risky alternative).

2 Model

In each period $t = 1, 2, 3, \dots$, a majority rule election is held between two candidates X and Y . The candidates compete by committing to *policies* $x_t, y_t \in \mathbb{R}$ that they implement if elected. Policies are transformed into *outcomes* by a *policy process*, ψ . Formally, a policy process is a function that maps from the policy space to the outcome space (also single dimensional) such that: $\psi : \mathbb{R} \rightarrow \mathbb{R}$.

The electorate consists of an odd number of voters. The voters care about outcomes (and indirectly about policies). Voter i 's ideal outcome is o_i and voters are ordered such that $o_i < o_j$ for $i < j$. Denote the median voter by m and set $o_m = 0$. Voters are impatient and discount the future entirely. The per period utility of voter i for policy p given outcome $\psi(p)$ is:

$$u_i(p) = -(o_i - \psi(p))^2.$$

Candidates have the same utility function over outcomes as do voters, although they may or may not be patient. The ideal outcomes for candidates X and Y are $-d$ and $d > 0$. Candidates are also motivated by rents from office (ego or otherwise), which delivers a fixed benefit of $\kappa > 0$.

The true policy process ψ is determined randomly by Nature prior to period 1. To focus on learning, the same policy process is in effect for all periods. I model ψ as a Brownian motion of drift μ and variance σ^2 .⁹ The challenge for the citizens in finding good policies is that they understand ψ imperfectly, knowing the parameters of the motion but not the realized path.¹⁰

Voters and candidates possess symmetric information about ψ and learn about ψ as policies are experimented with and outcomes observed. Citizens begin the game with knowledge of one point in the mapping: the status quo policy and outcome, (sq, o^{sq}) . They observe a new point in the mapping each time an experimental policy is implemented, such that at election t they know up to t distinct points in the mapping. Figure 1 depicts one possible realization of the Brownian motion passing through the status quo point.

The Brownian motion is *partially invertible*. By observing points in the mapping citizens infer valuable information about other policies and refine their beliefs, but the information revealed is incomplete and for so-far untried policies outcomes remain uncertain. The Brownian motion possesses the Markov property, implying that beliefs depend only on the nearest known point in either direction. On the flanks beliefs depend only on end-point and are *open ended*. If policy r is the right-most known point,

⁹Note that although Brownian motions are normally associated with movement through time, time plays no role here; the policy instrument serves as the independent variable.

¹⁰In Section 3.4 I relax the assumption that voters know the parameters of the motion.

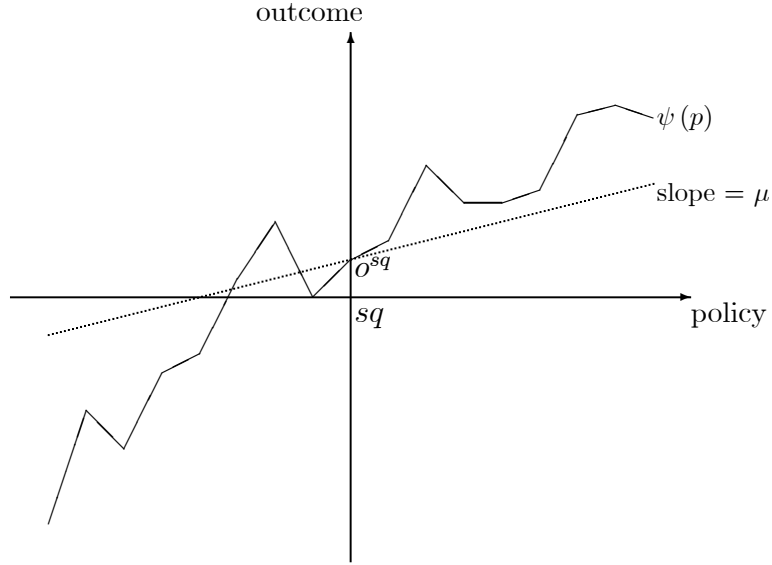


Figure 1: A Realized Path of Brownian Motion ψ : $\mu > 0$, σ^2 , and $\psi(sq) = \sigma^{sq}$

beliefs for policies $p > r$ are distributed normally with:

$$\text{Expected Outcome: } E\psi(p) = \psi(r) + \mu(p - r), \quad (1)$$

$$\text{Variance: } \text{var}(\psi(p)) = |p - r| \sigma^2. \quad (2)$$

The drift parameter μ measures the expected rate of change and the variance the “noisiness” of the policy process. The amount citizens learn about a policy p decreases in the distance of p from r and I say that the process is *proportionally invertible*.

Beliefs on the left flank are similarly open-ended. If the left-most known policy is l , beliefs for all policies $p < l$ depend only on policy l , and the following expressions are the analogues of Equations 1-2.

$$\text{Expected Outcome: } E\psi(p) = \psi(l) + \mu(p - l), \quad (3)$$

$$\text{Variance: } \text{var}(\psi(p)) = |p - l| \sigma^2. \quad (4)$$

At the first election beliefs are open-ended on either side of sq , as reflected in the constant drift line in Figure 1.

Policies between two known points in the mapping are said to lay on a *Brownian bridge*, with beliefs determined by the value at both ends of the bridge. For policies $p \in$

$[q_1, q_2]$ on the bridge between q_1 and q_2 , beliefs are distributed normally with moments:

$$\text{Expected Outcome} : E\psi(p) = o^{q_1} + \frac{p - q_1}{q_2 - q_1} (o^{q_2} - o^{q_1}), \quad (5)$$

$$\text{Variance} : \text{var}(\psi(p)) = \frac{(p - q_1)(q_2 - p)}{q_2 - q_1} \sigma^2. \quad (6)$$

Expected outcomes on the bridge are given by the straight line between the two ends and are independent of the drift μ . The variance is concave over the domain, reaching its peak halfway between the ends of the bridge and equaling zero (obviously) at the ends. (Note that $\frac{d\text{var}(\psi(p))}{dp} = \pm\sigma^2$ at the ends of the bridge.)

The appeal of the Brownian motion as a model of the policy process is that it captures several key features of policy making in practice. Partial invertibility allows for meaningful experimentation, learning, and prediction, and proportional invertibility captures the intuition of Lundblom (1959) that greater uncertainty is incurred the more policy is moved from what is known.

Moreover, the non-monotonicity of Brownian paths captures the basic risk of policy making: that in trying to make outcomes better, changes to policy may actually make things worse. The outcome of an experimental policy could overshoot the intended outcome or even move the outcome in an unintended direction.¹¹ The volatility of policy making – and how much citizens learn from observing outcomes – is determined by the parameters of the motion and the ratio $\frac{\sigma^2}{|\mu|}$ provides a simple measure of the *complexity* of the underlying issue. The larger the ratio the less precise are citizens' beliefs upon learning a point in the mapping and the more complex the issue.

A final advantage of the Brownian motion is tractability. The simplicity of Equations 1-6 combines easily with quadratic utility as the mean and variance are sufficient to determine expected utility. For any policy p , expected utility reduces to:

$$Eu_i(p) = - [E\psi(p) - o_i]^2 - \text{var}(\psi(p)), \quad (7)$$

and this mean-variance representation of utility plays a central role in the analytic results of the paper.¹²

I restrict attention to equilibria in which voters use weakly dominant strategies, requiring that they vote for the candidate that maximizes their utility given the history at the time of the election. I do not allow abstention (although I relax this in Section 4.2). To break ties, I suppose an indifferent voter supports the candidate with greatest variance and mixes equally otherwise, although any non-degenerate tie-breaking rule

¹¹These possibilities are consistent with Walker's (1969, p. 890) observation that: "Uncertainty and the fear of unanticipated consequences has always been a formidable barrier to reform."

¹²The logic of Equation 7 can also be interpreted as an intuitive behavioral rule-of-thumb: Citizens evaluate policies first by their expected outcome, and then subtract from the evaluation the associated risk.

can be used without meaningfully changing the results.

An equilibrium consists of strategies for both voters and candidates. Although voters are sophisticated, their voting behavior is straightforward given the policy choices of the candidates. As such, and as is standard in models of electoral competition, I report only the strategies of candidates in describing equilibrium. Denote the equilibrium strategies of the candidates at time t by x_t^* and y_t^* .

3 Results

I begin with two preliminary results. Lemma 1 shows that uncertainty and experimentation is not sufficient to reduce the power of the median voter. Consequently, a repeated election analogue of Black's (1958) median voter theorem holds and policy convergence obtains.

Lemma 1 *The (essentially¹³) unique equilibrium is convergent: for a given policy history at election t , platforms satisfy $x_t^* = y_t^* = \arg \max_{p_t \in \mathbb{R}} [Eu_m(p_t)]$.*

The omnipotence of the median voter reduces policy choice to the equivalent of a single person decision problem (with ideal outcome zero). This setting is the classic benchmark of political competition. That it holds here shows that uncertainty over the policy process is by itself insufficient to upset the logic of two-candidate competition. Later in the paper I enrich the model of political competition such that Lemma 1 no longer holds and show how policy experimentation is affected.

Policy convergence in Lemma 1 is *within* period and not *across* periods, raising the question of whether policy choices also converge through time. The next result establishes a sufficient condition for policy stability, showing that the willingness to experiment is a use-it-or-lose-it proposition. If in any period a previously implemented policy is reused citizens learn nothing new about the policy process. Consequently, the same choice is again optimal and experimentation stops. Formally, I say that policy z is *stable* on the equilibrium path if for some t' , $x_t^* = y_t^* = z$ for all $t \geq t'$. Throughout this section I focus on the utility of the median voter and refer only to the strategy of candidate X under the understanding that Y 's platforms is the same.

Lemma 2 *If $x_{t'}^* = x_t^*$ for some $t' > t$ then policy x_t^* is stable.*

I now turn to the equilibrium strategy, which I present constructively beginning with the opening election.

¹³Uniqueness in Lemma 1 is qualified as for some non-generic realizations of ψ divergence may occur in equilibrium. For example, if through experimentation citizens were to learn that $\psi(x) = -\psi(sq)$ for some x then it may be that the candidates diverge in equilibrium with one candidate offering policy x and the other sq . Possibilities of this sort arise generally, although hereafter I ignore these non-generic possibilities.

3.1 The First Election

The first election presents the citizenry with a basic trade-off: accept the known but possibly less than satisfactory status quo, or move policy in the hope of achieving a better outcome but at the risk of making things worse. Put another way, should citizens trust the devil-they-know or the devil-they-don't-know? Proposition 1 provides the answer, showing when policy choice is conservative and when risk is undertaken, and in this case characterizes exactly the size and direction of the policy movement. Define $\alpha = \frac{\sigma^2}{2|\mu|}$ as half the complexity of the underlying issue.

Proposition 1 *The equilibrium strategy at $t = 1$ is:*

- (i) *Stable at $x_1^* = sq$ if $o^{sq} \in [-\alpha, \alpha]$.*
- (ii) *Experimental if $o^{sq} \notin [-\alpha, \alpha]$, where:*

$$x_1^* - sq = \begin{cases} \frac{\alpha - o^{sq}}{\mu} > 0 & \text{if } o^{sq} > \alpha, \\ \frac{-\alpha - o^{sq}}{\mu} < 0 & \text{if } o^{sq} < -\alpha. \end{cases} .$$

If the outcome from the status quo is good enough it is immediately stable and no experimentation occurs in equilibrium. The stable outcome may diverge from the median voter's ideal outcome, although this divergence may be thought reasonable in the sense that only outcomes close to the median's ideal point can prove stable.

The more interesting case is when the status quo is not good enough and the optimal response is to change policy and experiment. Experimentation leads to two questions: In which direction does policy move and how far does it move? With uncertainty open-ended on either side of sq , the direction of movement depends on μ and the value of o^{sq} . The equilibrium choice involves a trade-off between greater variance and a more centrist expected outcome. Rearranging the solution for x_1^* reveals that the expected outcome is set to equal $\pm\alpha$, regardless of the location of the status quo or the outcome it produces ($+\alpha$ if $o^{sq} > \alpha$ and $-\alpha$ if $o^{sq} < -\alpha$). The left-hand panel of Figure 2 depicts the situation when $o^{sq} > 0$ and $\mu < 0$.

The size of a policy change is increasing in the unattractiveness of the status quo outcome and decreasing in the complexity of the underlying issue (as $\alpha = \frac{\sigma^2}{2|\mu|}$). This relationship formalizes the intuition that the citizenry is more willing to engage in risky policy making the more dissatisfied they are with the current state of affairs and the more control they have over an issue (lower complexity).

An interesting aspect of equilibrium behavior is that despite policy choices being driven by the median voter, it is not the median voter who most likes the policy that is ultimately offered. Rather, for an experimental policy with expected outcome α it is the voter with ideal outcome α that possesses the highest expected utility (despite the implemented policy not being this voter's favorite). To a naive observer, therefore, it may appear that candidates exhibit a partisan bias in their platforms where one

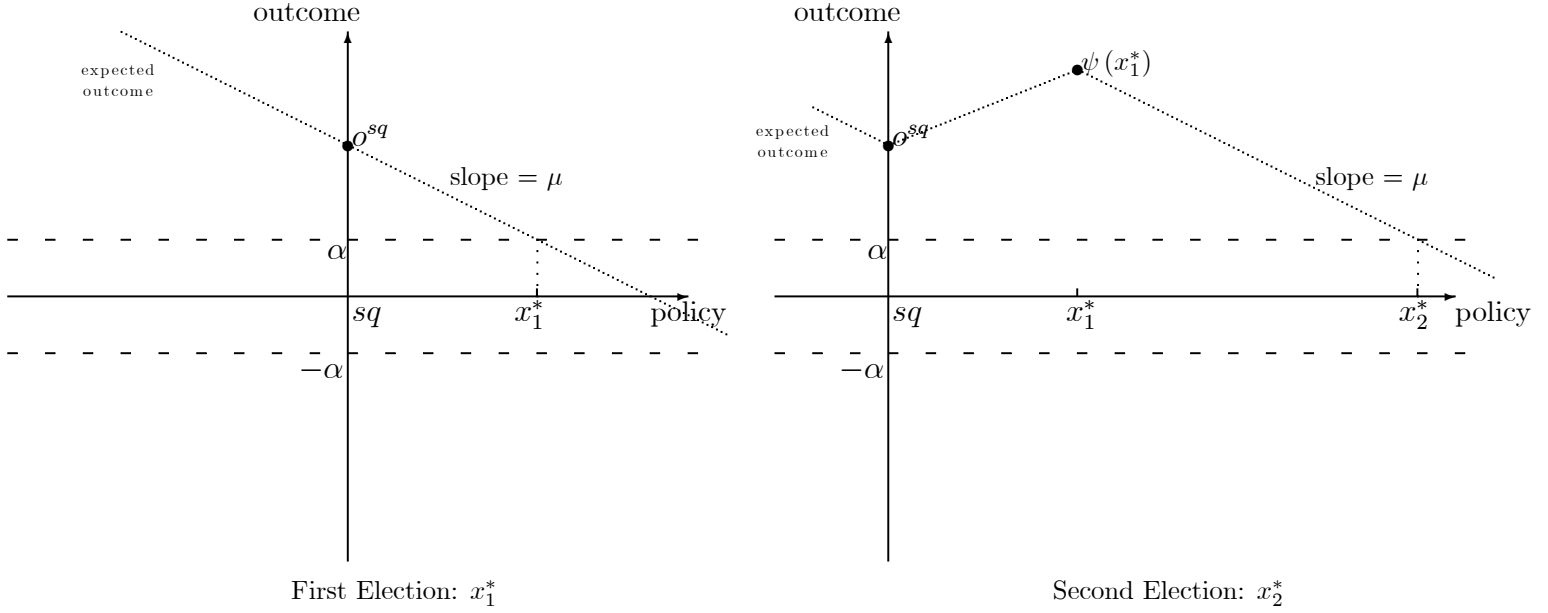


Figure 2: Monotonic Phase ($o^{sq} > \alpha, \mu < 0$)

doesn't exist. This property suggests care must be taken in interpreting “thermometer” scores, popular in the survey literature, that measure how much a voter likes a policy or candidate as higher scores do not imply a group of voters is the target of candidate policies.

3.2 The Second and Subsequent Elections

Following the first election the winning policy is implemented and the outcome realized. If policy at $t = 1$ was stable then learning ceases. If it was experimental voters now know two points in the policy process and update their beliefs and policy preferences accordingly. The structure of subsequent policy choices depends on whether the outcome from x_1^* is on the same or opposite side of 0 from o^{sq} . I begin with the *monotonic phase* in which $\psi(x_1^*)$ and o^{sq} (and all subsequent outcomes) are of the same sign. For ease of exposition, and without loss of generality, I describe equilibrium behavior for the case $\mu \leq 0$ and $o^{sq} \geq 0$.

Monotonicity

The case $\psi(x_1^*) \geq 0$ is depicted in the right-hand panel of Figure 2. With knowledge of two points in the policy process, beliefs are open-ended only on the flanks and between the known points a Brownian bridge forms. It is easy to see, however, that policies on the bridge are dominated by the end-point closer to 0 (policy sq in Figure 2) as this yields a more attractive expected outcome and with no uncertainty. If experimentation is optimal, therefore, it must be on a flank and, as is the case at the first election, the

right flank is preferable and the search for a good policy continues monotonically in the direction of the original movement.

Given the similarity of conditions in the second period to the first period, one may conjecture that the cut-points of Proposition 1 apply again at $t = 2$ and thereafter. Surprisingly, this conjecture is false. The first period calculations again apply *if* experimentation is optimal. What changes is that the willingness of citizens to experiment may diminish. In fact, policy stability can now be induced by sufficiently bad outcomes as well as outcomes that are sufficiently good.

The logic for candidate behavior at $t = 2$ extends to all subsequent periods in which policy has not stabilized and previous outcomes have been of the same sign. For this to be the case policy movements must have been monotonic and I refer to this as the *monotonic phase* of policy making.

Definition 1 *Policy making at election t is in the monotonic phase if $sq < x_1^* < \dots < x_{t-1}^*$ and $o^{sq}, \psi(x_1^*), \dots, \psi(x_{t-1}^*) \geq 0$.*

Define $\tau_t^* = \arg \min_{v' < t} [|\psi(x_{v'}^*)|, |o^{sq}|]$ as the most attractive outcome realized up to election t . Equilibrium behavior in the monotonic phase is described by the following.

Proposition 2 *In the monotonic phase at election $t \geq 2$, the equilibrium strategy is:*

(i) *Stable at:*

$$x_t^* = x_{t-1}^* \text{ if } \psi(x_{t-1}^*) \in [0, \alpha].$$

(ii) *Stable at:*

$$x_t^* = \tau_t^* \neq x_{t-1}^* \text{ if } \psi(x_{t-1}^*) > \delta_t,$$

where $\delta_t = \frac{\alpha^2 + \psi(\tau_t^*)^2}{2\alpha}$.

(iii) *Experimental with: $x_t^* > x_{t-1}^*$ if $\psi(x_{t-1}^*) \in (\alpha, \delta_t]$, where $\psi(x_{t-1}^*) + \mu(x_t^* - x_{t-1}^*) = \alpha$.*

Policy can now stabilize in two situations: when the most recent outcome is either “good enough” or when it is “bad enough.” The former possibility is the same as in Proposition 1, and in fact, the region of good enough stability is constant throughout the monotonic phase. In this case it is the most recent policy choice that proves stable.

The second type of stability is rather different and when it happens I say that policy making gets *stuck*. The stable policy in this case is τ_t^* (the previous most attractive policy) and not the most recent choice. For policy to backslide in this way to a previously chosen – and discarded – policy it is necessary for the outcome at x_{t-1}^* to have moved in the opposite direction to that anticipated. The importance of moving in the wrong direction is not that the policy will be chosen again, but that it reduces the expected

utility of further experimentation. Consequently, although experimentation at time t is always preferable to a bad outcome of policy x_{t-1}^* , it may not be preferred to less unattractive policies that had been previously implemented. For a sufficiently bad outcome at $t - 1$, therefore, policy making gets stuck. The critical value δ_t is that which equates the expected utility from experimenting to the sure outcome at τ_t^* .¹⁴

The possibility of getting stuck at an unattractive outcome raises the question of whether there is a bound on where policy can get stuck. Corollary 1 shows that the set of possibilities is bounded only by the outcome from the status quo (since sq can always be chosen). As the status quo outcome o^{sq} can be set arbitrarily, it is thus possible to construct paths such that policy making gets stuck at any arbitrarily inefficient outcome.

Corollary 1 *In the monotonic phase policy making can get stuck at any outcome in $[-o^{sq}, o^{sq}]$.*

The substantive difference between getting stuck and good-enough stability is best illustrated by the expected utility each delivers relative to that of the optimal choice in past periods. When policy stabilizes because it is good enough, the expected utility is greater than in past periods, reflecting the achievement of a good-enough outcome. In contrast, when policy gets stuck utility is lower than the expected utility in preceding periods, back at least to when the stable policy was first implemented, reflecting a disadvantageous realized set of policy outcomes.

Triangulation

The monotonic phase continues indefinitely until a policy proves stable (for any reason) or policy making over-shoots and an outcome is realized on the opposite side of the median voter's ideal point, establishing the other side of the political divide. In this case a Brownian bridge spans the median voter's ideal point and policy making transitions to the *triangulating phase*. The triangulating phase can begin as early as the second election and once it starts it cannot be reversed. Formally, the triangulating phase is defined as follows.

Definition 2 *Policy making at election $t \geq t^\Delta$ is in the triangulating phase if $sq < x_1^* < \dots < x_{t^\Delta-1}^*$ and $o^{sq}, \psi(x_1^*), \dots, \psi(x_{t-2}^*) > 0 > \psi(x_{t^\Delta-1}^*)$, and no previous policy has proven stable.*

The beginning of the triangulating phase marks the end of monotonic search. Hereafter experimentation must be on a *spanning bridge* (across zero) as all other experimental policies are dominated by known points. The left panel of Figure 3 depicts the

¹⁴It is important to note that although policy is stuck at an expected local minimum in the policy path, this is not due to the limitation of a hill-climbing algorithm. Voters evaluate all possible policies fully rationally. The realized path implies that optimal experimentation requires a large change in policy and such a bold move is costly in terms of risk. Consequently, the expected local minimum in the policy path corresponds to a global maximum in expected utility.

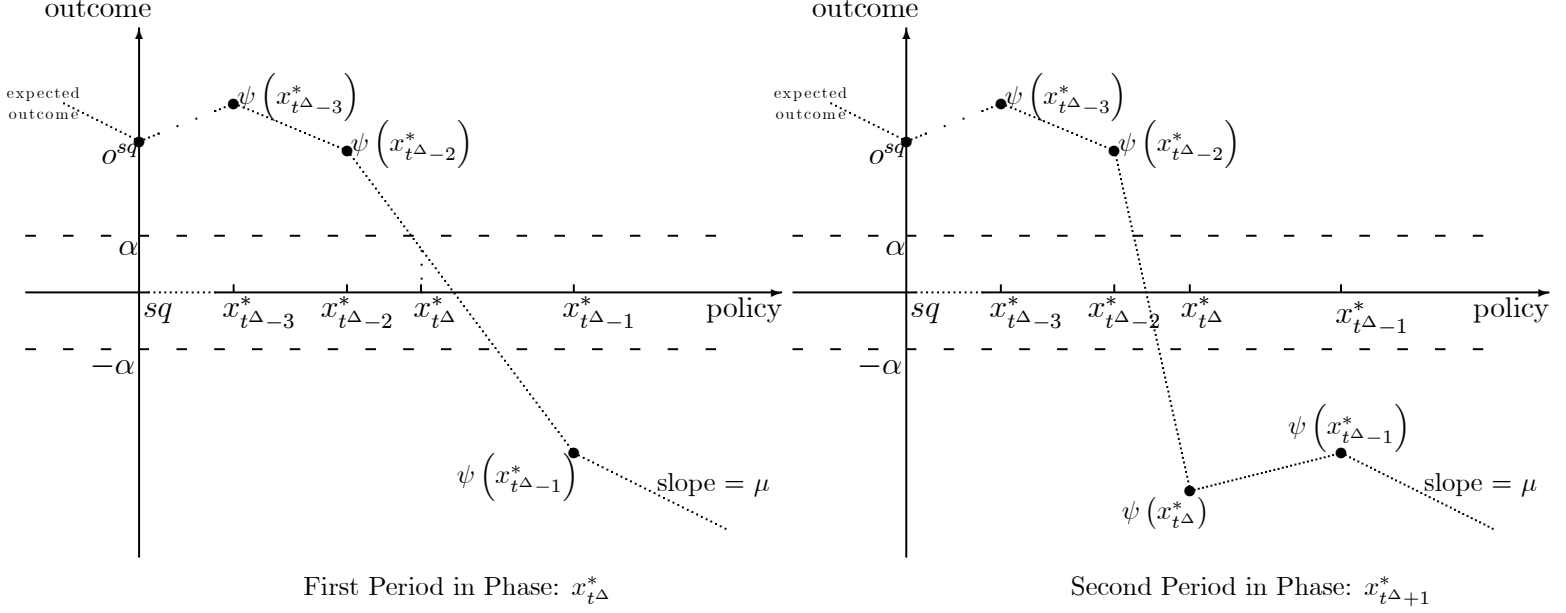


Figure 3: Triangulating Phase ($o^{sq} > \alpha, \mu < 0$).

situation at election t^Δ (and the dotted lines again represent expected outcomes for experimental policies).

Behavior at the beginning of the triangulating phase mimics that at the beginning of the monotonic phase in its simplicity: Stabilize at the most recent policy choice if it is good enough, otherwise continue experimenting. Despite the similarity, the trigger for stability and the degree of experimentation differ. For the bridge $\widehat{w \cdot z}$ between generic policies w and z define:

$$\alpha(\widehat{w \cdot z}) = \frac{\sigma^2}{-2 \left(\frac{\psi(z) - \psi(w)}{z - w} \right)},$$

where the bracketed term in the denominator is the slope of the bridge between w and z . Formally, $\alpha(\widehat{w \cdot z})$ generalizes α by replacing μ with the slope of the bridge. Equilibrium behavior upon first entering the triangulating phase is then as follows.

Proposition 3 *At election t^Δ in the triangulating phase the equilibrium strategy is:*

- (i) *Stable at $x_t^* = x_{t^\Delta-1}^*$ if $|\psi(x_{t^\Delta-1}^*)| \leq \alpha(\widehat{x_{t^\Delta-2}^* \cdot x_{t^\Delta-1}^*})$.*
- (ii) *Experimental otherwise, where $x_t^* \in (x_{t^\Delta-2}^*, x_{t^\Delta-1}^*)$ solves:*

$$E[\psi(x_{t^\Delta}^*)] = \alpha(\widehat{x_{t^\Delta-2}^* \cdot x_{t^\Delta-1}^*}) \left[1 - 2 \frac{x_{t^\Delta}^* - x_{t^\Delta-2}^*}{x_{t^\Delta-1}^* - x_{t^\Delta-2}^*} \right].$$

Two factors distinguish behavior here from the monotonic phase: the spanning bridge has slope steeper than μ , and the variance of outcomes is concave across the bridge.

Both factors make experimentation more lucrative than when uncertainty is open-ended, changing voters' tolerance for experimentation.

The boundary on stability (defined implicitly in part i) is strictly tighter than in the monotonic phase. The greater slope of the spanning bridge implies the marginal gain in expected outcome exceeds that when uncertainty is open-ended. Thus, the triangulating phase makes voters less willing to settle for any particular outcome. Also different from the monotonic phase is that the expected outcome of an experimental policy (defined implicitly in part ii) is more centrist than the stability boundary in part (i). This is the result of variance being concave across a bridge as then the marginal cost in variance is lower in the middle of a bridge than near the ends.

The upshot of a greater preference for risk is that policy making cannot get stuck at election t^Δ . As earlier policies did not prove stable at election $t^\Delta - 1$, they cannot prove stable at t^Δ when the pay-off from experimentation is greater; thus only the most recent policy can prove stable and only because it is good enough.

If an experimental policy is chosen at t^Δ the triangulating phase continues. The realization of $\psi(x_{t^\Delta}^*)$ breaks up the bridge between $x_{t^\Delta-2}^*$ and $x_{t^\Delta-1}^*$, as depicted in the right side panel of Figure 3, creating two new bridges, only one of which is spanning. As experimentation in the subsequent period must be on the spanning bridge the process repeats and a simple induction argument establishes that throughout the triangulating phase a unique spanning bridge exists.

Lemma 3 *In the triangulating phase one and only one Brownian bridge is spanning.*

The logic for behavior at election $t^\Delta + 1$ – the second period of the triangulating phase – extends to all later periods and is described in Proposition 4. For election $t > t^\Delta$, denote the endpoints of the unique spanning bridge by x_l^* and x_r^* (omitting dependence on t for simplicity) where by construction one of the ends is x_{t-1}^* , the most recently chosen policy. Recall that policy τ_t^* delivers the most centrist outcome of those observed at time t .

Proposition 4 *At election $t > t^\Delta$ in the triangulating phase the equilibrium strategy is:*

- (i) *Stable at $x_t^* = x_{t-1}^* \in \{x_l^*, x_r^*\}$ if $|\psi(x_{t-1}^*)| \leq \alpha \left(\widehat{x_l^* \cdot x_r^*} \right)$.*
- (ii) *Stable at $x_t^* = \tau_t^* \notin \{x_l^*, x_r^*\}$ if $|\psi(\tau_t^*)| < \frac{\sigma}{2} \sqrt{|x_r - x_l|}$ and $\psi(x_l^*) \approx -\psi(x_r^*)$.*
- (iii) *Experimental otherwise, where $x_t^* \in (x_l^*, x_r^*)$ solves:*

$$E[\psi(x_t^*)] = \alpha \left(\widehat{x_l^* \cdot x_r^*} \right) \left[1 - 2 \frac{x_t^* - x_l^*}{x_r^* - x_l^*} \right].$$

Proposition 4 differs from Proposition 3 in adding the possibility of getting stuck. The logic for getting stuck is similar to the monotonic phase but the trigger is different. In contrast to the monotonic phase, very bad outcomes make experimentation more

attractive, ensuring it continues. Instead, policy making gets stuck for moderately bad outcomes.

The logic of this result follows from the properties of experimentation on a bridge. If x_r^* delivers an outcome closer to zero than x_l^* ($|\psi(x_l^*)| > |\psi(x_r^*)|$), the optimal experimental policy is closer to x_r^* and delivers an outcome on the same side of zero. The payoff from the optimal policy is, however, decreasing in $|\psi(x_r^*)|$, despite the bridge getting steeper. In contrast, the utility of experimentation is increasing in $|\psi(x_l^*)|$ precisely because the bridge is getting steeper. As such, the utility of further experimentation is lower when the most recent outcome is moderately bad; more precisely, when $\psi(x_l^*) + \psi(x_r^*)$ is small.

Expected utility of experimenting is minimized when $\psi(x_l^*) = -\psi(x_r^*)$, which delivers expected utility of $-\frac{|x_r^* - x_l^*|}{4}\sigma^2$ that depends only on the width of the bridge (as the expected outcome of optimal experimentation is zero). This value provides a lower bound on experimenting and together with $\psi(\tau_l^*)$ a necessary condition for getting stuck. The precise bounds on getting stuck can be calculated but are not particularly illuminating and I do not consider them in detail.¹⁵

Good-enough stability is, of course, also possible. What voters deem good enough, however, is contracting throughout the triangulating phase. Driving this property is that the unique spanning bridge either gets steeper or policy stabilizes, thus if experimentation continues the marginal gain from experimenting increases and voters are less willing to settle. This property is described in Corollary 2.

Corollary 2 *The boundary on good-enough stability is strictly converging throughout the triangulating phase.*

Stability

The triangulating phase continues indefinitely until policy making stabilizes, as depicted in Figure 4. The flow-chart shows the sequence of phases if policy making were to evolve. Outstanding, however, is the question of whether policy making does move through all phases and ultimately stabilize. This question is non-trivial as, by Corollary 2, the bound on stability approaches zero and if it converges too quickly may not be reached with positive probability. Nevertheless, Proposition 5 confirms that a stable policy emerges in equilibrium almost surely.

Proposition 5 *With probability one, a stable policy appears on the equilibrium path.*

¹⁵Using the solution to optimal experimentation in (iii), policy making gets stuck if and only if:

$$|\psi(\tau_t^*)| < |E[\psi(x_t^*)]|^2 + \frac{(x_t^* - x_l^*)(x_r^* - x_t^*)}{x_r^* - x_l^*} \cdot \sigma^2.$$

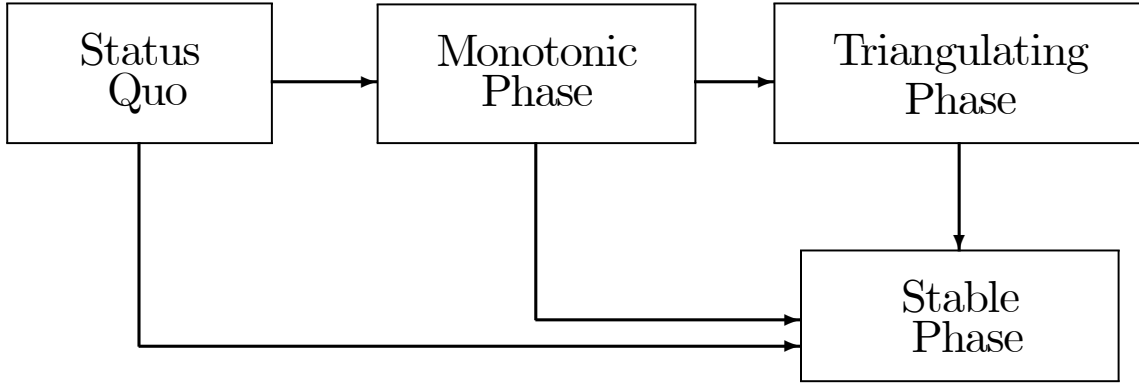


Figure 4: Timeline: Sequence of Phases.

This result confirms that along the equilibrium path learning eventually stops and policy settles down. However, as stability occurs in finite time, learning is incomplete and convergence of outcomes does not necessarily obtain. Almost surely, therefore, the policy that proves stable delivers an outcome divergent from the median voter’s ideal.

3.3 Simulations

Several questions of interest are not accessible analytically. In this section I offer simulations of the dynamic policy making process that provide insight to questions such as the following: How many periods pass before policy stabilizes? How close on average is the outcome of the stable policy to the preferences of the median voter? How often does policy making get stuck?¹⁶ The focus of the analysis is in varying the outcome of the status quo. For all simulations I fix $\mu = -1$ and $\sigma^2 = 4$, such that $\alpha = 2$.

Figure 5 reports properties of the policies that prove stable. I depict two related measures: the average utility loss for the median voter from the stable policy, and the distance of that policy from zero. The most striking feature is that the utility loss is not increasing in the unattractiveness of the status quo. In fact, the utility loss is non-monotonic in o^{sq} , at first increasing before slowly decreasing, with a precipitous drop between $o^{sq} = 5$ and $o^{sq} = 10$, and is relatively flat thereafter.

To a social planner who cares about all generations (adopting the interpretation that each generation chooses policy only once), this pattern implies that overall welfare is enhanced if the initial generation of voters faces a worse status quo policy. Intuitively, two factors drive this result. First, starting at an outcome o^{sq} just beyond the boundary α , voters have a higher probability of observing an outcome inside $[-\alpha, \alpha]$ but near the boundaries and stopping there. Secondly, the relative attractiveness of the sq for lower

¹⁶The simulations were performed via a Matlab program that is available from the author. For each o^{sq} the values reported are for 10,000 iterations.

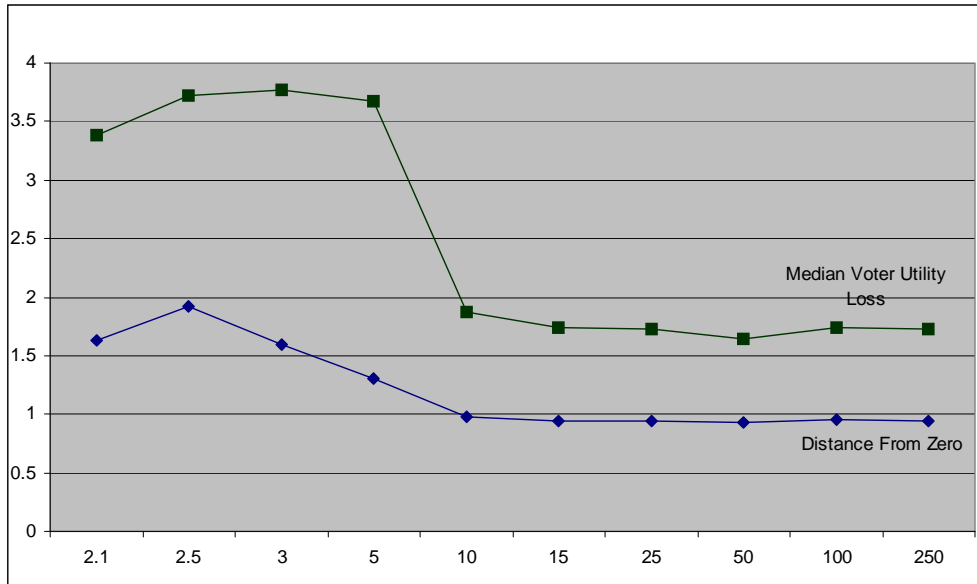


Figure 5: Properties of Stable Policy as a Function of o^{sq}

o^{sq}	2.1	2.5	3	5	10	15	25	50	100	250
Average	1.53	1.73	1.94	2.54	3.02	3.19	3.39	3.61	3.82	4.07
Min	1	1	1	1	1	1	1	1	1	1
Max	4	13	17	17	17	17	25	19	22	20
S.D	.56	.89	1.14	1.63	1.88	1.93	1.97	2.02	2.00	2.05
Monotonic	1.53	1.59	1.65	1.81	1.90	1.90	1.92	1.92	1.94	1.96
Triangulating	0.0004	0.13	0.28	0.73	1.12	1.29	1.47	1.70	1.88	2.11

Table 1: Number of Policy Changes Before Stability

o^{sq} increases the frequency with which policy making gets stuck at sq .

Figure 6 shows the relative frequency of events that trigger stability. For a less attractive o^{sq} the probability of entering the triangulating phase increases, and for a large o^{sq} the probability of a good-enough outcome being achieved in the triangulating phase is the predominant trigger of stability. Even in these cases, however, the probability of getting stuck is non-trivial and the sum across the monotonic and triangulating phases always exceeds 17% and ranges up to 50%.

Table 1 reports the number of times the policy is changed before stabilizing. The notable feature is that stability is achieved on average relatively quickly, although the range is broad. Driving this speed is that voters know the drift and variance of the policy process, allowing them to make dramatic changes to policy with relative confidence that outcomes will move in the intended direction. Consistent with this interpretation is that most of the increase for larger values of o^{sq} is due to an increase in the length of the triangulating phase. I consider this issue further in the following section.

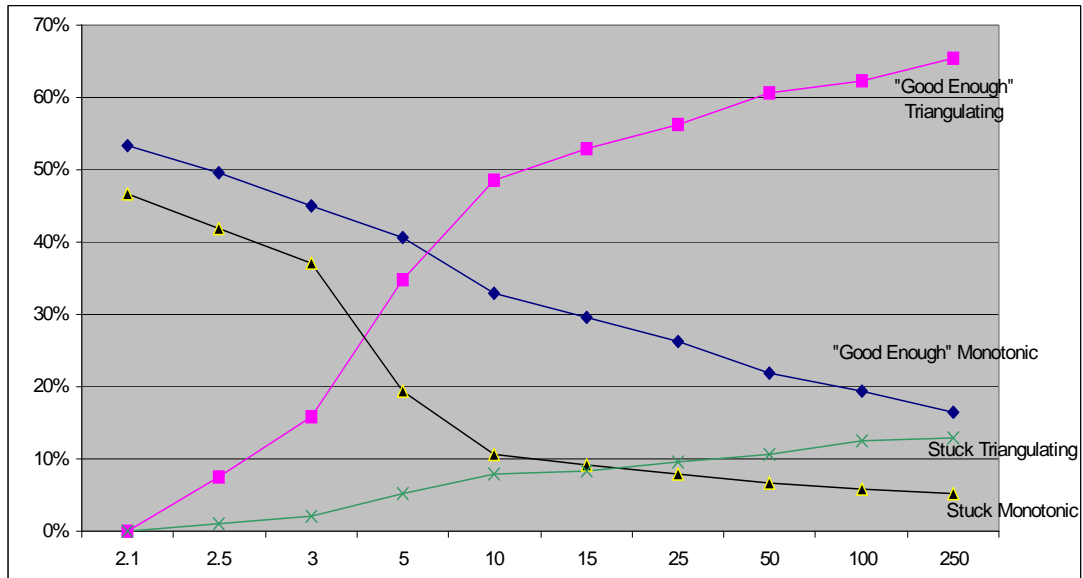


Figure 6: Percentage of Stabilizing Event as a Function of o^{sq}

3.4 Robustness

Although voters in this environment face considerable difficulty in finding good policies, their task is simplified by knowledge of the underlying parameters of the policy process. In practice, citizens may lack knowledge of these parameters as well as the realized path. I show here how this uncertainty can further constrain the willingness of voters to experiment with policy, but that nevertheless the basic structure of the previous analysis is robust to parameter uncertainty. For brevity I limit myself to the more interesting case of uncertainty over the drift parameter. Let citizens' prior beliefs ascribe equal probability to two values $\mu_1 < \mu_2$, where $\frac{\mu_1 + \mu_2}{2} = \mu < 0$, and $\alpha = \frac{\sigma^2}{2|\mu|}$. I present here only basic results and compare behavior with drift uncertainty to the case of no uncertainty (and drift μ).

I begin with the first election. With no uncertainty over drift the status quo is stable if $o^{sq} \leq \alpha$, otherwise the experimental policy chosen has an expected outcome of α . With uncertainty over drift the first of these properties holds whereas the second does not. Instead, citizens are less bold when they do experiment, producing an expected outcome that is strictly more divergent than α .

Corollary 3 *With drift uncertainty, the equilibrium strategy at $t = 1$ is:*

- (i) *Stable at $x_1^* = sq$ if $o^{sq} \in [-\alpha, \alpha]$.*
- (ii) *Experimental if $o^{sq} > \alpha$, where $E\psi(x_1^*) > \alpha$ and strictly increasing in o^{sq} .*

At the sq policy, the expected marginal gain of experimentation is $\frac{\mu_1 + \mu_2}{2}$, the same as when drift is known and this leads to the same stability cut-point. The equivalence

breaks down, however, for positive levels of experimentation. The gain from the steeper possible drift value (μ_1) is tempered by the fact that if this is the true drift then the expected outcome is already close to zero and the marginal gain is small. Although the reverse holds for the flatter drift μ_2 , the average of the two marginal gains leads to less experimentation.

In subsequent elections the effect of drift uncertainty is more subtle and substantial. Whereas with no drift uncertainty unfavorable outcomes are attributed to simple bad luck, they must now be interpreted for what they imply about drift. How extensively this inference problem affects behavior depends on whether uncertainty extends to the sign of the drift as well as the magnitude. I begin with $\mu_1 < \mu_2 < 0$ and behavior at the second election. For brevity I only partially characterize behavior.

Corollary 4 *With drift uncertainty and $\mu_1 < \mu_2 < 0$, the equilibrium strategy at $t = 2$ is:*

- (i) *Stuck at $x_2^* = sq$ if and only if $\psi(x_1^*) > \hat{\delta}_2 > \delta_2$, for some $\hat{\delta}_2$ and where δ_2 is as defined in Proposition 2.*
- (ii) *Experimental when $\psi(x_1^*) \in [\alpha - \varepsilon, \alpha + \varepsilon]$, for some $\varepsilon > 0$.*

Drift uncertainty produces contrasting effects on second period behavior, depending on the outcome realized after the first election. A bad outcome at $t = 1$ leads voters to assign more weight to μ_2 , rendering further experimentation less attractive and increasing voters' willingness to backslide to the *sq* policy. In contrast, a good outcome at $t = 1$ pushes more weight onto μ_1 , making further experimentation more attractive. Thus, policies that produce outcomes in a neighborhood of α – even those closer to zero – are not stable.¹⁷

Behavior is not so straightforward when $\mu_1 < 0 < \mu_2$, although the relevance of this case to ideological policy making is unclear. It implies that citizens can order policies according to expectations but cannot identify which end of the policy spectrum delivers liberal outcomes and which end delivers conservative outcomes. Under such extreme uncertainty, a sufficiently bad outcome may induce policy makers to reverse course and instead choose policies on the other side of the status quo (violating the precepts of the monotonic phase). This possibility shows the degree of uncertainty necessary to generate “course reversals” in policy dynamics. Many commentators, however, doubt the empirical relevance of such course reversals. Paul Krugman illuminates this view by analogizing a policy reversal in the following way:¹⁸

“A driver runs over a pedestrian; he looks back, realizes what he’s done.

¹⁷I cannot yet assert that the stability region at $t = 2$ is strictly tighter than $[0, \alpha]$ as an open question is whether the domain of good-enough stability is connected.

¹⁸Quoted from krugman.blogs.nytimes.com (March 3, 2008). Krugman attributes the anecdote to Jacob Frenkel.

“I’m so sorry,” he says. “Let me fix the damage.” So he backs up, running over the pedestrian a second time.”

Regardless of the type of drift uncertainty, its impact is limited to the monotonic phase of policy making (or its analogue) as experimentation in the triangulating phase is only on bridges, where the true drift is irrelevant to beliefs and policy choices.

4 The Structure of Political Competition

Driving the results of Section 3 is a simple model of political competition. In this section I enrich the model of political competition. In both variations considered the power of the median citizen is relaxed and the model no longer reduces to a single person decision problem. Although both variations add to the realism of the political model, their main contribution is to the broader normative question of how the design of a political system affects the level of experimentation and efficiency in policy making.

4.1 Candidate Divergence and Long-Term Efficiency

In practice candidates are unsure about the preferences of voters and I incorporate such uncertainty into the model. As is well known, this induces the candidates to diverge in their platform choice within elections. For brevity, I do not solve for the full dynamic path here (or in the next subsection), instead I focus on showing how divergence affects experimentation and learning.

4.1.1 Model Addendum

I retain the previous model with the following adjustments. Candidates care only about policy outcomes ($\kappa = 0$) and are equally impatient as voters. In addition to policy, voters evaluate candidates on a non-policy valence component. Specifically, voter i 's utility from policy p when offered by candidate $J \in \{X, Y\}$ is:

$$u_i^J(p) = -(o_i - \psi(p))^2 + \gamma_t^J.$$

Let $\gamma_t = \gamma_t^X - \gamma_t^Y$ be the difference in valence evaluations, where τ_t is distributed symmetrically and with full support over $[-\lambda, \lambda]$. The valence evaluation γ_t is common to all voters and a new γ_t is drawn independently each period. Valence is not observed by candidates when choosing their policy positions. Unless otherwise specified, set $\lambda = \infty$ such that both candidates have a positive probability of winning for any pair of platforms.

I say that an equilibrium is *convergent* if $x_t^* = y_t^*$, otherwise it is *divergent*.¹⁹ In an abuse of notation, I say that *stability is convergent* if the candidates stabilize at the same policy and *divergent* if they stabilize at distinct policies. Learning stops whether stability is convergent or divergent, but the policy implemented varies across elections in divergent stability according to the realization of the valence term.

4.1.2 Results

The median voter dislikes within-period divergence as her most preferred policy is no longer available. Lemma 4 shows this distaste extends at least to all the voters on one side of the median.

Lemma 4 *For any pair of platforms $x_t \neq y_t$ at election t , a strict majority of voters strictly prefer the convergent platforms $x_t^* = y_t^* = \arg \max_{p_t \in \mathbb{R}} [Eu_m(p_t)]$.*

Lemma 4 establishes that platform divergence does not improve policy making within period. The question I explore is whether divergence can improve policy making across periods.

For the monotonic phase I establish two results, showing that divergence encourages experimentation in the sense that policy making stabilizes for a strictly smaller set of histories. I begin with the requirement for good-enough stability. I retain the assumptions $\mu \leq 0$ and $o^{sq} \geq 0$. Denote by $z_t^* \in \{x_t^*, y_t^*\}$ the winning policy at election t .

Proposition 6 *At election t in the monotonic phase, policy making stabilizes at the most recent choice z_{t-1}^* if $\psi(z_{t-1}^*) \in [0, \alpha - d]$. If $d > \alpha$ good-enough stability cannot arise.*

The candidates converge in platforms if and only if they have the same most preferred policy (as otherwise deviating to their most preferred policy improves the expected outcome). Good-enough stability is obtainable, therefore, only if the most recent outcome is within α of the ideal outcomes of both candidates. Thus, the domain of outcomes that are good enough is strictly smaller than with convergent platforms. Indeed, if $d > \alpha$ no outcome satisfies the requirement and good-enough stability is not possible.

The condition for policy making to get stuck is also weakened. The candidate on the far side of the realized outcomes (in this case the candidate with ideal outcome $-d$) receives lower utility than the median voter from both experimenting further and from backtracking to a previous policy. Proposition 7 shows that the utility from backtracking to a known point decreases faster (as the expected outcome is further from $-d$) and this causes the necessary condition for getting stuck to weaken.

¹⁹Equilibrium existence (in possibly mixed strategies) is assured by appropriately truncating the policy space and noting the utility of voters and candidates is continuous in policies (as the full support of γ_t implies the probability of winning for each candidate is continuous).

Proposition 7 *Policy making gets stuck in the monotonic phase if and only if $\psi(z_{t-1}^*) > \hat{\delta}_t$, where:*

$$\hat{\delta}_t = \frac{\alpha^2 + \psi(\tau_t^*)^2 + d^2 + 2d(\psi(\tau_t^*) - \alpha)}{2\alpha} > \delta_t.$$

The impact of candidate divergence in the triangulating phase is more ambiguous. I begin by offering three conditions that are sufficient to rule out convergent stability. Proposition 8 extends the logic of Proposition 6 to the triangulating phase and follows from Corollary 2.

Proposition 8 *For $d > \alpha$ convergent stability occurs only when policy making gets stuck.*

Convergent stability is possible for $d < \alpha$ yet it remains difficult to obtain. The revelation of points in the mapping on either side of zero can give both candidates an attractive outside option. Proposition 9 provides conditions on these outside options (known points in the mapping) such that convergent stability is impossible. This and the following result apply to both good-enough stability and getting stuck. Define $\tau_t^{+*} = \arg \min_{t' < t} [\psi(x_{t'}^*), o^{sq}]$ and $\tau_t^{-*} = \arg \max_{t' < t} [\psi(x_{t'}^*)]$ as the most attractive outcomes realized up to election t on either side of zero (the previously defined τ_t^* is the policy among these two that is more attractive to the median voter).

Proposition 9 *If $d \geq \frac{\psi(\tau_t^{+*})}{2}$ convergent stability on an outcome less than zero is not possible. Similarly, convergent stability on an outcome greater than zero is not possible if $d \geq \frac{|\psi(\tau_t^{-*})|}{2}$.*

The next result considers the utility from experimentation and provides a condition on the width of the spanning bridge such that convergent stability is not possible.

Proposition 10 *Convergent stability cannot appear in the triangulating phase if $d^2 > \frac{\sigma^2}{4} (z_r^* - z_l^*)$, where z_l^* and z_r^* are the two ends of the spanning bridge.*

As the width of the spanning bridge can only narrow, an implication of this result is that once it is satisfied it is thereafter satisfied and convergent stability cannot obtain. This conclusion is in contrast to that when candidates converged within period in which case convergent stability is almost surely achieved.

The ambiguity for the triangulating phase comes from the possibility for divergent stability. I establish through example that divergent stability is possible. The example also demonstrates that divergent stability can obtain where policy would not stabilize with convergent platforms, demolishing the stronger claim that divergence induces more experimentation generally.

Example 1 Suppose $d = \alpha + 2\varepsilon$, for $\varepsilon > 0$ and small, and γ is distributed uniformly over $[-\lambda, \lambda]$ for λ large. If at election t :

$$\begin{aligned}\psi(\tau_t^{+*}) &= \alpha + \varepsilon, \\ \psi(\tau_t^{-*}) &= -\alpha - \varepsilon,\end{aligned}$$

the equilibrium is: $x_t^* = \tau_t^{+*}$, $y_t^* = \tau_t^{-*}$, and divergent stability obtains.

Although deviations exist that increase a candidate's chance of victory (to the median voter's ideal point, for example), the cost of doing so is a less desirable policy whenever the candidate wins. For λ sufficiently large the cost outweighs the benefit and divergent stability is supported.

4.2 Abstention and Experimentation

In this section I allow voters to abstain. The significance of abstention is that the identity of the *median voter* may differ from that of the *median citizen*, and may change from election to election. For simplicity, I assume a continuum of citizens (potential voters) distributed according to the density function f , where f is symmetric around zero and single-peaked, and I return to the baseline model (with no valence term).

I suppose that citizens abstain from voting when they are sufficiently *alienated* from all candidates. That is, a citizen votes for her favorite candidate if that candidate is sufficiently attractive, otherwise she abstains due to alienation. Formally, citizens have a *tolerance* level $v \geq 0$ such that they vote for their favorite candidate if the utility from that candidate weakly exceeds $-v$. If no candidate meets this threshold then a citizen abstains.²⁰

The threshold of tolerance implies that how much voters like a candidate, and not just who they like, matters to behavior. This distinction biases society toward known – and riskless – policies. The risk inherent in policy experiments lowers their appeal and dampens turnout. Figure 7 plots the expected utility across voters for a sq policy with outcome σ^{sq} and an experimental alternative $p > sq$ with expected outcome $E\psi(p)$. The expected utility curves have the same concavity but they differ in the maximum value.

The lower utility from an experimental policy implies that it draws support from a narrower section of the population. For sq to the left of experimental policy p (as in Figure 7), the sq policy gathers support from voters up to \sqrt{v} to the left of σ^{sq} , whereas policy p only gathers voters $\sqrt{v - (p - sq)\sigma^2}$ to the right of $E\psi(p)$. The experimental policy may nevertheless be majority preferred if voter density is greater

²⁰This behavior is known as *expressive* voting. Other voting theories could be used here and I chose expressive voting as it is simple and popular. The main message of this section – that abstention limits experimentation – emerges regardless of the voting theory used.

The policy process can also be enriched by relaxing *reversibility* and *determinism*. These assumptions comfort policy makers in that they know an outcome of a policy will be repeated and that they can return to that policy and achieve the same outcome should any policy experimentation go awry. In practice, however, outcomes are stochastic and history matters. Understanding how these features affect experimentation and learning are important open questions and should be accessible in this framework.

Perhaps the most interesting direction for future work is to further enrich the political environment. The baseline model restricts attention to a single polity and to learning-by-doing. Although trial-and-error is an unavoidable aspect of policy making, information about various policies can also be acquired from outside sources, such as from academics and lobbyists, or by observing the experience of different societies. The ability to learn across a federal system would, *ceteris paribus*, seem to increase efficiency, but the corresponding ability to free-ride on others' costly experimentation may undermine much of this effect. Similarly, the impact of outside experts on policy making is unclear if these experts have the incentive to strategically provide their information. The framework introduced here may offer some insight into how these trade-offs are resolved.

As noted previously, the baseline model reduces to a single person (the median voter) decision problem. This simplicity offers one advantage: the results are directly applicable to other settings. For example, the direction and distance of change are also important in the search for a job or a consumer product. To translate this model to labor market search, each policy may correspond to a particular job and outcomes represent the combination of pay and working conditions for that job. The job seeker has an ideal pay and working conditions combination, with outcomes above that level corresponding to a wage that is too low, and outcomes below that level to jobs requiring too many hours. The worker can order the jobs by expected outcome (as do voters with policies) but the realized job characteristics remain hidden until a job is tried.

Applied to this setting (with an ideal job outcome of zero), the results of previous sections describe the dynamics of impatient job search and explain why workers may sometimes get stuck in unattractive jobs. Moreover, interpreting each policy as a job suggests some natural extensions. For instance, employers may advertise by revealing (perhaps noisily) the characteristics of their job with the hope of enticing workers to join their firm. The logic of advertising in this sense is not straightforward, however, as if the advertisement creates a Brownian bridge it may induce a worker to experiment – and take a new job – but with a firm other than the one advertising.

This paper has introduced a novel model of policy making in a dynamic and uncertain environment. Amongst other things, the results provide an explanation for why unsatisfactory outcomes may persist in politics even when candidates are fully responsive to the preferences of the median voter. The importance of experimentation and learning to policy making has long been acknowledged in practice, yet formal analysis of how learning informs policy choices has received relatively little attention. The

analysis offered herein provides a framework to understand the tensions and risk in dynamic policy making, and to show how policy choice is fundamentally influenced by the history of choices and outcomes. Although many features of real policy making have been abstracted away, the framework offers enough tractability and flexibility to open up additional important and interesting questions. Regrettably these must be left for another time.

6 Appendix

Proof of Lemma 1: If candidate X offers policy p that delivers expected value $E\psi(p)$ and variance σ_p^2 , the utility of voter i is:

$$u_i(p) = -(o_i - E\psi(p))^2 - \sigma_p^2.$$

Differentiating with respect to the voter's ideal point:

$$\frac{du_i(p)}{do_i} = -2(o_i - E\psi(p)), \text{ and } \frac{d^2u_i(p)}{do_i^2} = -2.$$

As the same applies for candidate Y , single crossing holds and the median voter's preference is decisive.

As candidate utility functions are continuous in policy, holding the probability of victory constant, it follows by well-known results that in equilibrium both candidates must locate at the median voter's most preferred policy (as for $\kappa > 0$ the probability of victory is discontinuous in policy). As with probability one the median voter has a unique most preferred policy, the essentially unique equilibrium is convergent. ■

Proof of Lemma 2: By Lemma 1 the strategies of candidates depend only on the informational content of past play, a sufficient statistic for which is the set of unique winning policies. The lemma is then obvious by the stationarity of the problem. ■

Optimal Experimentation

Without loss of generality I assume hereafter that $\mu \leq 0$ and $\sigma^q \geq 0$. At any time t' , define the right and left-most policies implemented by x_{\min} and x_{\max} . Uncertainty is open ended on the flanks of x_{\min} and x_{\max} and by Brownian bridges elsewhere. I determine the optimal experimental policy for each situation in turn. Define $\alpha(w|z, \psi(z))$ as the value of $\psi(w)$ that solves $|\psi(w)| = \alpha(\widehat{w \cdot z})$ for the Brownian bridge between policies w and z ; thus x_{t-1}^* is stable at election t in the triangulating phase if and only if $|\psi(x_{t\Delta-1}^*)| < \alpha(x_{t\Delta-1}^*|x_{t\Delta-2}^*, \psi(x_{t\Delta-2}^*))$.

Open-Ended Uncertainty. Two properties are immediately clear: if $\psi(x_{\max}) \leq 0$ all experimental policies $z > x_{\max}$ are dominated by x_{\max} . Similarly all $w < x_{\min}$ are

dominated by x_{\min} if $\psi(x_{\min}) \geq 0$. Consider then $\psi(x_{\max}) \geq 0$ and policies $z \geq x_{\max}$. Expected utility for the median voter is:

$$Eu_m(z) = -[\psi(x_{\max}) + \mu(z - x_{\max})]^2 - (z - x_{\max})\sigma^2.$$

Differentiating:

$$\begin{aligned} \frac{dEu_m(z)}{dz} &= -2\mu[\psi(x_{\max}) + \mu(z - x_{\max})] - \sigma^2, \\ \frac{d^2Eu_m(z)}{dz^2} &= -2\mu^2 < 0. \end{aligned}$$

The second derivative ensures a unique maximum; solving the first order condition for an internal solution gives:

$$\psi(x_{\max}) + (z^* - x_{\max}) = \frac{\sigma^2}{-2\mu}, \quad (8)$$

and the expected outcome from policy z is set to the constant $\alpha = \frac{\sigma^2}{-2\mu}$. If $\psi(x_{\max}) \leq \alpha$ the corner solution is x_{\max} . (An analogous derivation holds for $z < x_{\min}$ when $\psi(x_{\min}) < 0$.) ■

Brownian bridge. Behavior on non-spanning bridges is straightforward: the optimal policy is the end(s) of the bridge closest to zero. Consider then a spanning bridge $\widehat{x_l x_r}$ where $\psi(x_l) > 0 > \psi(x_r)$ ($\psi(x_r) > 0 > \psi(x_l)$ is analogous) and suppose $|x_l| \leq |x_r|$. The median voter's expected utility for $z \in [x_l, x_r]$ is:

$$Eu_m(z) = -\left[\psi(x_l) + \frac{(z - x_l)}{(x_r - x_l)}(\psi(x_r) - \psi(x_l))\right]^2 - \frac{(z - x_l)(x_r - z)}{x_r - x_l}\sigma^2.$$

Differentiating:

$$\begin{aligned} \frac{dEu_m(z)}{dz} &= -2\frac{\psi(x_r) - \psi(x_l)}{(x_r - x_l)}\left[\psi(x_l) + \frac{(z - x_l)}{(x_r - x_l)}(\psi(x_r) - \psi(x_l))\right] - \frac{(x_r - z) - (z - x_l)}{x_r - x_l}\sigma^2, \\ \frac{d^2Eu_m(z)}{dz^2} &= -2\left[\frac{\psi(x_r) - \psi(x_l)}{x_r - x_l}\right]^2 + \frac{2}{x_r - x_l}\sigma^2. \end{aligned}$$

As $u_m(x_l) \geq u_m(x_r)$ by construction and the second derivative is independent of z , $\frac{d^2Eu_m(z)}{dz^2} \geq 0$ implies x_l is the optimal policy (and x_r also if $|x_l| = |x_r|$). Straightforward algebra establishes the second derivative is positive if and only if:

$$\alpha(\widehat{x_l x_r}) = \frac{\sigma^2}{-2\frac{\psi(x_r) - \psi(x_l)}{x_r - x_l}} \geq \frac{\psi(x_l) - \psi(x_r)}{2},$$

where the right-hand-side is the average distance from 0 of the two ends of the bridge. As by definition $\psi(x_l) \leq \frac{\psi(x_l) - \psi(x_r)}{2}$, a positive second derivative implies requires $|\psi(x_l)| \leq \alpha(x_l|x_r, \psi(x_r))$ (with the inequality strict for $|x_l| < |x_r|$).

Consider $\frac{d^2 Eu_m(z)}{dz^2} < 0$, noting this ensures a unique optimal policy. The end point x_l dominates experimenting iff $\frac{dEu_m(z)}{dz} \leq 0$ at $z = x_l$. By substituting $z = x_l$ into the first derivative and rearranging, this is true iff $\psi(x_l) \leq \alpha(x_l|x_r, \psi(x_r))$. For $\psi(x_l) > \alpha(x_l|x_r, \psi(x_r))$ a unique internal optimum exists and is found by rearranging $\frac{dEu_m(z)}{dz} = 0$, noting that the term in the square brackets is the expected value. ■

Thus, for any value of the second derivative the end point x_l is the optimal policy if and only if $\psi(x_l) \leq \alpha(x_l|x_r, \psi(x_r))$, otherwise experimenting on the bridge is preferred. I state here three properties of optimal experimentation on a bridge that later prove useful:

Property i: The optimal policy z^* produces: $E[\psi(z^*)|\widehat{x_l x_r}] \geq 0$, and is closer to x_l than to x_r ,

Property ii: The expected outcome of optimal policy z^* satisfies: $E[\psi(z^*)|\widehat{x_l x_r}] < \alpha(x_l|x_r, \psi(x_r))$,

Property iii: The stable boundary in period $t + 1$, conditional on the spanning bridge being $\widehat{x_z^* x_r}$, satisfies: $E[\psi(z^*)|\widehat{x_l x_r}] < \alpha(z^*|x_r, \psi(x_r)) < \alpha(x_l|x_r, \psi(x_r))$.

Property (i) follows from the fact that for any \tilde{z} such that $E[\psi(\tilde{z})|\widehat{x_l x_r}] < 0$ there exists a corresponding \tilde{z}' such that $|E[\psi(\tilde{z})|\widehat{x_l x_r}]| = E[\psi(\tilde{z}')|\widehat{x_l x_r}]$ and with lower variance.

To establish Property (ii), substitute $E[\psi(z)|\widehat{x_l x_r}] = \alpha(x_l|x_r, \psi(x_r))$ into $\frac{dEu_m(z)}{dz}$. Simplifying gives:

$$\begin{aligned} \frac{dEu_m(z)}{dz} \Big|_{\alpha(x_l|x_r, \psi(x_r))} &= -2 \frac{\psi(x_r) - \psi(x_l)}{(x_r - x_l)} \alpha(x_l|x_r, \psi(x_r)) - \frac{(x_r - z) - (z - x_l)}{x_r - x_l} \sigma^2 \\ &= -2 \frac{\psi(x_r) - \psi(x_l)}{(x_r - x_l)} \frac{(x_r - x_l)}{\psi(x_r) - \alpha(x_l|x_r, \psi(x_r)) - 2} - \frac{(x_r - z) - (z - x_l)}{x_r - x_l} \sigma^2 \\ &= \sigma^2 \left[\frac{\psi(x_r) - \psi(x_l)}{\psi(x_r) - \alpha(x_l|x_r, \psi(x_r))} - \frac{(x_r - z) - (z - x_l)}{x_r - x_l} \right] \\ &> 0, \end{aligned}$$

because for experimentation to be optimal $\psi(x_l) > \alpha(x_l|x_r, \psi(x_r))$ such that the first term is greater than one, and from Property (i) $z < \frac{x_r + x_l}{2}$ such that the second term is less than one. Thus, the optimal z^* is further to the right and $E[\psi(z^*)|\widehat{x_l x_r}] < \alpha(x_l|x_r, \psi(x_r))$ as the slope of the bridge is negative.

The first inequality of Property (iii) follows from the concavity of variance on a bridge: If $\psi(z^*) = E[\psi(z^*)|\widehat{x_l x_r}]$ then $\frac{dvar(\psi(z)|\widehat{x_l x_r})}{dz} \Big|_{z^*} = \sigma^2$ whereas on $\widehat{x_l x_r}$, the

previous period's bridge, $\frac{d\text{var}(\psi(z)|\widehat{x_l x_r})}{dz}|_{z^*} < \sigma^2$ (and the bridges have the same slope, offering the same gain in expected value). In a similar vein, the second inequality follows because the $t + 1$ bridge is narrower and steeper if $\alpha(x_l|x_r, \psi(x_r))$, $\psi(z^*) = \alpha(x_l|x_r, \psi(x_r))$. Formally:

$$\alpha(x_l|x_r, \psi(x_r)) > \frac{\sigma^2}{-2\frac{\psi(x_r) - \alpha(x_l|x_r, \psi(x_r))}{x_r - z^*}}$$

since $x_r - z^* < x_r - x_l$, implying $\alpha(z^*|x_r, \psi(x_r)) < \alpha(x_l|x_r, \psi(x_r))$. ■

Proof of Proposition 1: $x_{\min} = x_{\max} = sq$ at $t = 1$. The result follows from Equation 8 and the optimal response to uncertainty. ■

Proof of Proposition 2: The requirement $\psi(x_1^*), \dots, \psi(x_{t-1}^*) \geq 0$ implies the optimal policy is in the set $\tau_t^* \cup (x_{t-1}^*, \infty)$. If $\tau_t^* = x_{t-1}^*$ the problem is equivalent to period 1. Thus, x_{t-1}^* is stable if $x_{t-1}^* \leq \alpha$ and dominated by z^* from Equation 8 otherwise. So suppose $\tau_t^* \neq x_{t-1}^*$ and note that this implies by induction that $\psi(x_{t-1}^*) > \alpha$. As x_{t-1}^* is dominated by both the optimal experimental policy z^* and τ_t^* , equilibrium behavior requires a comparison of utility. The utility from τ_t^* is straightforward. The expected utility from z^* is:

$$\begin{aligned} Eu_m(z^*) &= -\left[\frac{\sigma^2}{-2\mu}\right]^2 - \left[\frac{\sigma^2}{-2\mu} - \psi(x_{\max})\right] \sigma^2 \\ &= \frac{1}{2} \frac{\sigma^2}{\mu} \left[\frac{1}{2} \frac{\sigma^2}{\mu} + 2\psi(x_{\max})\right], \end{aligned}$$

which is strictly decreasing in $\psi(x_{\max})$. The result follows by setting $\frac{1}{2} \frac{\sigma^2}{\mu} \left[\frac{1}{2} \frac{\sigma^2}{\mu} + 2\delta_t\right] = -[\psi(\tau_t^*)]^2$. ■

Proof of Corollary 1: As the support of the Brownian motion is \mathbb{R} , a realization $\psi(x_1^*) > \delta_2$ occurs with positive probability for any o^{sq} , and policy making gets stuck. For any other $o \in [-o^{sq}, o^{sq}]$ a path can be constructed such that $\psi(\tau_t^*) = o$ and the same argument goes through. ■

Proof of Proposition 3: The optimal choice on the spanning bridge is given by the properties above. As by the definition of the triangulating phase $\psi(x_{t\Delta-1}^*) < 0 < E\psi(x_{t\Delta-1}^*)$ at $t - 1$, the bridge is of steeper slope than μ . There exists, therefore, a $\hat{z} \in (x_{t\Delta-2}^*, x_{t\Delta-1}^*)$ such that $E[\psi(\hat{z})|x_{t\Delta-2}^* \widehat{x_{t\Delta-1}^*}] = \alpha$ and $\text{var}(\psi(\hat{z})|x_{t\Delta-2}^* \widehat{x_{t\Delta-1}^*}) < (\hat{z} - x_{t\Delta-2}^*) \sigma^2 < (x_{t\Delta-1}^* - x_{t\Delta-2}^*) \sigma^2$ by the properties of variance on a Brownian bridge. As $x_{t\Delta-1}^*$ dominated all $z \leq x_{t\Delta-2}^*$ at time $t - 1$, policy \hat{z} dominates them also, implying the optimal policy is on the spanning bridge. ■

Proof of Lemma 3: The result holds by induction. It is true at time t^Δ , the first period of the triangulating phase. For any $t > t^\Delta$ suppose it is true. The policy x_t^* is either experimental on the unique spanning bridge or stable. Stability does not change beliefs and uniqueness holds at time $t + 1$. Denote the unique spanning bridge by $\widehat{x_l^t x_r^t}$, for $\psi(\widehat{x_l^t}) > 0 > \psi(x_r^t)$ and $x_t^* \in (x_l^t, x_r^t)$. Then $\psi(x_t^*) > 0$ implies $\widehat{x_l^t x_r^t}$ is spanning and $\widehat{x_l^t x_r^t}$ is not. The reverse holds for $\psi(x_t^*) < 0$, and uniqueness holds at $t + 1$. The induction argument is complete. ■

Proof of Proposition 4: From property (iii) for bridges proved above, the threshold for stability of either end of a spanning bridge is decreasing through time. This implies that only the most recently formed end at x_{t-1}^* can prove stable, as claimed in parts (i) and (ii) of the proposition. The condition for stability in part (i) is proved above (and is the same as in Proposition 3), and optimal experimentation on the bridge in part (iii) is as before.

If part (i) does not apply then stability can occur elsewhere only if $\tau_t^* \notin \{x_l^*, x_r^*\}$ and dominates experimentation on the spanning bridge. For fixed x_l and x_r and assuming $|\psi(x_l)| \leq |\psi(x_r)|$, the expected utility of optimal experimentation is strictly decreasing in $|\psi(x_l)|$ and strictly increasing in $|\psi(x_r)|$. If $x_l = x_{t-1}^*$ (the most recent choice) then, for fixed $\psi(x_r)$, utility is minimized at $\psi(x_l) = -\psi(x_r)$, in which case optimal experimentation implies $z^* = \frac{x_l + x_r}{2}$, such that the expected outcome is 0 and:

$$\begin{aligned} Eu_m(z^*) &= -\frac{\left(\frac{x_e + x_l}{2} - x_e\right) \left(x_l - \frac{x_e + x_l}{2}\right)}{x_l - x_e} \sigma^2 \\ &= -\frac{\sigma^2}{4} (x_l - x_r). \end{aligned}$$

Thus, τ_t^* is stable iff $|\psi(\tau_t^*)| < \frac{\sigma}{2} \sqrt{|x_r - x_l|}$ and $\psi(x_l) + \psi(x_r)$ is in some neighborhood of zero. ■

Proof of Corollary 2: Established above as Property (iii) for experimentation on spanning bridges. ■

Proof of Proposition 5: As $E\psi(x_t^*) = \alpha$ in the monotonic phase, the realization satisfies $\psi(x_t^*) < \alpha$ with probability $\frac{1}{2}$. Thus, the monotonic phase ends each period with at least probability $\frac{1}{2}$, and with probability one it eventually ends. If policy stabilizes the result holds, so consider the probability that experimentation in the triangulating phase continues indefinitely.

Consider election t and the spanning bridge $\widehat{x_l^t x_r^t}$ with $|\psi(x_r^t)| > \psi(x_l^t) > 0$. By Properties (i)-(iii) above, $0 < E[\psi(z^*) | \widehat{x_l^t x_r^t}] < \alpha(z^* | x_r, \psi(x_r)) < \alpha(x_l | x_r, \psi(x_r))$, and the probability that $x_r^{t+1} - x_l^{t+1} > \frac{x_r^t - x_l^t}{2}$ and the triangulating phase continues is strictly less than $\frac{1}{2}$. Thus, with probability 1 the triangulating phase ends or the

width of the spanning bridge $x_r^t - x_l^t$ approaches 0. As the variance is bounded by $(x_r^t - x_l^t) \sigma^2$ this also implies that the slope of the spanning bridge is with probability 1 by some finite value. Therefore, with probability 1 there is some t such that $\frac{d^2 Eu_m(z)}{dz^2} = -2 \left[\frac{\psi(x_r) - \psi(x_l)}{x_r - x_l} \right]^2 + \frac{2}{x_r - x_l} \sigma^2 > 0$, which from the derivation of optimal experimentation implies policy stability and the result follows. ■

Proof of Corollary 3: The first order condition for optimal experimentation becomes:

$$\begin{aligned} \frac{dEu_m(z)}{dz} &= -\mu_1 [o^{sq} + \mu_1 (z - sq)] - \mu_2 [o^{sq} + \mu_2 (z - sq)] - \sigma^2, \\ \frac{d^2 Eu_m(z)}{dz^2} &= -\mu_1^2 - \mu_2^2 < 0. \end{aligned}$$

Without drift uncertainty, the optimal experimental policy z^* delivers expected outcome α . Substituting this into the first derivative and setting $\mu_1 = \mu - \nu$ and $\mu_2 = \mu + \nu$ gives:

$$\begin{aligned} \left. \frac{dEu_m(z)}{dz} \right|_{z^*} &= -(\mu + \nu) [\alpha + (z - sq) \nu] - (\mu - \nu) [\alpha - (z - sq) \nu] - \sigma^2, \\ &= -2(z^* - sq) \nu^2 < 0. \end{aligned}$$

Thus, the optimal experimental policy under drift uncertainty is in (sq, z^*) when $z^* > sq$, and equal to z^* when $z^* = sq$. ■

Proof of Corollary 4: By stochastic dominance and $E\psi(x_1^*) < o^{sq}$, a realization $\psi(x_1^*) > o^{sq}$ induces posterior beliefs to place more weight on μ_2 . Similarly, a realization $\psi(x_1^*) < E\psi(x_1^*)$ puts more weight on μ_1 . Both results then follow from straightforward algebra. ■

Proof of Lemma 4: From the proof of Lemma 1, for any policy the concavity of $u_i(p)$ in o_i is -2; thus, the concavity of the expectation over any two policies x_t and y_t is also -2. By definition the median voter prefers x_t^* to a lottery over x_t and y_t and x_t^* is a maximum. Equal concavity of utility in o_i over both lotteries implies all voters to one side of the median also strictly prefer x_t^* . ■

Proof of Proposition 6: $\lambda = \infty$ and the full support of γ_t imply that each candidate wins with strictly positive probability for any policy p . As candidates care only about policy outcomes, convergent stability is possible only if both candidates share the same ideal policy as otherwise deviating to the ideal strictly improves a candidate's expected outcome. The result then follows from Proposition 2. ■

Proof of Proposition 7: For $o^{sq} > 0$, policy τ_t^* is candidate X 's ideal iff:

$$(\psi(\tau_t^*) + d)^2 < \alpha^2 + \frac{[\psi(z_{t-1}^*)]^2 - (-d + \alpha)}{-\mu} \cdot \sigma^2$$

Substituting $\frac{\sigma^2}{-\mu} = 2\alpha$ and rearranging gives the required condition. That the constraint for candidate Y is weaker can be checked by substituting $-d$ for d and noticing the final term in the definition of $\hat{\delta}_t$ is then negative. ■

Proof of Proposition 8: The result follows from Proposition 6 and Corollary 2. ■

Proof of Proposition 9: $d \geq \frac{\psi(\tau_t^{+*})}{2}$ implies candidate Y can always obtain a utility of at least $-\left(\frac{\psi(\tau_t^{+*})}{2}\right)^2$; as any policy with an outcome (certain or not) on the other side of zero delivers strictly less utility, such a policy cannot be optimal for candidate Y . The result follows as convergent stability obtains only if candidates share an ideal policy (see the proof of Proposition 6). For candidate X the condition $d \geq \frac{|\psi(\tau_t^{-*})|}{2}$ is analogous. ■

Proof of Proposition 10: If $\psi(\tau_t^{+*}) \in [0, d]$ and $\psi(\tau_t^{-*}) \in [-d, 0]$ convergent stability is not possible. So suppose $\psi(\tau_t^{+*}) > d$, in which case a policy on the spanning bridge delivers an expected outcome of d . Y 's utility from this policy is weakly greater than $-\frac{\sigma^2}{4}(z_r^* - z_l^*)$, whereas any policy with expected outcome less than zero delivers utility strictly less than $-d^2$. An analogous argument holds for candidate X if $\psi(\tau_t^{-*}) < -d$ and the result follows. ■

Proof of Example 1: Fixing $x_t^* = \tau_t^{-*}$, the probability candidate Y wins election for any y is:

$$f(x_t^*, y) = \frac{1}{2} + \frac{[\psi(\tau_t^{-*})]^2 - E[\psi(y)]^2 - \text{var}[\psi(y)]}{2\lambda}.$$

And Y 's utility is:

$$u^Y(x_t^*, y) = -f(x_t^*, y) ([d - E[\psi(y)]]^2 + \text{var}[\psi(y)]) - (1 - f(x_t^*, y)) (d + \psi(x_t^*))^2.$$

Differentiating:

$$\begin{aligned} \frac{du^Y(x_t^*, y)}{dy} &= -([d - E[\psi(y)]]^2 + \text{var}[\psi(y)]) \frac{df(x_t^*, y)}{dy} \\ &\quad - f(x_t^*, y) \left(-2[d - E[\psi(y)]] \frac{dE[\psi(y)]}{dy} + \frac{d\text{var}[\psi(y)]}{dy} \right) + \frac{df(x_t^*, y)}{dy} (d + \psi(x_t^*))^2. \end{aligned}$$

For λ sufficiently large $\frac{df(x_t^*, y)}{dy}$ is arbitrarily small and the sign is determined by the middle term. Across the spanning bridge $p \in (\tau_t^{+*}, \tau_t^{-*})$, $[d - E[\psi(p)]] > 0$ and $\frac{dE[\psi(p)]}{dp} < 0$. Policies on the half of the bridge $\tilde{p} \in \left(\frac{\tau_t^{+*} + \tau_t^{-*}}{2}, \tau_t^{-*}\right)$ are strictly dominated by the reflection about $\frac{\tau_t^{+*} + \tau_t^{-*}}{2}$ (as \tilde{p} delivers the same probability of winning with a less attractive expected outcome). And on the half of the bridge closer to Y , $p \in \left(\tau_t^{+*}, \frac{\tau_t^{+*} + \tau_t^{-*}}{2}\right)$, the variance satisfies $\frac{dvar[\psi(p)]}{dp} > 0$ and the middle term is bounded away from zero for fixed ε . Thus $y = \tau_t^{+*}$ is optimal and the result follows. ■

Proof of Proposition 11: Voter i 's expected utility over the two policies is:

$$u_i(\tau_t^*) = - (o_i - o^{\tau_t^*})^2, \quad \text{and} \quad u_i(p) = - (o_i - E\psi(p))^2 - \sigma_p^2.$$

For $v > \alpha^2$ and assuming the intervals of support don't overlap, the measure of support for each alternative is:

$$V(\tau_t^*) = \int_{o^{\tau_t^*} - \sqrt{v}}^{o^{\tau_t^*} + \sqrt{v}} f(x) dx, \quad \text{and} \quad V(p) = \int_{E\psi(p) - \sqrt{v - \sigma_p^2}}^{E\psi(p) + \sqrt{v - \sigma_p^2}} f(x) dx.$$

Suppose the election is tied at tolerance level v' . As f is single peaked, this implies that p 's interval of support is closer to the median citizen at 0: formally, without loss of generality, if $o^{\tau_t^*} + \sqrt{v} < 0$ and $E\psi(p) > 0$, then $|E\psi(p) - \sqrt{v - \sigma_p^2}| < |o^{\tau_t^*} + \sqrt{v}|$ and $|E\psi(p) + \sqrt{v - \sigma_p^2}| < |o^{\tau_t^*} - \sqrt{v}|$ (and that $V(\tau_t^*)$ doesn't span zero).

Differentiating vote shares with respect to voter tolerance:

$$\begin{aligned} \frac{dV(\tau_t^*)}{dv} &= \frac{1}{2\sqrt{v}} [f(o^{\tau_t^*} + \sqrt{v}) + f(o^{\tau_t^*} - \sqrt{v})], \\ \frac{dV(p)}{dv} &= \frac{1}{2\sqrt{v - \sigma_p^2}} \left[f\left(E\psi(p) + \sqrt{v - \sigma_p^2}\right) + f\left(E\psi(p) - \sqrt{v - \sigma_p^2}\right) \right], \end{aligned}$$

which gives $\frac{dV(p)}{dv} > \frac{dV(\tau_t^*)}{dv}$. Thus, policy τ_t^* wins over p for all tolerance levels less than v' . ■

References

- [1] Aghion, Philippe, Patrick Bolton, Christopher Harris, and Bruno Jullien. 1991. "Optimal Learning by Experimentation." *Review of Economic Studies* 58: 621-654.
- [2] Bala, Venkatesh, and Sanjeev Goyal. 1998. "Learning From Neighbours." *Review of Economic Studies* 65: 595-621.

- [3] Baron, David. 1996. "A Dynamic Theory of Collective Goods Programs." *American Political Science Review* 90 (June): 316-330.
- [4] Barro, Robert J. 1973. "The Control of Politicians: An Economic Model." *Public Choice* 14: 19-42.
- [5] Bendor, Jonathan. 1995. "A Model of Muddling Through." *American Political Science Review* 89 (December): 819-840.
- [6] Black, D. 1958. *The Theory of Committees and Elections* (Cambridge University Press: Cambridge, U.K.).
- [7] Callander, Steven. 2007. "A Theory of Policy Expertise." Working paper.
- [8] Calvert, Randall L. 1985. "Robustness of the Multidimensional Voting Model: Candidates' Motivations, Uncertainty and Convergence." *American Journal of Political Science* 29: 1056-1070.
- [9] Duggan, J. 2000. "Repeated Elections With Asymmetric Information." *Economics and Politics* 12 (July): 109-135.
- [10] Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50: 5-26.
- [11] Fernandez, Raquel, and Dani Rodrik. 1990. "Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty." *American Economic Review* 81(5): 1146-55.
- [12] Friedman, Benjamin M. 1984. "Lessons From the 1979-82 Monetary Policy Experiment." *American Economic Review Papers and Proceedings* 74: 382-387.
- [13] Gilligan, Thomas, and Keith Krehbiel. 1987. "Collective Decision Making and Standing Committees: An Informational Rationale for Restrictive Amendment Procedures." *Journal of Law, Economics, and Organization* 3 (2): 287-335.
- [14] Kollman, Kenneth, John Miller, and Scott Page. 2000. "Decentralization and the Search for Policy Solutions." *Journal of Law, Economics, and Organization* 16 (April): 102-128.
- [15] Lundblom, Charles. 1959. "The Science of 'Muddling Through'." *Public Administration Review* 19: 79-88.
- [16] Majumdar, Sumon, and Sharun W. Mukand. 2004. "Policy Gambles." *American Economic Review* 94(September): 1207-22.

- [17] Merton, Robert K. 1936. "The Unanticipated Consequences of Purposive Social Action." *American Sociological Review* 1 (December): 894-904.
- [18] Piketty, Thomas. 1995. "Social Mobility and Redistributive Politics." *Quarterly Journal of Economics* CX (August): 551-584.
- [19] Roemer, John E. 1994. "The Strategic Role of Party Ideology When Voters are Uncertain About How the Economy Works." *American Political Science Review* 88 (June): 327-335.
- [20] Strulovici, Bruno. 2008. "Voting and Experimentation." Working paper.
- [21] Walker, Jack L. Jr. 1969. "The Diffusion of Innovations Among the American States." *American Political Science Review* 63 (September): 880-899.
- [22] Wittman, Donald. 1983. "Candidate Motivation: A Synthesis of Alternative Theories." *American Political Science Review* 77: 142-157.