

A Theory of Information Cascades in Sequential Voting*

S. Nageeb M. Ali[†]
Stanford University

Navin Kartik[‡]
UC San Diego

PRELIMINARY DRAFT. COMMENTS WELCOME.

March 8, 2006

Abstract

This paper develops an informational but rational theory of cascade behavior when voting is sequential. We analyze a two-candidate election in which voters are uncertain about the realization of a state variable that affects the utility of all voters. Voters vote in an exogenously fixed sequence, observing the history of votes at each point. We show that there is an (essentially) strict equilibrium with *posterior-based voting*: each voter votes for the candidate who she believes to be better, taking into account the information revealed in prior votes. In this equilibrium, herding on a candidate occurs with positive probability; moreover, as the population size grows, the probability of a herd converges to one. The dynamics of information cascades are analyzed and applied to the concept of momentum in elections. Our results help understand and have implications for sequential voting mechanisms such as Presidential Primaries and roll-call voting.

Keywords: sequential voting, information aggregation, bandwagons, momentum, information cascades, herd behavior

J.E.L. Classification: C7, D72

*We thank Susan Athey, Doug Bernheim, Jon Levin, Marc Meredith, and Joel Sobel for helpful discussions.

[†]Email: nageeb@gmail.com; Web: <http://www.stanford.edu/~nageeb>; Address: 579 Serra Mall, Stanford, CA 94305.

[‡]Email: nkartik@ucsd.edu; Web: <http://econ.ucsd.edu/~nkartik>; Address: Department of Economics, 9500 Gilman Drive, La Jolla, CA 92093-0508.

“...when New Yorkers go to vote next Tuesday, they cannot help but be influenced by Kerry’s victories in Wisconsin last week. Surely those Wisconsinites knew something, and if so many of them voted for Kerry, then he must be a decent candidate.”

— Duncan Watts in *Slate Magazine*

1 Introduction

Many elections take place over time. The most prominent example of this lies at the heart of the American Presidential selection process: the Primaries. A series of elections by which a party selects its candidate for the general Presidential election, the Primaries are held sequentially across states over a few months. On a smaller scale, but also explicitly sequential, are the roll-call voting mechanisms used by city councils and Congressional bodies. A more subtle example is the general U.S. Presidential election itself, where the early closing of polls in some states introduces a temporal element into voting.

In contrast, most models of voting are static. The distinction between simultaneous and sequential voting mechanisms is not just of theoretical interest, but also relevant to policy. It is often suggested that allowing voters to condition their choices on the acts of previous voters can give rise to *momentum* effects: the fact that a particular alternative is leading in initial voting rounds may induce some voters to select it who would have otherwise voted differently (see, for example, Bartels [5]). While the empirical evidence on momentum in U.S. Presidential Primaries is particularly suggestive but not definitive, it is clear that a belief in momentum effects has shaped electoral policy and strategy. Candidates compete heavily in the initial battleground states, eager to begin a bandwagon that will propel them to ultimate victory. Campaign funds and media attention are often disproportionately devoted to these early states.¹ This has led to the front-loading of elections: several states have moved their Primaries to earlier dates in order to ensure that their election is held near the beginning of the sequence (Morton and Williams [23]).² Yet, while this concept of momentum plays a significant role in political discourse, its cause has remained controversial and mysterious. The issue is important because the efficiency and equity of a sequential voting procedure may depend on the nature of momentum effects.

In this paper, we develop a theory of rational voting behavior that generates information cascades in sequential voting. While it is possible—even likely—that factors such as campaign

¹Bartels [5] and Gurian [19] document how a substantial share of the media’s attention and candidates’ campaign resources are devoted to New Hampshire even though that state accounts for merely 4 out of 538 total electoral votes.

²For instance, since 1977, New Hampshire law has stated that its primary is to be the first in the nation. As a result, the state has had to move its primary, originally in March, earlier in the year to remain the first. New Hampshire’s primary was held on February 20 in 1996, February 1 in 2000, and then January 27 in 2004 to compete with front-loaded primaries in other states.

finance or the media play a role in electoral momentum, we direct our attention to the most fundamental aspect of sequential voting: the inferences formed by voters based upon the history of prior votes, and the rational behavior induced by these inferences. We formalize an informational theory of voting momentum that is built on the simple logic that if initial voters vote informatively, their decisions provide useful information to later voters. Thus, if voters have preferences with a common value component such as choosing the best or most electable presidential candidate, then winning the support of initial voters can provide information that may induce greater support from future voters.

At first glance, this idea may seem like a mere application of the theory of rational herds in sequential decision-making, initiated by Banerjee [2] and Bikhchandani, Hirshleifer and Welch [7]. However, the typical environments studied in this literature share the important feature that a player's payoffs do not directly depend on the actions of others. Since there are no payoff complementarities, the *only* externality in these models is informational. Consequently, it is straightforward that in these "economic" settings, the optimal action for a player is independent of how her choice affects the decisions of future players: all that a player cares about is extracting information from the observed history of decisions. That is, in these models, each player optimally behaves like a backward-looking Bayesian, and has no strategic reason to consider the impact of her choice on the choices of subsequent actors. In contrast, an election is a game *with* payoff interdependencies: a voter's payoffs are determined not only by her own action or vote, but also by the votes of those before and after her. Thus, an individual's vote may affect her payoffs in two different ways. First, a vote changes the likelihood of a particular candidate winning in the event that all other votes are tied. Second, a vote can signal information to future voters, and therefore induce different (distributions over) actions of future voters. Given the direct action externalizes and the associated incentive to signal information, the optimal choice of a voter is no longer independent of how her vote influences subsequent voters. In particular, a voter may wish to vote *against* the candidate she currently believes to be better, so as to communicate her private information to future voters and/or prevent future voters from herding. Due to this forward-looking incentive, the possibility of rational herds in a sequential voting environment raises significantly different issues from the standard herding literature.

The main contribution of this paper is to derive a theory of information cascades in sequential voting in a fully rational model with standard preferences. Specifically, we consider a sequential version of a well-established voting environment (Feddersen and Pesendorfer [13, 14]). There are two candidates, and a finite population of voters who vote in an exogenously fixed sequence, each observing the entire history of prior votes. Voters are uncertain about the realization of a binary state variable that affects their utility. Each voter receives a private binary signal that contains some information about the state. There are two kinds of voters: Neutrals and Partisans. Neutrals desire to elect the "correct" or better candidate, which depends on the unknown state. Partisans, on the other hand, wish to elect their exogenously preferred

candidate regardless of the state (there are thus two kinds of Partisans, one for each candidate). Whether a voter is Neutral or Partisan is her private information.

We show that there is a Perfect Bayesian equilibrium with *Posterior-based Voting* (PBV) strategies: Partisans vote for their preferred candidate, and each Neutral voter votes sincerely for the candidate who she believes to be better at the time of casting her vote, fully taking into account the information revealed in prior votes. In other words, each voter does in fact behave like a backward-looking Bayesian, and this is optimal behavior given the strategies of other voters. This strategy profile leads to herd behavior by later voters with high probability in large elections, converging to probability one as the population size grows to infinity.³ Importantly, PBV is an (essentially) strict equilibrium; thus its existence is not due to any ad-hoc resolution of indifference among players, and the equilibrium is robust to small perturbations of the model. In the equilibrium, voters who herd strictly prefer to do so even when accounting for the possibility that their vote may change the outcome.

PBV is the appropriate generalization of *sincere voting* (Austen-Smith and Banks [1]) when moving from simultaneous to sequential voting. In voting sincerely, a voter uses all available information—her prior and signal in the standard simultaneous voting environment—to form her expectation of which candidate is better for her, and votes for this candidate. Importantly, there is no presumption of conditioning on being pivotal. In our sequential voting environment, the available information to a voter is not only her private signal, but also the history of votes. Thus, the profile of PBV strategies constitutes an especially attractive strategy profile, since it features sincerity by all voters. That this profile is an equilibrium means that even if a voter were to condition on being pivotal, such sincere voting is optimal given the play of other voters.

Voting games typically possess multiple equilibria. Due to the sheer complexity of the space of all strategy profiles in our sequential environment, it is not feasible to provide a complete characterization of all equilibria. However, we prove that for generic parameters, within a broad class of strategies profiles—so-called Cut Point Voting profiles (Callander [9])—every equilibrium induces herding with arbitrarily large probability in all sufficiently large elections. A precise definition of what behavior is permitted in a Cut Point Voting profile is deferred to Section 4, but we note here that the class of Cut Point Voting profiles subsumes PBV, informative voting by Neutrals (voting for the candidate favored by one’s private signal alone), all Neutrals voting for the same candidate independent of history or private signal, and a host of other behavior. While we believe PBV is a particularly salient equilibrium because it features all voters playing sincerely, our result says that even if voters coordinated on some other Cut Point Voting equilibrium—should any exist—herd behavior would nevertheless ensue with high probability in large elections with generic parameters.

The notion that herd behavior can be the outcome in a fully rational sequential voting model

³Throughout, we use the terms information cascade and herd behavior interchangeably. Smith and Sorensen [26] have noted a difference, but the notions coincide in our setting.

is perhaps surprising from the standpoint of political economy. Much of the early informal literature in political science explains momentum effects by invoking assumptions such as a psychological desire to vote for the eventual winner (e.g. Bartels [4, 5]); this has recently been formalized by Callander [9]. While we do not wish to enter a debate about the plausibility of this assumption, we believe it is important to establish a non-behaviorally-based benchmark.⁴ In an important contribution, Dekel and Piccione [12] study a model of strategic voting and show that under some assumptions, any symmetric equilibrium of their simultaneous voting game is outcome-equivalent to an equilibrium of the corresponding sequential voting game. Unfortunately, this has sometimes been misinterpreted as implying that conditioning on being pivotal implies that no information can be extracted from a history of observed votes.⁵ Our PBV equilibrium explicitly shows this is not the case; in particular, in the PBV equilibrium we derive, strategic voters *do* learn from previous voters *and* they act fully rationally, as if they are pivotal.

The plan for the paper is as follows. In the remainder of this section, we discuss the application of our theory to Presidential Primaries in more detail. Section 2 lays out the model, and Section 3 derives the main results about PBV strategies and equilibrium. We discuss implications, limitations, and extensions in Section 4. A discussion of related literature is postponed until Section 5. Section 6 concludes. All formal proofs are relegated to the Appendix.

1.1 The Application to Presidential Primaries

The notion of momentum in elections has received the greatest attention in the U.S. Presidential Primaries process. It has often been argued persuasively that early victories makes a difference in the prospects for any candidate. This is reflected not only in the disproportionate amount of campaign funds that go towards convincing early voters, but also in data on voting behavior and polls. In one particularly striking conclusion, Bartels [5] estimates that if the Primaries had been held on the same day in all states, both Republican and Democrat parties' nominees may have been very different in various years, and moreover, this may have led to different Presidents altogether.

Our model provides an understanding of how herds in Primaries may form, regardless of whether one believes that voters are truly strategic, so long as they incorporate the information revealed in prior elections into their decision-making. In the PBV equilibrium we derive, candi-

⁴It is difficult to empirically distinguish between momentum arising because voters believe a candidate will be the eventual *winner*, or all things considered, simply the *better* candidate.

⁵For example, in a recent paper on costly voting, Battaglini [6, p. 455] states that “strategic voters do not learn from previous voters because they act as if they are pivotal ...”. Similarly, Feddersen and Pesendorfer [15], in explaining Dekel and Piccione’s result, asserted that “in symmetric environments, this informational difference [between sequential and simultaneous voting] is irrelevant, because all strategic voters condition their vote on being pivotal.” Such statements cannot generally be true without referring to a voter’s belief over the play of other voters.

dates become increasingly attractive to neutral voters through their vote lead. Since candidates always receive votes from their partisan supporters, the task facing a neutral voter is to infer how much of a candidate’s lead has emerged from prior neutral voters. Thus, the information that is revealed in early elections affects how subsequent voters view each candidate. A particularly important feature of our model is that we permit partisanship to be asymmetric across candidates: candidate L may have a greater level of partisan support than candidate R . A greater level of partisan support for one candidate mitigates the impact of a vote for that candidate on future voters’ behavior in our PBV equilibrium. A straightforward extension naturally leads to some signaling dynamics: if a candidate wins in a state where he was anticipated to have little partisan support, this indicates that many neutral voters voted for him.⁶ Since neutrals vote on the basis of their information (so long as a herd has not started), a “surprise” victory improves the candidate’s standing with all remaining neutrals. By the same token, a victory in a state where a candidate already has much partisan support—such as his home state—does little to improve his position with all remaining neutrals.

We believe that these signaling dynamics are illustrative of momentum effects in Presidential Primaries. Analyses of the media and voter perceptions have emphasized that Primaries are often presented in a manner similar to a horse-race: greater attention is paid to how candidates perform relative to expectations, and less so to policy issues (Bartels [5] and Robinson and Sheehan [25]). In such a context, surprise victories gain a large amount of coverage, and is believed to have a large impact on subsequent votes, while anticipated victories receive less attention from the media and voters.

The impact of surprise victories is exemplified in Carter’s victory in the 1976 Democratic Primaries. Hamilton Jordan, who would later become Jimmy Carter’s White House Chief of Staff, outlined the importance of a surprise victory in New Hampshire in a memorandum two years before Carter’s campaign: “...a strong surprise in New Hampshire should be our goal, which would have a tremendous impact on successive primaries...” (Bartels [5]). This assessment is believed to be correct. For example, simulating Carter’s national popularity prior to his victory in New Hampshire, Bartels [5] predicts that Carter would have lost the 1976 nomination to George Wallace had the Primaries been simultaneously held in all states. Our model suggests how this surprise victory aided Carter. Carter, being a candidate from the South, was not believed to have much partisan support in New Hampshire. Thus, a victory there would indicate that a large number of neutral voters found reasons to vote for this candidate, and therefore, serves as a strong signal to voters in other states that Carter may be the right candidate. Indeed, having won New Hampshire, Carter demonstrated to voters, the media, and financiers that he could win without his partisan support in the South. As Carter’s press secretary, Jody Powell, put it: “...there’s no better way to do that than to go way the hell off

⁶Strictly speaking, our model does not permit the probability of partisanship to vary across voters (states), but this is straightforward to incorporate at the cost of notational burden.

yonder to New Hampshire and win a primary. . .” (Bartels [5, p. 202]).

History also indicates that victories in a candidate’s home state or where partisan support is highly likely have less of an impact on future voter behavior. This was not only the case with Tom Harkin of Iowa, who was unable to capitalize on his victory in the 1992 Iowa caucuses, but more perhaps famously, is also believed to have affected Walter Mondale of Minnesota in the 1984 Primaries. Expected to have support in the neighboring state, Mondale’s victory in the Iowa caucuses was largely overshadowed by Gary Hart’s performance. That Hart was able to garner 16.5% of the votes in Iowa generated a burst of interest from voters and the media. By performing better than anticipated, Gary Hart was suddenly perceived as a candidate that stood out.

While the informational model here contributes to the understanding of electoral dynamics, we do not wish to suggest that it captures the whole story. In our model, we abstract away from many important institutional details such as campaign finance and media attention. Having said that, it is difficult to shed light on why both financial and media resources are devoted to the first few elections without making specific assumptions about voting behavior. Insofar as the purpose of many elections—especially Primaries—is to provide and aggregate information, we believe that these institutional details should be embedded in an informational model similar to the one below. Thus, this paper may be viewed as a first step towards understanding the role that these institutions play in voting behavior.

2 Model

We consider a voting game with a finite population of n voters, n odd. Voters vote for one of two candidates, L or R , in a fixed sequential order, one at a time. Without loss of generality, we label the voters $1, \dots, n$, where a lower numbered voter votes earlier in the sequence. Each voter observes the entire history of votes when it is her turn to vote. The winner of the election, denoted $W \in \{L, R\}$, is selected by simple majority-rule. The state of the world, $\omega \in \{L, R\}$, is unknown, but individuals share a common prior over the possible states, and $\pi > \frac{1}{2}$ is the ex-ante probability of state L . Prior to voting, each voter i receives a private signal, $s_i \in \{l, r\}$, drawn from a Bernoulli distribution with precision γ (i.e., $\Pr(s_i = l | \omega = L) = \Pr(s_i = r | \omega = R) = \gamma$). We assume that individual signals are independently drawn conditional on the state, and that $\gamma > \pi$, which means that each signal is more informative than the prior.⁷

In addition to being privately informed about her signal, a voter also has private information about her preferences: she is either an L -partisan (L_p), a Neutral (N), or an R -partisan (R_p). We denote this *preference type* of voter i by t_i . Each voter’s preference type is drawn independently from the same distribution, which assigns probability $\tau_L \in (0, \frac{1}{2})$ to preference

⁷Our analysis will carry over with obvious changes to cases where the signal precision is asymmetric across states of the world.

type L_p , probability $\tau_R \in (0, \frac{1}{2})$ to R_p , and probability $\tau_N \equiv 1 - \tau_L - \tau_R$ to the Neutral type, N .⁸ The uncertainty over the true state makes Neutral voters uncertain over which is the better candidate, whereas for $C \in \{L, R\}$, a C -partisan prefers candidate C independent of the state. Payoffs for voter i are defined by the function $u(t_i, W, \omega)$ as follows:

$$\begin{aligned} u(L_p, W, \omega) &= \mathbf{1}_{\{W=L\}} \text{ for } \omega \in \{L, R\} \\ u(R_p, W, \omega) &= \mathbf{1}_{\{W=R\}} \text{ for } \omega \in \{L, R\} \\ u(N, L, L) &= u_i(N, R, R) = 1 \\ u(N, L, R) &= u_i(N, R, L) = 0 \end{aligned}$$

Therefore, regardless of the state of the world, a Partisan wishes that her preferred candidate is elected.⁹ A Neutral, on the other hand, would like the election to select the better candidate (i.e., the appropriate candidate for each state). It is important to observe that an individual vote affects one's payoffs only indirectly through the vote's influence on the winner, W .

Denote by $G(\pi, \gamma, \tau_L, \tau_R; n)$ the sequential voting game defined above with prior π , signal precision γ , preference type parameters τ_L and τ_R , and n voters. Throughout the subsequent analysis, we use the term *equilibrium* to mean a (weak) Perfect Bayesian equilibrium of this game (Fudenberg and Tirole [17]). Let $h^i \in \{L, R\}^{i-1}$ be the realized history of votes when it is voter i 's turn to act; denote $h^1 = \phi$. A pure strategy for voter i is a map $v_i : \{L_p, N, R_p\} \times \{L, R\}^{i-1} \times \{l, r\} \rightarrow \{L, R\}$. We say that a voter i *votes informatively* following a history h^i if $v_i(N, h^i, l) = L$ and $v_i(N, h^i, r) = R$. Beliefs play a key role in the discussion, hence it is helpful to denote the posterior probability that voter i places on state L by $\mu_i(h^i, s_i)$.

3 Posterior-Based Voting

3.1 Definition and Dynamics

We start the analysis by introducing *Posterior-Based Voting* (PBV) strategy profiles and studying the induced dynamics, postponing to the next subsection the question of whether such a profile constitutes an equilibrium. Let $\mathbf{v} = (v_1, \dots, v_n)$ denote a strategy profile. In the usual fashion, $\mathbf{v}_{-i} = (v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n)$ denotes a profile of strategies for all players except i .

Definition 1. *A strategy profile, \mathbf{v} , satisfies (or is) Posterior-Based Voting (PBV) if for every*

⁸While formally, our model does not allow for the pure common values setting where $\tau_L = \tau_R = 0$, the limit of the equilibria we describe is an equilibrium of the game when $\tau_L = \tau_R = 0$; we return to this point in Section 4.2.

⁹Since heterogeneous preferences and heterogeneous priors are essentially interchangeable in voting models such as this one, the Partisan types may be interpreted as individuals whose prior is so strong on one candidate that no amount of available information can overturn it.

voter i , type t_i , history h^i , signal s_i , and for any $W, W' \in \{L, R\}$,

1. $\mathbb{E}_\omega[u_i(t_i, W, \omega)|h^i, s_i; \mathbf{v}_{-i}] > \mathbb{E}_\omega[u_i(t_i, W', \omega)|h^i, s_i; \mathbf{v}_{-i}] \Rightarrow v_i(t_i, h^i, s_i) = W$
2. $\mathbb{E}_\omega[u_i(t_i, L, \omega)|h^i, s_i; \mathbf{v}_{-i}] = \mathbb{E}_\omega[u_i(t_i, R, \omega)|h^i, s_i; \mathbf{v}_{-i}] \Rightarrow v_i(t_i, h^i, L) = L$ and $v_i(t_i, h^i, R) = R$

Strictly speaking, PBV is a property of a strategy *profile*. We will sometimes simply say PBV instead of a PBV strategy profile. Moreover, we also abuse terminology by referring to a PBV strategy: this refers to a strategy for a player that is part of a PBV profile. Let us discuss each part of the above definition.

The first part of the definition requires that given the history of votes and her private signal, if a voter believes that electing candidate L (R) will yield strictly higher utility than electing candidate R (L), then she should vote for candidate L (R). In other words, in any PBV profile, every agent votes for the candidate that maximizes her expected utility given all currently available information, where beliefs about the state of the world are updated taking as given the strategies employed by other voters (in particular, the strategies of previous voters). Since Partisan voters have a preference ordering over candidates that is independent of the state of the world, Partisans must vote for their preferred candidate in a PBV profile. If a Neutral voter's posterior is $\mu_i(h^i, s_i) \neq \frac{1}{2}$, then she votes for the candidate she believes to be better.

Part two of the definition is a tie-breaking rule. It requires that when a Neutral voter has posterior $\mu_i(h^i, s_i) = \frac{1}{2}$, she vote informatively by “voting her signal” (or more precisely, votes as if she had only her signal available and not the prior history of votes). In doing so, she reveals her signal to future voters. There are two points to be emphasized. First, this tie-breaking convention is hostile towards herding because it requires a Neutral to reveal her signal when indifferent. Since our objective is to show that PBV leads to herding, we are stacking the deck against us; the interpretation of our results are correspondingly made stronger by choosing this tie-breaking rule. Second, any specification of how to break ties only matters for a non-generic constellation of parameters $(\pi, \gamma, \tau_L, \tau_R)$. That is, for a generic tuple, $(\pi, \gamma, \tau_L, \tau_R)$, when PBV is played, it will never be the case that there is a Neutral voter with posterior $\mu_i(h^i, s_i) = \frac{1}{2}$. We defer a further discussion of this point to Remark 1 in the Appendix.

The above discussion implies the behavior of voter i in the PBV profile is summarized as follows:

$$\begin{aligned}
 v_i(L_p, h^i, s_i) &= L \\
 v_i(R_p, h^i, s_i) &= R \\
 v_i(N, h^i, s_i) &= \begin{cases} L & \text{if } \mu_i(h^i, s_i) > \frac{1}{2} \text{ or } \{\mu_i(h^i, s_i) = \frac{1}{2} \text{ and } s_i = l\} \\ R & \text{if } \mu_i(h^i, s_i) < \frac{1}{2} \text{ or } \{\mu_i(h^i, s_i) = \frac{1}{2} \text{ and } s_i = r\} \end{cases}
 \end{aligned}$$

PBV is sophisticated insofar as voters infer as much as possible from the past history, taking into account the strategies of preceding players. However, since PBV does not involve any computations regarding the implications of one's vote on future votes, it can represent behavior without a forward-looking component. *A priori*, this makes it unclear whether PBV can be optimal strategic behavior. In particular, were a voter to condition on being pivotal (affecting the outcome of the election), then she would certainly have to consider how her vote affects the votes of those after her.¹⁰ We postpone this issue to Section 3.2, instead focussing for the moment on the voting dynamics induced by PBV.

Consider a voter $i > 1$ who observes a history h^i where k voters voted L and $i - k - 1$ voters voted R , and knows (or assumes) that preceding Neutrals voted informatively and preceding Partisans voted their bias. In voting according to PBV, the voter faces a simple Bayesian inference problem: conditional on the observed history and private signal, what is the probability that the state is L ? If the state were L , the probability that a preceding voter would vote for L is $\tau_L + (1 - \tau_L - \tau_R)\gamma$, where the first term is the probability that a voter is an L -partisan, and the second term is the probability that a voter is Neutral and received an l signal. Similar computations for the remaining combinations of probabilities of voting for each candidate in each state lead to the the following likelihood ratio:

$$\frac{\Pr(h^i|\omega = L)}{\Pr(h^i|\omega = R)} = \left(\frac{\tau_L + \tau_N\gamma}{\tau_L + \tau_N(1 - \gamma)} \right)^k \left(\frac{\tau_R + \tau_N(1 - \gamma)}{\tau_R + \tau_N\gamma} \right)^{i-k-1} \quad (1)$$

Since $\gamma > 1 - \gamma$, it is clear from the above expression that the likelihood ratio is strictly increasing in k . Thus, seeing a greater number of votes for L strictly raises Neutral voter i 's belief that L is the better candidate, given the postulated beliefs about preceding voter behavior. Observe that the presence of partisanship adds noise to the public history: when τ_L and τ_R are both approximately 0, the likelihood ratio is approximately $\left(\frac{\gamma}{1-\gamma}\right)^{2k+1-i}$, which is as large as it can possibly be since this reflects a perfectly informative public history. On the other hand, when τ_L and τ_R are both approximately 0.5, the public history is close to pure noise: the likelihood ratio is approximately 1.

We introduce two examples to illustrate these ideas and introduce PBV dynamics.

Example 1. Consider the PBV strategy profile of a sequential voting game with 5 voters where $\pi = \frac{2}{3}$, $\tau_L = \tau_R = \tau = \frac{1}{4}$ and $\gamma = \frac{3}{4}$. Voter 1, if neutral, votes informatively since her signal overturns her prior. Suppose she votes L (and therefore must have received signal l). If voter 2 receives signal l , her posterior that the state is L is given by $\mu_2(L, l) = \frac{10}{11}$. If voter 2 receives an r signal, the posterior is $\mu_2(L, r) = \frac{10}{19}$. In either case, $\mu_2 > \frac{1}{2}$, and thus, regardless of

¹⁰It may be useful to think of the analogy with strategic bidder in a 1st price common value auction. There, a rational bidder must account for the *winner's curse*: given a conjecture on how others bid, a rational bidder determines the optimal bid by conditioning on the event that her bid changes the winner of the auction.

her signal, a Neutral voter 2 votes for candidate L . Let us now consider the position of voter 3. If she observes a history of (L, L) , she concludes that voter 2 was either an L -partisan or a Neutral; if she observes a history of (L, R) , she concludes that voter 2 must have been an R -partisan. In either case, voter 2's vote is uninformative about s_2 . Thus, voter 3 is essentially in the same position as voter 2, and by the previous logic, will vote L independent of her signal if she is Neutral. By induction, it follows that all future Neutrals vote for Candidate L .

In the above example, learning ceases once candidate L has a positive vote-lead. From an informational standpoint, this is clearly inefficient: given the payoff complementarities, subsequent neutrals would wish to communicate their information especially if their signals favor candidate R . This is however impossible: if a Neutral voter 2 voted R to reveal her signal, future voters would ignore the message, instead believing that voter 2 must have been an R -partisan. Thus, once a herd begins, future voters are no longer “willing to listen.”

Example 2. Consider the same parameters from Example 1, except that $\tau_L = \tau_R = \tau = \frac{1}{3}$. As we assumed there, suppose voter 1 has voted L . If voter 2 is Neutral and receives an l signal, her posterior is $\mu_2(h^2, l) = \frac{42}{47} > \frac{1}{2}$ and hence she votes L . On the other hand, if she receives signal r , then $\mu_2(h^2, r) = \frac{14}{29} < \frac{1}{2}$, and therefore, she would vote R . In other words, a Neutral voter 2 votes informatively. Now consider the voter 3. If she observes a history of (L, R) , her posterior based on the public history is equal to her prior (the vote lead for L is 0), and therefore, she is essentially in the position of voter 1, and votes informatively. On the other hand, consider the history (L, L) . If a Neutral voter 3 were to receive an l signal, she would vote L because her posterior is $\mu_3((L, L), l) \approx 0.92$. If instead, voter 3 were to receive a signal for r , then $\mu_3((L, L), r) \approx 0.57$, and so she would still vote L . Thus, regardless of signal, voter 3's vote is uninformative, and by induction, it follows that all future Neutrals vote for Candidate L .

Since the only difference between Example 1 and Example 2 is that the latter has a higher probability of partisanship, the examples illustrate that an increase in partisanship leads to greater noise in the public history; consequently, a higher vote lead is required to induce a Neutral voter to discard her own information.

We now provide a general characterization of the dynamics induced when voters play PBV. The behavior of Partisans is simple, since they always vote for their preferred candidate. Neutrals shall initially vote informatively: in particular, voter 1 always votes informatively if she is Neutral. We will prove that there exist sequences $n_L(i)$ and $n_R(i)$ such that once the vote lead for candidate $C \in \{L, R\}$ after some history h^i is $n_C(i)$, all future Neutrals vote uninformatively (i.e. independent of their private signals) for candidate C , effectively herding on this candidate. It is important to emphasize that for voter j to herd in this manner does not require the vote lead for candidate C to be $n_C(j)$ when it is voter j 's turn to vote—it only requires that there have been some history h^i with $i \leq j$ where the vote lead after h^i for candidate $C \in \{L, R\}$ was

$n_C(i)$. A Neutral j votes informatively if and only for every history $\{h^i\}_{i \leq j}$, it was not the case that the vote lead for candidate $C \in \{L, R\}$ after history i was $n_C(i)$. To understand this logic, consider the first instance of a history h^i in which the vote lead for candidate $C \in \{L, R\}$ is in fact $n_C(i)$. In this case, voter i votes uninformatively for candidate C if Neutral, and uninformatively for her preferred candidate if Partisan. As far as voter $i + 1$ is then concerned, the informationally-relevant public history is only h^i , rather than h^{i+1} , since i 's vote contains no information about the state of the world. Inductively, this applies to all future voters, and we conclude that all learning ceases in the system after history h^i .

To make these notions precise, two state variables are useful. For any history h^i , the *vote lead*, $\Delta(h^i)$, is the vote lead for candidate L , and is defined recursively as follows:

$$\Delta(h^1) = 0, \text{ and for all } i > 1, \Delta(h^i) = \Delta(h^{i-1}) + (\mathbf{1}_{\{v_{i-1}=L\}} - \mathbf{1}_{\{v_{i-1}=R\}}) \quad (2)$$

Positive values of Δ indicate that candidate L has more votes than R , whereas negative values indicate that R leads L . The second state variable, called the *phase*, summarizes whether learning is still ongoing in the system (denoted phase 0), or herding has started on one of the candidates (denoted phase L or R). The phase mapping is thus $\Psi : h^i \rightarrow \{L, 0, R\}$, and defined by the following transition mapping:

$$\Psi(h^1) = 0, \text{ and for all } i > 1, \Psi(h^i) = \begin{cases} \Psi(h^{i-1}) & \text{if } \Psi(h^{i-1}) \in \{L, R\} \\ L & \text{if } \Psi(h^{i-1}) = 0 \text{ and } \Delta(h^i) = n_L(i) \\ R & \text{if } \Psi(h^{i-1}) = 0 \text{ and } \Delta(h^i) = -n_R(i) \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Clearly, the system starts in the learning phase, and transitions into a herding phase when the vote lead for candidate $C \in \{L, R\}$ hits the threshold $n_C(i)$ for the first time. Moreover, herding phases are absorbing.

The threshold sequences, $n_L(i)$ and $n_R(i)$, are determined by explicitly considering posteriors. For example, $n_L(i)$ is the smallest vote lead for candidate L such that if voter i were to receive a signal r , and assumes that all prior Neutrals voted informatively and all prior Partisans voted for their preferred candidates, her signal in favor of Candidate R would be strictly overturned by the history in favor of Candidate L . That is, the threshold $n_L(i)$ is the unique integer less than or equal to $(i - 1)$ that solves

$$\Pr(\omega = L | \Delta(h^i) = n_L(i) - 2, s_i = r) \leq \frac{1}{2} < \Pr(\omega = L | \Delta(h^i) = n_L(i), s_i = r) \quad (4)$$

If it is the case that a signal r outweighs even that history h^i where every prior vote was cast in favor of Candidate L , we denote $n_L(i) = i$.

Similarly, the threshold $n_R(i)$ is the unique¹¹ integer less than or equal to i that solves

$$\Pr(\omega = L | \Delta(h^i) = -n_R(i) + 2, s_i = l) \geq \frac{1}{2} > \Pr(\omega = L | \Delta(h^i) = -n_R(i), s_i = l) \quad (5)$$

If it is the case that a signal r outweighs even that history h^i where $\Delta(h^i) = -(i-1)$, we denote $n_R(i) = i$.

Since these thresholds play a crucial role in our analysis, it is worth emphasizing their interpretation. The threshold $n_L(i)$ means that if a Neutral voter i observes a history h^i with candidate L has a vote lead of $n_L(i) - 2$ votes, then i 's posterior is (strictly) higher on state L than R if and only if $s_i = l$. On the other hand, if we switch an R vote to an L vote in h^i , and obtain a vote lead that is greater by 2 votes, i.e. $n_L(i)$, then i 's posterior is strictly higher on state L regardless of her private signal. A similar intuition applies to $n_R(i)$. Using the thresholds $\{n_L(i)\}_{i=1}^n$ and $\{-n_R(i)\}_{i=1}^n$, we can define the phase mapping Ψ for every history.

We summarize the foregoing discussion with the following result. All formal proofs are in the Appendix.

Proposition 1. *Every game $G(\pi, \gamma, \tau_L, \tau_R; n)$ has a unique PBV strategy profile. For each $i \leq n$, there exist thresholds, $n_L(i) \leq i$ and $n_R(i) \leq i$, such that if voters play PBV in the game $G(\pi, \gamma, \tau_L, \tau_R; n)$, then a Neutral voter i votes*

1. *informatively if $\Psi(h^i) = 0$;*
2. *uninformatively for L if $\Psi(h^i) = L$;*
3. *uninformatively for R if $\Psi(h^i) = R$;*

where Ψ is as defined in (3). The thresholds $n_L(i)$ and $n_R(i)$ are independent of the population size n .

[Figure illustrating PBV dynamics to be added here.]

In general, when voters play the PBV profile, a herd develops if and only if there is a history h^i such that $\Psi(h^i) \neq 0$. To study how likely this is, assume without loss of generality that the true state is R . We define the *public likelihood ratio* as the ratio of the public belief that the state is L versus state R after any history: $\lambda(h^i) = \frac{\Pr(\omega=L|h^i)}{\Pr(\omega=R|h^i)}$. Fixing play according to PBV, the realized path of play is governed by the draw of preference-types and signals. Consequently, the public likelihood ratios can be viewed as a stochastic process, which we denote $\langle \lambda_i \rangle$, where each λ_i is the likelihood ratio when it is voter i 's turn to act. It is well-known (e.g. Smith and Sorensen [26]) that this stochastic process is a martingale conditional on the true state, R . By the Martingale Convergence Theorem, the process $\langle \lambda_i \rangle$ converges almost surely to a random

¹¹Implicit in our definitions of $n_L(i)$ and $n_R(i)$ is that these be vote leads that are logically possible given the various configurations of $(i-1)$ votes.

variable, λ_∞ .¹² That is, the public likelihood ratio must eventually converge or settle down. Since PBV is informative so long as $\Psi = 0$, this convergence requires that $\Psi \in \{L, R\}$ in the limit, i.e. herds eventually occur with probability 1. This intuition underlies the following result for our finite voter game.

Theorem 1. *For every $(\pi, \gamma, \tau_L, \tau_R)$ and for every $\varepsilon > 0$, there exists $\bar{n} < \infty$ such that for all $n > \bar{n}$, if voters play PBV, then $\Pr[a \text{ herd develops in } G(\pi, \gamma, \tau_L, \tau_R; n)] > 1 - \varepsilon$.*

3.2 PBV Equilibrium

PBV as a behavioral rule entails that a voter votes sincerely for who she believes to be better on the basis of the history she observes and her private signal. It is not predicated on the hypothesis that a voter understands the game strategically, and considers how she should vote in the event that her vote changes the outcome of the game. We believe that such a behavioral prescription is attractive from a descriptive standpoint: insofar as many voters may simply vote for the candidate they *actually* prefer on the basis of public and private information—as opposed to hypothetically *would* prefer if pivotal—PBV accommodates their behavior. However, the theory would be unattractive if any strategic voter has an incentive to deviate from PBV when all other voters are following it. As noted earlier, since strategic voters have forward-looking incentives in our environment, whereas PBV is explicitly only backward-looking, the strategic optimality of PBV is far from obvious.

To this end, we now establish that PBV is optimal behavior for a voter even when she does condition on being pivotal, given that all other voters are following PBV. Consequently, PBV is an equilibrium, and therefore accommodates both sincere and strategic voting. In fact, we prove an even stronger result. Say that the *election is undecided* at history h^i if both candidates still have a chance to win the election given the history h^i ; in other words, it must be true for each candidate that if he receives all future votes, then he wins the election. We will show that when all other voters use PBV, it is uniquely optimal for a voter to follow PBV at any history where the election is undecided (of course, if the election is already decided, a voter is indifferent over how she votes). Accordingly, we refer to PBV as an *essentially strict* equilibrium.¹³

To show that PBV is an essentially strict equilibrium, the following three behaviors need to be shown as uniquely optimal when all other voters are playing PBV, and the election remains

¹²The careful reader will note that we are discussing an infinite sequence, even though the game consists of a finite number of voters. For the discussion, consider simply extending the sample path of realized likelihood ratios one voter at a time by adding a voter with a new draw of a preference-type and signal (recall these i.i.d conditional on the true state) who plays PBV.

¹³A strict equilibrium of a game is one in which a deviation to any other strategy makes a player strictly worse off. A sequential voting game cannot possess any strict equilibria because following any history where a candidate has captured a majority of the votes, all actions yield identical payoffs; only histories where the election remains undecided are strategically relevant. Therefore, “essentially strict equilibrium” as we define it is the appropriate modification of the usual notion of strict equilibrium.

undecided:

1. Partisans always vote for their preferred candidate.
2. In the herding phase for L (R), Neutrals vote uninformatively for L (R).
3. In the learning phase, Neutrals vote informatively.

Throughout the subsequent discussion, we discuss optimality for one voter facing a history where the election is undecided, implicitly assuming that all others are playing PBV. Note that since PBV prescribes Partisans to vote for their preferred candidate independent of history or private signal, every voter who is voting at a history where the election is undecided is pivotal with positive probability.

It is straightforward to establish that it is strictly optimal for Partisans to vote for their preferred candidate, conditional on others following the PBV strategy profile. Since other Partisans are assumed to vote for her preferred candidate independent of history, no player can influence the behavior of any other Partisan voter. Thus, a Partisan voter is pivotal only to the extent that she influences the behavior of future Neutral voters, or in the event that all other votes excluding hers are tied. An important observation is that in a PBV profile, the play of Neutrals is weakly monotonic in the voting history: a larger vote lead for a particular candidate induces at least as many Neutrals (and on expectation, strictly more when in the learning phase) to vote for that candidate. Hence, a Partisan voter can not influence any future voter in a desirable direction by voting against her preferred candidate; on the other hand, by voting against her preferred candidate, she will be strictly worse off in those events when the votes of all other voters are tied. Since the election is undecided by hypothesis, these two facts combine to imply that it is strictly optimal for a Partisan to vote for her preferred candidate.¹⁴

Lemma 1. *If all other agents are playing PBV and the election is undecided at the current history, it is strictly optimal for a Partisan to play PBV.*

Now we discuss the optimal behavior for Neutrals. Let us first consider the case where a herd has already begun. Suppose that it is a Neutral voter i 's turn to vote when there is a herd on candidate L , i.e. the phase is $\Psi(h^i) = L$ (the logic when the herd is for R is symmetric). Since the L phase is absorbing, voter i knows that all subsequent voting after her is completely uninformative: Partisans vote for their preferred candidates and Neutrals vote for L . This implies that conditioning on being pivotal does not change i 's posterior beliefs about the state of the world. Consequently, Neutral voter i strictly prefers to vote on the

¹⁴Interestingly, we note that eliminating weakly-dominated strategies is not sufficient to guarantee that Partisans vote for their preferred candidate, unlike the case of a simultaneous election. In fact, there exist equilibria (in undominated strategies) where a Partisan voter strictly prefers to vote against her preferred candidate. While we do not suggest that such equilibria are plausible, they serve to emphasize the differences between simultaneous and sequential voting environments.

basis of her posterior, $\mu_i(h^i, s_i)$. Crucially, the public belief for voter i before her signal is observed is the same as that of voter j 's ($j \leq i$) when the phase transitioned from $\Psi(h^{j-1}) = 0$ to $\Psi(h^j) = L$, since the votes starting with j are uninformative about the state of the world. That is, $\mu_i(h^j, s_i) = \mu_j(h^j, s_i)$. By the definition of when the phase transitions (3), we have $\Delta(h^j) = n_L(j)$, which implies by the construction of $n_L(j)$ (4) that $\mu_j(h^j, s_i) > \frac{1}{2}$, for any $s_i \in \{l, r\}$. Thus, Neutral voter i strictly prefers to vote for L regardless of her signal.

Lemma 2. *If all other agents are playing PBV and the election is undecided at the current history, h^i , it is strictly optimal for a Neutral voter i to vote for candidate C if $\Psi(h^i) = C$, for all $C \in \{L, R\}$.*

It remains only to establish that it is optimal for a Neutral voter to vote informatively when votes can still communicate information, i.e. when in the learning phase, $\Psi = 0$. In an undecided history, strategic deviations in the learning phase could significantly influence the behavior of future Neutral voters. Thus, with two possible signals and many possible vote leads, establishing that it is optimal for Neutral voters to vote informatively requires showing that multiple incentive compatibility conditions are satisfied. Observe that while in the learning phase, the public history is not sufficiently strong so as to overturn the private signal received by a Neutral voter. That is, for all h^i where $\Psi(h^i) = 0$, voter i has posterior $\mu_i(h^i, s_i) \geq \frac{1}{2}$ if $s_i = l$ and $\mu_i(h^i, s_i) \leq \frac{1}{2}$ if $s_i = r$. Therefore, by voting informatively, a Neutral voter i is voting in the direction favored by her actual posterior. The cost of doing this is that she may be pushing future voters closer towards herding. This is costly to a Neutral insofar as herding suppresses potentially valuable information possessed by later Neutrals. Thus, it is conceivable that a Neutral may gain by voting *against* a leading candidate, *even if* her private signal (and her posterior) favors this candidate. Therefore, when receiving a signal in favor of the leading candidate, a Neutral voter is faced with the trade-off between voting informatively and inducing future Neutral voters to vote informatively. This trade-off is most stark for the Neutral voter i who faces $\Delta(h^i)$ equaling $n_L(i+1) - 1$ (or $-n_R(i+1) + 1$), since voting for candidate L (or R) immediately starts a herd, ending social learning altogether. Would it be better to suppress her own information so as to preserve the value of her successors' information?

We shall show that it is strictly optimal to vote informatively, sacrificing (if necessary) the value of her successors' information. The key idea underlying the strict optimality of this action for voter i is that if i votes against her current posterior or equivalently, against her signal,¹⁵ then there can never be enough public information—in the entire sequence of other voters' votes—that is sufficient to shift i 's belief in favor of the candidate she votes for. That is to say, given that other voters are playing according to the PBV strategy profile, and voter i does not know the types of those who vote after her, the value of her information ultimately outweighs that of any subsequent profile of votes that changes the outcome following her deviation. To

¹⁵i.e. if i votes L or R when $s_i = r$ or $s_i = l$ respectively

illustrate this logic, assume the symmetric partisanship case where $\tau_L = \tau_R$ (we establish the result for the asymmetric case in the Appendix). In this case, the thresholds $n_L(j)$ and $n_R(j)$ do not vary with voters, and can be denoted simply as constants $n_L > 0$ and $n_R > 0$. Consider the case where voter i is Neutral and $\Psi(h^i) = 0$ with $\Delta(h^i) = n_L - 1$ (the logic is symmetric when $\Delta(h^i) = -n_R + 1$). It is only when $s_i = l$ and accordingly supposed to vote L that the Neutral voter i faces the tradeoff between voting informatively thereby starting a herd, versus voting uninformatively but at least temporarily preventing a herd. The incentives can be examined by focussing on the events in which i 's vote is pivotal, since in any other event, voter i is indifferent between her choices.

Being pivotal in a sequential voting game introduces subtleties relative to a simultaneous voting game. Since by hypothesis all other voters are playing PBV, we can assess the impact of voter i 's vote for any vector of realizations of preference types and signals amongst the future voters. The set of type-signal realizations in which voter i is pivotal, denoted as Piv_i , consist of all those vectors where $v_i = L$ results in a different electoral winner from $v_i = R$. Furthermore, let us also denote those events where a cascade for candidate $C \in \{L, R\}$ forms as ξ^C ; and denote the events where no cascade forms as $\tilde{\xi}$. Since voter i is pivotal with positive probability (because the election is as yet undecided), and by hypothesis has received signal l , we can write:

$$\begin{aligned} \Pr(\omega = L | Piv_i, h^i, l) = & \Pr(\xi^L | Piv_i, h^i, l) \Pr(\omega = L | \xi^L, Piv_i, h^i, l) \\ & + \Pr(\xi^R | Piv_i, h^i, l) \Pr(\omega = L | \xi^R, Piv_i, h^i, l) \\ & + \Pr(\tilde{\xi} | Piv_i, h^i, l) \Pr(\omega = L | \tilde{\xi}, Piv_i, h^i, l) \end{aligned}$$

It needs to be shown that the above conditional probability strictly exceeds $\frac{1}{2}$. This means that conditional on her observed history, private signal, and being pivotal, voter i 's belief that the state is L is strictly higher than her belief that the state is R . Our argument proceeds by demonstrating that her posterior conditional on being pivotal and on each of the three events, ξ^L , ξ^R or $\tilde{\xi}$, is at least $\frac{1}{2}$ (with one strictly inequality). Therefore, regardless of how future voting unfolds, conditional on being pivotal and what she has observed, voter i believes that it is strictly optimal to vote L .

Consider the first term in the above summation. Since voting is uninformative once an L -cascade begins, $\Pr(\omega = L | \xi^L, Piv_i, h^i, l) = \Pr(\omega = L | \xi^L, h^i, l)$. Regardless of how voter i votes, this probability strictly exceeds $\frac{1}{2}$ since an L -cascade forms only when candidate L commands a lead of at least $n_L - 1$ votes from non- i voters in the learning phase, and moreover, $s_i = l$. By the definition of n_L , $\Pr(\omega = L | \xi^L, Piv_i, h^i, l) > \frac{1}{2}$.

Similarly, consider the event in which i is pivotal and an R -cascade begins. Since voting is uninformative once an R -cascade begins, $\Pr(\omega = L | \xi^R, Piv_i, h^i, l) = \Pr(\omega = L | \xi^R, h^i, l)$. Since $\Delta(h^i) = n_L - 1$, the event ξ^R can happen only if i votes R (if i votes L , an L -cascade forms immediately) and moreover, candidate R subsequently gains a lead of n_R votes in the learning

phase. This requires that after i 's votes, there are $n_L + n_R - 2$ additional votes cast for R in the learning phase. Given that L has an informative vote lead of $n_L - 1$ prior to i 's vote, conditioning on ξ^R in effect reveals a *net* (i.e. subtracting any informative L votes) total of $n_R - 1$ informative votes for R , ignoring i 's own vote. But by definition of n_R , i 's belief given her own signal $s_i = l$ and $n_R - 1$ informative votes for R is weakly in favor of L . In other words, $\Pr(\omega = L | \xi^R, Piv_i, h^i, l) \geq \frac{1}{2}$.

Finally, consider the event that no cascade occurs and i is pivotal. This can happen only if voter i votes R , the vote lead at the end of the election lies in $\{-n_R + 1, \dots, n_L - 1\}$, and $\Psi(h^{n+1}) = 0$. This implies that the *net* total of additional votes for R must be no greater than $n_L + n_R - 2$, and could be strictly less; otherwise, an R cascade would start. Since voter i faces a history where L commands a lead of $n_L - 1$ in the learning phase, the net total of informative votes for R is no greater than $n_R - 1$, and could be strictly less. By the same logic as earlier using the definition of n_R , it follows that $\Pr(\omega = L | \tilde{\xi}, Piv_i, h^i, l) > \frac{1}{2}$.

We have just argued that conditional on her observed history, signal, and being pivotal, voter i believes candidate L to be the better candidate with probability at least $\frac{1}{2}$ regardless of how she is pivotal, and moreover, in some pivotal events that happen with positive probability (viz., either an L -cascade or no cascade) this belief is strictly greater than $\frac{1}{2}$. Since voting L leads to a strictly higher probability of L winning the election, it is strictly optimal for voter i to vote L , even though such a choice marks the end of informative voting. Such an argument is generalized to the case where $\tau_L \neq \tau_R$ in the Appendix; in such a game, the thresholds shift in favor of the candidate less favored by the partisan bias. While the non-uniformity of thresholds introduces challenges, the basic logic of the argument remains unchanged.

The above argument applies only to a voter who could start a cascade by vote informatively. Because the tradeoff between voting informatively and inducing others to vote informatively is most stark at the boundaries of the learning phase, it can also be established that if the incentive conditions are met for histories h^i where $\Delta(h^i) \in \{n_L(i+1) - 1, -n_R(i+1) + 1\}$, then they are met for all other vote leads in the learning phase (Lemma 5 in the Appendix). In sum, we can state the following.

Lemma 3. *If all other agents are playing PBV and the election is undecided at the current history, h^i , it is strictly optimal for a Neutral voter i to vote informatively when $\Psi(h^i) = 0$ and $\Delta(h^i) \in \{-n_R(i+1) + 1, \dots, n_L(i+1) - 1\}$.*

The three preceding Lemmas combine to imply that PBV is strictly optimal for any voter who must vote at a history where the election is undecided, given that all other voters are playing PBV. This allows us to state the main result. Recall that an equilibrium is essentially strict if any strategy that yields a player the same expected utility as the equilibrium strategy (given that all others are playing their prescribed strategies) agrees with the equilibrium strategy at all possible histories for the player where the election is undecided.

Theorem 2. *The PBV strategy profile is an essentially strict equilibrium.*

There are three points we would like to emphasize about this equilibrium. First, its (essential) strictness implies that we are not relying on any convenient resolution of indifference when the election remains yet undecided to establish the equilibrium. Thus, strategic voters follow PBV not because they are powerless or indifferent to change the outcome, but instead because deviations yield strictly worse expected payoffs. Of course, this implies that the equilibrium is robust to small perturbations of the model. Second, because Partisan voters always vote for their preferred candidates in the PBV equilibrium, every information set is reached with positive probability. Therefore, off-the-equilibrium-path beliefs play no role in our analysis, and PBV is a Sequential Equilibrium (Kreps and Wilson [22]). Third, the robustness of PBV to any parameter configuration—in particular, in settings with asymmetric partisanship—is quite remarkable. The simultaneous strategic voting literature (e.g. Feddersen and Pesendorfer [13, 14]) has focussed on equilibria in large elections that require a large degree of strategic voting (by Neutrals) to offset asymmetric Partisan bias and effectively induce expected symmetry in equilibrium. In large voting games, such equilibria demand either complex coordination and/or complex mixing strategies. The relative simplicity of PBV works in its favor: because a greater partisan base makes it more difficult for a candidate to establish his quality to Neutrals, the thresholds adjust in ways that surprisingly make voting on the basis of one’s actual posterior strictly optimal.

4 Discussion

4.1 Other Equilibria

Theorems 1 and 2 show that rational behavior in a sequential voting game can produce herd behavior through PBV, where each voter votes for who she actually prefers based on her observed history and private signal. In this section, we examine the issue of other equilibria besides PBV. Given the size of the strategy space in a sequential voting game, we cannot completely characterize all equilibria; instead, we proceed in two steps. First, we discuss equilibria within the class of *Cut Point Voting* (CPV) strategy profiles introduced by Callander [9]. While this class of strategy profiles entails pure strategy and certain monotonicity and symmetry restrictions, it permits a wide range of interactive behavior. We prove that any CPV equilibrium leads to herding with large probability in large elections. Thereafter, we further broaden the discussion to non-CPV profiles and consider the possibility of efficient equilibria (which cannot involve herding). We argue that although such equilibria exist and are appealing in terms of their efficiency, their plausibility is dubious on grounds of complexity.

The class of CPV strategy profiles is a generalization of PBV. CPV behavior for a Neutral is determined by a combination of the private signal and the weight of the public history,

subsuming PBV as a special case. To define a CPV profile, let $\mu(h^i) \equiv \Pr(\omega = L|h^i)$, so that $\mu(h^i)$ denotes the public belief following history h^i .

Definition 2. A strategy profile, \mathbf{v} , is a Cut Point Voting (CPV) strategy profile if there exists $0 \leq \mu_* \leq \mu^* \leq 1$ such that for every voter i , history h^i , and signal s_i ,

$$v_i(N, h^i, s_i) = \begin{cases} L & \text{if } \mu(h^i) > \mu^* \text{ or } \{\mu(h^i) \geq \mu_* \text{ and } s_i = l\} \\ R & \text{if } \mu(h^i) < \mu_* \text{ or } \{\mu(h^i) \leq \mu^* \text{ and } s_i = r\} \end{cases}$$

That is, in a CPV strategy profile, Neutrals vote according to their signals alone if and only if the public belief when it is their turn to vote lies within $[\mu_*, \mu^*]$. If the public belief is outside this range, a Neutral votes for one of the candidates independent of her private signal. Note that CPV places no restriction on Partisan behavior.¹⁶ Denote a CPV profile with belief thresholds μ_* and μ^* as $CPV(\mu_*, \mu^*)$. The belief thresholds define the extent to which a CPV profile weighs past history relative to the private signal. $CPV(0, 1)$ corresponds to informative voting (by Neutrals) where history never influences play, whereas $CPV(1 - \gamma, \gamma)$ corresponds to Posterior-based Voting by Neutrals. Similarly, $CPV(0, 0)$ and $CPV(1, 1)$ represent strategy profiles where every Neutral votes uninformatively for candidate L and R respectively. Therefore, CPV captures a variety of behavior for Neutrals, including PBV and Informative Voting as special cases.

Say that an equilibrium is a CPV equilibrium if the equilibrium strategy profile is a CPV profile. While we are unable to derive a tight characterization of what non-PBV but CPV profiles—if any—constitute equilibria, we can nevertheless extend the conclusion of Theorem 1 to any CPV equilibria.

Theorem 3. For every $(\pi, \gamma, \tau_L, \tau_R)$ such that $\tau_L \neq \tau_R$, and for every $\varepsilon > 0$, there exists $\bar{n} < \infty$ such that for all $n > \bar{n}$, if voters play a CPV equilibrium, $\Pr[a \text{ herd develops in } G(\pi, \gamma, \tau_L, \tau_R; n)] > 1 - \varepsilon$.

Therefore, for generic parameter configurations, large elections lead to herds with arbitrarily high probability within the class of CPV equilibria. The main step of the argument for Theorem 3 shows that when $\tau_L \neq \tau_R$, any CPV equilibrium in a large election must have belief thresholds μ_* and μ^* that are bounded away from 0 and 1 respectively. This implies the Theorem because any such “interior” CPV must result in herds with arbitrarily high probability in sufficiently large elections for the same reason as outlined in the discussion preceding Theorem 1.

In the knife-edge case of symmetric partisanship, $\tau_L = \tau_R$, informative voting by Neutrals ($CPV(0, 1)$) coupled with Partisans voting for their preferred candidates is an equilibrium for any population size. In fact, it can be shown that in this symmetric case, this equilibrium is the only CPV equilibrium besides PBV. However, existence of this equilibrium is extremely

¹⁶It is straightforward to show, however, that if Neutrals are playing CPV, then it is uniquely optimal for a Partisan to vote for her preferred candidate at any undecided history (Lemma 8 in the Appendix).

non-generic: for any, even arbitrarily small, difference in partisanship, this profile is no longer an equilibrium in all sufficiently large voting populations (Lemma 9 in the Appendix). This stands in stark contrast to PBV which is an equilibrium independent of the parameter configuration. Moreover, not only are the parameters which support the informative voting equilibrium non-generic, they are not particularly compelling economically, since there is no reason to generally expect two candidates in an election to possess the same level of partisan support.

To understand the full import of Theorem 3, it is useful to consider exactly what restrictions CPV places with respect to arbitrary strategy profiles. There are five substantive restrictions. First, every CPV profile is in pure strategies. Second, for any voter i , if two distinct histories induce the same public belief when it is i 's turn to vote, then i 's behavior when Neutral (as a function of his private signal) is identical following these two histories. Third, CPV profiles are weakly monotonic in private signals: for any history, if a Neutral voter i were to vote for candidate L (R) when receiving signal r (l), then she must also vote for candidate L (R) after receiving signal l (r). Fourth, CPV profiles are weakly monotonic in the public belief: an increase in the public belief in favor of one candidate must not induce any Neutral voter to switch her vote from that candidate to the other. Fifth, CPV strategy profiles are symmetric across voters: two Neutral voters vote identically if they receive the same private signal and face histories with the same public belief. Formally, we show in Proposition XXX in the Appendix that any strategy profile that satisfies these five restrictions is a CPV profile.

We view restricting attention to equilibria that satisfy CPV's requirements of pure strategies, and weak monotonicity in signals and public beliefs as fairly compelling on the basis of how one would expect a voter to behave in a large election. We are also comfortable with the requirement that a voter should behave identically in any two histories that induce the same public belief. It is less clear whether the requirement of symmetry across voters with respect to public beliefs is particularly reasonable (or unreasonable) in an asymmetric setting like sequential voting. Dropping this last restriction would lead to a more general class of profiles, *asymmetric CPV*, where the belief thresholds can be voter specific. Whether all such equilibria must generically result in herding in large elections is an open question.

The above discussion and CPV in particular is motivated by focussing on equilibria on the basis of behavioral plausibility; an alternative is to focus on efficiency. Given the presence of Partisan types, *ex interim* Pareto efficiency has little bite in our setting unlike in pure common value elections. Instead, we consider informational efficiency. The benchmark introduced by Feddersen and Pesendorfer [13, 14], is that of *Full Information Equivalence* (FIE): FIE is attained by a sequence of equilibria as the population size $n \rightarrow \infty$ if the outcome of the election for each n is identical to that which would have been chosen were all information public with probability converging to 1. In such a case, in spite of information being held privately by voters, the first best is asymptotically attainable.

The logic of Feddersen and Pesendorfer [13, 14] and Dekel and Piccione [12] can be combined

to show that FIE is attainable in our sequential voting model. The argument runs as follows. For any parameter set $(\pi, \gamma, \tau_L, \tau_R)$, minor modifications of Feddersen and Pesendorfer [13, 14] imply that the simultaneous voting variant of our model can achieve FIE through a sequence of symmetric equilibria where Partisans vote their bias, and Neutrals play strategies that put positive probability on voting both L and R . For any n , consider a strategy profile in the sequential voting game where every voter at every history plays the same (history-independent) strategy as she would in the above construction. Since all voters are acting independently of history, the events in which a voter is pivotal is identical in both the simultaneous and sequential games; therefore, the strategy profile is an equilibrium of the sequential game, and FIE is attainable.

For this reason, Dekel and Piccione [12] have argued that there is no reason to believe that a sequential voting game will necessarily be less efficient than its simultaneous analog. However, this is true only to the extent that we believe that these efficient equilibria are plausible; if instead, the structure of sequential voting makes certain inefficient equilibria more salient (in terms of behavioral plausibility), then there is reason to expect differences between sequential and simultaneous voting. We provide two reasons below to question the plausibility of the symmetric mixed equilibrium that approximates informational efficiency in large elections.

If partisanship is asymmetric—as is generically true—and the population is sufficiently large, the symmetric equilibrium that approximates efficiency features precise randomization by Neutrals when receiving a signal in favor of the candidate with greater partisan support. For any Neutral voter to be indifferent at these information sets requires a precise level of uncertainty over the actions of others; it is difficult to see how such uncertainty could be generated and focal. In contrast, the simplicity of a pure strategy profile such as PBV where it is believed that all other voters are simply voting sincerely is attractive.¹⁷

Moreover, the symmetric equilibrium that approximates efficiency hinges on it being commonly known that all voting behavior is unaffected by the history of votes. Were a Neutral voter to admit the possibility that her vote would influence that of future voters, her incentives to play according to the symmetric equilibrium may dissipate altogether.¹⁸ In practice, it seems generally accepted that the vote lead influences voting behavior. Indeed, Bartels [5] and Popkin [24] argue that voters keep careful track of how candidates have performed relative to expectations when deliberating how to vote. This suggests it may be counterfactual to focus on an equilibrium where it is commonly believed that history does not influence behavior.

¹⁷At this stage, we do not know whether there is an efficient asymmetric equilibrium in pure strategies in either the simultaneous or sequential voting versions of our model. Should one exist, it would be subject to our second critique, but not that of randomization.

¹⁸A voter might admit that the behavior of a future voter is influenced by the public history for various reasons. First, she may believe that the future voter is strategically unsophisticated, and would vote sincerely. Second, there may be higher order uncertainty: if the future voter herself presumes than an even later voter votes on the basis of history, then it may be strategically optimal for the future voter to vote on the basis of history, and so forth.

4.2 Pure Common Value Elections

Most of the prior papers on sequential voting games have considered pure common value environments where every voter is Neutral (e.g. Fey [16], Witt [27], and Callander [9]). While we believe that partisanship is a genuine aspect of political economy, it is instructive to compare and contrast our results with this earlier research.

Fix a set of (π, γ, n) , i.e. a prior, signal precision, and population size. Consider a sequence of our sequential voting games where the probability of partisanship, (τ_L, τ_R) , is vanishing. It can be checked that for all sufficiently small (τ_L, τ_R) , the PBV equilibrium involves an L -herd beginning once $\Delta(h^i) \geq 1$, and an R -herd beginning once $\Delta(h^i) \leq -2$. Moreover, the limit game with no partisanship ($\tau_L = \tau_R = 0$) also has a PBV equilibrium with these thresholds.

Proposition 2. *Assume $\tau_L = \tau_R = 0$. Then there is an equilibrium where all voters use PBV. In this equilibrium, a transition to the herding phase begins once $\Delta(h^i) \in \{1, -2\}$.*

Therefore, the PBV equilibrium correspondence is upper hemi-continuous at $\tau_L = \tau_R = 0$. However, there is an important difference between the PBV equilibrium without partisanship and the PBV equilibria in games with strictly positive probability of partisanship. When τ_L and τ_R are arbitrarily small but strictly positive, Partisans vote for their preferred candidates regardless of history in the PBV equilibrium; thus, every possible voting history can occur on the equilibrium path. In particular, even after a cascade begins, there is positive probability that any future vote may go “against the herd” — this happens if a future voter is Partisan for the requisite candidate. On the other hand, in the PBV equilibrium when $\tau_L = \tau_R = 0$, once a cascade begins, every vote thereafter is for the leading candidate — this is because all voters are necessarily Neutral. Consequently, once a cascade has started, a vote for the losing candidate is an off-the-equilibrium-path action. This implies that some of the beliefs that sustain the equilibrium are necessarily off-the-equilibrium-path beliefs, and a theory of “reasonable” beliefs now becomes necessary: if voters see a vote going against a herd, how should they interpret it, and given their interpretation, would voters still wish to herd?

A natural place to begin understanding this game would be to investigate the implications of imposing standard beliefs-based refinements for signaling games such as the *Intuitive Criterion*, *D1* (Cho and Kreps [11]), or *Divinity* (Banks and Sobel [3]). However, none of these refinements have bite in the pure common values sequential voting game where $\tau_L = \tau_R = 0$. To see why, consider the even stronger refinement criterion of *Never a Weak Best Response* (Kohlberg and Mertens [21]). Informally, if future voters interpret a deviation from a herd as being equally likely to come from a voter with signal $s_i = l$ as from a voter with signal $s_i = r$, then future voters should not update their beliefs at all based on i 's vote, and hence it is a weak best response for voter i to deviate from the herd regardless of her signal. Given this, the belief that a deviation is equally likely to come from either signal-type of voter i survives Never a

Weak Best Response, which is the strongest of standard dominance-based belief refinements.¹⁹

In contrast, the aforementioned papers that consider common value sequential voting (Fey [16], Wit [27], and Callander [9]) impose the following belief restriction: if a voter i votes for R once an L -cascade has begun, it must be believed that $s_i = r$; similarly, if i votes for L once an R -cascade has begun, it must be believed that $s_i = l$. Such a belief restriction implies that it is always possible for a voter to reveal a contrarian signal once in a herd; accordingly, we label this belief restriction as *Perpetual Revelation*. Under this condition, Fey [16] and Wit [27] show that because of the signaling motive inherent in common value sequential voting, at least one voter with a signal that opposes the herd would always wish to deviate out of the herd and reveal her signal to future voters. That is, Perpetual Revelation is sufficient to halt momentum by inducing anti-herding for at least one voter. It is for this reason that Callander [9] augments the model of Fey [16] and Witt [27] by introducing a warm-glow effect for voters that vote for the winning candidate. This warm-glow effect, however small, is sufficient to overpower the anti-herding nature of perpetual revelation so long as there are infinitely many voters.

The limit of PBV equilibria as $\tau_L, \tau_R \rightarrow 0$ violates Perpetual Revelation. This is clearly seen by observing that in the PBV equilibrium for $\tau_L, \tau_R > 0$, any contrarian vote once a cascade has begun must (correctly) be attributed to a Partisan, and hence reveals no information about the voter’s private signal. Since this is true for any $\tau_L, \tau_R > 0$, it also holds in the limit PBV equilibrium for $\tau_L = \tau_R = 0$. Perpetual Revelation precludes such beliefs, but we do not find this violation troubling because as discussed, Perpetual Revelation is not implied by any standard belief refinement, and to this extent must be considered ad-hoc. We stress that there is nothing within the pure common values model that justifies imposing Perpetual Revelation.

In contrast, taking the limit as $\tau_L, \tau_R \rightarrow 0$ of our model justifies the notion that when that confronted with an “unexpected” vote in the pure common values setting, subsequent voters treat it as uninformative about the state of the world. Since partisanship is widely thought to be a genuine aspect of political economy, it seems reasonable that voters should include partisan preferences within the set of possibilities they consider when seeing an unexpected vote. Whether justifications for other beliefs such as Perpetual Revelation can be reasonably formalized, and how these interact with the possibility of partisanship is a topic for future research.²⁰

¹⁹It may be useful to think of an analogy with a cheap talk situation, since once a herd has begun, it is effectively costless (when $\tau_L = \tau_R = 0$) for a voter to vote either way, since all future voters are Neutrals. Therefore, a vote has no intrinsic cost for a voter, and its “meaning” or informative content is purely endogenous. For the same reason as “babbling” cannot be eliminated by standard refinements in cheap talk games, so too is the herding outcome not refined away here.

²⁰We note that one informal justification for Perpetual Revelation may stem from the notion that some voters are *Gurus*: their information is sufficiently superior to the rest of the population. This is in the spirit of the unbounded signals analyzed by Smith and Sorensen [26] in a non-voting sequential decisions setting. However, perhaps counterintuitively, simply allowing for Gurus in the pure common-values model will not by itself eliminate the herding equilibrium, unlike in Smith and Sorensen [26]. The reason goes back to the point that the value of a deviation in the voting context depends on the inferences it induces in subsequent voters. Hence if subsequent

5 Related Literature

This paper lies at the intersection of the literature on rational herding and the information aggregation approach to elections. Economic models of herding or information cascades originating from Bikchandani et al. [7] and Banerjee [2] typically consider sequential decision-making environments where each player’s payoff depends only on her own action and an unknown state. However, each agent possesses an informative signal about the state. Consequently, there is a *pure information externality* in the sense that no agent cares about the action of any other agent aside from extracting the information it may reveal about the state.²¹ In contrast, strategic models of voting starting with Austen-Smith and Banks [1] and Feddersen and Pesendorfer [13] possess *action externalities* in the sense that a player’s payoff depends on the profile of actions chosen by all players. To date however, most of the extant papers in this area focus on simultaneous voting; exceptions are discussed below. The model in this paper combines sequential decision-making with both information and action externalities.

We have already touched upon the important contribution of Dekel and Piccione [12]. They show that in symmetric environments, every symmetric (and responsive) equilibrium of a simultaneous election remains an equilibrium when the election is made sequential. While their model does not literally apply to our setting because of the presence of Partisans,²² the conclusion that existence of an efficient symmetric equilibrium in the simultaneous implies the same in the sequential game holds. Our contribution is to derive an equilibrium—Posterior-based Voting—that has no counterpart in simultaneous voting; moreover, we have argued that this is a particularly salient equilibrium in the sequential game because it embodies sincere voting throughout.

The only paper we are aware of that has demonstrated rich equilibrium momentum effects in a common value election with strategic voting is Callander [9] (cf. Fey [16] and Witt [27]). Two crucial assumptions he makes are that all voters derive a warm-glow benefit from voting for the eventual winner, and there are an infinite number of voters. Since we make neither of these assumptions, our analysis is distinct. While his framework permits a number of interesting comparisons between sequential and simultaneous elections (Callander [10]), the assumption that voters effectively have a taste for conformity renders the herding result less appealing from a theoretical perspective. Moreover, most of the herding equilibria he discusses do not feature sincere voting, unlike our PBV.

In a recent paper, Battaglini [6] shows that if there are costs to voting and voters can voters would ignore a deviation, then the deviation has no value, regardless of how strong a voter’s private information may be—this is a key contrast with the case of “economic” decision-making. That said, interacting the possibility of Gurus with partisanship may hold interesting insights.

²¹For this reason, though momentum effects in Presidential Primaries was cited as an important example of information cascades by Bikchandani et al. [7], such a conclusion was undoubtedly premature, since their analysis did not account for inherent action interdependencies in voting.

²²Formally, Axiom 2 of Dekel and Piccione [12] would fail.

abstain, the unique symmetric equilibrium of a simultaneous election does strictly better than any equilibrium of a sequential election.^{23,24} While this result is certainly interesting, the issue is orthogonal to momentum, as Battaglini [6] himself asserts. For a large class of voting costs, it would generally be the case that so long as the population size is sufficiently large, many early voters would optimally choose to abstain and free-ride of later voters. Thus, applied to Primaries, this implies that the initial rounds should be largely uninformative and momentum should begin in an *interim* round. To the contrary, it seems that empirically, early rounds of the Primaries are informative and the most fertile for bandwagon effects (Bartels [5]).

Abstracting away from questions of strategic voting and information aggregation, Klumpp and Polborn [20] have studied momentum in a model where two candidates lobby to win Primaries by allocating a fixed amount of campaign funds across voters/states, and the winner is selected randomly according to Tullock’s success function. When Primaries are held sequentially, such an environment engenders momentum: the winner of an earlier primary has stronger incentives to lobby for later victories. Since their model focuses attention on the behavior of candidates rather than voters, our approach is complementary to theirs. In particular, their approach cannot explain the significance of surprise victories (or the non-significance of expected victories), whereas such dynamics emerge from the informational perspective offered here.

6 Conclusion

This paper has proposed an informational theory of cascades in sequential voting environments. We developed a model of a binary election where a proportion of the voters seek to elect the better candidate, and the remainder have partisan preferences. Our main result is there exists an (essentially) strict equilibrium which leads to herding with high probability in large elections. This equilibrium features Posterior-based Voting, where each Neutral voters votes for whichever candidate she thinks is better given all currently available information at the time of casting her vote, whereas Partisans vote for their preferred candidate. In this sense, all voters play sincerely, and this is fully rational behavior given the strategies of all other voters. The induced dynamics in this equilibrium captures notions of candidate momentum in sequential voting, including the ebbs, flows, and performance relative to expectations in a dynamic election.

²³A rough intuition for this is that in the symmetric (mixed) equilibrium of the simultaneous election, it is unknown who else will vote and in which way, and this provides some incentives for voters to vote. In the sequential election, on the other hand, voters may choose to abstain once a particular candidate accumulates a sufficiently large vote lead margin.

²⁴We should note that in contrast, Gershkov and Szentes [18] show that from a mechanism design perspective, a voting mechanism that proceeds sequentially does better than simultaneous voting when information acquisition is costly. This is, however, a mechanism design result, and thus inherently different from both our focus and that of the other papers under discussion. In particular, the sequential mechanism of Gershkov and Szentes [18] is not in fact sequential voting as the term is generally used in the voting literature, because in their mechanism voters do not know their position in the sequence, nor observe other votes.

The results in our paper raise various issues that deserve further study. We shall highlight these below.

The discussion in Section 4 concerning full information equivalence may suggest that in terms of aggregating information, simultaneous voting dominates sequential voting in large elections. There are two caveats to note. First, while it is true that simultaneous voting can achieve full informational equivalence asymptotically, this requires voters to coordinate on the “right” equilibria. As is well known, simultaneous voting games possess many equilibria, including those where no information is aggregated at all. It may be that every equilibrium of the sequential voting game aggregates some information, and hence at least in a “worst case analysis”, sequential voting can dominate simultaneous voting. Since we do not have a characterization of all equilibria in our sequential voting game, this is an open question.

More importantly, real world sequential voting mechanisms may feature many benefits over simultaneous counterparts that are outside our model. For example, when thinking about Presidential Primaries, one natural point of departure is that candidates face greater constraints in campaigning across states that hold primaries on the same day than across states whose primaries are sequenced. Formally, this may be modeled as a constraint on the informativeness of signals that are obtained by voters who vote simultaneously. If sequencing can increase the informativeness of signals, this may provide a rationale for greater information aggregation in sequential voting mechanisms.

We have focussed our general analysis on the case of elections with only two candidates. Given the nature of the winnowing process in the U.S. Presidential Primaries, it is important to understand the dynamics of sequential voting with more than two candidates. A particularly intriguing possibility is that an appropriately generalized form of Posterior-based Voting may be an equilibrium of our model when there are three or more candidates. We plan to investigate this in future work.

This paper has abstracted away from the role of institutions, and concentrated on voters as being the sole players. Certainly, in practice, there are other forces involved in dynamic elections, many of which are strategic in nature, such as the media, campaign finance contributors, and so forth. By examining the potential for sequential voting alone to create herding, we have a benchmark to understand the role of different institutions in electoral momentum. Our model can thereby serve as a building block to assess the impact of these institutional forces.

Finally, we note that our analysis may have implications for social learning with payoff and information externalities more generally outside the confines of voting. Many environments of economic interest—such as dynamic coordination games, timing of investments, or network choice—feature both decision-making, private information that has social value, and payoff interdependencies. Our results suggest that information cascades may survive in various environments despite incentives to signal one’s information to successors.

7 Appendix

7.1 Proofs for Section 3.1

We begin with preliminaries that formally construct the thresholds $n_L(i)$ and $n_R(i)$ for each constellation of parameters $(\pi, \gamma, \tau_L, \tau_R)$, and each index i . Define the functions

$$f(\tau_L, \tau_R) \equiv \frac{\tau_L + (1 - \tau_L - \tau_R)\gamma}{\tau_L + (1 - \tau_L - \tau_R)(1 - \gamma)}$$

where the domain is $\tau_L, \tau_R \in [0, \frac{1}{2})$. It is straightforward to verify that f is continuous and strictly exceeds 1 over its domain.

For a fixed $(\pi, \gamma, \tau_L, \tau_R)$, define $\{n_L(i)\}_{i=1}^\infty$ as follows. For each positive integer i and any integer k where $|k| < i$ and $i - k$ is odd, define the function $g_i(k) = (f(\tau_L, \tau_R))^k \left(\frac{f(\tau_L, \tau_R)}{f(\tau_R, \tau_L)} \right)^{\frac{i-k-1}{2}}$. Observe that $g_i(k)$ is strictly increasing in k . For all i such that $g_i(i-1) \leq \frac{(1-\pi)\gamma}{\pi(1-\gamma)}$, set $n_L(i) = i$. If $g_i(i-1) > \frac{(1-\pi)\gamma}{\pi(1-\gamma)}$, we shall set $n_L(i)$ to be the unique integer that solves:

$$g_i(n_L(i) - 2) \leq \frac{(1-\pi)\gamma}{\pi(1-\gamma)} < g_i(n_L(i)) \quad (6)$$

Given that $g_i(-(i-1))$ is strictly less than $\frac{(1-\pi)\gamma}{\pi(1-\gamma)}$, and the strict monotonicity of $g_i(k)$ in k , a unique solution exists to the above equation.

Similarly, we define $\{n_R(i)\}_{i=1}^\infty$ as follows. For all i such that $g_i(-(i-1)) \geq \frac{(1-\pi)(1-\gamma)}{\pi\gamma}$, set $n_R(i) = i$. If $g_i(-(i-1)) < \frac{(1-\pi)(1-\gamma)}{\pi\gamma}$, set $n_R(i)$ to be the unique integer that solves:

$$g_i(-n_R(i) + 2) \geq \frac{(1-\pi)(1-\gamma)}{\pi\gamma} > g_i(-n_R(i)) \quad (7)$$

As before, given the monotonicity of $g_i(k)$ in k , and that $g_i(i-1) = (f(\tau_L, \tau_R))^{i-1} \geq 1 > \frac{(1-\pi)(1-\gamma)}{\pi\gamma}$, a unique solution exists to the above equation.

We use these values of n_L and n_R to define $\Psi(\cdot)$ as in equation (3) from the text, and now use this to prove Proposition 1.

Proposition 1 on pp. 12

Proof. The claim is obviously true for Voter 1 as $\Psi(h^1) = 0 \in (-n_R(1), n_L(1))$, and by construction, a PBV strategy involves a Neutral Voter 1 to vote informatively. To proceed by induction, assume that the claim about behavior is true for all Neutral voter $j < i$.

Case 1: $\Psi(h^i) = 0$: All preceding neutrals have voted informatively. It is straightforward to see that the the posterior $\mu(h^i, s_i) = \mu(\tilde{h}^i, s_i)$ if $\Delta(h^i) = \Delta(\tilde{h}^i)$ and $\Psi(h^i) = \Psi(\tilde{h}^i) = 0$ (i.e., so long as all preceding neutrals have voted informatively, only vote lead matters, and not the actual sequence). As such, we can define $\tilde{\mu}_i(\Delta, s_i) = \mu(h^i, s_i)$ where $\Delta = \Delta(h^i)$. Note

that

$$\tilde{\mu}_i(\Delta, l) = \frac{\pi\gamma g_i(\Delta)}{\pi\gamma g_i(\Delta) + (1-\pi)(1-\gamma)}$$

Simple manipulation shows that $\tilde{\mu}_i(\Delta, l) \geq \frac{1}{2} \Leftrightarrow g_i(\Delta) \geq \frac{(1-\pi)(1-\gamma)}{\pi\gamma}$. This latter inequality holds since by hypothesis, $\Psi(h^i) = 0$, and therefore, $\Delta \geq -n_R(i) + 1$. If $\tilde{\mu}_i(\Delta, l) > \frac{1}{2}$, then Condition 1 of the PBV definition requires that Neutral voter i vote L given $s_i = l$; if $\tilde{\mu}_i(\Delta, l) = \frac{1}{2}$, then Condition 2 of the PBV definition requires that Neutral voter i vote L given $s_i = l$.

Similarly, using Bayes' rule,

$$\tilde{\mu}_i(\Delta, r) = \frac{\pi(1-\gamma)g_i(\Delta)}{\pi(1-\gamma)g_i(\Delta) + (1-\pi)\gamma}$$

Simple manipulation shows that $\tilde{\mu}_i(\Delta, r) \leq \frac{1}{2} \Leftrightarrow g_i(\Delta) \leq \frac{(1-\pi)\gamma}{\pi(1-\gamma)}$. The latter inequality holds since by hypothesis, $\Psi(h^i) = 0$, and therefore, $\Delta \leq n_L - 1$. If $\tilde{\mu}_i(\Delta, r) < \frac{1}{2}$, then Condition 1 of the PBV definition requires that Neutral voter i vote R given $s_i = r$; if $\tilde{\mu}_i(\Delta, r)$ then Condition 2 of the PBV definition requires that Neutral voter i vote R given $s_i = r$.

Case 2: $\Psi(h^i) = L$. Then all Neutrals who voted prior to the first time Ψ took on the value L voted informatively, whereas no voter voted informatively thereafter. Let $j \leq i$ be such that $\Psi(h^j) = L$ and $\Psi(h^{j-1}) = 0$; therefore, $\Delta(h^j) = n_L(j)$. Then, $\mu(h^j, s_j) = \tilde{\mu}_j(n_L(j), s_j)$. Since all voting after that of $(j-1)$ is uninformative, $\mu(h^i, s_i) = \mu(h^j, s_i) = \tilde{\mu}_j(n_L(j), s_i)$. A simple variant of the argument in Case 1 implies that $\tilde{\mu}_j(n_L(j), l) > \frac{1}{2}$, and therefore Condition 1 of the PBV definition requires that Neutral voter i vote L given $s_i = l$. Consider now $s_i = r$. Since $g_j(n_L(j)) > \frac{(1-\pi)\gamma}{\pi(1-\gamma)}$, it follows that $\tilde{\mu}_j(n_L(j), r) > \frac{1}{2}$, and therefore Condition 1 of the PBV definition requires that Neutral voter i vote L even following $s_i = r$.

Case 3: $\Psi(h^i) = R$. Then all Neutrals who voted prior to the first time Ψ took on the value R voted informatively, whereas no voter voted informatively thereafter. Let $j \leq i$ be such that $\Psi(h^j) = R$ and $\Psi(h^{j-1}) = 0$; therefore, $\Delta(h^j) = -n_R(j)$. Then, $\mu(h^j, s_j) = \tilde{\mu}_j(-n_R(j), s_j)$. Since all voting after that of $(j-1)$ is uninformative, $\mu(h^i, s_i) = \mu(h^j, s_i) = \tilde{\mu}_j(-n_R(j), s_i)$. A simple variant of the argument in Case 1 implies that $\tilde{\mu}_j(-n_R(j), r) < \frac{1}{2}$, and therefore Condition (1) of the PBV definition requires that Neutral voter i vote R given $s_i = r$. Consider now $s_i = l$. Since $g_j(-n_R(j)) < \frac{(1-\pi)(1-\gamma)}{\pi\gamma}$, it follows that $\tilde{\mu}_j(-n_R(j), l) < \frac{1}{2}$, and therefore Condition 1 of the PBV definition requires that Neutral voter i vote R even following $s_i = l$. \square

Remark 1. As promised in the text, we argue here that the tie-breaking Condition (2) of the PBV definition only matters for a non-generic set of parameters $(\pi, \gamma, \tau_L, \tau_R)$. Observe that from the proof of Proposition 1, the posterior of voter i having observed a history h^i and private signal s_i is $\frac{1}{2}$ if and only if $\Psi(h^i) = 0$ and $g_i(\Delta(h^i)) \in \left\{ \frac{(1-\pi)(1-\gamma)}{\pi\gamma}, \frac{(1-\pi)\gamma}{\pi(1-\gamma)} \right\}$. For any particular (π, γ, τ_L) , this occurs for at most a countable collection of τ_R . Therefore, for a given (π, γ) ,

the set $\Gamma_{\pi,\gamma} \equiv \left\{ (\tau_L, \tau_R) \in \left(0, \frac{1}{2}\right)^2 : g_i(\Delta) \in \left\{ \frac{(1-\pi)(1-\gamma)}{\pi\gamma}, \frac{(1-\pi)\gamma}{\pi(1-\gamma)} \right\} \text{ for some positive integer } i \text{ and } \Delta \text{ where } |\Delta| \leq 1 \right\}$ is isomorphic to a 1-dimensional set. Thus, the need for tie-breaking arises only for a (Lebesgue) measure zero set of parameters (π, γ, τ) .

Theorem 1 on pp. 13

Proof. The proof consists of two steps: first, we show that there must almost surely be a herd in the limit as the population size $n \rightarrow \infty$; second, we show that this implies the finite population statement of the Theorem. Assume without loss of generality that the true state is R . (If the true state is L , one proceeds identically, but using the the inverse of the likelihood ratio λ_i).

Step 1: As discussed in the text, by the Martingale Convergence Theorem for non-negative random variables (Billingsley [8, pp. 468–469]), $\lambda_i \xrightarrow{a.s.} \lambda_\infty$ with $Support(\lambda_\infty) \subseteq [0, \infty)$. Define $\bar{\Lambda} \equiv [0, \underline{b}] \cup [\bar{b}, \infty)$ and $\Lambda = [0, \underline{b}) \cup (\bar{b}, \infty)$, where \underline{b} (resp. \bar{b}) is the likelihood ratio such that the associated public belief that the state is L causes the the posterior upon observing signal l (resp. r) to be exactly $\frac{1}{2}$. Note that by their definitions, $\underline{b} < \frac{1}{2} < \bar{b}$. To prove that there must almost surely be a herd in the limit, it needs to be shown that eventually $\langle \lambda_i \rangle \in \Lambda$ almost surely.²⁵

We claim that $Support(\lambda_\infty) \subseteq \bar{\Lambda}$. To prove this, fix some $x \notin \bar{\Lambda}$ and suppose towards contradiction that $x \in Support(\lambda_\infty)$. Since voting is informative when $\lambda_i = x$, the probability of each vote is continuous in the likelihood ratio around x . Moreover, the updating process on the likelihood ratio following each vote is also continuous around x . Thus, Theorem B.2 of Smith and Sorensen [26] applies, implying that for both possible votes, either (i) the probability of the vote is 0 when the likelihood ratio is x ; or (ii) the updated likelihood ratio following the vote remains x . Since voting is informative, neither of these two is true—contradiction.

The argument is completed by showing that $\Pr(\lambda_\infty \in \Lambda) = 1$. Suppose not, towards contradiction. Then since $Support(\lambda_\infty) \subseteq \bar{\Lambda}$, it must be that $\Pr(\lambda_\infty \in \{\underline{b}, \bar{b}\}) > 0$. Without loss of generality, assume $\Pr(\lambda_\infty = \underline{b}) > 0$; the argument is analogous if $\Pr(\lambda_\infty = \bar{b}) > 0$. Observe that if $\lambda_m < \underline{b}$ for some m , then by definition of \underline{b} and PBV, $\lambda_{m+1} = \lambda_m < \underline{b}$ and this sequence of public likelihood ratios converges to a point less than \underline{b} . Thus $\Pr(\lambda_\infty = \underline{b}) > 0$ requires that for any $\varepsilon > 0$, eventually $\langle \lambda_i \rangle \in [\underline{b}, \underline{b} + \varepsilon)$ with strictly positive probability. But notice that by the definition of \underline{b} , if $\lambda_i = \underline{b}$ then voter i votes informatively under PBV and thus if $\lambda_i = \underline{b}$, either $\lambda_{i+1} < \underline{b}$ (if $v_i = R$) or $\lambda_{i+1} = \frac{1}{2}$ (if $v_i = L$). By continuity of the updating process in the public likelihood ratio on the set $[\underline{b}, \bar{b}]$, it follows that if $\varepsilon > 0$ is chosen small enough, then $\lambda_i \in [\underline{b}, \underline{b} + \varepsilon)$ implies that $\lambda_{i+1} \notin [\underline{b}, \underline{b} + \varepsilon)$. This contradicts the requirement that for any $\varepsilon > 0$, eventually $\langle \lambda_i \rangle \in [\underline{b}, \underline{b} + \varepsilon)$ with strictly positive probability.

Step 2: Since $\lambda_i \xrightarrow{a.s.} \lambda_\infty$, λ_i converges in probability to λ_∞ , i.e. for any $\delta, \eta > 0$, there exists $\bar{n} < \infty$ such that for all $n > \bar{n}$, $\Pr(|\lambda_n - \lambda_\infty| \geq \delta) < \eta$. Since $\Pr(\lambda_\infty \in \Lambda) = 1$, for any $\varepsilon > 0$,

²⁵To be clear, when we say that $\langle \lambda_i \rangle$ eventually lies (or does not lie) in some set S almost surely, we mean that with probability one there exists some $k < \infty$ that for all $i > k$, $\lambda_k \in (\notin) S$.

we can pick $\delta > 0$ small enough such that $\Pr(\lambda_\infty \in [0, \underline{b} - \delta) \cup (\bar{b} + \delta, \infty)) > 1 - \frac{\varepsilon}{2}$. Pick $\eta = \frac{\varepsilon}{2}$. With these choices of δ and η , the previous statement implies that there exists $\bar{n} < \infty$ such that for all $n > \bar{n}$, $\Pr(\lambda_n \in \Lambda) > 1 - \varepsilon$, which proves the theorem. \square

7.2 Proofs for Section 3.2

We need various intermediate steps to needed to prove the results.

Definition 3. (*Winning Prob.*) For a history h^i , let $P(\Psi(h^i), \Delta(h^i), n - i + 1, \omega)$ be the probability with which L wins given the phase $\Psi(h^i)$, the vote lead $\Delta(h^i)$, the number of voters who have not yet voted ($n - i + 1$), and the true state is ω .

Note that once $\Psi(h^i) \in \{L, R\}$, all agents are voting uninformatively, and therefore, $P(\Psi(h^i), \cdot)$ is independent of state. For the subsequent results, let K denote $n - i$, and Δ denote $\Delta(h^i)$.

Lemma 4. For all h^i , $P(\Psi(h^i, L), \Delta + 1, K, \omega) \geq P(\Psi(h^i, R), \Delta - 1, K, \omega)$ for all $\omega \in \{L, R\}$. The inequality is strict if $K > \Delta - 1$.

Proof. Consider any realized profile of preference types and signals of the remaining K voters given true state ω (conditional on the state, this realization is independent of previous voters' types/signals/votes). In this profile, whenever a voter i votes for L given a vote lead $\Delta - 1$, he would also vote L given a vote lead $\Delta + 1$. Thus, if the type-signal profile is such that L wins given an initial lead of $\Delta - 1$, then L would also win given an initial lead of $\Delta + 1$. Since this applies to an arbitrary type-signal profile (of the remaining K voters, given state ω), it follows that $P(\Psi(h^i, l), \Delta + 1, K, \omega) \geq P(\Psi(h^i, r), \Delta - 1, K, \omega)$. That the inequality is strict if $K > \Delta - 1$ follows from the fact that with positive probability, the remaining K voters may all be Partisans, with exactly Δ more R -partisans than L -partisans. In such a case, L wins given initial informative vote lead $\Delta + 1$, whereas R wins given initial informative vote lead $\Delta - 1$. \square

Lemma 1 on pp. 14

Proof. We will begin by showing that an L -partisan always votes L if others are playing PBV strategies. By voting L , an L -partisan's utility is:

$$\mu(h^i, s_i) P(\Psi(h^i, L), \Delta + 1, K, L) + (1 - \mu(h^i, s_i)) P(\Psi(h^i, L), \Delta + 1, K, R)$$

If she voted R , her utility is

$$\mu(h^i, s_i) P(\Psi(h^i, R), \Delta - 1, K, L) + (1 - \mu(h^i, s_i)) P(\Psi(h^i, R), \Delta - 1, N, R)$$

It follows from Lemma 4 that the L -partisan voter i strictly prefers to vote L when the election is undecided (i.e. $K > \Delta - 1$).

The same arguments apply *mutatis mutandis* to see that R -partisans strictly prefer to vote R when the election is undecided. \square

To show that following PBV is optimal for a Neutral voter (conditional on others following PBV strategies), we need to describe the inferences a Neutral voter makes conditioning on being pivotal. Following standard notation, let $(t_{-i}, s_{-i}) = ((t_1, s_1), \dots, (t_{i-1}, s_{i-1}), (t_{i+1}, s_{i+1}), \dots, (t_n, s_n))$; this is a profile of type and signal realizations for all other voters apart from i . Given that other players are playing PBV, for any realized profile (t_{-i}, s_{-i}) , a vote by voter i deterministically selects a winner because PBV does not involve mixing. For a vote by player i , $V_i \in \{L, R\}$, denote the winner of the election $x(V_i; (t_{-i}, s_{-i})) \in \{L, R\}$. Then, we can denote the event in which a voter is pivotal as $Piv_i = \{(t_{-i}, s_{-i}) : x(L; (t_{-i}, s_{-i})) \neq x(R; (t_{-i}, s_{-i}))\}$. By arguments identical to Lemma 4, for a given profile (t_{-i}, s_{-i}) , if a subsequent voter after i votes L following $V_i = R$, then she would also do so following $V_i = L$. Therefore,

$$(t_{-i}, s_{-i}) \in Piv_i \Rightarrow [x(L, (t_{-i}, s_{-i})) = L \text{ and } x(R, (t_{-i}, s_{-i})) = R] \quad (8)$$

Let $U(V_i|h^i, s_i)$ denote a Neutral Voter i 's expected utility from action $V \in \{L, R\}$ when she faces a history h^i and has a private signal, s_i . If $\Pr(Piv_i|h^i, s_i) = 0$, then no action is sub-optimal for Voter i . If $\Pr(Piv_i|h^i, s_i) > 0$, i 's vote changes her expected utility if and only if her vote is pivotal. Therefore, in such cases,

$$U(V|h^i, s_i) > U(\tilde{V}|h^i, s_i) \Leftrightarrow U(V|h^i, s_i, Piv_i) > U(\tilde{V}|h^i, s_i, Piv_i) \text{ for } V \neq \tilde{V}$$

It follows from equation (8) that $U(L|h^i, s_i, Piv_i) = \Pr(\omega = L|h^i, s_i, Piv_i)$ and $U(R|h^i, s_i, Piv_i) = 1 - \Pr(\omega = L|h^i, s_i, Piv_i)$. Therefore, if $\Pr(\omega = L|h^i, l, Piv_i) > \frac{1}{2}$, it is strictly optimal for a Neutral Voter i to vote for L , and if $\Pr(\omega = L|h^i, l, Piv_i) < \frac{1}{2}$, it is strictly optimal for a Neutral Voter i to vote R .

Lemma 2 on pp. 15

Proof. The proof mirrors the intuition laid out in the text. Consider a history, h^i , where $\Psi(h^i) = L$. Then, since all future Neutral voters uninformatively vote for Candidate L , conditioning on being pivotal does not change i 's posterior belief about ω . Therefore, $\Pr(\omega = L|h^i, s_i, Piv_i) = \Pr(\omega = L|h^i, s_i)$, which by construction strictly exceeds $\frac{1}{2}$ for all s_i (since $\Psi(h^i) = L$). Therefore, a Neutral voter strictly prefers to vote L . An analogous argument applies when $\Delta(h^i) = -n_R$. \square

Lemma 3 on pp. 3

We shall first use an intermediary lemma to simplify the verification of incentive constraints.

Lemma 5. Consider any h^i where $\Psi(h^i) = 0$ and $\Delta(h^i) = \Delta$. Then if Neutral Voter $(i+1)$ votes informatively when $\Delta(h^{i+1}) \in \{\Delta-1, \Delta+1\}$, then Neutral Voter i shall vote informatively when $\Delta(h^i) = \Delta$.

Proof. We prove that it is optimal for i to vote L given signal $s_i = l$; a similar logic holds for optimality of voting R with signal r . By hypothesis, $\Psi(h^i, v_i) = 0$ for $v_i \in \{L, R\}$. It is necessary and sufficient that

$$\tilde{\mu}_i(\Delta, l) [P(0, \Delta+1, K, L) - P(0, \Delta-1, K, L)] - (1 - \tilde{\mu}_i(\Delta, l)) [P(0, \Delta+1, K, R) - P(0, \Delta-1, K, R)] \geq 0$$

Define the state-valued functions $p(\cdot)$ and $q(\cdot)$

$$p(\omega) = \begin{cases} \tau_L + (1 - \tau_L - \tau_R) \gamma & \text{if } \omega = L \\ \tau_L + (1 - \tau_L - \tau_R) (1 - \gamma) & \text{if } \omega = R \end{cases}$$

$$q(\omega) = \begin{cases} \tau_R + (1 - \tau_L - \tau_R) (1 - \gamma) & \text{if } \omega = L \\ \tau_R + (1 - \tau_L - \tau_R) \gamma & \text{if } \omega = R \end{cases}$$

Since voter $i+1$ votes informatively if Neutral (because both $\Delta+1$ and $\Delta-1$ are non-herd leads), the probability that $i+1$ votes L and R in state ω is $p(\omega)$ and $q(\omega)$ respectively. Noting the recursive relation $P(\Psi(h^i), \Delta, K+1, \omega) = p(\omega) P(\Psi(h^i, l), \Delta+1, K, \omega) + q(\omega) P(\Psi(h^i, r), \Delta-1, K, \omega)$ it follows that the above inequality holds if and only if

$$0 \leq \tilde{\mu}_i(\Delta, l) \left[\begin{aligned} & (P(\Psi(h^i, l, l), \Delta+2, K-1, L) - P(0, \Delta, K-1, L)) p(L) \\ & + (P(0, \Delta, K-1, L) - P(\Psi(h^i, r, r), \Delta-2, K-1, L)) q(L) \end{aligned} \right]$$

$$- (1 - \tilde{\mu}_i(\Delta, l)) \left[\begin{aligned} & (P(\Psi(h^i, l, l), \Delta+2, K-1, R) - P(0, \Delta, K-1, R)) p(R) \\ & + (P(0, \Delta, K-1, R) - P(\Psi(h^i, r, r), \Delta-2, K-1, R)) q(R) \end{aligned} \right]$$

Dividing by $p(R)(1 - \tilde{\mu}_i(\Delta, l))$, the above is equivalent to

$$0 \leq \left(\begin{aligned} & \frac{\tilde{\mu}_i(\Delta, l)}{1 - \tilde{\mu}_i(\Delta, l)} \frac{p(L)}{p(R)} [P(\Psi(h^i, l, l), \Delta+2, K-1, L) - P(0, \Delta, K-1, L)] \\ & - [P(\Psi(h^i, l, l), \Delta+2, K-1, R) - P(0, \Delta, K-1, R)] \end{aligned} \right)$$

$$+ \left(\begin{aligned} & \frac{\tilde{\mu}_i(\Delta, l)}{1 - \tilde{\mu}_i(\Delta, l)} \frac{q(L)}{p(R)} [P(0, \Delta, K-1, L) - P(\Psi(h^i, r, r), \Delta-2, K-1, L)] \\ & - \frac{q(R)}{p(R)} [P(0, \Delta, K-1, R) - P(\Psi(h^i, r, r), \Delta-2, K-1, R)] \end{aligned} \right)$$

We now argue that each of the two lines of the RHS above is positive.

1. Since $\frac{\tilde{\mu}_i(\Delta, l)}{1 - \tilde{\mu}_i(\Delta, l)} = \frac{\pi\gamma}{(1-\pi)(1-\gamma)} g_i(\Delta)$ and $\frac{p(L)}{p(R)} = f(\tau_L, \tau_R)$, it follows that

$$\frac{\tilde{\mu}_i(\Delta, l)}{1 - \tilde{\mu}_i(\Delta, l)} \frac{p(L)}{p(R)} = \frac{\pi\gamma}{(1-\pi)(1-\gamma)} g_{i+1}(\Delta+1)$$

$$= \frac{\tilde{\mu}_{i+1}(\Delta+1, l)}{1 - \tilde{\mu}_{i+1}(\Delta+1, l)}$$

Since IC holds for voter $i+1$ with vote lead $\Delta+1$, $\frac{\tilde{\mu}_{i+1}(\Delta+1,l)}{1-\tilde{\mu}_{i+1}(\Delta+1,l)} \geq \frac{P(\Psi(h^i,l,l),\Delta+2,K-1,R)-P(0,\Delta,K-1,R)}{P(\Psi(h^i,l,l),\Delta+2,K-1,L)-P(0,\Delta,K-1,L)}$, which proves that the first line of the desired RHS is positive.

2. Using the previous identities,

$$\begin{aligned} \frac{\tilde{\mu}_i(\Delta,l)}{1-\tilde{\mu}_i(\Delta,l)} \frac{q(L)}{q(R)} &= \frac{\pi\gamma}{(1-\pi)(1-\gamma)} g_{i+1}(\Delta-1) \\ &= \frac{\tilde{\mu}_{i+1}(\Delta-1,l)}{1-\tilde{\mu}_{i+1}(\Delta-1,l)} \end{aligned}$$

Since IC holds for voter $i+1$ with vote lead $\Delta-1$, $\frac{\tilde{\mu}_{i+1}(\Delta-1,l)}{1-\tilde{\mu}_{i+1}(\Delta-1,l)} \frac{q(L)}{q(R)} \geq \frac{P(0,\Delta,K-1,R)-P(\Psi(h^i,r,r),\Delta-2,K-1,R)}{P(0,\Delta,K-1,L)-P(\Psi(h^i,r,r),\Delta-2,K-1,L)}$, and thus the second line of the desired RHS is positive. □

This Lemma implies that if the IC's hold when $\Delta(h^i) \in \{-n_R(i+1), n_L(i+1)-1\}$, they hold for all intermediary values of Δ . Thus, we shall now proceed to verify that the incentive conditions indeed do hold at these thresholds.

Lemma 6. *Consider any h^i where $\Psi(h^i) = 0$ and $\Delta(h^i) = \Delta$. Then, if $P(\Psi(h^i,l), \Delta+1, K, L) \geq P(\Psi(h^i,l), \Delta+1, K, R)$ and $P(\Psi(h^i,r), \Delta-1, K, L) \geq P(\Psi(h^i,r), \Delta-1, K, R)$, then $P(0, \Delta, K+1, L) \geq P(0, \Delta, K+1, R)$.*

Proof. Simple manipulations yield

$$\begin{aligned} &P(0, \Delta, K+1, L) - P(0, \Delta, K+1, R) \\ &= p(L)P(\Psi(h^i,l), \Delta+1, K, L) + q(L)P(\Psi(h^i,r), \Delta-1, K, L) \\ &\quad - [p(R)P(\Psi(h^i,l), \Delta+1, K, R) + q(R)P(\Psi(h^i,r), \Delta-1, K, R)] \\ &\geq p(L)P(\Psi(h^i,l), \Delta+1, K, L) + q(L)P(\Psi(h^i,r), \Delta-1, K, L) \\ &\quad - [p(L)P(\Psi(h^i,l), \Delta+1, K, R) + q(L)P(\Psi(h^i,l), \Delta-1, K, R)] \\ &= p(L)[P(\Psi(h^i,l), \Delta+1, K, L) - P(\Psi(h^i,l), \Delta+1, K, R)] \\ &\quad + q(L)[P(\Psi(h^i,r), \Delta-1, K, L) - P(\Psi(h^i,r), \Delta-1, K, R)] \\ &\geq 0 \end{aligned}$$

where the first inequality uses the fact that $p(L) \geq p(R)$ and $P(\Delta+1, K, R) \geq P(\Delta-1, K, R)$; and the last inequality uses the hypotheses of the Claim. □

Corollary 1. *For all h^i , $P(\Psi(h^i), \Delta, K+1, L) \geq P(\Psi(h^i), \Delta, K+1, R)$.*

Proof. Base Step: The Claim is true when $K=0$. To see this, first note that Δ must be even for $P(\Psi(h^i), \Delta, 1, \omega)$ to be well-defined. If $\Delta \neq 0$ (hence $|\Delta| \geq 2$), then $P(\Psi(h^i), \Delta, 1, L) = P(\Psi(h^i), \Delta, 1, R)$. For $\Delta=0$, we have $P(0, 1, L) = p(L) > p(R) = P(0, 1, R)$.

Inductive Step: For any $K \geq 2$, the desired inequality trivially holds if $\Delta \in \{-n_R, n_L\}$ because $P(n_L, K, L) = P(n_L, K, R)$ and $P(-n_R, K, L) = P(-n_R, K, R)$. So it remains to consider only $\Delta \in \{-n_R + 1, \dots, n_L - 1\}$. Assume inductively that $P(\Delta + 1, K - 1, L) \geq P(\Delta + 1, K - 1, R)$ and $P(\Delta - 1, K - 1, L) \geq P(\Delta - 1, K - 1, R)$. [The Base Step guaranteed this for $K = 2$.] Using the previous Lemma, it follows that $P(\Delta, K, L) \geq P(\Delta, K, R)$ for all $\Delta \in \{-n_R + 1, \dots, n_L - 1\}$. \square

Proposition 3. *Consider history h^i such that $\Delta(h^i) = \Delta = n_L(i + 1) - 1$ and $\Psi(h^i) = 0$. Then if all other voters are playing PBV, and a neutral voter i receives signal r , it is strictly optimal for her to vote R . Analogously, if $\Delta(h^i) = -n_R(i + 1) + 1$, and if all other voters are playing PBV, and a neutral voter i receives signal l , it is strictly optimal to vote L .*

Proof. If $K < \Delta - 1$, then L will win the election regardless of her action, and thus, voting R is not sub-optimal. Consider the more interesting case where $K > \Delta - 1$.²⁶ Then for the voter to vote truthfully, it must be that

$$\begin{aligned} & \tilde{\mu}_i(\Delta, r) P(0, \Delta - 1, K, L) + (1 - \tilde{\mu}_i(\Delta, r)) (1 - P(0, \Delta - 1, K, R)) \\ & \geq \tilde{\mu}_i(\Delta, r) P(L, \Delta + 1, K, L) + (1 - \tilde{\mu}_i(\Delta, r)) (1 - P(L, \Delta + 1, K, R)) \\ \Rightarrow & \frac{\tilde{\mu}_i(\Delta, r)}{1 - \tilde{\mu}_i(\Delta, r)} \leq \frac{P(L, \Delta + 1, K, R) - P(0, \Delta - 1, K, R)}{P(L, \Delta + 1, K, L) - P(0, \Delta - 1, K, L)} \end{aligned}$$

Note that $P(0, \Delta - 1, K, L) \geq P(0, \Delta - 1, K, R)$ by Corollary 1, and by definition, $P(L, \Delta + 1, K, R)$ equals $P(L, \Delta + 1, K, L)$. Therefore, the right-hand side of the equation is bounded below by 1. Also, since $\Psi(h^i) = 0$, $\mu(h^i, r) = \tilde{\mu}_i(\Delta, r) < \frac{1}{2}$, the LHS is bounded above by 1, establishing our strict inequality. An analogous argument applies to prove the case where $\Delta(h^i) = -n_R + 1$ and $s_i = l$. \square

The remaining step is to now establish that if a Neutral Voter i is at the information set where $\Psi(h^i) = 0$, and $(\Delta(h^i), s_i) \in \{(n_L(i + 1) - 1, l), (-n_R(i + 1) + 1, r)\}$, it is strictly optimal for her to vote informatively even though this ends the *learning* phase. Consider the information set where $\Psi(h^i) = 0$, $\Delta(h^i) = n_L(i + 1) - 1$, and $s_i = l$. By the discussion in 31, it suffices to show that $\Pr(\omega = L | h^i, l, Piv_i) > \frac{1}{2}$. For any i , and for any $k > i$, let ξ_k^Ψ be the set of types $\{(t_j, s_j)\}_{j \neq i}$ that is consistent with history h^i , induces $(\Psi(h^{k-1}), \Psi(h^k)) = (0, \Psi)$ where $\Psi \in \{L, R\}$ after $v_i = R$, and where i 's vote is pivotal. Let $K^\Psi = \{k > i : \xi_k^\Psi \neq \emptyset\}$. Denote by ξ_Δ^0 the set of types $\{(t_j, s_j)\}_{j \neq i}$ that are consistent with h^i , induces $\Psi(h^n) = 0$ and $\Delta(h^{n+1}) = \Delta < 0$ after $v_i = R$, and where i 's vote is pivotal. Let $K_\Delta^0 = \{\Delta : \xi_\Delta^0 \neq \emptyset\}$. Observe that since the event $(h^i, Piv_i) = \cup_\Psi (\cup_{k \in K^\Psi} \xi_k^\Psi) \cup (\cup_{\Delta \in K_\Delta^0} \xi_\Delta^0)$, by the definition of

²⁶Note that it is not possible for $k = n_L - 2$ since this would imply the possibility of a herd, something ruled out in a model with an odd number of voters.

conditional probability

$$\Pr(\omega = L|h^i, l, Piv_i) = \sum_{\Psi \in \{L, R\}} \sum_{k \in K^\Psi} \Pr(\xi_k^\Psi | h^i, l, Piv_i) \Pr(\omega = L | \xi_k^\Psi, l) + \sum_{\Delta \in K_\Delta^0} \Pr(\xi_\Delta^0 | h^i, l, Piv_i) \Pr(\omega = L | \xi_\Delta^0)$$

We shall argue that $\Pr(\omega = L|h^i, l, Piv_i) > \frac{1}{2}$ by showing that $\Pr(\omega = L|\xi_k^L, l) > \frac{1}{2}$ for each $k \in K^L$, $\Pr(\omega = L|\xi_k^R, l) \geq \frac{1}{2}$ for each $k \in K^R$, and $\Pr(\omega = L|\xi_\Delta^0, l) \geq \frac{1}{2}$ for each $\Delta \in K_\Delta^0$.

Consider $k \in K^L$: this denotes a set of types such that an L -herd begins after the vote of voter $(k-1)$. Since only votes in the learning phase are informative, $\Pr(\omega = L|\xi_k^L, l) = \Pr(\omega = L|l, \Psi(h^{k-1}) = 0, \Delta(h^k) = n_L(k))$. Given that $v_i = R$, the informational content of this event is equivalent to a history \tilde{h}^{k-1} where $\Delta(\tilde{h}^{k-1}) = n_L(k) + 1$, and all neutrals are assumed to have voted informatively. Therefore,

$$\Pr(\omega = L|\xi_k^L, l) = \frac{\pi\gamma g_{k-1}(n_L(k) + 1)}{\pi\gamma g_{k-1}(n_L(k) + 1) + (1-\pi)(1-\gamma)}$$

Observe that $g_{k-1}(n_L(k) + 1) > g_k(n_L(k)) > \frac{(1-\pi)\gamma}{\pi(1-\gamma)}$. Therefore, $\Pr(\omega = L|\xi_k^L, l) >$

$$\frac{\gamma^2}{\gamma^2 + (1-\gamma)^2} > \frac{1}{2}.$$

Now consider $k \in K^R$: this denotes a set of types such that an R -herd begins after the vote of voter $(k-1)$. As before, only votes in the learning phase contain information about the state of the world; thus, $\Pr(\omega = L|\xi_k^R, l) = \Pr(\omega = L|l, \Psi(h^{k-1}) = 0, \Delta(h^k) = -n_R(k))$. Given that $v_i = R$, the informational content is equivalent to a history \tilde{h}^{k-1} where $\Delta(\tilde{h}^{k-1}) = -n_R(k) + 1$, and all neutrals are assumed to have voted informatively. Therefore,

$$\Pr(\omega = L|\xi_k^R, l) = \frac{\pi\gamma g_{k-1}(-n_R(k) + 1)}{\pi\gamma g_{k-1}(-n_R(k) + 1) + (1-\pi)(1-\gamma)}$$

As by assumption, $\Delta(h^{k-1}) = -n_R(k) + 1$ and $\Psi(h^{k-1}) = 0$, it must be that $g_{k-1}(-n_R(k) + 1) \geq \frac{(1-\pi)(1-\gamma)}{\pi\gamma}$. Therefore, $\Pr(\omega = L|\xi_k^R, l) \geq \frac{1}{2}$.

Now consider the event $\Delta \in K_\Delta^0$: this denotes a set of types such that no herd ever begins and the final vote lead is $\Delta < 0$. Therefore,

$$\Pr(\omega = L|\xi_\Delta^0, l) = \frac{\pi\gamma g_n(\Delta + 1)}{\pi\gamma g_n(\Delta + 1) + (1-\pi)}$$

Since $\Delta(h^n) \in \{\Delta - 1, \Delta + 1\}$ and $\Psi(h^n) = 0$, we have $g_n(\Delta + 1) \geq \frac{(1-\pi)(1-\gamma)}{\pi\gamma}$, and therefore $\Pr(\omega = L|\xi_\Delta^0, l) \geq \frac{1}{2}$.

We use the above three facts to deduce that $\Pr(\omega = L|h^i, l, Piv_i) > \frac{1}{2}$: observe that

$$\sum_{\Psi \in \{L, R\}, k \in K^\Psi} \Pr(\xi_k^\Psi | h^i, l, Piv_i) + \sum_{\Delta \in K_\Delta^0} \Pr(\xi_\Delta^0 | h^i, l, Piv_i) = 1$$

and also $\sum_{k \in K^L} \Pr(\xi_k^L | h^i, l, Piv_i) > 0$. Therefore, $\Pr(\omega = L | h^i, l, Piv_i)$ is a convex combination of numbers that are bounded below by $\frac{1}{2}$, with strictly positive weight on a number that strictly exceeds $\frac{1}{2}$.

An analogous argument can be made to ensure optimality at the information set where $\Delta(h^i) = -n_R(i+1) + 1$ and $s_i = r$.

7.3 Proofs for Section 4

Theorem 3 on pp. 19

We argue through a succession of lemmas that there exist $\bar{\mu}^* < 1$ and $\underline{\mu}_* > 0$ such that when $\tau_L \neq \tau_R$, in a large enough election, a CPV (μ_*, μ^*) is an equilibrium only if $\underline{\mu}_* \leq \mu_* < \mu^* \leq \bar{\mu}^*$. This suffices to prove the Theorem, because then, the arguments of Theorem 1 apply with trivial modifications. Note that in all the lemmas below, it is implicitly assumed when we consider a particular voter's incentives that she is at an undecided history.

For any CPV (μ_*, μ^*) , we can define threshold sequences $\{\tilde{n}_L(i)\}_{i=i}^\infty$ and $\{\tilde{n}_R(i)\}_{i=i}^\infty$ similarly to $\{n_L(i)\}_{i=i}^\infty$ and $\{n_R(i)\}_{i=i}^\infty$, except using the belief threshold μ^* (resp. μ_*) in place of the PBV threshold γ (resp. $1-\gamma$). That is, for all i such that $g_i(i-1) \leq \frac{(1-\pi)\mu^*}{\pi(1-\mu^*)}$, set $\tilde{n}_L(i) = i$. If $g_i(i-1) > \frac{(1-\pi)\mu^*}{\pi(1-\mu^*)}$, we set $\tilde{n}_L(i)$ to be the unique integer that solves $g_i(\tilde{n}_L(i) - 2) \leq \frac{(1-\pi)\mu^*}{\pi(1-\mu^*)} < g_i(\tilde{n}_L(i))$. For all i such that $g_i(-(i-1)) \geq \frac{(1-\pi)(\mu_*)}{\pi(1-\mu_*)}$, set $\tilde{n}_R(i) = i$. If $g_i(-(i-1)) < \frac{(1-\pi)(\mu_*)}{\pi(1-\mu_*)}$, set $\tilde{n}_R(i)$ to be the unique integer that solves $g_i(-\tilde{n}_R(i) + 2) \geq \frac{(1-\pi)(\mu_*)}{\pi(1-\mu_*)} > g_i(-\tilde{n}_R(i))$. Given these threshold sequences \tilde{n}_L and \tilde{n}_R , we define the phase mapping $\tilde{\Psi} : h^i \rightarrow \{L, 0, R\}$ in the obvious way that extends the PBV phase mapping Ψ . We state without proof the following generalization of Proposition 1.

Proposition 4. *Fix a parameter set $(\pi, \gamma, \tau_L, \tau_R, n)$. For each $i \leq n$, if voters play CPV (μ_*, μ^*) in the game $G(\pi, \gamma, \tau_L, \tau_R; n)$, there exist sequences $\{\tilde{n}_L(i)\}_{i=i}^\infty$ and $\{\tilde{n}_R(i)\}_{i=i}^\infty$ satisfying $|\tilde{n}_C(i)| \leq i$ such that a Neutral voter i votes*

1. *informatively if $\tilde{\Psi}(h^i) = 0$;*
2. *uninformatively for L if $\tilde{\Psi}(h^i) = L$;*
3. *uninformatively for R if $\tilde{\Psi}(h^i) = R$;*

where $\tilde{\Psi}$ is the phase mapping with respect to \tilde{n}_L and \tilde{n}_R . The thresholds $\tilde{n}_L(i)$ and $\tilde{n}_R(i)$ do not depend on the population size, n .

Lemma 7. *There exists $\bar{\mu}^* < 1$ and $\underline{\mu}_* > 0$ such that in any CPV (μ_*, μ^*) ,*

1. *if $\mu^* > \bar{\mu}^*$ then $\tilde{n}_L(i) > n_L(i)$ for all i such that $n_L(i) < i$;*
2. *if $\mu_* < \underline{\mu}_*$, then $-\tilde{n}_R(i) < -n_R(i)$ for all i such that $-n_R(i) > -i$.*

Proof. We give the argument for part (1); it is similar for part (2). Define $\bar{\mu}^*$ by the equality $\frac{\bar{\mu}^*}{1-\bar{\mu}^*} = \frac{\gamma}{1-\gamma} f(\tau_L, \tau_R) f(\tau_R, \tau_L)$. It is straightforward to compute from the definition of $g_i(\cdot)$ that for any k (such that $|k| < i$ and $i-k$ is odd), $g_i(k-2) f(\tau_L, \tau_R) f(\tau_R, \tau_L) = g_i(k)$. Suppose $\mu^* > \bar{\mu}^*$ and there is some i with $\tilde{n}_L(i) \leq n_L(i)$. By the definitions of $n_L(i)$ and $\tilde{n}_L(i)$, and the monotonicity of $g_i(k)$ in k ,

$$\begin{aligned} g_i(n_L(i) - 2) &= g_i(n_L(i)) [f(\tau_L, \tau_R) f(\tau_R, \tau_L)]^{-1} \\ &\geq g_i(\tilde{n}_L(i)) [f(\tau_L, \tau_R) f(\tau_R, \tau_L)]^{-1} \\ &> \frac{(1-\pi)\mu^*}{\pi(1-\mu^*)} [f(\tau_L, \tau_R) f(\tau_R, \tau_L)]^{-1} \\ &> \frac{(1-\pi)\gamma}{\pi(1-\gamma)} \end{aligned}$$

contradicting the definition of $n_L(i)$ which requires that $g_i(n_L(i) - 2) \leq \frac{(1-\pi)\gamma}{\pi(1-\gamma)}$. \square

Lemma 8. *If all Neutral voters play according to a CPV profile, it is uniquely optimal for an L-partisan to vote L and an R-partisan to vote R.*

Proof. This follows from the weak monotonicity imposed by CPV; trivial modifications to the argument in Lemma 1 establish this result. \square

Lemma 9. *In a large enough election, CPV (0, 1) is not an equilibrium unless $\tau_L = \tau_R$.*

Proof. Suppose all voters play CPV strategy (0, 1). Without loss of generality assume $\tau_L > \tau_R$; the argument is analogous if $\tau_L < \tau_R$. Let $\varsigma_t(n)$ denote the number of voters of preference-type $t \in \{L, R, N\}$ when the electorate size is n . Denote $\tau_N = 1 - \tau_L - \tau_R$. Suppose voter 1 is Neutral and has received signal l . She is pivotal if and only if amongst the other $n-1$ voters, the number of L votes is exactly equal the number of R votes. Let $\varsigma_{N,s}(n)$ denote the number of Neutrals who have received signal $s \in \{l, r\}$. Under the CPV profile (0, 1), voter 1 is pivotal if and only if $\varsigma_{N,r}(n) - (\varsigma_{N,l}(n) - 1) = \varsigma_L(n) - \varsigma_R(n)$. By the Weak Law of Large Numbers, for any $\varepsilon > 0$ and any $t \in \{L, R, N\}$, $\lim_{n \rightarrow \infty} \Pr\left(\left|\frac{\varsigma_t(n)}{n} - \tau_t\right| < \varepsilon\right) = 1$. Consequently, since $\tau_L > \tau_R$, for any $\varepsilon > 0$ and $k > 0$, there exists \bar{n} such that for all $n > \bar{n}$, $\Pr(\varsigma_L(n) - \varsigma_R(n) > k) > 1 - \varepsilon$. Thus, denoting Piv_1 as the set of preference-type and signal realizations where the Neutral voter 1 with $s_i = l$ is pivotal, we have that for any $\varepsilon > 0$ and $k > 0$, there exists \bar{n} such that for all $n > \bar{n}$, $\Pr(\varsigma_{N,r}(n) - \varsigma_{N,l}(n) > k | Piv_1) > 1 - \varepsilon$. Since $\Pr(\omega = L | \varsigma_{N,r}(n), \varsigma_{N,l}(n))$ is strictly decreasing in $\varsigma_{N,r}(n) - \varsigma_{N,l}(n)$, it follows that by considering k large enough in the previous statement, we can make $\Pr(\omega = L | Piv_1) < \frac{1}{2}$ in large enough elections. Consequently, in large enough elections, voter 1 strictly prefers to vote R when she is Neutral and has received $s_i = l$, which is a deviation from the the CPV strategy (0, 1). \square

Lemma 10. *In a large enough election, CPV (μ_*, μ^*) is not an equilibrium if either $\mu_* > \frac{1}{2}$ or $\mu^* < \pi$.*

Proof. If $\mu_* > \pi$, then the first voter votes uninformatively for R if Neutral, and consequently, all votes are uninformative. Thus, conditioning on being pivotal adds no new information to any voter. Since $\mu_1(h^1, l) > \pi > \frac{1}{2}$ (recall that $h^1 = \phi$), voter 1 has an incentive to deviate from the CPV strategy and vote L if she is Neutral and receives signal $s_1 = l$.

If $\mu_* \in (\frac{1}{2}, \pi]$, let h^{k+1} be a history of k consecutive R votes. It is straightforward that for some integer $k \geq 1$, $\mu(h^k) \geq \mu_* > \mu(h^{k+1})$. Since an R -herd has started when it is voter $k+1$'s turn to vote, conditioning on being pivotal adds to information to voter $k+1$. Suppose voter $k+1$ is Neutral and receives $s_{k+1} = l$. Then since an R -herd has started, she is supposed to vote R . But since $\mu_{k+1}(h^{k+1}, l) > \mu(h^k) \geq \mu_* > \frac{1}{2}$, she strictly prefers to vote L .

If $\mu_* < \pi$, the argument is analogous to the case of $\mu_* > \pi$, noting that $\mu_1(h^1, r) < \frac{1}{2}$ because $\gamma > \pi$. \square

Lemma 11. *In a large enough election, CPV $(\mu_*, 1)$ is not an equilibrium for any $\mu_* \in (0, \pi]$.*

Proof. Suppose CPV $(\mu_*, 1)$ with $\mu_* \in (0, \pi]$ is an equilibrium. Consider a Neutral voter m with signal $s_m = r$ and history h^m such that $\mu(h^m) \geq \mu_*$ but $\mu(h^{m+1}) < \mu_*$ following $v_m = R$. (To see that such a configuration can arise in a large enough election, consider a sequence of consecutive R votes by all voters.) Voter m is supposed to vote R in the equilibrium. We will show that she strictly prefers a deviation to voting L in a large enough election.

Claim 1: If the true state is R , then following $v_m = L$, the probability of an R -herd converges to 1 as the electorate size $n \rightarrow \infty$. *Proof:* Recall that the likelihood ratio stochastic process $\lambda_i \xrightarrow{a.s.} \lambda_\infty$ (where the domain can be taken as $i = m+1, m+2, \dots$). Since voter i votes informatively if and only if $\lambda_i \geq \frac{\mu_*}{1-\mu_*}$, the argument used in proving Theorem 1 allows us to conclude that $Support(\lambda_\infty) \subseteq [0, \frac{\mu_*}{1-\mu_*}]$ and $\Pr(\lambda_\infty = \frac{\mu_*}{1-\mu_*}) = 0$. Consequently, there is a herd on R eventually almost surely in state R .

Claim 2: $\Pr(Piv_m | \omega = R)$ converges to 0 as the electorate size $n \rightarrow \infty$. *Proof:* To be explicit, we use superscripts to denote the electorate size n , e.g. we write Piv_m^n instead of Piv_m . Denote

$$\begin{aligned} X^n &= \{(t_{-m}, s_{-m}) \in Piv_m : L\text{-herd after } v_m = L, R\text{-herd after } v_m = R\} \\ Y^n &= \{(t_{-m}, s_{-m}) \in Piv_m : \text{no herd after } v_m = L, R\text{-herd after } v_m = R\} \\ Z^n &= \{(t_{-m}, s_{-m}) \in Piv_m : R\text{-herd after } v_m = L \text{ and } v_m = R\} \end{aligned}$$

We have $Piv_m^n = X^n \cup Y^n \cup Z^n$; hence it suffices to show that $\Pr(X^n) \rightarrow 0$, $\Pr(Y^n) \rightarrow 0$, and $\Pr(Z^n) \rightarrow 0$. That $\Pr(X^n) \rightarrow 0$ and $\Pr(Y^n) \rightarrow 0$ follows straightforwardly from Claim 1. To show that $\Pr(Z^n) \rightarrow 0$, let Ψ_k^n denote the phase after voter k has voted, i.e. when it is voter $k+1$'s turn to vote. For any n , consider the set of $\{(t_j, s_j)\}_{j=m+1}^n$ such that after $v_m = L$, $\Psi_n^n \neq L$; denote this set Ξ^n . Partition this into the sets that induce $\Psi_n^n = 0$ and $\Psi_n^n = R$, denoted $\Xi^{n,0}$ and $\Xi^{n,R}$ respectively. Clearly, $Z^n \subseteq \Xi^{n,R}$. For any ε , for

large enough n , regardless of m 's vote, $\Pr(\Psi_n^0 = 0) < \varepsilon$ by Claim 1, and thus, $\Pr(\Xi^{n,0}) < \varepsilon$. Now consider any $n' > n$. $Z^{n'} \subseteq \Xi^n$ because if there is a L -herd following $v_m = L$ with electorate size n , there cannot be an R -herd following $v_m = L$ with electorate size n' . Thus, $\Pr(Z^{n'}) = \Pr(\Xi^{n,0}) \Pr(Z^{n'} | \Xi^{n,0}) + \Pr(\Xi^{n,R}) \Pr(Z^{n'} | \Xi_n^R) < \varepsilon + \Pr(\Xi^{n,R}) \Pr(Z^{n'} | \Xi_n^R)$ for large enough n . We have $\Pr(Z^{n'} | \Xi^{n,R}) = \frac{\Pr(Z^{n'} \cap \Xi^{n,R})}{\Pr(\Xi^{n,R})}$. It is straightforward to see that $\Pr(Z^{n'} \cap \Xi^{n,R}) \rightarrow 0$ as $n' \rightarrow \infty$, using the fact that $\tau_L < 1 - \tau_L$ and invoking the Weak Law of Large Numbers similarly to Lemma 9. Note that $\Pr(\Xi^{n,R})$ is bounded away from 0 because if sufficiently many voters immediately after m are R -partisans, then an R -herd will start regardless of m 's vote. This proves that $\Pr(Z^{n'}) \rightarrow 0$.

Claim 3: If the true state is L , then following $v_m = L$, the probability that L wins is bounded away from 0 as the electorate size $n \rightarrow \infty$. *Proof:* Define $\xi(h^i) = \frac{\Pr(\omega=R|h^i)}{\Pr(\omega=L|h^i)}$; this generates a stochastic process $\langle \xi_i \rangle$ ($i = m+1, m+2, \dots$) which is a martingale conditional on state L , and thus $\langle \xi_i \rangle \xrightarrow{a.s.} \xi_\infty$. Note that $\xi_{m+1} < \frac{1-\mu^*}{\mu^*}$ since $\mu(h^m) \geq \mu^*$ and $v_m = L$. Since voter i votes informatively if and only if $\xi_i \in \left(0, \frac{1-\mu^*}{\mu^*}\right]$, the argument used in proving Theorem 1 allows us to conclude that $\text{Support}(\xi_\infty) \subseteq \{0\} \cup \left[\frac{1-\mu^*}{\mu^*}, \infty\right)$ and $\Pr(\xi_\infty = \frac{1-\mu^*}{\mu^*}) = 0$. Suppose towards contradiction that $0 \notin \text{Support}(\xi_\infty)$. This implies $\mathbb{E}[\xi_\infty] > \frac{1-\mu^*}{\mu^*}$. By Fatou's Lemma (Billingsley [8, p. 209]), $\mathbb{E}[\xi_\infty] \leq \lim_{n \rightarrow \infty} \mathbb{E}[\xi_n]$; since for any $n \geq m+1$, $\mathbb{E}[\xi_n] = \xi_{m+1}$, we have $\frac{1-\mu^*}{\mu^*} < \mathbb{E}[\xi_\infty] \leq \xi_{m+1} < \frac{1-\mu^*}{\mu^*}$, a contradiction. Thus, $0 \in \text{Support}(\xi_\infty)$, and it must be that $\Pr(\xi_\infty = 0) > 0$. The claim follows from the observation that for any history sequence where $\xi_i(h^i) \rightarrow 0$ it must be that $\Delta(h^i) \rightarrow \infty$.

Consider the expected utility for voter m from voting R or L respectively, conditional on being pivotal: $EU_m(v_m = R | \text{Piv}_m) = \Pr(\omega = R | \text{Piv}_m)$ and $EU_m(v_m = L | \text{Piv}_m) = \Pr(\omega = L | \text{Piv}_m)$. Thus, she strictly prefers to vote L if and only if $\Pr(\omega = L | \text{Piv}_m) > \Pr(\omega = R | \text{Piv}_m)$, or equivalently, if and only if $\Pr(\text{Piv}_m | \omega = L) > \Pr(\text{Piv}_m | \omega = R) \frac{1-\mu_m(h^m, r)}{\mu_m(h^m, r)}$. By Claim 2, $\Pr(\text{Piv}_m | \omega = R)$ converges to 0 as electorate grows. On the other hand, $\Pr(\text{Piv}_m | \omega = L)$ is bounded away from 0, because by Claim 3, the probability that L wins following $v_m = L$ is bounded away from 0, whereas if $v_m = R$, a R -herd starts and thus the probability that R wins converges to 1 as the electorate size grows. Therefore, in a large enough election, $\Pr(\text{Piv}_m | \omega = L) > \Pr(\text{Piv}_m | \omega = R) \frac{1-\mu_m(h^m, r)}{\mu_m(h^m, r)}$, and it is strictly optimal for m to vote L following his signal $s_m = r$, which is a deviation from the CPV strategy. \square

Lemma 12. *In a large enough election, CPV $(0, \mu^*)$ is not an equilibrium for any $\mu^* \in [\pi, 1)$.*

Proof. Analogous to Lemma 11, it can be shown here that in a large enough election there is a voter who when Neutral is supposed to vote L with signal l , but strictly prefers to vote R . \square

Lemma 13. *In a large enough election, CPV (μ_*, μ^*) is not an equilibrium if $\mu^* \in (\bar{\mu}^*, 1)$ and $\mu_* \in (0, \frac{1}{2}]$.*

Proof. Fix an equilibrium CPV (μ_*, μ^*) with $\mu^* \in (\bar{\mu}^*, 1)$ and $\mu_* \in (0, \frac{1}{2}]$. By Lemma 7, $\tilde{n}_L(i) > n_L(i)$ for all i . Consider a Neutral voter m with signal $s_m = r$ and history h^m such that $\mu(h^m) \geq \mu_*$ but $\mu(h^{m+1}) < \mu_*$ following $v_m = R$. (To see that such a configuration can arise in a large enough election, consider a sequence of consecutive R votes by all voters.) Voter m is supposed to vote R in the equilibrium. We will show that she strictly prefers a deviation to voting L in a large enough election.

First, note that by following the argument of Theorem 1, it is straightforward to show that regardless of m 's vote, a herd arises with arbitrarily high probability when the electorate size n is sufficiently large. Define X^n, Y^n , and Z^n as in Lemma 11, where n indexes the electorate size. Plainly, $\Pr(Y^n) \rightarrow 0$. The argument of Claim 2 in Lemma 11 implies with obvious modifications that $\Pr(Z^n) \rightarrow 0$. Finally, $\Pr(X^n) \rightarrow 0$ because there exists $m' > m$ such that if $v_i = L$ for all $i \in \{m+1, \dots, m'\}$, then $\Psi_{m'}^n = L$, and $\Pr(v_i = L \text{ for all } i \in \{m+1, \dots, m'\}) \geq (\tau_L)^{m'-m} > 0$. Since $Piv_m^n = X^n \cup Y^n \cup Z^n$, we conclude that as $n \rightarrow \infty$, $\Pr(X^n | Piv_m^n) \rightarrow 1$, whereas $\Pr(Y^n | Piv_m^n) \rightarrow 0$ and $\Pr(Z^n | Piv_m^n) \rightarrow 0$. Consequently, for any $\varepsilon > 0$, there exists \bar{n} such that for all $n > \bar{n}$,

$$|EU_m(v_m = L | X^n, s_m = r) - EU_m(v_m = L | Piv_m^n, s_m = r)| < \varepsilon$$

and

$$|EU_m(v_m = R | X^n, s_m = r) - EU_m(v_m = R | Piv_m^n, s_m = r)| < \varepsilon$$

It therefore suffices to show that for any $n > m$, $EU_m(v_m = L | X^n, s_m = r) > EU_m(v_m = R | X^n, s_m = r)$, or equivalently, $\Pr(\omega = L | X^n, s_m = r) > \Pr(\omega = R | X^n, s_m = r)$. For any $k \in \{m+1, \dots, n\}$, define $X_k^n = \{(t_{-m}, s_{-m}) \in Piv_m^n : \Psi_{k-1}^n = 0 \text{ but } \Psi_k^n = L \text{ after } v_m = L, \Psi_k^n \text{ after } v_m = R\}$. Clearly, this requires $\tilde{n}_L(k+1) < k+1$. For $i \neq j$, $X_i^n \cap X_j^n = \emptyset$, but $X^n = \cup_{k=m+1}^n X_k^n$, and thus $\Pr(\omega | X^n, s_m = r) = \cup_{k=m+1}^n \Pr(\omega | X_k^n, s_m = r) \Pr(X_k^n | X^n)$. It therefore suffices to show that for any $k \in \{m+1, \dots, n\}$, $\Pr(\omega = L | X_k^n, s_m = r) > \Pr(\omega = R | X_k^n, s_m = r)$. Given that $v_m = L$, the informational content of X_k^n is equivalent to a history h^{k+1} where $\Delta(h^{k+1}) = \tilde{n}_L(k+1) - 2$, and all neutrals are assumed to have voted informatively. Therefore,

$$\Pr(\omega = L | X_k^n, s_m = r) = \frac{\pi \gamma g_{k+1} (\tilde{n}_L(k+1) - 2)}{\pi \gamma g_{k+1} (\tilde{n}_L(k+1) - 2) + (1 - \pi)(1 - \gamma)}$$

Since $\tilde{n}_L(k+1) < k+1$ and $\tilde{n}_L(i) > n_L(i)$ for all i , it must be that $\tilde{n}_L(k+1) - 2 \geq n_L(k+1)$. Consequently,

$$\begin{aligned} \Pr(\omega = L | X_k^n, s_m = r) &\geq \frac{\pi \gamma g_{k+1} (n_L(k+1))}{\pi \gamma g_{k+1} (n_L(k+1)) + (1 - \pi)(1 - \gamma)} \\ &> \frac{1}{2} \end{aligned}$$

where the second inequality is by the definition of $n_L(k+1)$. □

Lemma 14. *In a large enough election, CPV (μ_*, μ^*) is not an equilibrium if $\mu^* \in [\pi, 1)$ and $\mu_* \in (0, \underline{\mu}_*)$.*

Proof. Analogous to Lemma 13, it can be shown here that in a large enough election there is a voter who when Neutral is supposed to vote L with signal l , but strictly prefers to vote R . \square

Proposition 2 on pp. 22

Proof. We shall begin by describing the PBV strategy profile and then proceed to establish that it is an equilibrium. Observe that for $\tau_L = \tau_R = 0$, for every i , $g_i(k) = \left(\frac{\gamma}{1-\gamma}\right)^k$. As $\frac{\gamma}{1-\gamma} > \frac{(1-\pi)\gamma}{\pi(1-\gamma)} > 1 > \frac{1-\gamma}{\gamma} > \frac{(1-\pi)(1-\gamma)}{\pi\gamma} > \left(\frac{1-\gamma}{\gamma}\right)^2$. Therefore,

$$(n_L(i), n_R(i)) = \begin{cases} (1, 3) & \text{if } i \text{ is even} \\ (2, 2) & \text{if } i \text{ is odd} \end{cases}$$

Therefore, PBV prescribes that if $\Psi(h^{i-1}) = 0$, and $\Delta(h^i) \in \{1, -2\}$, then $\Psi(h^i) \neq 0$. To show that this strategy profile is an equilibrium, we shall consider the incentives in the learning and herding phases separately.

Observe that if $\Psi(h^i) = L(R)$, voting for $R(L)$ occurs with zero-probability on the path of play. We shall consider a belief-restriction that specifies that any off-path vote is simply ignored and does not affect the public belief. Given this belief-restriction, it can be shown using Theorem 2 (particularly Lemmas 2, 3) that PBV is an equilibrium. \square

References

- [1] D. Austen-Smith and J. S. Banks. Information aggregation, rationality, and the Condorcet jury theorem. *The American Political Science Review*, 90(1):34–45, March 1996.
- [2] A. Banerjee. A simple model of herd behavior. *Quarterly Journal of Economics*, 107(3):797–817, September 1992.
- [3] J. S. Banks and J. Sobel. Equilibrium selection in signaling games. *Econometrica*, 55(3):647–661, May 1987.
- [4] L. M. Bartels. Candidate choice and the dynamics of the presidential nominating process. *American Journal of Political Science*, 31(1):1–30, 1987.
- [5] L. M. Bartels. *Presidential Primaries and the Dynamics of Public Choice*. Princeton University Press, Princeton, NJ, 1988.
- [6] M. Battaglini. Sequential voting with abstention. *Games and Economic Behavior*, 51:445–463, 2005.
- [7] S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026, October 1992.
- [8] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, third edition edition, 1995.
- [9] S. Callander. Bandwagons and momentum in sequential voting. Kellogg-MEDS, working paper, 2004.
- [10] S. Callander. Vote timing and information aggregation. Kellogg-MEDS, working paper, 2004.
- [11] I.-K. Cho and D. M. Kreps. Signaling games and stable equilibria. *Quarterly Journal of Economics*, 102:179–221, 1987.
- [12] E. Dekel and M. Piccione. Sequential voting procedures in symmetric binary elections. *Journal of Political Economy*, 108:34–55, 2000.
- [13] T. Feddersen and W. Pesendorfer. The swing voter’s curse. *The American Economic Review*, 86(3):408–424, 1996.
- [14] T. Feddersen and W. Pesendorfer. Voting behavior and information aggregation in elections with private information. *Econometrica*, 65:1029–1058, 1997.

- [15] T. Feddersen and W. Pesendorfer. Elections, information aggregation, and strategic voting. *Proceedings of the National Academy of Sciences of the United States of America*, 96:10572–10574, 1999.
- [16] M. Fey. Informational cascades and sequential voting. University of Rochester Working Paper., 2000.
- [17] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, Cambridge, MA, 1991.
- [18] A. Gershkov and B. Szentes. Optimal voting schemes with costly information acquisition. Hebrew U., 2004.
- [19] P. Gurian. Resource allocation strategies in presidential nomination campaigns. *American Journal of Political Science*, 30(4):802–821, 1986.
- [20] T. A. Klumpp and M. K. Polborn. Primaries and the new hampshire effect. *Journal of Public Economics*, forthcoming, 2005.
- [21] E. Kohlberg and J.-F. Mertens. On the strategic stability of equilibria. *Econometrica*, 54(5):1003–1037, September 1986.
- [22] D. Kreps and R. Wilson. Sequential equilibria. *Econometrica*, 50:863–894, 1982.
- [23] R. B. Morton and K. C. Williams. *Learning By Voting*. The University of Michigan Press, Ann Arbor, 2001.
- [24] S. L. Popkin. *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. The University of Chicago Press, Chicago, 1991.
- [25] M. J. Robinson and M. A. Sheehan. *Over the Wire and on TV*. Russell Sage Foundation, New York, 1983.
- [26] L. Smith and P. Sorensen. Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398, 2000.
- [27] J. Wit. Herding behavior in a roll-call voting game. Department of Economics, University of Amsterdam, 1997.