Social Learning in Elections*

S. Nageeb Ali and Navin Kartik†
University of California, San Diego

April 2008

Abstract

Elections with sequential voting, such as presidential primaries, are widely thought to feature social learning and momentum effects, where the choices of early voters influence the behavior of later voters. Momentum may take time to build, and can depend on how candidates perform in each stage relative to expectations. This paper develops a rational theory of behavior in sequential elections that is consistent with these phenomena. We analyze an election with two candidates in which some voters are uncertain about which candidate is more desirable. Voters obtain private signals and vote in a sequence, observing the history of votes at each point. For a wide class of aggregation rules, we show that even strategic and forward-looking voters can herd on a candidate with positive probability, and such a "bandwagon" can occur with probability approaching one in large electorates.

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

*An earlier version of this paper was circulated under the title "A Theory of Momentum in Sequential Voting." We thank Susan Athey, Doug Bernheim, Vince Crawford, Eddie Dekel, Nir Jaimovich, Fahad Khalil, Jon Levin, Mark Machina, Marc Meredith, Roger Myerson, Marco Ottaviani, Joel Sobel, and numerous seminar and conference audiences for helpful comments and discussions. Chulyoung Kim provided outstanding research assistance. Ali is grateful to the Institute for Humane Studies, the John M. Olin Foundation, and Stanford's Economics Department for financial support. Kartik thanks the National Science Foundation for support, and the Institute for Advanced Study at Princeton for hospitality and support.

†snali@ucsd.edu and nkartik@ucsd.edu respectively. 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, February 24, 2004

1 Introduction

Many elections take place over time. The most prominent example is the U.S. presidential primary system, where voters cast ballots over the course of a few months. On a smaller scale, but also explicitly sequential, are the roll-call voting mechanisms used by city councils, Congressional bodies, and various committees in organizations. 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.

Sequential elections offer early voters the opportunity to communicate their views to future voters through their votes. Indeed, it is often suggested that in a sequential election, later voters are heavily influenced by the choices of previous voters. Such history dependence can result in *momentum* effects: the very fact that a particular alternative is leading in initial voting rounds may induce some later voters to select it who would have otherwise voted differently. While some scholars have viewed momentum as capturing efficient information aggregation (Mayer and Busch, 2004), others have criticized it on the grounds that early voters emerge as being too influential (Palmer, 1997).

Despite the prominent role that momentum plays in political rhetoric, relatively little is understood about how or why voters take into account the decisions of early voters. One explanation that has been suggested is that of "cue-taking," where voters learn the relative merits of each alternative from the choices of other voters, and vote accordingly (Bartels, 1988). This notion is reminiscent of the observational social learning literature initiated by Banerjee (1992) and Bikhchandani et al. (1992), but there is a theoretical challenge in interpreting electoral momentum in this vein. Since elections are collective processes, a voter's payoffs are determined by the electorate's entire profile of votes, and her own decision affects her payoffs only in the event that her vote is pivotal, i.e. changes the outcome of the election. Consequently, the strategic incentives in a sequential election differ from those in traditional social learning models. In particular, unlike the standard herding environment, an individual voter cares about the decisions of those who vote after her. Accordingly she may find little reason to conform with the choices of those who precede her since she recognizes that her vote is pivotal when many future votes are cast for the currently trailing alternative. She must therefore account for both the informational content of being pivotal and the potentially complex effects her vote has on subsequent voters. The strategic and informational issues become even more subtle when voters may be motivated by either private or common values.

The main contribution of this paper is to demonstrate that purely information-based momentum in sequential elections can emerge as the outcome of strategic behavior. We consider a sequential version of the canonical informational voting environment (Feddersen and Pesendorfer, 1996, 1997). A finite population of voters choose among two alternatives, voting in an exogenously fixed sequence and observing the history of prior votes. The winner is decided by some monotonic voting rule, such as simple majority or supermajority with one alternative acting as a status quo. There are two kinds of voters: Neutrals and Partisans. Neutrals desire to elect the "correct" or better alternative, which depends on the realization of an unknown state variable. Partisans, on the other hand, wish to elect their exogenously preferred alternative regardless of the state. Whether a voter is Neutral or Partisan is her private information. Each voter also receives a private binary signal that contains some information about the unknown state.

In this setting, we characterize a Perfect Bayesian equilibrium that features social learning. In this equilibrium, *Posterior-Based Voting* (PBV), a voter uses all currently available information—her prior, signal, and the observed history of votes—to form her expectation of which alternative is better for her, and votes for that alternative. Such behavior is inherently dependent on history, and in some histories, voters ignore their private information altogether, generating a "bandwagon" on one of the alternatives. The social learning exhibited by this equilibrium captures aspects of the behavior suggested by informal accounts of cue-taking. In PBV, each individual voter responds to her private information and the voting history exactly as she would in a standard herding environment. Partisan voters always vote for their preferred candidates while Neutral voters initially reveal their information through their vote. Since each voter's preference type is privately known, subsequent voters cannot perfectly infer whether a vote reflects partisan interests or information about the state of the world. Using Bayes rule, voters form posterior beliefs about the relative merits of each alternative, and once the informational content of the voting history swamps the private information of any individual voter, Neutral voters ignore their private information altogether. In large elections, PBV results in this form of herding by voters with high probability.

While the qualitative features of PBV are similar to those identified in standard herding models, the possibility for social learning has novel implications for elections.

In terms of electoral outcomes, herding leads to inefficient information aggregation for Neutral voters: even though a large electorate collectively has enough information to make what would be very likely the right decision for Neutrals, the votes of the early voters set the course for the election. The information and private preferences of early voters therefore has a substantial impact on the outcome of the election, and elections can be fragile to shocks that directly affect only relatively few voters. Indeed, an early voter whose vote counts for less than those of other individual voters—even when her vote is a straw vote—can nevertheless play a powerful rule through her informational influence on subsequent voters.

The history-dependent behavior we identify generates rich electoral dynamics, where momentum can ebb and flow. Moreover, in forming assessments of each alternative, Neutral voters judge the voting history relative to each alternative's partisan base. Consequently, even a trailing alternative may possess momentum if that alternative is believed to have relatively weak partisan support. By the same token, an alternative that commands the stronger partisan base does not gain momentum simply from leading in the election, but needs to accumulate a sufficiently large lead. A consequence of the social learning in our electoral model, therefore, is that a surprisingly good performance by one alternative can trump an anticipated vote-lead by the other. This phenomenon of judging relative to expectations mirrors voting patterns and political rhetoric about "needing to beat the odds" in the Presidential Primaries.[1]

That alternatives are judged relative to expectations under PBV generates interesting incentives for political candidates in electoral competition. We show that the ex-ante likelihood of victory for a candidate can strictly decrease when there is small increase in her partisan base. This is because an increase in a candidate's partisan base causes future Neutrals to discount the information content of any vote lead the candidate accumulates during the election. At certain critical partisanship levels, this negative effect in the social learning component of PBV can dominate the positive direct effect of an increase in partisan base.

Our analysis demonstrates that social learning and momentum can arise when each voter acts to have the election result in her preferred alternative. We view this as pro-

[1]As one example, prior to the 1984 Democratic primaries, Walter Mondale had gained strong partisan support from labor and interest groups, and was expected to "seize the nomination by force of organization" (Orren, 1985, p. 57). Yet, while he was successful in winning the first caucus in Iowa, his victory with close to 50% of its votes was largely overshadowed by Gary Hart's ability to garner about 17% of the votes. Even though Hart's performance in Iowa was far inferior to Mondale's on an absolute scale, Hart had performed much better than expected, and this garnered him momentum in subsequent elections (Bartels, 1988). Similarly, analyzing the 2004 Democratic Primaries, Knight and Schiff (2007) find that John Kerry benefited substantially from his surprising victories in early states.

viding some insight into the theoretical underpinnings for such phenomena. A surprising aspect of our PBV equilibrium is that apparent myopia is entirely consistent with fully strategic behavior, since each voter votes for the alternative she currently believes is best for her. Although this equilibrium is robust to many small perturbations in the game and uncertainty in the size of the electorate, we also illustrate by example that PBV may fail to be an equilibrium in more complex sequential voting games. Accordingly, we expect that in richer electoral environments, social learning will take the form of more elaborate history-dependent behavior, and in this sense, the current paper may be viewed as a benchmark.

The possibility for social learning in sequential elections complements the important results of Dekel and Piccione (2000). They show that a class of symmetric sequential voting games possess equilibria in which voting behavior is independent of history. Specifically, there are equilibria where voters ignore the public history and respond to their private information exactly as they would in the symmetric equilibrium of the corresponding simultaneous election. The insight that emerges from Dekel and Piccione's result is that the payoff-interdependencies inherent in sequential voting can support behavior that ignores the timing structure.[2] On the other hand, our results demonstrate that voters can be influenced by the information revealed in prior votes even when they are strategically motivated to be forward-looking. Combined, these findings emphasize that what is crucial to a voter's optimal behavior are her beliefs about how others in the election are acting: in PBV equilibrium, the belief that others are voting contingent on the history makes it optimal for a voter to herself vote in a history-dependent fashion.

We build upon two distinct literatures: the strategic voting approach to information aggregation pioneered by Austen-Smith and Banks (1996), Feddersen and Pesendorfer (1996), and Myerson (1998); and the literature on observational social learning, especially Bikhchandani et al. (1992) and Smith and Sorensen (2000). More directly related to this paper are two prior studies of social learning in elections. In unpublished work, Wit (1997) and Fey (2000) analyze a version of PBV in a simple majority rule, common values environment. Our paper differs along several dimensions. The possibility for private values requires a different method of analysis because voters have to account for the presence of partisans in the inferences they draw from the voting history and in their assessment of their strategic incentives. A dividend of our methods is that we can study arbitrary monotonic voting rules, demonstrating the capacity for social

[2] Dekel and Piccione (2000) show that a sequential voting game can also possess history-independent asymmetric equilibria in which initial voters vote informatively and later voters ignore their private information in every history. Which such equilibria would appear to involve herding, they are qualitatively distinct from PBV insofar as no voter is influenced by the choices of prior voters.

learning in a wide class of electoral environments. PBV also generates rich dynamics and comparative statics in elections with a confluence of private and common values that are impossible under pure common values. Lastly, we should note that while the focus here is on the existence of equilibria with social learning and their implications for elections, Wit (1997) and Fey (2000) argue that belief-based refinements preclude such possibilities in their models; this issue simply does not arise in our framework, as we discuss in Section 4.

Sequential elections have also been studied from other perspectives. Callander (2007) derives bandwagon equilibria in an infinite voter model in which each voter has exogenous preferences to vote for the winner in addition to a common value component. We believe that it is important to study a setting when voters do not have exogenous preferences to coordinate their choice, both for theoretical reasons and because the empirical evidence on whether voters possess conformist motivations is inconclusive.[3] The models also generate different insights: for instance, unique to the current paper is that the vote lead an alternative must accumulate to trigger a bandwagon can fluctuate as the election progresses, bandwagons can begin on a trailing alternative, and the onset of a bandwagon does not guarantee victory for either alternative. In addition, the technical analysis in our setting with a finite number of voters and the presence of private values is substantially different from that of an infinite population with a preference for conformity.

Iaryczower (2007) studies strategic voting when there are two separate multi-member committees, one of which votes after the other. Klumpp and Polborn (2006) use a contest model of competition between candidates to develop a campaign finance model of momentum. Morton and Williams (1999, 2001) have discussed how momentum may arise from coordination motives among voters when there are more than two candidates.[4] While not focusing on momentum, Battaglini (2005) and Battaglini et al. (2007) investigate the implications of voting costs on behavior in simultaneous and sequential elections.

The plan for the remainder of the paper is as follows. Section 2 lays out the model, and Section 3 derives the main results about PBV strategies and equilibrium. We discuss various implications and extensions of our analysis in Section 4. Section 5 concludes. All formal proofs are deferred to the Appendix.

[3] See the discussion in Bartels (1988, pp. 108–112). Kenney and Rice (1994) attempt to test the strength of various explanations for momentum, including both the preference for conformity theory and an informational theory similar to the one proposed here. They find some support for both, although neither is statistically significant; their methodology highlights the difficulties involved in such an exercise.

[4] Also in a setting with more than two alternatives, Piketty (2000) develops a two-period model where voters in the first period try to communicate their political beliefs so as to influence voting behavior in the second period.

2 Model

We consider a voting game with a finite population of n voters. Voters vote for one of two alternatives, L or R, in a fixed sequential order, one at a time. We label the voters $1, \ldots, n$, where without loss of generality, a lower numbered voter votes earlier in the sequence. Each voter i observes all preceding votes when it is her turn to cast her vote V_i. Let $V = (V_1, ..., V_n)$ be a profile of cast votes. A unique winner of the election, denoted $W \in \{L, R\}$, is determined by a *deterministic*, *monotone*, *non-trivial* voting rule, $\Im$.[5] This large class of voting rules includes majority, unanimity, or any anonymous threshold voting rule, as well as non-anonymous voting rules that provide asymmetric voting weights or veto power to particular voters. The generality in this regard is helpful in interpreting the model more broadly, for example, applying to settings where agents or interest groups must sequentially choose to endorse an alternative, and endorsements translate into outcomes in a manner that reflects asymmetric power.

Voters are uncertain about the realized state of the world, $\omega \in \{L, R\}$, that can affect their payoffs, and share a common prior over the possible states. Let $\pi \geq \frac{1}{2}$ be the ex-ante probability of state L. Before voting, each voter i receives a private signal, $s_i \in \{l, r\}$, drawn from a Bernoulli distribution with precision γ (i.e., $\Pr(s_i = l|\omega = L) = \Pr(s_i = r|\omega = R) = \gamma$), with $\gamma \in (\pi, 1)$.[6] Individual signals are drawn independently conditional on the state.

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 $\tau_L > 0$ to preference type L_p, probability $\tau_R > 0$ to R_p, and probability $1 - \tau_L - \tau_R > 0$ to the Neutral type, N. The preference ordering over alternatives is state dependent for Neutrals, but state independent for Partisans. Specifically, given a preference type $t_i \in \{L_p, N, R_p\}$, electoral winner $W \in \{L, R\}$, and state of the world $\omega \in \{L, R\}$, voter i's payoffs are defined by the function $u(t_i, W, \omega)$ as follows, where $\mathbf{1}_{\{X\}}$ denotes the

[5]The class of *deterministic* voting rules are those that map a profile of votes to a winner. A deterministic voting rule $\Im$ is *monotone* if for any profile of votes V where $\Im(V) = L$ and $V_i = R$ for some i, then $\Im\left(\tilde{V}\right) = L$ if $\tilde{V}_i = L$ and $\tilde{V}_{-i} = V_{-i}$. A deterministic, monotone voting rule $\Im$ is *non-trivial* if $\Im(L, \ldots, L) = L$ and $\Im(R, \ldots, R) = R$.

[6]This implies that any individual's signal is more informative than the prior but not perfectly informative. Our analysis generalizes with obvious changes to cases where the signal precision is asymmetric across states of the world.

indicator function on the event X:

$$\begin{aligned} u\left(C_p, W, \omega\right) &= \mathbf{1}_{\{W=C\}} \text{ for } C \in \{L, R\}, \\ u\left(N, W, \omega\right) &= \mathbf{1}_{\{W=\omega\}}. \end{aligned}$$

Therefore, a voter of preference type C_p ($C \in \{L, R\}$) is a Partisan for alternative C, desiring this alternative to be elected regardless of the state of the world. A Neutral voter, on the other hand, would like to elect alternative $C \in \{L, R\}$ if and only if that alternative is the better one, i.e. if the state $\omega = C$. Note that each voter cares about her individual vote only instrumentally, through its influence on the winner of the election.

Denote by $G\left(\pi, \gamma, \tau_L, \tau_R; n, \Im\right)$ the sequential voting game defined above with prior π, signal precision γ, preference type parameters τ_L and τ_R, n voters, and voting rule $\Im$. Throughout the subsequent analysis, we use the term *equilibrium* to mean a (weak) Perfect Bayesian equilibrium of this game (Fudenberg and Tirole, 1991). 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 $v_i : \{L_p, N, R_p\} \times \{L, R\}^{i-1} \times \{l, r\} \to \{L, R\}$. We say that a voter i *votes informatively* following a history h^i if $v_i\left(N, h^i, l\right) = L$ and $v_i\left(N, h^i, r\right) = R$. The posterior probability that voter i places on state L is denoted by $\mu_i(h^i, s_i)$.

We now clarify the role of two modeling choices.

Partisans. The Partisan types here are analogous to those in a number of papers in the literature, such as Feddersen and Pesendorfer (1996, 1997) and Feddersen and Sandroni (2006). Nevertheless, Partisans are not necessary for the existence of a history-dependent equilibrium. We analyze the game without Partisans—pure common value elections—in Section 4 and show that Posterior-Based Voting remains an equilibrium of that game. While some of the sequential voting literature has restricted attention to the case of pure common values, we believe that the presence of partisanship is relevant both theoretically and in practice, since partisans introduce private values into the electoral setting.[7] The presence of Partisans allows our model to generate rich momentum effects, including the feature that an alternative's performance is judged relative to expectations in a non-trivial way.[8]

[7]For example, in the context of presidential primaries, while some voters in a party hope to nominate the party candidate who is more electable in the general presidential election, there are others who may not be so sophisticated and simply wish to select a particular party candidate without considering the general election.

[8]Although we formalize Partisans as preferring an alternative regardless of the state of the world, this is purely for expositional purposes. It leaves our analysis unchanged to model Partisans and Neutrals as sharing the same state-dependent rankings over alternatives, but simply specifying that Partisan

Information Structure. The binary information structure studied here is the canonical focus in both the voting literature (e.g. Austen-Smith and Banks, 1996; Feddersen and Pesendorfer, 1996) and the social learning literature (e.g. Bikhchandani et al., 1992). Due to its prevalence and the complexities of studying history-dependent equilibria in sequential voting, this important benchmark is our focus here. Nonetheless, the issue of richer information structures is certainly important, and we discuss this in Section 4.

3 Posterior-Based Voting

We begin the analysis by introducing *Posterior-Based Voting* (PBV) and characterizing its induced dynamics. This characterization allows us to demonstrate that such behavior is an equilibrium in Section 3.2.

3.1 Definition and Dynamics

Let $\mathbf{v} = (v_1, \ldots, v_n)$ denote a strategy profile and $\mathbf{v}^i = (v_1, \ldots, v_{i-1})$ denote a profile of strategies for all players preceding i.

Definition 1. *A strategy profile, $\mathbf{v}$, satisfies (or is) Posterior-Based Voting (PBV) if for every voter i, type t_i, history h^i, signal s_i, and for any $W, W' \in \{L, R\}$,*

1. $\mathbb{E}_\omega[u(t_i, W, \omega)|h^i, s_i; \mathbf{v}^i] > \mathbb{E}_\omega[u(t_i, W', \omega)|h^i, s_i; \mathbf{v}^i] \Rightarrow v_i(t_i, h^i, s_i) = W,$
2. $\mathbb{E}_\omega[u(t_i, L, \omega)|h^i, s_i; \mathbf{v}^i] = \mathbb{E}_\omega[u(t_i, R, \omega)|h^i, s_i; \mathbf{v}^i] \Rightarrow \begin{cases} v_i(t_i, h^i, l) = L \\ v_i(t_i, h^i, r) = R. \end{cases}$

PBV is a property of a strategy *profile*. A PBV strategy refers to a strategy for a player that is part of a PBV profile.

The first part of the definition requires that given the history of votes and her private signal, if a voter believes that the selection of alternative L (R) will yield strictly higher utility than selecting alternative R (L), then she votes for alternative L (R). In other words, in a PBV profile, each voter updates her beliefs about alternatives using all currently available information (taking as given the strategies of previous voters), and then votes for the alternative she currently believes to be best for her. This coincides with rational behavior in a traditional herding environment without payoff interdependencies.

preferences (or priors) be sufficiently "biased" in favor of one alternative. In fact, the Partisan bias need not depend on the population size, so that sufficient information can change even Partisans' views about their preferred candidate. It is also worth emphasizing that Partisans are not "dominant-strategy types" as in, for example, Smith and Sorensen (2000); fn. 12 elaborates on this point.

Since Partisan voters have a preference ordering over alternatives that is independent of the state of the world, the definition immediately implies that Partisans vote for their preferred alternative in a PBV profile, independent of signal and history. Whenever a Neutral voter's posterior $\mu_i(h^i, s_i)$ differs from $\frac{1}{2}$, she votes for the alternative she believes to be strictly better.

Part two of the definition is a tie-breaking rule that requires that when a Neutral voter has posterior $\mu_i(h^i, s_i) = \frac{1}{2}$, she votes informatively and reveals her signal to future voters.[9] This tie-breaking rule does not play a significant role in our analysis because any choice of how to break ties applies only for a non-generic constellation of parameters $(\pi, \gamma, \tau_L, \tau_R)$. Remark 1 in Appendix A formalizes this point.

PBV is sophisticated insofar as voters infer as much as possible from the past history, taking into account the strategies of preceding players. However, the behavior in PBV is nevertheless apparently myopic. Since voters are influenced by the voting history in the PBV profile, a strategic voter who conditions on being pivotal should account for how her vote affects the decisions of those after her. The requirement of PBV, on the other hand, is simply that a voter cast her vote for the alternative she would wish to select based on what she currently knows without conditioning on pivotality. This distinction between myopic and strategic reasoning makes it unclear, *a priori*, that PBV can be an equilibrium. Indeed, PBV is a sequential analogue of what has been described as *Sincere* or *Naive* voting in simultaneous environments (Austen-Smith and Banks, 1996; Feddersen and Pesendorfer, 1997). We prove that PBV is in fact an equilibrium later in this section; for the moment, however, we illustrate the dynamics induced by PBV using the following example, which will be routine to readers well-versed with observational social learning.

Example 1. *The voting rule $\Im$ is simple majority rule, $n = 9$, $\pi = \frac{1}{2}$, $\gamma = \frac{3}{4}$, and $\tau_L = \tau_R = \frac{1}{4}$. Since Partisan voters always vote for their preferred alternative in PBV, we focus on the behavior of Neutrals. A Neutral Voter 1 votes informatively, since her signal is more informative than her prior. When it is Voter 2's turn to act, she does not know whether Voter 1 is Neutral or Partisan, but she Bayesian updates on the true state using her own signal and Voter 1's vote (under the presumption that Voter 1 has behaved according to PBV). A simple computation shows that regardless of Voter 1's choice, a Neutral Voter 2 will also vote informatively under PBV. Similarly, a Neutral Voter 3 votes informatively in every history that she faces.*

Suppose that Voter 4 observes history (L, L, L). Now, based on the information re-

[9] Note that the tie from the standpoint of PBV does not imply that the voter is strategically indifferent between her choices.

vealed in the voting history, Voter 4 believes it strictly more likely that L is the state of the world even when she obtains an r signal. Therefore, in PBV, a Neutral Voter 4 votes for L regardless of her private information. Hence, regardless of her preference type, Voter 4 ignores her signal, voting R if and only if she is an R-partisan. Since Voter 5 presumes PBV behavior by others, she recognizes that after the first three voters vote L, Voter 4's vote reveals no information about the state. By the same reasoning as that of Voter 4, a Neutral Voter 5 and all subsequent Neutral voters uninformatively vote L. □

The Example highlights the general pattern in a PBV profile: Partisans always vote for their preferred candidates, while Neutral voters initially vote informatively. Once the information contained in the public history swamps that from a private signal, Neutral voters begin herding. There are several points to be emphasized about the nature of these herds more generally. First, even after a herd begins for an alternative, Partisans continue to vote for their preferred alternative. Thus, it is always possible to see votes "contrary" to the herd, and any such contrarian vote is correctly inferred as having come from a Partisan. The possibility for these contrarian votes also means that a herd forming on an alternative does not imply victory for that alternative. Second, once herding begins, the public belief about the state of the world remains fixed because all subsequent voting is uninformative. Third, it is possible for a herd to form on an alternative that is *trailing* because the informational content of the voting history is not limited to merely whether an alternative is leading, but also how that alternative is performing relative to the ex-ante distribution of private preferences. This is illustrated by the following example.

Example 2. *Let* $n = 9$, $\pi = \frac{2}{3}$, $\gamma = \frac{3}{4}$, $\tau_L = 0.1$, *and* $\tau_R = 0.45$. *Suppose voters play the PBV profile, and voter* 6 *observes the history* $h^6 = (R, R, R, L, L)$. *Straightforward calculations show that given this history, voters 1 through 5 must have voted informatively if Neutral. Due to the relatively small partisan support for L, the two L votes from voters* 4 *are* 5 *are sufficient to overturn the impact of the preceding 3 votes for R in terms of influencing voter 6's posterior. In fact, even if voter* 6 *receives an r signal, she believes that it is more likely that alternative L is better. Thus, regardless of signal, voter* 6*'s vote is uninformative, and by induction, all future Neutrals vote for alternative L. The herd on L has formed even when L is trailing in the election.* □

For our equilibrium analysis, we need a characterization of voting dynamics in terms of the history. It is convenient to use two state variables that summarize the impact of history on behavior. For any history, h^i, the *vote lead* for alternative L, $\Delta(h^i)$, is defined as follows:

$$\Delta(h^i) = \sum_{j=1}^{i-1} \left(\mathbf{1}_{\{v_j=L\}} - \mathbf{1}_{\{v_j=R\}}\right). \tag{1}$$

The second state variable, called the *phase*, summarizes whether Neutral voters are continuing to learn about the quality of the alternatives (denoted phase 0), or learning has terminated in a herd for one of the alternatives (denoted phase L or R). The phase map is thus $\Psi : h^i \to \{L, 0, R\}$, defined by

$$\Psi\left(h^1\right) = 0; \forall i > 1, \Psi\left(h^i\right) = \begin{cases} \Psi\left(h^{i-1}\right) & \text{if } \Psi\left(h^{i-1}\right) \in \{L, R\} \\ L & \text{if } \Psi\left(h^{i-1}\right) = 0 \text{ and } \Delta\left(h^i\right) = n_L\left(i\right) \\ R & \text{if } \Psi\left(h^{i-1}\right) = 0 \text{ and } \Delta\left(h^i\right) = -n_R\left(i\right) \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The sequences $n_L(i)$ and $n_R(i)$ in the phase map equation (2) are determined by explicitly considering posteriors. For example, assuming that all prior Neutrals voted informatively, $n_L(i)$ is the smallest vote lead for alternative L such that at a history h^i with $\Delta(h^i) = n_L(i)$, the public history in favor of L outweighs a private signal r. Therefore, the threshold $n_L(i)$ is the unique integer less than or equal to $i-1$ that solves

$$\Pr\left(\omega = L | \Delta\left(h^i\right) = n_L\left(i\right) - 2, s_i = r\right) \le \frac{1}{2} < \Pr\left(\omega = L | \Delta\left(h^i\right) = n_L\left(i\right), s_i = r\right). \tag{3}$$

If it is the case that a history h^i with $\Delta(h^i) = i-1$ is outweighed by signal r, we set $n_L(i) = i$. Similarly, the threshold $n_R(i)$ is the unique integer less than or equal to i that solves

$$\Pr\left(\omega = L | \Delta\left(h^i\right) = -n_R\left(i\right) + 2, s_i = l\right) \ge \frac{1}{2} > \Pr(\omega = L | \Delta\left(h^i\right) = -n_R\left(i\right), s_i = l), \tag{4}$$

where again, implicitly, it is assumed that all prior Neutrals voted informatively. If it is the case that a signal l outweighs even that history h^i where $\Delta(h^i) = -(i-1)$, we set $n_R(i) = i$.

We summarize with the following characterization result (all proofs are in the Appendix).

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

1. *informatively if $\Psi(h^i) = 0$;*

2. *uninformatively for* $C \in \{L, R\}$ *if* $\Psi(h^i) = C$,

where Ψ *is as defined in (2). The thresholds* $n_L(i)$ *and* $n_R(i)$ *are independent of the population size,* n.

The above result demonstrates that PBV behavior can be characterized by thresholds on the vote lead. The threshold sequences $\{n_L(i)\}_{i=1}^n$ and $\{-n_R(i)\}_{i=1}^n$ serve as triggers for when the electoral process transitions from learning to herding for L and R respectively. So long as the vote lead never reaches either threshold, Neutral voters continue to vote informatively. The first time that the vote lead reaches a threshold marks the onset of herding, after which all subsequent voting is uninformative. For the parameters in Example 1, herding begins the first time that either alternative has 3 more votes in the learning phase than the other. The thresholds in Example 2 are more intricate, varying across the two alternatives and from one voter to the next.

3.2 PBV Equilibrium

We now establish that PBV is an equilibrium of the sequential voting game. In fact, we prove a stronger result. Say that the *election is undecided* at history h^i if, for each alternative, there exists some sequence of votes of the remaining $n - i + 1$ voters that results in the alternative winning the election. An equilibrium is *strict* if conditional on others following their equilibrium strategies, it is uniquely optimal for a voter to follow her equilibrium strategy at any undecided history.[10]

Theorem 1. *The PBV strategy profile is an equilibrium, and generically, is strict.*

There are three points we wish to to emphasize about PBV equilibrium. First, strictness for generic parameters implies that its existence does not rely upon how voter indifference is resolved when the election remains undecided. Given that others are playing PBV, a strategic voter follows PBV not because she is indifferent between or powerless to change the outcome, but rather because deviations yield strictly worse expected payoffs. Strictness also implies that the equilibrium is robust to many small perturbations of the model. Second, because Partisan voters always vote for their preferred alternatives in the PBV equilibrium, every information set is reached with positive probability.

[10]This definition is non-standard, but is the appropriate modification of the usual definition for voting games. Usually, a strict equilibrium of a game is one in which a deviation to any other strategy makes a player strictly worse off (Fudenberg and Tirole, 1991). A sequential voting game (with $n \geq 3$) cannot possess any strict equilibria in this sense, because after any history where a candidate has captured sufficiently many votes to win the election, all actions yield identical payoffs. That is, only histories where the election remains undecided are strategically relevant to voters. Our definition of strictness restricts attention to these histories.

Therefore, off-the-equilibrium-path beliefs play no role in our analysis, and PBV is a Sequential Equilibrium (Kreps and Wilson, 1982). Third, behavior in PBV equilibrium is genuinely history-dependent: there is generally no outcome-equivalent equilibrium in the simultaneous election counterpart of our model.[11]

In light of the strategic elements inherent in voting, it may be surprising that PBV is an equilibrium. In PBV, each voter votes as if her choice alone selects the electoral outcome. Such behavior, *prima facie*, appears very different from a voting equilibrium, which necessarily involves voters assessing the optimal choice in the event that they are pivotal. In the sequential voting environment here, a voter can be pivotal in two distinct ways: first, the voter can be a tie-breaker for the election, and second, the voter can influence subsequent voters and thereby cause a domino effect that changes the electoral outcome. In equilibrium, voters have to account for all such possibilities and the strategic and informational implications of being pivotal in every possible way. We illustrate the important issues and discuss why PBV is an equilibrium by continuing Example 1.

Example 1 Continued. Recall that $\Im$ is simple majority rule, $n = 9$, $\pi = \frac{1}{2}$, $\gamma = \frac{3}{4}$, and $\tau_L = \tau_R = \frac{1}{4}$. Consider a voter acting when the election is undecided, and assume that all others are following PBV. For PBV to be a (strict) equilibrium requires the following three kind of behaviors to be (uniquely) optimal, depending on the voter's preference type and the phase of the election:

(i) Partisans vote for their preferred alternative;

(ii) In the herding phase for L (R), Neutrals vote uninformatively for L (R);

(iii) In the learning phase, Neutrals vote informatively.

Point (i) above is a consequence of a monotonicity property of PBV: when others play PBV, a Partisan voter can never increase the probability of a future L vote, say, by herself voting R rather than L.[12] So consider point (ii). Since voting in the herding phase is uninformative, a Neutral voter in this phase learns nothing new about the state of the world when conditioning on being pivotal, and therefore strictly prefers to vote

[11] Therefore, the uninformative voting that occurs once a herd has been triggered in PBV equilibrium is distinct from the uninformative voting in the asymmetric equilibria identified by Dekel and Piccione (2000, Theorem 2), where voters who vote uninformatively do so regardless of history.

[12] This monotonicity does not hold in an arbitrary strategy profile. Thus, eliminating weakly-dominated strategies is not sufficient to guarantee that Partisans vote for their preferred candidate, unlike the case of a simultaneous election. It is in fact possible to construct sequential voting equilibria (in undominated strategies) where a Partisan votes against her preferred candidate.

on the basis of her current posterior. By definition of PBV, the posterior in a herding phase for L (R) favors L (R) regardless of the private signal.[13]

It remains, then, to consider optimality of point (iii) above, i.e. that a Neutral voter should vote informatively in the learning phase. Such a voter faces the following tradeoffs. The benefit of voting informatively is voting for who she believes to be the better candidate and simultaneously communicating this information to subsequent voters (for, if she does not vote informatively, subsequent voters will misperceive her signal). On the other hand, the cost of voting informatively is that it may push future voters towards herding and/or push the election towards being decided, suppressing valuable information possessed by later Neutrals. As an illustration of this tension, consider the decision problem faced by Voter 3 who observes two prior votes for L, and receives a signal in favor of l. To resolve the tradeoff that she faces between voting informatively (which triggers a herd on L) and a deviation to an R vote, we must assess exactly how her deviation can change the electoral outcome.

The set of type-signal realizations in which Voter 3 is pivotal, denoted as Piv, consist of all those vectors where her voting L results in L winning the election and her voting R results in R winning.[14] Denote the set of type-signal profiles where $v_3 = R$ results in a herd for alternative $C \in \{L, R\}$ as ξ^C; let the set of type-signal profiles that induce no herding after $v_3 = R$ be denoted $\tilde{\xi}$. As previous noted, in this example, a herd begins once an alternative accumulates 3 more votes than the other; in other words, it takes at least 3 L (R) votes in the learning phase to overturn an r (l) signal. The argument proceeds by showing that Voter 3's posterior conditional on being pivotal and on each of the three mutually exclusive and exhaustive events, ξ^L, ξ^R, or $\tilde{\xi}$, is strictly greater than $\frac{1}{2}$.

First, consider the event $\xi^R \cap Piv$: this is when by voting R, Voter 3 would cause a future R-herd and R to win. Because voting is uninformative once the herd begins, conditioning on this event reveals no more information than just the fact that an R herd begins in the future, and so $\Pr\left(\omega = L | Piv, \xi^R, h^3, l\right) = \Pr\left(\omega = L | \xi^R, h^3, l\right)$.[15] A future

[13] It might be useful to contrast this logic with that of the symmetric equilibrium of Dekel and Piccione (2000, Theorem 1). There, a voter would consider subsequent votes to be informationally equivalent to prior votes, and hence does learn something beyond the current posterior when conditioning on being pivotal. In PBV, on the other hand, a voter assesses the information content of subsequent votes on the basis of her observed history: in a herding phase, she correctly judges subsequent votes as being uninformative about the state.

[14] In general, there are strategy profiles where i can be pivotal in a way that $v_i = L$ results in R winning, whereas $v_i = R$ results in L winning. This is not possible in PBV because as we noted earlier, PBV features a weak monotonicity of subsequent votes in i's vote.

[15] This property is special to cases where partisanship is symmetric, because in such cases the vote leads that trigger herds are uniform across voters. In general, being pivotal will also reveal some information about when a herd begins, which may in turn reveal information about the state.

R-herd can happen only if following Voter 3's vote for R, alternative R subsequently gains a lead of 3 votes in the learning phase. This requires that following 3's vote, R receives an additional 4 votes over L in the learning phase. Since L commands a vote lead of 2 prior to Voter 3's vote, conditioning on ξ^R in effect reveals to Voter 3 a *net* total of 2 votes for R over L in the learning phase (note that Voter 3 does not count her own vote in this calculation). Since it takes a minimum of 3 votes for R in the learning phase to overturn an l signal, this net total of 2 votes for R is not sufficient to convince Voter 3 that R is at least as likely as L to be the state of the world. Hence, Voter 3's posterior conditional on this future contingent event ascribes strictly higher probability to state L than R.

Now consider the other possible pivotal events: a future L-herd ($\xi^L \cap Piv$) and no herding ($\tilde{\xi} \cap Piv$). In both of these cases, amongst voters $4, \ldots, 9$, alternative R can receive at most 4 votes over alternative L; otherwise, an R-cascade would start. By the same logic as before, it follows that Voter 3 ascribes strictly greater probability to L.

Therefore, conditional on her observed history, signal, and being pivotal—regardless of how pivotality obtains—Voter 3 believes alternative L to be the better alternative with probability 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 3 to vote L, even though such a choice ends the learning phase. Based on a Neutral voter's incentives to vote informatively even when that triggers a herd (like Voter 3 above), it is possible to show that all other incentive compatibility conditions in the learning phase are satisfied. $\square$

While the example helps clarify as to how PBV is an equilibrium, it is special and omits a number of points that are addressed in the formal proof of Theorem 1. One issue is the strictness of the equilibrium for generic parameters. Another is that when one alternative has a stronger partisan base than the other, the demands of pivotality become more complex. In simultaneous elections, as demonstrated by Feddersen and Pesendorfer (1996), this would give Neutral voters an incentive to deviate from informative voting and vote against the bias induced by asymmetric partisanship. In PBV, the herding thresholds adjust to the asymmetric partisan bias: in particular, the margins of victory that trigger herding are no longer uniform but vary across voters and alternatives. We establish that this adjustment in these thresholds preserves the incentives for Neutral voters to vote informatively in the learning phase. Finally, in simultaneous elections, equilibria and notions of pivotality are generally sensitive to the choice of voting rule (e.g. Austen-Smith and Banks, 1996; Feddersen and Pesendorfer, 1998). We show that PBV—a strategy profile invariant to the voting rule—is an equilibrium regardless of the monotonic voting rule used.

3.3 The Likelihood of Herds

In PBV, Neutral voters reveal their information through their votes until the voting history is more informative than the private signal that any individual voter obtains; Neutral voters then begin herding. Analogous to results in the standard herding environment, we can show that herds eventually occur with probability 1 in large elections, as a consequence of the bounded signal precision ($\gamma < 1$) and the martingale convergence theorem applied to the evolution of public beliefs in the electorate.[16]

Theorem 2. *For every* $(\pi, \gamma, \tau_L, \tau_R, \Im)$ *and for every* $\varepsilon > 0$*, there exists* $\overline{n} < \infty$ *such that for all* $n > \overline{n}$*, if voters play PBV, then* $\Pr[\Psi(h^n) \neq 0$ *in* $G(\pi, \gamma, \tau_L, \tau_R; n)] > 1 - \varepsilon$*.*

It is also worth noting that that because herds can occur with positive probability in large elections, the probability that the elected candidate matches the state of the world is bounded away from 1 in large elections in PBV equilibrium.

4 Discussion

This section explores some implications of PBV equilibrium, and some extensions. We will often specialize the class of voting rules to so-called "quota rules," or q-rules for short, i.e. a voting rule such that L is the winner if and only if it receives at least a fraction $q \in (0, 1)$ of the total number of votes (R can be thought of as a status quo). For notational clarity, we denote such voting rules by their threshold q rather than the more general notation $\Im$.

4.1 Comparative Statics

An interesting implication of the social learning in PBV is that political candidates may not always wish to increase the size of their expected partisan base: the ex-ante probability of victory for an alternative can strictly decrease when its expected partisan base increases.[17] To see why, note that under PBV, there are two ex-ante effects that a small increase in, say, τ_L has, holding all other parameters fixed. This has a direct effect

[16] One may wonder whether herds will eventually occur with probability 1 even while the election remains undecided. In general, this depends on the voting rule. In the special case of simple majority rule, this is in fact the case.

[17] Although perhaps counter-intuitive, this comports well with some scholars' views on presidential primaries. For example, in describing the 1976 Democratic primaries, Bartels (1988, p.3) writes, "The key to success, it appeared, *was not to enter the campaign with a broad coalition of political support,* but to rely on the dynamics of the campaign itself, particularly its earliest public phases, to generate that support." (emphasis added)

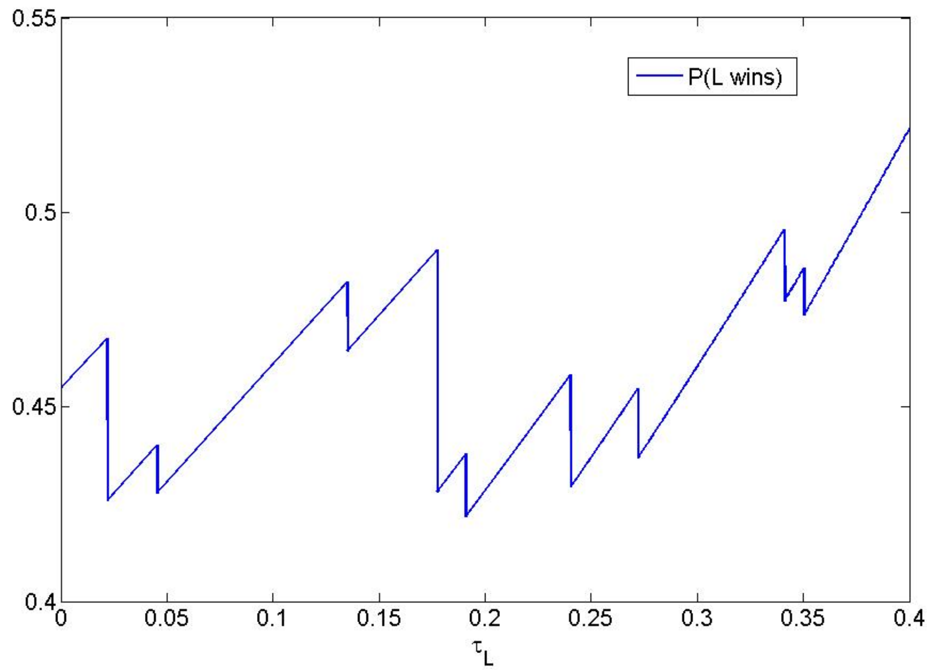

Figure 1: Probability that L wins as a function of L-partisanship under simple majority rule, $n = 7$, $\gamma = 0.7$, $\pi = 0.6$, and $\tau_R = 0.4$.

of increasing the probability of any vote being for L, since L-partisans always vote for L whereas Neutrals do not always do so. On the other hand, there can be an indirect effect on Neutrals through the social learning channel: the thresholds for herding can become less favorable to L, i.e. $n_L(\cdot)$ can increase for one or more voters. This indirect effect happens because an increase in τ_L makes Neutral voters discount votes for L during the learning phase. Herding thresholds only change at certain critical points of partisanship, but at such points, their effect locally swamps the direct effect noted earlier, because they have a discontinuous effect on subsequent votes whereas the direct effect is continuous. Figure 1 illustrates this by plotting the probability that L wins as a function of τ_L for a particular set of parameters.

A similar implication can be drawn concerning the probability that the elected candidate matches the state of the world, i.e. the ex-ante utility of Neutrals, as a function of partisanship. In this case, the direct effect of increasing partisanship for one candidate, say τ_L, is to reduce the ex-ante utility of Neutrals; at certain levels, however, there is an indirect positive social learning effect which leads to reduced herding—more informative voting from Neutrals—along some equilibrium paths. Figure 2 illustrates that the social learning effect can locally dominate the direct partisanship effect, so that the welfare of Neutrals can improve with a local increase in partisanship.

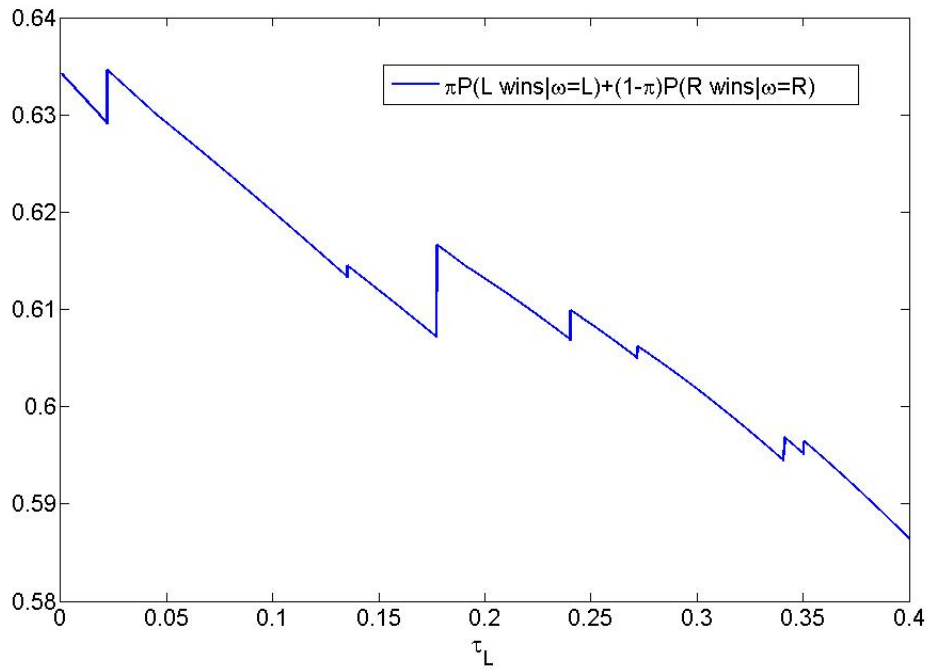

Figure 2: Probability that elected candidate matches the state of the world as a function of L-partisanship under simple majority rule, $n = 7$, $\gamma = 0.7$, $\pi = 0.6$, and $\tau_R = 0.4$.

4.2 Other Equilibria

Characterizing all equilibria seems intractable for general sequential voting games. This is particularly the case since we have been quite permissive about the class of voting rules. We offer below some limited results about other equilibria for the case of elections aggregated by quota rules.

A natural way to generalize PBV is to vary the thresholds of beliefs that induce herding: this is the set of *Cut-point Voting* (CPV) strategy profiles introduced by Callander (2007). We defer the formal definition of CPV profiles to the Supplementary Appendix; however, the class includes any strategy profile that can be defined by two beliefs thresholds, $\mu^* \in [0, 1]$ and $\mu_* \in [0, \mu^*]$, with a Neutral voter voting L if her posterior on $\omega = L$ is greater than some μ^*, voting R if her posterior is less than some μ_*, and voting her signal when the posterior is in $[\mu_*, \mu^*]$. Plainly, PBV is the particular case where $\mu_* = \mu^* = \frac{1}{2}$. We can show that for generic parameters, any equilibrium in the class of CPV profiles involves herding with probability 1 in large elections (Theorem 4 in the Supplementary Appendix);[18] however, there is no guarantee that there generally exists

[18]Note that the class of CPV profiles permits history-independent behavior, since the thresholds can be $\mu^* = 1$ and $\mu_* = 0$.

any CPV equilibrium apart from PBV. In the special case in which the expected partisan base for the alternatives is equal ($\tau_L = \tau_R$), for almost all values of π and γ, the only history-dependent CPV equilibrium in large elections is in fact PBV (Proposition 8 in the Supplementary Appendix).

Dekel and Piccione (2000) have shown that symmetric sequential q-rule voting games possess history-independent equilibria that are equivalent in outcomes to symmetric equilibria of otherwise identical simultaneous voting games. In our model, for any parameter set $(\pi, \gamma, \tau_L, \tau_R, n, q)$, the simultaneous voting analog of our model possesses a symmetric equilibrium where each Partisan voter votes for her preferred alternative, and each Neutral votes for an alternative on the basis of a signal-dependent probability. In the sequential voting game, consider a strategy profile where, independent of history, every voter plays the same strategy as in the above construction. The insight of Dekel and Piccione (2000) is that because all voters are acting independently of history, and each vote is pivotal with positive probability, the events in which a voter is pivotal is identical in both the simultaneous and sequential games; therefore, since the profile is symmetric and an equilibrium of the simultaneous game, it is also an equilibrium of the sequential game.

The existence of the history-independent equilibrium is certainly an important theoretical benchmark. For example, it implies that there are equilibria which achieve asymptotic full-information equivalence (Feddersen and Pesendorfer, 1997). We interpret our results as complementary: we show that social learning and momentum effects *can* arise from rational behavior in sequential elections owing to informational concerns, leading to impediments in information aggregation; whereas Dekel and Piccione (2000) show that such effects do not *have to* arise.

4.3 Equivalence of Voting Rules

As we have noted, in our environment, PBV is an equilibrium for every monotonic voting rule. This provides an interesting contrast with the insight of Dekel and Piccione (2000) that strategic voting can be unaffected by the timing of a voting game. In their history-independent equilibria, conditioning on being pivotal negates any usefulness from observing the history of votes; necessarily such equilibria are sensitive to the choice of voting rule. In the history-dependent equilibrium we have constructed, the opposite occurs: regardless of the voting rule, if others vote according to PBV, conditioning on being pivotal does not contain more payoff-relevant information than the public history and one's private signal.

This raises the question of how changes in the voting rule affect the electoral outcome when voters vote according to PBV. While different rules may yield different (distributions over) electoral outcomes under PBV, many voting rules are in fact asymptotically equivalent. To state this formally, let $G(\pi, \gamma, \tau_L, \tau_R; n, q)$ denote a sequential election aggregated according to the quota rule q. The following result shows that these voting rules can be partitioned into three classes such that all rules within any class are asymptotically ex-ante equivalent: they elect the same winner with probability approaching 1 in large voting games.

Theorem 3. *Fix any parameters $\pi, \gamma, \tau_L, \tau_R$, and assume that for any n, q, voters play PBV in the game $G(\cdot; n, q)$. For any $\varepsilon > 0$, there exists $\bar{n}$ such that for all $n > \bar{n}$,*

(a) $|Pr(L \text{ wins in } G(\cdot; n, q) - Pr(L \text{ wins in } G(\cdot; n, \tilde{q})| < \varepsilon$ for all $q, \tilde{q} \in (\tau_L, 1 - \tau_R)$;

(b) $\Pr(L \text{ wins in } G(\cdot; n, q)) > 1 - \varepsilon$ for all $q \in [0, \tau_L)$;

(c) $\Pr(L \text{ wins in } G(\cdot; n, q)) < \varepsilon$ for all $q \in (1 - \tau_R, 1]$.

Parts (b) and (c) of Theorem 3 are obvious consequences of the presence of Partisans, since in large elections they are sufficient to determine the winner under "extreme" quota rules. The interesting result is part (a) of the Theorem: all "interior" quota rules—where outcomes are not determined asymptotically by partisanship alone —are nevertheless asymptotically equivalent under PBV. The equivalence relies on the fact that for all such voting rules, once Neutrals begin herding on an alternative, that alternative wins with probability approaching one in large electorates. Therefore, all of these interior rules are asymptotically equivalent once a herd begins on a particular alternative. Asymptotic ex-ante equivalence of these voting rules then follows from the observation that the probability of a herd beginning on a particular alternative is independent of the voting rule (since PBV behavior is defined independently of the voting rule), and by Theorem 2, this probability approaches 1 in large games.[19]

4.4 Extensions

4.4.1 Pure Common Value Elections

We have thus far studied a model where some voters are Neutral, and others are Partisan.[20] Some element of private values is surely an important aspect of elections in

[19] It is worth nothing that the symmetric, history-dependent equilibria of Dekel and Piccione (2000) would also entail a form of asymptotic equivalence across quota rules, but the reason is very different. In those equilibria, voting rule equivalence emerges because asymptotically, there is full-information equivalence for each rule; in contrast, in PBV, the voting rule equivalence arises even without full-information equivalence.

[20] As noted earlier, although we have formalized Partisans as those without a common value element to their preferences whatsoever, this is only for expositional convenience. Partisans may in fact have a

practice, and this leads to rich effects in our model such as judging relative to expectations, herds on trailing candidates, and comparative statics. Nevertheless, we now briefly discuss how our results extend to a game of pure common values, where it is common knowledge that all voters are Neutral.

Consider an election $G(\pi, \gamma, \tau_L, \tau_R; n, \Im)$ in which $\tau_L = \tau_R = 0$. In this game, let PBV^0 refer to the strategy profile that is the limit of PBV as $\tau_L, \tau_R \to 0$. Theorem 1 extends as follows.[21]

Proposition 2. *If $\tau_L = \tau_R = 0$, there is an equilibrium where voters play PBV^0.*

The PBV^0 equilibrium of the pure common values game differs from PBV in the game with positive partisan probabilities in one notable respect. In the PBV^0 equilibrium, once a herd begins on an alternative, no vote is ever cast for the other alternative. Consequently, some vote profiles are off-the-equilibrium-path. In PBV^0, any vote that is contrary to a herd is simply ignored by subsequent voters. While PBV^0 satisfies standard requirements such as sequentiality and extensive-form perfectness, Fey (2000) has suggested a refinement that precludes the associated off-the-equilibrium-path beliefs. Since our fundamental interest is in elections with a confluence of private and common values—where, recall, there are no votes off the equilibrium path, and moreover, PBV is generically a strict equilibrium—the issue of whether there is a "right" equilibrium in the pure common values game is besides our scope.

4.4.2 Population Uncertainty

As argued by Myerson (1998, 2000), it may be unrealistic to assume that in large elections, each voter knows exactly how many other voters there are in the game. Population uncertainty can have significant implications for voting behavior in sequential elections: based on the history, a voter can update her beliefs about how many others are participating, and at least, set a lower bound on the number of other voters. Nevertheless, in spite of these complexities, PBV remains an equilibrium of our model when augmented with population uncertainty.

Formally, we define an election with population uncertainty as follows. Suppose there is a countably infinite set of available voters, indexed by $i = 1, 2, \ldots$. Nature first

common value component to their preferences; it is simply that their preferences (or priors) are biased in favor of one alternative, although not necessarily to the extent that sufficient information cannot change their views. Specifically, under complete information, an L-Partisan needs at least three net signals in favor of alternative R to prefer electing alternative R over L (and analogously for an R-Partisan), independently of the population size. In contrast, a Neutral needs only one net signal in favor of an alternative to prefer that alternative being elected.

[21]For the particular case of simple majority rule, versions of this result appear in Fey (2000) and Wit (1997).

draws the size of the electorate, n, according to probability measure η with (possibly unbounded) support on the natural numbers. The draw is unobserved by any agent. A voter is selected to vote if and only if her index i is no larger than n. All those who are selected to vote do so sequentially, in a roll-call order, only observing the history of votes, and receiving no information about the numbers of voters to follow. The rest of the game, in terms of preferences, information, and how outcomes are determined, remains as before. Denote a game with population uncertainty η and aggregated according to the quota rule, q, as $G^{\eta}(\pi, \gamma, \tau_L, \tau_R; q)$.[22]

The logic for why PBV remains an equilibrium is straightforward. Consider the decision faced by a voter i who observes history h^i, signal s_i, and assumes that all other voters who are selected to vote will behave according to PBV. Voter i can assess her best response by conditioning on every possible realization of population size that is weakly larger than i. By Theorem 1, for every such realization, voter i would wish to vote according to PBV. Since the behavior prescribed by PBV is independent of the population size, the result follows.[23]

Proposition 3. *For every game $G^{\eta}(\pi, \gamma, \tau_L, \tau_R; q)$, the PBV strategy profile is an equilibrium.*

4.4.3 Richer Environments

Since our simple model is the canonical framework for both social learning and informational elections, we view it as an important benchmark. That PBV is, generically, a strict equilibrium of the model implies that it remains an equilibrium even if the setting is slightly perturbed in many ways. However, the simplicity of history-dependence in PBV—myopic behavior—need not persist in some (significantly) richer environments. While a detailed study of more complex history-dependent behavior is beyond the scope of this paper, we can provide an example of how PBV may fail. In what follows, a forward-looking Neutral voter with sufficiently inferior information is hesitant to end the election and prefers to delegate that choice to a better-informed successor.

[22] This formulation of population uncertainty is quite general, encompassing the kinds of uncertainty that have been considered in simultaneous elections, such as Poisson distributions (Myerson, 1998) and binomial distributions (Feddersen and Pesendorfer, 1996). By restricting to quota rules here, we ensure that the outcome of the election is well-defined regardless of the actual number of voters.

[23] It is interesting to note that in sequential voting games with a sufficient degree of population uncertainty, there generally cannot exist a symmetric equilibrium in which all voters who are selected to vote play the same strategy. The reason is that by learning about (lower bounds on) the population size, the incentives for voters at different stages of the election are quite different. In particular, a symmetric equilibrium of the simultaneous election with population uncertainty does not generally remain an equilibrium of the corresponding sequential election.

Example 3. *Consider a simple majority rule election with* 3 *voters where* $\pi \simeq \frac{1}{2}$, $\tau_L \simeq 0$, *and* $\tau_R = \frac{1}{2}$. *Suppose that a Neutral voter can be one of two types : a Guru (G) who obtains a perfectly informative signal, or a Follower (F) who obtains a signal of precision* $\gamma = \frac{3}{5}$. *Conditional on being Neutral, a voter is a Guru with probability* $\tau_G = \frac{4}{5}$. *Consider the PBV strategy profile, and suppose that Voter* 1 *has voted* R, *and Voter* 2 *is a Neutral Follower with signal* r. *PBV clearly prescribes that she vote for* R, *ending the election and giving Voter 2 an expected payoff of* $\frac{8}{11}$. *If she deviates and votes for* L, *she leaves the election undecided. However, by voting* L, *Voter* 2 *will induce a Neutral Follower Voter 3 to herd on* L. *Given the low probability of any voter being a Follower and the high likelihood of Gurus, Voter 2's expected payoff from voting for* L *is* $\frac{87}{110}$, *which is strictly greater than her payoffs from voting* R. □

5 Conclusion

We have shown that informational social learning and momentum can emerge as the outcome of strategic voting. In the context of presidential primaries, Bartels (1988) and Popkin (1991) argue that voters keep careful track of how alternatives have performed relative to expectations when deliberating how to vote, and that the information provided to voters during the primaries is dominated by *horse-race* statistics that document alternatives' performance in preceding states. More recently, Knight and Schiff (2007) examine the reaction of voters in several states in daily opinion polls to primary results during the 2004 Democratic Primaries, and find evidence of momentum and voters' judging candidates relative to expectations in the spirit of the results provided here. Our analysis identifies an informational rationale for voters to track alternatives' performance, and to assess the quality of an alternative relative to its expected support.

The social learning in PBV equilibrium permits application of some rich insights that have been derived in other contexts to electoral mechanisms. Most fundamentally, it delivers an understanding of why sequential elections can be fragile to the realized preferences, information, and actions of a few early voters. It also provides a way to think about how the release of public information can generate *momentum shifts*. To illustrate, suppose that at each point in time, there is a very small probability that a perfect signal about the state of world may be released publicly.[24] If this information is released during a herding phase for L but reveals the state to the R, then all Neutral voters will suddenly shift their votes from L to R.

[24] It is straightforward to check that PBV remains an equilibrium when such a perturbation is made to a generic parametrization of our model.

The parsimonious model studied here departs from real-world sequential elections in a number of ways, and we think it is plausible that various issues outside our scope are also relevant to issues of electoral momentum and, more generally, history-dependent voting. We have restricted attention to binary elections. 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 alternatives. Coordination among voters when faced with multiple alternatives may also play a role in political momentum, as has been emphasized by Morton and Williams (2001). Our model also treats voting as entirely sequential, one voter at a time, largely for purposes of tractability. Though there are elections of this form—for example, roll-call voting mechanisms used in city councils and legislatures—there are many dynamic elections, such as the primaries, that feature a mixture of simultaneous and sequential voting. To what extent such games possess history-dependent equilibria with similar qualitative features is a significant question for future research. Finally, we have abstracted away from the role of institutions, and concentrated on voters as being the sole players. In practice, there are other forces involved in some 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 momentum, this paper may be viewed as providing a benchmark to understand the impact of different institutions on sequential elections.

Appendix A: Proofs

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. For any history h^i, define the *public likelihood ratio*, $\lambda(h^i) \equiv \frac{\Pr(\omega=L|h^i)}{\Pr(\omega=R|h^i)}$. $\lambda(h^i)$ captures how informative the history h^i is, given the behavior of preceding voters. Under PBV, a Neutral Voter i votes informatively so long as $\lambda(h^i) \in \left[\frac{1-\gamma}{\gamma}, \frac{\gamma}{1-\gamma}\right]$, votes L regardless of her signal if $\lambda(h^i) > \frac{\gamma}{1-\gamma}$, and votes for R regardless of her signal if $\lambda(h^i) < \frac{1-\gamma}{\gamma}$. To characterize this behavior in terms of the vote lead $\Delta(h^i)$, define the function

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

where the domain is $\{(\tau_L, \tau_R) : \tau_L + \tau_R \leq 1,\ \tau_L > 0, \tau_R > 0\}$. It is straightforward to verify that f strictly exceeds 1 over its domain.

For each positive integer i and any integer k where $|k| < i$ and $i-k$ is odd, define the function $g_i(k) = (f(\tau_L, \tau_R))^k \left(\frac{f(\tau_L,\tau_R)}{f(\tau_R,\tau_L)}\right)^{\frac{i-k-1}{2}}$. Note that for a history h^i where $\Delta(h^i) = k$ and all prior Neutrals voted informatively and Partisans voted for their preferred alternatives, $g_i(k) = \frac{\Pr(h^i|\omega=L)}{\Pr(h^i|\omega=R)}$; thus $g_i(k) = \left(\frac{1-\pi}{\pi}\right)\lambda(h^i)$. Plainly, $g_i(k)$ is strictly increasing in k.

For a given $(\pi, \gamma, \tau_L, \tau_R)$, define $\{n_L(i)\}_{i=1}^{\infty}$ as follows. 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 set $n_L(i)$ to be the unique integer that solves

$$g_i(n_L(i) - 2) \leq \frac{(1-\pi)\gamma}{\pi(1-\gamma)} < g_i(n_L(i)). \quad (5)$$

Since $g_i(-(i-1))$ is strictly less than $\frac{(1-\pi)\gamma}{\pi(1-\gamma)}$, and $g_i(k)$ is strictly increasing in k, a unique solution exists to (5).

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

$$g_i(-n_R(i) + 2) \geq \frac{(1-\pi)(1-\gamma)}{\pi\gamma} > g_i(-n_R(i)). \quad (6)$$

As before, since $g_i(k)$ is strictly increasing in k, and $g_i(i-1) = (f(\tau_L, \tau_R))^{i-1} \geq 1 > \frac{(1-\pi)(1-\gamma)}{\pi\gamma}$, a unique solution exists to (6).

These values of $n_L(i)$ and $n_R(i)$ define $\Psi(\cdot)$ as in equation (2) from the text.

A.1 Proposition 1

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 voting 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 posterior $\mu(h^i, s_i) = \mu\left(\tilde{h}^i, s_i\right)$ if $\Delta(h^i) = \Delta\left(\tilde{h}^i\right)$ and $\Psi(h^i) = \Psi\left(\tilde{h}^i\right) = 0$ (i.e., so long as all preceding Neutrals have voted informatively, only vote lead matters, and not the actual sequence). Thus, we can define $\tilde{\mu}_i(\Delta, s_i) = \mu(h^i, s_i)$ where $\Delta = \Delta(h^i)$. By Bayes' rule,

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

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: $\Psi(h^i) = R$. This is analogous to Case 2 above, and therefore omitted.

A.2 Tie-breaking does not arise generically

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*

$$\Gamma_{\pi,\gamma} \equiv \left\{(\tau_L, \tau_R) \in \left(0, \frac{1}{2}\right)^2 : g_i(\Delta) \in \left\{\frac{(1-\pi)(1-\gamma)}{\pi\gamma}, \frac{(1-\pi)\gamma}{\pi(1-\gamma)}\right\} \textit{ for some } i \in \mathbb{Z}^+ \textit{ and } |\Delta| \leq i\right\}$$

is isomorphic to a 1-dimensional set. Thus, the need for tie-breaking arises only for a set of parameters $(\pi, \gamma, \tau_L, \tau_R)$ *of (Lebesgue) measure* 0*.*

A.3 Theorem 1

We split the proof into three parts, paralleling the steps in the discussion in the text following Theorem 1. Throughout, to prove that PBV is an equilibrium, we assume that the relevant history is undecided since all actions at a decided history yield the same payoffs. The first subsection below establishes the strict optimality of Partisans' behavior (Proposition 4); the second subsection establishes strict optimality of Neutrals' behavior in herding phases (Proposition 5); the third subsection deals with Neturals in the learning phase (Proposition 6).

As a preliminary step, we define a useful piece of notation and state an important monotonicity result. 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 players 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 $\Delta(h^i)$. Throughout, for any history h^j and alternative C, we use the notation $\Psi(h^j, C)$ to indicate $\Psi(h^{j+1})$ where h^{j+1} is the history following h^j and $v_j = C$.

The ensuing Lemma establishes an intuitive monotonicity in the winning probability as a function of the vote lead.

Lemma 1. *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 $\Delta - 1$. □

A.3.1 Partisans' behavior

Proposition 4. *If all other players are playing PBV and the election is undecided at the current history, it is strictly optimal for a Partisan to vote for her preferred alternative.*

Proof. Consider an L-partisan (the argument is analogous for an R-partisan). The expected utility from voting L is

$$\mu\left(h^i, s_i\right) P\left(\Psi\left(h^i, L\right), \Delta+1, K, L\right) + \left(1 - \mu\left(h^i, s_i\right)\right) P\left(\Psi\left(h^i, L\right), \Delta+1, K, R\right), \quad (7)$$

while the expected utility from voting R is

$$\mu\left(h^i, s_i\right) P\left(\Psi\left(h^i, R\right), \Delta-1, K, L\right) + \left(1 - \mu\left(h^i, s_i\right)\right) P\left(\Psi\left(h^i, R\right), \Delta-1, K, R\right). \quad (8)$$

Since $K > \Delta - 1$ when the election is undecided, Lemma 1 implies that (7) > (8), which proves the result. □

A.3.2 Neutrals in the Herding Phase

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. As usual, let a profile of type and signal realizations for all other voters apart from i be denoted

$$(t_{-i}, s_{-i}) \equiv ((t_1, s_1), ..., (t_{i-1}, s_{i-1}), (t_{i+1}, s_{i+1}), ..., (t_n, s_n)).$$

Given that other players are playing PBV, for any realized profile (t_{-i}, s_{-i}), i's vote deterministically selects a winner because PBV does not involve mixing. For a vote by Voter i, $V_i \in \{L, R\}$, denote the winner of the election $x\left(V_i; (t_{-i}, s_{-i})\right) \in \{L, R\}$. Then, denote the event in which Voter i is pivotal as $Piv_i = \{(t_{-i}, s_{-i}) : x\left(L; (t_{-i}, s_{-i})\right) \neq x\left(R; (t_{-i}, s_{-i})\right)\}$. By arguments identical to Lemma 1, 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,

$$Piv_i = \{(t_{-i}, s_{-i}) : x\left(L, (t_{-i}, s_{-i})\right) = L \text{ and } x\left(R, (t_{-i}, s_{-i})\right) = R\}. \tag{9}$$

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

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

It follows from equation (9) that $U\left(L | h^i, s_i, Piv_i\right) = \Pr\left(\omega = L | h^i, s_i, Piv_i\right)$ and $U\left(R | h^i, s_i, Piv_i\right) = 1 - \Pr\left(\omega = L | h^i, s_i, Piv_i\right)$. Therefore, if $\Pr\left(\omega = L | h^i, l, Piv_i\right) > \frac{1}{2}$, it is strictly optimal for a Neutral Voter i to vote for L, and if $\Pr\left(\omega = L | h^i, l, Piv_i\right) < \frac{1}{2}$, it is strictly optimal for a Neutral Voter i to vote R.

Proposition 5. *If all other players 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 alternative C if $\Psi(h^i) = C$, for all $C \in \{L, R\}$.*

Proof. Consider a history, h^i, where $\Psi\left(h^i\right) = L$. Since all future Neutral voters vote uninformatively for L, $\Pr\left(\omega = L | h^i, s_i, Piv_i\right) = \Pr\left(\omega = L | h^i, s_i\right)$, which by construction strictly exceeds $\frac{1}{2}$ for all s_i (since $\Psi\left(h^i\right) = L$). Therefore, a Neutral voter strictly prefers to vote L. An analogous argument applies when $\Psi(h^i) = R$. $\square$

A.3.3 Neutrals in the Learning Phase

Proposition 6. *If all other players are playing PBV and the election is undecided at the current history, h^i, it is (generically, strictly) optimal for a Neutral Voter i to vote informatively when $\Psi\left(h^i\right) = 0$ and $\Delta(h^i) \in \{-n_R\left(i+1\right) + 1, \ldots, n_L\left(i+1\right) - 1\}$.*

We prove this Proposition in a series of steps. Lemma 2 below shows that if the incentive constraints hold for certain voters at certain histories of the learning phase, then they hold for all other possible histories in the learning phase. This simplifies the verification of many incentive constraints to that of a few important constraints. We then verify that those constraints also hold in Lemmas 3, 6, and 7.

Lemma 2. *Consider any h^i where $\Psi(h^i) = 0$ and $\Delta(h^i) = \Delta$. Then if it is incentive compatible for Neutral Voter $(i+1)$ to vote informatively when $\Delta(h^{i+1}) \in \{\Delta - 1, \Delta + 1\}$, then it is incentive compatible for Neutral Voter i to vote informatively when $\Delta(h^i) = \Delta$. Moreover, if the incentive compatibility condition for Neutral Voter $(i+1)$ holds strictly at least in one of the two cases when $\Delta(h^{i+1}) \in \{\Delta - 1, \Delta + 1\}$, then it holds strictly for Neutral Voter i.*

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. It is necessary and sufficient that

$$\begin{array}{c} \tilde{\mu}_i(\Delta, l)\left[P(0, \Delta+1, K, L) - P(0, \Delta-1, K, L)\right] \\ -(1-\tilde{\mu}_i(\Delta, l))\left[P(0, \Delta+1, K, R) - P(0, \Delta-1, K, R)\right] \end{array} \geq 0. \tag{10}$$

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

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

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

$$P\left(\Psi\left(h^i\right), \Delta, K+1, \omega\right) = p(\omega) P\left(\Psi\left(h^i, L\right), \Delta+1, K, \omega\right) + q(\omega) P\left(\Psi\left(h^i, R\right), \Delta-1, K, \omega\right),$$

it follows that the above inequality holds if and only if

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

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

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

We now argue that each of the two lines of the right hand side above is non-negative.

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

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

   Since IC holds for Voter $i+1$ with vote lead $\Delta+1$, observe that if the election is undecided for $i+1$, $\frac{\tilde{\mu}_{i+1}(\Delta+1,l)}{1-\tilde{\mu}_{i+1}(\Delta+1,l)} \geq \frac{P\left(\Psi\left(h^i,L,L\right),\Delta+2,K-1,R\right)-P(0,\Delta,K-1,R)}{P(\Psi(h^i,L,L),\Delta+2,K-1,L)-P(0,\Delta,K-1,L)}$, which proves that the first line of the desired right hand side is non-negative. If the election is decided for $i+1$ with vote lead $\Delta+1$, then the first line of the desired right hand side is exactly 0.

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 $\Delta-1$, observe that if the election is undecided for $i-1$, then $\frac{\tilde{\mu}_{i+1}(\Delta-1,l)}{1-\tilde{\mu}_{i+1}(\Delta-1,l)} \frac{q(L)}{q(R)} \geq \frac{P(0,\Delta,K-1,R)-P\left(\Psi\left(h^i,R,R\right),\Delta-2,K-1,R\right)}{P(0,\Delta,K-1,L)-P(\Psi(h^i,R,R),\Delta-2,K-1,L)}$, and thus the second line of the desired right hand side is non-negative. If the election is decided for $i+1$ with vote lead $\Delta-1$, then the second line of the desired right hand side is exactly 0.

Observe that if incentive compatibility holds strictly for Voter $i+1$ in either one of the two cases, then at least one of the two lines of the right hand side is strictly positive, and consequently inequality (10) must hold strictly. □

By the above Lemma, we are left to only check the incentive conditions for a Neutral Voter i with undecided history h^i such that $\Psi(h^i)=0$, but Voter $i+1$ will not vote informatively when Neutral if either $v_i = L$ or $v_i = R$. This possibility can be divided into two cases:

1. either i's vote causes the phase to transition into a herding phase; or
2. i is the final voter ($i = n$) and $\Delta(h^n) = 0$.

Lemma 3 below deals with the latter case; Lemmas 6 and 7 concern the former. (Lemmas 4 and 5 are intermediate steps towards Lemma 6.)

Lemma 3. *If there exists a history h^n such that $\Psi(h^n) = 0$ and Voter n is pivotal, then it is incentive compatible for Voter n to vote informatively. For generic parameters of the game, the incentive compatibility conditions hold strictly.*

Proof. Since n is the final voter and pivotal, $\Pr(\omega = L|h^n, s_n, Piv_n) = \Pr(\omega = L|h^n, s_n)$. Since $\Psi(h^n) = 0$, $\Pr(\omega = L|h^n, l) \geq \frac{1}{2} \geq \Pr(\omega = L|h^n, r)$. Therefore, voting informatively is incentive compatible. Recall from Remark 1 that $\Pr(\omega = L|h^n, s_n) = \frac{1}{2}$ for some $s_n \in \{l, r\}$ only if $(\pi, \gamma, \tau_L, \tau_R)$ is such that $(\tau_L, \tau_R) \in \Gamma_{\pi,\gamma}$, which is a set of (Lebesgue) measure 0. If $(\tau_L, \tau_R) \notin \Gamma_{\pi,\gamma}$, then given that $\Psi(h^n) = 0$, $\Pr(\omega = L|h^n, l) > \frac{1}{2} > \Pr(\omega = L|h^n, r)$, and therefore for generic parameters, voting informatively is strictly optimal for Voter n. □

Lemma 4. *Consider any h^i where $\Psi(h^i) = 0$ and $\Delta(h^i) = \Delta$. Then, $P(\Psi(h^i, l), \Delta + 1, K, L) \geq P(\Psi(h^i, l), \Delta + 1, K, R)$ and $P(\Psi(h^i, r), \Delta - 1, K, L) \geq P(\Psi(h^i, r), \Delta - 1, K, R)$ implies $P(0, \Delta, K + 1, L) \geq P(0, \Delta, K + 1, R)$.*

Proof. Simple manipulations yield

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

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 Lemma. □

Lemma 5. *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 $\Delta \in \{-n_R + 1, \ldots, n_L - 1\}$. 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 $\Delta \in \{-n_R + 1, \ldots, n_L - 1\}$. □

Lemma 6. *Consider history h^i such that $\Delta(h^i) = n_L(i+1) - 1$ 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. Consider h^i such that $\Delta(h^i) = n_L(i+1) - 1 = \Delta$, and $\Psi(h^i) = 0$. For it to be strictly optimal for the voter to vote informatively, it must be that

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

which is equivalent to

$$\frac{\tilde{\mu}_i(\Delta, r)}{1 - \tilde{\mu}_i(\Delta, r)} < \frac{P(L, \Delta + 1, K, R) - P(0, \Delta - 1, K, R)}{P(L, \Delta + 1, K, L) - P(0, \Delta - 1, K, L)}. \tag{11}$$

By Lemma 5, $P(0, \Delta - 1, K, L) \geq P(0, \Delta - 1, K, R)$, and by definition, $P(L, \Delta + 1, K, R) = P(L, \Delta + 1, K, L)$. Therefore, the right-hand side of (11) is bounded below by 1. Since $\Psi(h^i) = 0$, $\mu(h^i, r) = \tilde{\mu}_i(\Delta, r) < \frac{1}{2}$, the left-hand side of (11) is strictly less than 1, establishing the strict inequality. An analogous argument applies to prove the case where $\Delta(h^i) = -n_R(i+1) + 1$ and $s_i = l$. □

Lemma 7. *Consider history h^i such that $\Delta(h^i) = n_L(i+1) - 1$ and $\Psi(h^i) = 0$. Then if all other voters are playing PBV, and a Neutral Voter i receives signal l, it is optimal for her to vote L. 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 r, it is optimal to vote R. For generic parameters of the game, the optimality is strict.*

Proof. Consider the information set where $\Psi(h^i) = 0$, $\Delta(h^i) = n_L(i+1) - 1$, and $s_i = l$. It suffices to show that $\Pr(\omega = L|h^i, l, Piv_i) \geq \frac{1}{2}$. 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 $\left(\Psi\left(h^{k-1}\right), \Psi\left(h^k\right)\right) = (0, \Psi)$ where $\Psi \in \{L, R\}$ after $v_i = R$, and where i's vote is pivotal. Let $K^\Psi = \left\{k > i : \xi_k^\Psi \neq \emptyset\right\}$. 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 $K_\Delta^0 = \left\{\Delta : \xi_\Delta^0 \neq \emptyset\right\}$. Observe that since the event $(h^i, Piv_i) = \cup_\Psi \left(\cup_{k \in K^\Psi} \xi_k^\Psi\right) \cup \left(\cup_{\Delta \in K_\Delta^0} \xi_\Delta^0\right)$, by the definition of conditional probability

$$\begin{aligned}\Pr\left(\omega = L|h^i, l, Piv_i\right) &= \sum_{\Psi \in \{L,R\}} \sum_{k \in K^\Psi} \Pr\left(\xi_k^\Psi|h^i, l, Piv_i\right) \Pr\left(\omega = L|\xi_k^\Psi, l\right) \\ &\quad + \sum_{\Delta \in K_\Delta^0} \Pr\left(\xi_\Delta^0|h^i, l, Piv_i\right) \Pr\left(\omega = L|\xi_\Delta^0, l\right).\end{aligned}$$

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

Consider $k \in K^L$: ξ_k^L denotes a set of signal-type realizations that induce an L-herd after the vote of Voter $(k-1)$ (and meet the aforementioned conditions). Since only votes in the learning phase are informative,

$$\Pr\left(\omega = L|\xi_k^L, l\right) = \Pr\left(\omega = L|l, \Psi\left(h^{k-1}\right) = 0, \Delta\left(h^k\right) = n_L(k)\right).$$

Given that $v_i = R$, the informational content of this event is equivalent to a history $\tilde{h}^{k-1}$ where $\Delta\left(\tilde{h}^{k-1}\right) = n_L(k) + 1$, and all Neutrals are assumed to have voted informatively. Therefore,

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

Observe that $g_{k-1}(n_L(k) + 1) > g_k(n_L(k)) > \frac{(1-\pi)\gamma}{\pi(1-\gamma)}$. Therefore, $\Pr\left(\omega = L|\xi_k^\Psi, l\right) > \frac{\gamma^2}{\gamma^2 + (1-\gamma)^2} > \frac{1}{2}$.

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

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

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

Now consider the event $\Delta \in K_\Delta^0$: ξ_Δ^0 denotes a set of signal-type realizations that induce no herd and a final vote lead of $\Delta < 0$. Therefore,

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

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

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

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

Therefore, $\Pr\left(\omega = L|h^i, l, Piv_i\right)$ is a convex combination of numbers that are bounded below by $\frac{1}{2}$.

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

We complete the proof by explaining why the incentive constraints are satisfied strictly for generic parameters of the game. From the arguments above, indifference arises only if there exists some $k \leq n$ and history h^k such that $\Pr\left(\omega = L|h^k, s_k\right) = \frac{1}{2}$. Recall from Remark 1 that this can hold only if $(\pi, \gamma, \tau_L, \tau_R)$ is such that $(\tau_L, \tau_R) \in \Gamma_{\pi,\gamma}$, which is a set of (Lebesgue) measure 0. If $(\tau_L, \tau_R) \notin \Gamma_{\pi,\gamma}$, then for every k, $\Psi\left(h^k\right) = 0$ implies that $\Pr\left(\omega = L|h^k, l\right) > \frac{1}{2} > \Pr\left(\omega = L|h^k, r\right)$. Therefore, for generic parameters of the game, following PBV is strictly optimal for Voter n regardless of history. □

Lemmas 2, 3, 6, and 7 establish Proposition 6: conditional on all others playing according to PBV, it is optimal for Neutrals to vote informatively in the learning phase. Observe that generic parameters of the game yield strict optimality of the incentive conditions in Lemmas 3 and 7, and therefore, by Lemma 2, all the incentive conditions in the learning phase hold strictly generically.

A.4 Theorem 2

The proof of Theorem 2 consists of two steps: first, we show that there must almost surely be a herd in the limit as the population size $n \to \infty$; 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 inverse of the likelihood ratio λ_i.)

Step 1: By the Martingale Convergence Theorem for non-negative random variables (Billingsley, 1995, 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 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 (2000) 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 $\Pr\left(\lambda_\infty \in \{\underline{b},\bar{b}\}\right) > 0$. Without loss of generality, assume $\Pr(\lambda_\infty = \underline{b}) > 0$; the argument is analogous if $\Pr\left(\lambda_\infty = \bar{b}\right) > 0$. 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

[25] To 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 $i > k$, $\lambda_k \in (\notin)\ S$.

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

A.5 Theorem 3

We consider a PBV strategy profile and consider two threshold rules q and q' where $\tau_L < q < q' < 1 - \tau_R$. Given a profile of n votes, let S_n denote the total number of votes cast in favor of L.

Pick $\varepsilon > 0$. From Theorem 2, we know that there exists $\bar{k}$ such that for all $k \geq \bar{k}$, $\Pr\left(\Psi\left(h^k\right) = L\right) + \Pr\left(\Psi\left(h^k\right) = R\right) > 1 - \frac{\varepsilon}{2}$. Pick any $k \geq \bar{k}$. By the Weak Law of Large Numbers, for every $\kappa > 0$, $\lim_{n\to\infty} \Pr\left(\left|\frac{S_n}{n} - (1 - \tau_R)\right| < \kappa | \Psi\left(h^k\right) = L\right) = 1$ and $\lim_{n\to\infty} \Pr\left(\left|\frac{S_n}{n} - \tau_L\right| < \kappa | \Psi\left(h^k\right) = R\right) = 1$. Pick some $\kappa < \min\left\{(1 - \tau_R) - q', q - \tau_L\right\}$. There exists some $\bar{n} > k$ such that for all $n \geq \bar{n}$, $\Pr\left(\left|\frac{S_n}{n} - (1 - \tau_R)\right| < \kappa | \Psi\left(h^k\right) = L\right) > 1 - \frac{\varepsilon}{2}$ and $\Pr\left(\left|\frac{S_n}{n} - \tau_L\right| < \kappa | \Psi\left(h^k\right) = R\right) > 1 - \frac{\varepsilon}{2}$. Observe that by the choice of κ, $\left|\frac{S_n}{n} - (1 - \tau_R)\right| < \kappa$ implies that L wins under both rules q and q' whereas $\left|\frac{S_n}{n} - \tau_L\right| < \kappa$ implies that L loses under both rules q and q'. For any $n \geq \bar{n}$, we have

$$\begin{aligned}
& |Pr(\text{L wins in } G\left(\pi, \gamma, \tau_L, \tau_R; n, q\right) - Pr(\text{L wins in } G\left(\pi, \gamma, \tau_L, \tau_R; n, q'\right)| \\
< & \left|\sum_{x \in \{L,R\}} \Pr\left(\Psi\left(h^k\right) = x\right) \Pr\left(\frac{S_n}{n} \in (q, q'] \,|\Psi\left(h^k\right) = x\right)\right| + \frac{\varepsilon}{2} \\
< & \ \varepsilon,
\end{aligned}$$

which proves part (a) of the Theorem.

For part (b), consider any $q < \tau_L$. The probability with which a voter votes L is at least τ_L. Therefore, invoking the Weak Law of Large Numbers, $\lim_{n\to\infty} \Pr\left(\frac{S_n}{n} < q\right) = 0$. The argument is analogous for part (c).

A.6 Proposition 2

The logic parallels that of Theorem 1. We omit the verification that Neutral voters are better off voting informatively in the learning phase since this is a straightforward translation of Lemma 6. To verify incentives in the herding phase, observe that with the off-equilibrium path belief specification that votes contrary to the herd are ignored, no voter is strictly better off by voting contrary to the herd.

Appendix B (Supplementary, not for publication): Cut-Point Voting

In this Appendix, we consider the class of *Cut-Point Voting* (CPV) strategy profiles introduced by Callander (2007). While this class does entail some restrictions, it covers a range of strategy profiles, generalizing PBV to permit behavior ranging from fully informative to uninformative voting (for Neutrals). We prove that for generic parameters, any equilibrium within this class leads to herding with high probability in large elections. This result is of interest because it suggests that our conclusions concerning momentum are more general beyond the PBV equilibrium, with the important caveat that we do not know whether non-PBV but CPV equilibria generally exist. Indeed, Proposition 8 shows that in the special case where $\tau_L = \tau_R$ and for generic values of (π, γ), PBV is the only history-dependent CPV equilibrium of large voting games.

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 exist* $0 \leq \mu_* \leq \mu^* \leq 1$ *such that for every Voter* i*, history* h^i*, and signal* s_i*,*

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

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 $[\mu_*, \mu^*]$; otherwise, a Neutral votes for one of the alternatives independently of her private signal. Denote a CPV profile with belief thresholds μ_* and μ^* as $CPV(\mu_*, \mu^*)$. These 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 PBV. Similarly, $CPV(0,0)$ and $CPV(1,1)$ represent strategy profiles where every Neutral votes uninformatively for alternative L and R respectively. Therefore, CPV captures a variety of behavior for Neutrals.

A CPV equilibrium is an equilibrium whose 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 show that generically, large elections lead to herds with high probability within the class of CPV equilibria.

Theorem 4. *For every* $(\pi, \gamma, \tau_L, \tau_R)$ *such that* $\tau_L \neq \tau_R$, *and for every* $\varepsilon > 0$, *there exists* $\overline{n} < \infty$ *such that for all* $n > \overline{n}$, *if voters play a CPV equilibrium,* $\Pr[$*a herd develops in* $G(\pi, \gamma, \tau_L, \tau_R; n)] > 1 - \varepsilon$.

Proof. We argue through a succession of lemmas that there exist $\overline{\mu}^* < 1$ 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 $\underline{\mu}_* < \mu_* < \mu^* < \overline{\mu}^*$. This suffices to prove the Theorem, because then, the arguments of Theorem 2 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 $CPV(\mu_*, \mu^*)$, 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 thresholds sequences $\tilde{n}_L$ and $\tilde{n}_R$, we define the phase mapping $\tilde{\Psi} : h^i \to \{L, 0, R\}$ in the obvious way that extends the PBV phase mapping Ψ. We state without proof the following generalization of Proposition 1.

Proposition 7. *Fix a parameter set* $(\pi, \gamma, \tau_L, \tau_R, n)$. *For each* $i \leq n$, *if voters play* $CPV(\mu_*, \mu^*)$ *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 8. *There exists* $\bar{\mu}^* < 1$ *and* $\underline{\mu}_* > 0$ *such that in any* $CPV(\mu_*, \mu^*)$,

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

Proof. Define $\bar{\mu}^*$ by the equality $\frac{\bar{\mu}^*}{1-\bar{\mu}^*} = \frac{\gamma}{1-\gamma} f(\tau_L, \tau_R) f(\tau_R, \tau_L)$ and define $\underline{\mu}_*$ by $\frac{\underline{\mu}_*}{1-\underline{\mu}_*} = \frac{1-\gamma}{\gamma} \left(f\left(\tau_L, \tau_R\right) f\left(\tau_R, \tau_L\right)\right)^{-1}$. We give the argument for part (1); it is similar for part (2). 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 $\tilde{n}_L(i) \leq n_L(i) < i$. By the definitions of $n_L(i)$ and $\tilde{n}_L(i)$, and the monotonicity of $g_i(k)$ in k,

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

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

Lemma 9. *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 4 establish this result. □

Lemma 10. *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 $t \in \{L, R, N\}$, $\lim_{n\to\infty} \Pr\left(\left|\frac{\varsigma_t(n)}{n} - \tau_t\right| < \varepsilon\right) = 1$. Consequently, since $\tau_L > \tau_R$, for any $\varepsilon > 0$ and $k > 0$, there exists $\bar{n}$ such that for all $n > \bar{n}$, $\Pr\left(\varsigma_L\,(n) - \varsigma_R\,(n) > k\right) > 1 - \varepsilon$. 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 $n > \bar{n}$, $\Pr\left(\varsigma_{N,r}\,(n) - \varsigma_{N,l}\,(n) > k | Piv_1\right) > 1 - \varepsilon$. Since $\Pr\left(\omega = L | \varsigma_{N,r}\,(n)\,, \varsigma_{N,l}\,(n)\right)$ 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 $\Pr\left(\omega = L | Piv_1\right) < \frac{1}{2}$ 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 CPV strategy $(0, 1)$. $\square$

Lemma 11. *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 $\mu_* \in \left(\frac{1}{2}, \pi\right]$, let h^{k+1} be a history of k consecutive R votes. It is straightforward that for some integer $k \geq 1$, $\mu\left(h^k\right) \geq \mu_* > \mu\left(h^{k+1}\right)$. 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 $\mu_{k+1}\left(h^{k+1}, l\right) > \mu\left(h^k\right) \geq \mu_* > \frac{1}{2}$, 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 12. *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.

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

<u>Claim 2</u>: $\Pr(Piv_m | \omega = R)$ converges to 0 as the electorate size $n \to \infty$. *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 $\Pr(X^n) \to 0$, $\Pr(Y^n) \to 0$, and $\Pr(Z^n) \to 0$. That $\Pr(X^n) \to 0$ and $\Pr(Y^n) \to 0$ follows straightforwardly from Claim 1. To show that $\Pr(Z^n) \to 0$, 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, $\Pr(\Psi_n^n = 0) < \varepsilon$ 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, $\Pr\left(Z^{n'}\right) = \Pr(\Xi^{n,0}) \Pr\left(Z^{n'} | \Xi^{n,0}\right) + \Pr\left(\Xi^{n,R}\right) \Pr\left(Z^{n'} | \Xi_n^R\right) < \varepsilon + \Pr\left(\Xi^{n,R}\right) \Pr\left(Z^{n'} | \Xi_n^R\right)$ for large enough n. We have $\Pr\left(Z^{n'} | \Xi^{n,R}\right) = \dfrac{\Pr\left(Z^{n'} \cap \Xi^{n,R}\right)}{\Pr\left(\Xi^{n,R}\right)}$. It is straightforward to see that $\Pr\left(Z^{n'} \cap \Xi^{n,R}\right) \to 0$ as $n' \to \infty$, using the fact that $\tau_L < 1 - \tau_L$ and invoking the Weak Law of Large Numbers similarly to Lemma 10. Note that $\Pr\left(\Xi^{n,R}\right)$ 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 $\Pr\left(Z^{n'}\right) \to 0$. ‖

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 $n \to \infty$. *Proof*: Define $\xi(h^i) = \frac{\Pr(\omega = R | h^i)}{\Pr(\omega = L | h^i)}$; this generates a stochastic process $\langle \xi_i \rangle$ $(i = m+1, m+2, \ldots)$ which is a martingale conditional on state L, and thus $\langle \xi_i \rangle \overset{a.s.}{\to} \xi_\infty$. 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 2 allows us to conclude that $Support(\xi_\infty) \subseteq \{0\} \cup \left[\frac{1-\mu_*}{\mu_*}, \infty\right)$ and $\Pr\left(\xi_\infty = \frac{1-\mu_*}{\mu_*}\right) = 0$. Suppose towards contradiction that $0 \notin Support(\xi_\infty)$. This implies $\mathbb{E}[\xi_\infty] > \frac{1-\mu_*}{\mu_*}$. By Fatou's Lemma (Billingsley, 1995, p. 209), $\mathbb{E}[\xi_\infty] \leq \lim_{n \to \infty} \mathbb{E}[\xi_n]$; 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, $0 \in Support(\xi_\infty)$, and it must be that $\Pr(\xi_\infty = 0) > 0$. The claim follows from the observation that for any history sequence where $\xi_i(h^i) \to 0$ it must be that $\Delta(h^i) \to \infty$. ‖

Consider the expected utility for Voter m from voting R or L respectively, conditional on being pivotal: $EU_m(v_m = R|Piv_m) = \Pr(\omega = R|Piv_m)$ and $EU_m(v_m = L|Piv_m) = \Pr(\omega = L|Piv_m)$. Thus, she strictly prefers to vote L if and only if $\Pr(\omega = L|Piv_m) > \Pr(\omega = R|Piv_m)$, or equivalently, if and only if $\Pr(Piv_m|\omega = L) > \Pr(Piv_m|\omega = R)\frac{1-\mu_m(h^m,r)}{\mu_m(h^m,r)}$. By Claim 2, $\Pr(Piv_m|\omega = R)$ converges to 0 as electorate grows. On the other hand, $\Pr(Piv_m|\omega = L)$ 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, $\Pr(Piv_m|\omega = L) > \Pr(Piv_m|\omega = R)\frac{1-\mu_m(h^m,r)}{\mu_m(h^m,r)}$, and it is strictly optimal for m to vote L following his signal $s_m = r$, which is a deviation from the CPV strategy. □

Lemma 13. *In a large enough election, $CPV(0, \mu^*)$ is not an equilibrium for any $\mu^* \in [\pi, 1)$.*

Proof. Analogous to Lemma 12, 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. □

Lemma 14. *In a large enough election, $CPV(\mu_*, \mu^*)$ is not an equilibrium if $\mu^* \in [\bar{\mu}^*, 1)$ and $\mu_* \in (0, \frac{1}{2}]$.*

Proof. Fix an equilibrium $CPV(\mu_*, \mu^*)$ with $\mu^* \in [\bar{\mu}^*, 1)$ and $\mu_* \in (0, \frac{1}{2}]$. By Lemma 8, $\tilde{n}_L(i) > n_L(i)$ 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 2, 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 12, where n indexes the electorate size. Plainly, $\Pr(Y^n) \to 0$. The argument of Claim 2 in Lemma 12 implies with obvious modifications that $\Pr(Z^n) \to 0$. Finally, $\Pr(X^n) \nrightarrow 0$ because there exists $m' > m$ such that if $v_i = L$ for all $i \in \{m+1, \ldots, m'\}$, then $\Psi^n_{m'} = L$, and $\Pr(v_i = L \text{ for all } i \in \{m+1, \ldots m'\}) \geq (\tau_L)^{m'-m} > 0$. Since $Piv^n_m = X^n \cup Y^n \cup Z^n$, we conclude that as $n \to \infty$, $\Pr(X^n|Piv^n_m) \to 1$, whereas $\Pr(Y^n|Piv^n_m) \to 0$ and $\Pr(Z^n|Piv^n_m) \to 0$. Consequently, for any $\varepsilon > 0$, there exists $\bar{n}$ such that for all $n > \bar{n}$,

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

and

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

Therefore, it suffices to show that for any $n > m$,

$$EU_m\left(v_m = L|X^n, s_m = r\right) > EU_m\left(v_m = R|X^n, s_m = r\right),$$

or equivalently, $\Pr\left(\omega = L|X^n, s_m = r\right) > \Pr\left(\omega = R|X^n, s_m = r\right)$. For any $k \in \{m+1, \ldots, n\}$, define

$$X_k^n = \left\{(t_{-m}, s_{-m}) \in Piv_m^n : \ \Psi_{k-1}^n = 0 \text{ but } \Psi_k^n = L \text{ after } v_m = L, \ \Psi_n^n \text{ after } v_m = R\right\}$$

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 $\Pr\left(\omega|X^n, s_m = r\right) = \cup_{k=m+1}^n \Pr\left(\omega|X_k^n, s_m = r\right)\Pr\left(X_k^n|X^n\right)$. It therefore suffices to show that for any $k \in \{m+1, \ldots, n\}$, $\Pr\left(\omega = L|X_k^n, s_m = r\right) > \Pr\left(\omega = R|X_k^n, s_m = r\right)$. Given that $v_m = L$, the informational content of X_k^n is equivalent to a history h^{k+1} where $\Delta\left(h^{k+1}\right) = \tilde{n}_L(k+1) - 2$, and all Neutrals are assumed to have voted informatively. Therefore,

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

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,

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

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

Lemma 15. *In a large enough election, $CPV(\mu_*, \mu^*)$ is not an equilibrium if $\mu^* \in [\pi, 1)$ and $\mu_* \in \left(0, \underline{\mu}_*\right]$.*

Proof. Analogous to Lemma 14, 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. □

While the above theorem shows that for generic constellation of parameters, a $CPV(\mu_*, \mu^*)$ profile is an equilibrium in large elections if and only if $\underline{\mu}_* < \mu_* < \mu^* < \overline{\mu}^*$, we do not know whether there exists any CPV equilibrium that meets that restriction other than

PBV. In the special case in which $\tau_L = \tau_R$, we now show that for generic values of (π, γ), the only *history-dependent* CPV equilibrium of large elections is indeed PBV.[26]

Proposition 8. *Suppose that $\tau_L = \tau_R = \tau$. Then for every q-rule and for almost all (π, γ, τ), in a large enough election, the only history-dependent CPV equilibrium is PBV.*

Proof. Consider a history-dependent CPV equilibrium with thresholds μ_* and μ^*. We will argue that for large enough n, this must coincide with PBV for $\tau_L = \tau_R$ and generic (π, γ, τ). For the profile to be history-dependent requires either $\mu_* \neq 0$ or $\mu^* \neq 1$. All the other arguments of Theorem 4 apply, and therefore $\underline{\mu}_* < \mu_* < \mu^* < \overline{\mu}^*$. This establishes that there exists i with herding thresholds $\tilde{n}_L(i) < i$ and $\tilde{n}_R(i) < i$. The hypothesis that $\tau_L = \tau_R = \tau$ implies that $g_i(\Delta) = (f(\tau, \tau))^{\Delta}$, independent of i. Therefore, $\tilde{n}_L(i) = \tilde{n}_L(j)$ for all voters i, j such that $|i - j|$ is even and $i, j > \tilde{n}_L(i)$; and analogously for the sequence $\{\tilde{n}_R(i)\}_{i=1}^{\infty}$.

<u>Step 1</u>: We first argue that for each i, $\tilde{n}_L(i) \geq n_L(i)$. Suppose towards contradiction that there exists i such that $\tilde{n}_L(i) < n_L(i)$. For generic values of (π, γ, τ), since g is monotonic in its argument, $g_i(\tilde{n}_L(i)) \leq g_i(n_L(i) - 2) < \frac{(1-\pi)\gamma}{\pi(1-\gamma)}$.[27] Consider undecided history h^i in which $\tilde{\Psi}(h^{i-1}) = 0$ and $\Delta(h^i) = \tilde{n}_L(i)$; the CPV profile prescribes that all subsequent Neutral voters vote L regardless of signal. Suppose that Voter i is Neutral and receives signal r: as in Lemma 5, since all subsequent voters are voting uninformatively, $\Pr(\omega = L | h^i, r, Piv_i) = \Pr(\omega = L | h^i, r)$ which can be verified as less than $\frac{1}{2}$ because $g_i(\tilde{n}_L(i)) < \frac{(1-\pi)\gamma}{\pi(1-\gamma)}$. Therefore, conditioning on being pivotal, Voter i would prefer to deviate from the CPV and vote R. An analogous argument establishes that $\tilde{n}_R(i) \geq n_R(i)$.

<u>Step 2</u>: The proof is completed by showing that there cannot be an i with either $\tilde{n}_L(i) > n_L(i)$ or $\tilde{n}_R(i) > n_R(i)$. Suppose otherwise: in particular, that there is a Voter i^* with $\tilde{n}_L(i^*) > n_L(i^*)$ (the argument is analogous for the R thresholds, hence omitted). Let $k = \max\{i : n_L(i) = i\}$ denote the voter with the highest index who never herds for L in any history in a PBV profile (it is straightforward to show such a k exists). Let $J = \{i : i > k \text{ and } \tilde{n}_L(i) = n_L(i)\}$. By Step 1, we can restrict attention to the case where for any i such that $i > k$ and $i \notin J$, $\tilde{n}_L(i) > n_L(i)$. There are two exclusive and exhaustive cases to consider.

First, assume $J = \emptyset$. Consider the choice of Voter m given signal $s_m = r$ and history h^m such that $\tilde{\Psi}(h^m) = 0$ and $\tilde{\Psi}(h^m, R) = R$. An argument analogous to that in Lemma

[26]To be clear, a pure strategy profile is history-dependent if there exist some voter i, type t_i, signal s_i, and histories h^i, $\tilde{h}^i$ such that both histories are undecided and $v_i(t_i, s_i, h^i) \neq v_i\left(t_i, s_i, \tilde{h}^i\right)$.

[27]The role that genericity plays is in establishing the strictness of the second inequality. Given a (π, γ), $g_i(k) = (f(\tau, \tau))^k$ is in $\left\{\frac{(1-\pi)\gamma}{\pi(1-\gamma)}, \frac{(1-\pi)(1-\gamma)}{\pi\gamma}\right\}$ for only countably many τ.

14 establishes that for sufficiently large populations, Voter m would strictly prefer to deviate from the CPV prescription and vote L.

Second, assume $J \neq \emptyset$. Since $\tilde{n}_L(i) = \tilde{n}_L(i+2)$ for any $i > k$, it must be that if $j \in J$ then $j+1 \notin J$; moreover $\tilde{n}_L(j+2m+1) = \tilde{n}_L(j+1) > n_L(j+1) = n_L(j+2m+1)$ for all positive integers m. We also note that if $j \in J$, $g_{j+1}(\tilde{n}_L(j)+1) = (f(\tau,\tau))^{\tilde{n}_L(j)+1} > \frac{(1-\pi)\mu^*}{\pi(1-\mu^*)}$, and consequently $\tilde{n}_L(j+1) \leq \tilde{n}_L(j)+1$. Similarly, $n_L(j+1) \geq n_L(j)-1$, and therefore, $\tilde{n}_L(j+1) = n_L(j)+2$ since $\tilde{n}_L(j) = n_L(j)$. Thus, for $j \in J$, $\tilde{\Psi}(h^{j+1}) = L$ implies that $\tilde{\Psi}(h^j) = L$, and similarly $\Psi(h^j) = L$ implies that $\Psi(h^{j-1}) = L$. For $j' > k$ and $j' \notin J$, let $\bar{\Delta}$ denote $n_L(j')$. The foregoing establishes that under PBV, L-herds can begin only on those voters with indices $j' > k$ and not in the set J and are triggered when the vote lead reaches $\bar{\Delta}$; however under the alternative CPV, L-herds begin only on those voters with indices $j \in J$ and are triggered when the vote lead reaches $\bar{\Delta}+1$. Let us consider the incentives of Voter m, h^m such that $\left(\tilde{\Psi}(h^m), \Delta(h^m)\right) = \left(0, -\tilde{\Delta}\right)$, and $\tilde{\Psi}(h^m, R) = R$, and $s_m = r$ in the alternative CPV profile. Suppose that Voter m deviates and votes L, and consider the event in which this deviation triggers a future L-herd. An L-herd begins if and only if following the deviation, there are on net $\left(\tilde{\Delta}-1\right)+\left(\bar{\Delta}+1\right)$ more votes for L; since $\Delta(h^m) = -\tilde{\Delta}$, an L-herd following a deviation reveals an excess of $\bar{\Delta}$ votes for L. Voter m's posterior conditional on her signal, deviation, and the subsequent L-herd is therefore $\frac{\pi(1-\gamma)(f(\tau,\tau))^{\bar{\Delta}}}{\pi(1-\gamma)(f(\tau,\tau))^{\bar{\Delta}}+(1-\pi)\gamma} > \frac{1}{2}$, since by construction, a vote lead of $\bar{\Delta}$ suffices to trigger an L-herd in a PBV profile. Therefore, conditional on a deviation inducing a future L-herd, Voter m believes that it is more probable that $\omega = L$. Using arguments from Lemma 14, Voter m then strictly prefers to vote L in sufficiently large elections, contrary to the prescription of CPV. □

References

Austen-Smith, D. and J. S. Banks (1996): "Information Aggregation, Rationality, and the Condorcet Jury Theorem," *The American Political Science Review*, 90, 34–45. 4, 8, 9, 15

Banerjee, A. (1992): "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107, 797–817. 1

Bartels, L. M. (1988): *Presidential Primaries and the Dynamics of Public Choice*, Princeton, NJ: Princeton University Press. 1, 3, 5, 16, 23

Battaglini, M. (2005): "Sequential Voting with Abstention," *Games and Economic Behavior*, 51, 445–463. 5

Battaglini, M., R. Morton, and T. Palfrey (2007): "Efficiency, Equity, and Timing in Voting Mechanisms," *American Political Science Review*, 101, 409–424. 5

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, 992–1026. 1, 4, 8

Billingsley, P. (1995): *Probability and Measure*, New York: Wiley, 3rd ed. 36, 42

Callander, S. (2007): "Bandwagons and Momentum in Sequential Voting," *Review of Economic Studies*, 74, 653–684. 5, 18, 38

Dekel, E. and M. Piccione (2000): "Sequential Voting Procedures in Symmetric Binary Elections," *Journal of Political Economy*, 108, 34–55. 4, 13, 14, 19, 20

Feddersen, T. and W. Pesendorfer (1996): "The Swing Voter's Curse," *The American Economic Review*, 86, 408–424. 2, 4, 7, 8, 15, 22

——— (1997): "Voting Behavior and Information Aggregation in Elections with Private Information," *Econometrica*, 65, 1029–1058. 2, 7, 9, 19

——— (1998): "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting," *The American Political Science Review*, 92, 23–35. 15

Feddersen, T. and A. Sandroni (2006): "Ethical Voters and Costly Information Acquisition," *Quarterly Journal of Political Science*, 1, 287–311. 7

Fey, M. (2000): "Informational Cascades and Sequential Voting," University of Rochester Working Paper. 4, 5, 21

Fudenberg, D. and J. Tirole (1991): *Game Theory*, Cambridge, MA: MIT Press. 7, 12

Iaryczower, M. (2007): "Strategic Voting in Sequential Committees," Caltech, working paper. 5

Kenney, P. J. and T. W. Rice (1994): "The Psychology of Political Momentum," *Political Research Quarterly*, 47, 923–938. 5

Klumpp, T. A. and M. K. Polborn (2006): "Primaries and the New Hampshire Effect," *Journal of Public Economics*, 31, 1073–1114. 5

Knight, B. and N. Schiff (2007): "Momentum in Presidential Primaries," Brown University, working paper. 3, 23

Kreps, D. and R. Wilson (1982): "Sequential Equilibria," *Econometrica*, 50, 863–894. 13

Mayer, W. G. and A. E. Busch (2004): *The Front-Loading Problem in Presidential Nominations*, Washington, D.C.: Brookings Institution Press. 1

Morton, R. B. and K. C. Williams (1999): "Information Asymmetries and Simultaneous versus Sequential Voting," *American Political Science Review*, 93, 51–67. 5

——— (2001): *Learning By Voting*, Ann Arbor: The University of Michigan Press. 5, 24

Myerson, R. (1998): "Extended Poisson Games and the Condorcet Jury Theorem," *Games and Economic Behavior*, 25, 111–131. 4, 21, 22

——— (2000): "Large Poisson Games," *Journal of Economic Theory*, 94, 7–45. 21

Orren, G. R. (1985): "The Nomination Process: Vicissitudes of Candidate Selection," in *The Elections of 1984*, ed. by M. Nelson, Washington DC: Congressional Quarterley Press. 3

Palmer, N. A. (1997): *The New Hampshire Primary and the American Electoral Process*, Westport: Praeger Publishers. 1

Piketty, T. (2000): "Voting as Communicating," *Review of Economic Studies*, 67, 169–91. 5

Popkin, S. L. (1991): *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*, Chicago: The University of Chicago Press. 23

Smith, L. and P. Sorensen (2000): "Pathological Outcomes of Observational Learning," *Econometrica*, 68, 371–398. 4, 8, 36

Wit, J. (1997): "Herding Behavior in a Roll-Call Voting Game," Department of Economics, University of Amsterdam. 4, 5, 21