The Strategic Disadvantage of Voting Early*

Eddie Dekel[†] and Michele Piccione[‡]

January 18, 2012

Abstract

The objective of this paper is to study a model of voting with multiple candidates in which the voters choose the time in which they cast their votes. By moving late a voter can condition his vote on which candidates are still viable. By moving early the voter can decrease the set of viable opponents, i.e., the field of candidates. It turns out that the latter factor is harmful: while decreasing the field increases the likelihood that future voters vote for one's favorite candidate, it also increases the likelihood that they vote for one of the remaining viable opponents, and the latter effect is dominant in some situations.

In particular, if voters vote for their favorite candidates as long as the probability of that candidate winning is strictly positive, then early voting has a strategic disadvantage and all equilibria are equivalent to simultaneous voting. Voting in this manner is an equilibrium when one's favorite candidate is significantly better than all the others. On the other hand, when some other candidate is almost as good then any Markov, symmetric and anonymous equilibrium must involve sequential voting (and differ from simultaneous voting).

In general in sequential voting environments, because of the disadvantage of narrowing the field of candidates highlighted here, optimal voting strategies can be counter intuitive and involve voting for one's least favorite candidate so as to keep the field large.

*We gratefully acknowledge ESRC Grant R000222823 for financial support.
[†]Northwestern and Tel Aviv University
[‡]London School of Economics and Political Science

1 Introduction

We study a simple model of voting in which the voters choose the time in which they cast their votes. Our main objective is to study how the preferences of the voters affect the temporal structure of voting. We highlight a particular strategic disadvantage of "early voting" in the presence of multiple candidates: narrowing down the field of competitors encourages subsequent voters to vote either for early voters's preferred candidate or for her opponents, and the latter negative effect can dominate the former positive one.

Several institutions present instances of sequential voting, with either exogenous or endogenous timing. In many legislative bodies, voting takes place sequentially. US presidential primaries are a particularly interesting example of sequential voting since each state decides the date of its own election.

Momentum or bandwagons effects are generally held to be the main manifestation of the difference between simultaneous and sequential voting. Such effects underscore the advantages of early voters over later voters in the determination of the election outcome. Momentum effects can arise from two distinct forces that interact in sequential voting. The first is strategic. As votes are cast sequentially, some candidates may find their chances of winning significantly reduced, and some voters may decide to shift their votes in favour of candidates that are more likely to succeed. The second is informational. Early voting can signal the quality of the candidates to later voters.

In this paper, we abstract from issues pertinent to the signalling of information about the valuation of candidates by assuming that the preferences of the voters are independent. We study an election with three candidates in which voting consists of two stages. In the first, "timing game," stage, prior to the preferences being realized, voters choose simultaneously the period in which they will cast their votes. In the second, "voting game," stage, the preferences are realized privately and voting takes place according to the order chosen in the first stage. In the second-stage voting game voters in later periods observe the votes expressed by the voters in earlier periods. The candidate that obtains the majority of votes wins the election and ties are broken by a fair lottery.

We focus on the trade-off between two incentives that guide the strategic considerations of voters with respect to the temporal dimension of voting. On the one hand, a later voter has lower chances of wasting a vote as she acquires better knowledge about the likelihood of victory of the candidates. On the other hand, an early voter can influence the outcome of the election by changing the probability of victory of some candidates in later periods. In particular, voting early influences the field of candidates and can keep one's candidate viable and make others unviable. We demonstrate a novel and surprising insight into the nature of sequential voting: if later voters persist in voting for their favorite candidate despite low chances of victory, early voting has a **strategic disadvantage**. In particular, if voters vote for their preferred candidates as long as these candidates have a positive chance of winning, all equilibria in the first stage are equivalent to simultaneous voting. The reason is that, with such voting behavior, voting for one's favorite candidate, say A, in the first period can only help by making another candidate, say B, unviable and thus inducing voters for whom B was the favorite and A the second best to switch to A. But voters from whom B was the favorite and for whom C was the second best would be induced to switch to C. While these two forces might seem to cancel one another, it turns out that the losses to C are more significant than the gains to A. (While we do not develop this, it is clear that with more candidates the costs would be even greater.) Thus shrinking the field turns out to be bad when voters continue to vote for their favorite candidate so long as she might win. On the other hand, if voters tend to give up on candidates well before they have no hope of winning, then the opposite holds: it can be beneficial to vote early. We then observe that voting for one's favorite as long as she can possibly win is optimal if one's second-best candidate is relatively bad, and giving up on one's favorite is optimal if one's second-best is almost as good as one's favorite. Thus we identify the environments of preferences in which one would expect sequential or simultaneous voting to arise.

The paper is organized as follows. In the next section, we briefly discuss the related literature. The model and results are stated in Sections 3 and 4. In section 5, we discuss possible extensions. The Appendix contains some of the proofs.

2 Related Literature

The literature on sequential voting has mostly dealt with the issue of information aggregation in binary elections. Dekel and Piccione (2000) study a model in which voting is sequential and show that, in symmetric environments, the symmetric equilibria in simultaneous voting are also equilibria in any sequential structure. Battaglini (2005) shows that, with abstention and costly voting, the above inclusion fails and the set of equilibria in simultaneous and sequential voting can be disjoint. Fey (1996) and Wit (1997) study a two-signal, common-value environment in which herd-cascade equilibria exist. Morton and Williams (1999) find theoretical and laboratory evidence that later voters use the information transmitted by earlier voters. Callander (2007) shows the existence of "bandwagons" when the number of voters is infinite and voters have a desire to conform with the majority. Battaglini et al. (2007) compare the equity, information aggregation, and efficiency of simultaneous and sequential voting rules when voting is costly and information is incomplete. Ali and Kartik (2010) construct equilibria in which voters vote sincerely and that exhibit momentum effects.

Knight and Schiff (2010) study polling data from US presidential elections and find that early voters have significantly more influence than late voters. Deltas et al. (2010) provide a framework for analyzing the trade-off between learning about the candidates' quality (valence) and coordination in primary elections. They simulated various types of sequential elections using estimated structural parameters and showed that sequential elections in which candidates remain in the race yield the highest expected valence.

The theoretical literature on multi-stage voting also includes Sloth (1993), who studies sequential voting with perfect information and shows that the subgame perfect equilibria of roll-call voting are closely related to sophisticated equilibria of simultaneous voting, and Bag et al. (2009) who show that a class of voting procedures based on repeated ballots is Condorcet consistent.

3 The Model

The model has N, $N \geq 4$, voters and 3 candidates, denoted by A, B, C, whom voter i values as (v_A^i, v_B^i, v_C^i). We assume that each (v_A^i, v_B^i, v_C^i) is i.i.d. across voters with probability $f(\cdot)$ on $[0,1]^3$. The function $f(\cdot)$ is symmetric across the candidates and assigns zero probability to ties. In one special case that we will consider in this paper, each voter's preferences are a random selection of permutations of $(1, x, 0)$, for fixed $x \in (0,1)$. We will refer to this case as to the x-model.

Voting consists of two **stages**: a timing, and then a voting, stage. In the timing stage, before the preferences are realized, voters choose simultaneously the **period** $t \in \{1, 2\}$ in which they will cast their votes.[1] This decision is assumed to be irrevocable. In the second, voting, stage, preferences are realized privately and voting takes place according to the order decided in the timing stage. Voters know the timing stage decisions of all voters and, at their time of voting in the second stage, know the earlier votes. The election is won by the candidate that obtains the majority of votes. Ties are decided by a fair lottery.

To define the voting stage strategies of the voters, let $\Omega = \{1, 2\}^N$ be the set of all possible timing-stage outcomes, i.e., specifications of who votes when. For $\omega \in \Omega$, let $t^i(\omega)$ be the period in which player i votes, $\mathcal{H}^2(\omega) = \{A, B, C\}^{|\{i:t^i(\omega)=1\}|}$ be the set of possible realizations of votes in period 1, and $\mathcal{H}^1(\omega)$ the empty history. A *voting-stage* strategy for player i is a collection $s^i = \{s^i(\omega)\}_{\omega\in\Omega}$ where each $s^i(\omega)$ maps $\mathcal{H}^{t^i(\omega)}(\omega) \times [0, 1]^3$ to Δ, the set of probability distributions over $\{A, B, C\}$.

4 Persistent Voting

In this section, we study the equilibrium behavior in the timing stage fixing the voting-stage strategies in a special manner: voters vote for their favorite candidate so long as it is possible she might win. Formally, given a realization of votes in the first period, a candidate J is said to be *second-period viable* if J wins the election with strictly positive probability when all voters in the second period vote for J. A voting-stage strategy is said to be *persistent* if, when voting in the first period it votes for the candidate with the highest valuation and, when voting in the second period, it votes for the **viable** candidate with the highest valuation. The game in which the set of voting strategies is restricted to persistent strategies is called a *P-voting game*.

The assumption that voting is persistent is consistent with equilibrium behavior when preferences are such that the difference in valuations between the most preferred and the second most preferred candidate is sufficiently large for all realizations. For example, in the x-model persistence is consistent with equilibrium behavior when x is small. Since the number of voters and candidates is finite, the set of possible realizations of votes is also finite. Hence, the increase in the probability of victory of a viable candidate from receiving one extra vote is bounded from below by a strictly positive number.

[1] The number of periods is chosen solely for simplicity. See Section 6.

4.1 Characterization of Persistent Voting

The following example shows that sequential voting can be an equilibrium in a P-voting game, and that this equilibrium is distinct from a simultaneous-move equilibrium (which also exists in this case, and is strictly worse from the perspective of one player). Consider the x-model and suppose that $N = 5$ and that 4 voters vote in the first period. The behavior of a voter in the last period differs from the behavior in simultaneous voting and affects the outcome only when the candidate with the highest valuation obtains zero votes in the first period and the other two candidates get 2 votes each. The expected utility of a first-period voter conditional upon such realizations is $\frac{1}{2} + \frac{x}{4}$. This is also the expected utility of all voters when they move simultaneously in the last period. However, the utility of the second-period voter with a persistent voting strategy is higher than in simultaneous voting as the candidate valued x wins with certainty when the most preferred candidate is not viable. To see that this is indeed an equilibrium, note that a first-period voter is indifferent between voting in the first or in the second period as, by moving to the second period, voting will be equivalent in outcome to simultaneous voting.[2]

The next theorem shows that this example is very special. If $N \geq 6$, sequential voting cannot be an equilibrium outcome.

Theorem 1 *Suppose that $N \geq 6$. Any P-voting game has an equilibrium in which all voters choose to vote in the second period. Moreover, all equilibria are equivalent in outcome to this equilibrium.*

Proof: See Appendix.

The intuition behind this result is simple despite its long proof. First, the events in which a candidate ceases to be voted for are exactly those in which that candidate, say candidate C, has no chance of winning the election. Therefore, one cannot save one's favorite candidate by voting early. The question remains whether one can help coordination on one's favorite candidate by voting early. By voting early one can make another candidate unviable. Indeed, conditional upon the event of C, say, becoming unviable, voting is effectively binary in that only two candidates can win the election. However in a binary election, the utility of a voter is decreasing in the number

[2] The proof of the following theorem will imply that there cannot exist an example in which the voters in the first-period strictly prefer voting in the first period.

of voters as her influence gets diluted. Having some voters switch from C to A or B with equal probabilities is then equivalent to increasing the number of voters in a binary election. Hence, a voter is better off by voting in later periods and allowing candidates that are not viable to receive votes.

5 Voting early is advantageous: large x

In this section, we will investigate whether sequential voting is an equilibrium when voters in the voting stage switch their vote away from their preferred candidate when she is still viable. Define n_J^1 to be the total number of votes received by candidate J in period 1, $J = A, B, C$. Given a profile of voting-stage strategies s, let $s(\omega)$ denote the N-tuple $(s^1(\omega), ..., s^N(\omega))$. The profile s is *symmetric across voters* if the strategies depend only on the number of voters in each period and are identical for voters in the same period, i.e., if $s(\Upsilon\omega) = \Upsilon s(\omega)$ for any permutation Υ and any $\omega \in \Omega$. Given a profile s that is symmetric across voters, let $\pi_s^t(N^1)$ be the expected payoff of a voter of period t given that N^1 voters vote in period 1. Given a profile of voting-stage strategies s, define a *timing-stage Nash equilibrium* (induced by s) as a Nash equilibrium of the game in which the strategy space in the voting stage is restricted to the singleton s. We now state an elementary existence result.

Proposition 1 *Given any profile s of voting-stage strategies that is across voters, the game has a timing-stage Nash equilibrium in pure strategies. Moreover, if $\pi_s^1(X+1) > \pi_s^2(X)$, there exists a timing-stage Nash equilibrium in which the number of voters in the first period is strictly greater than X.*

Proof: To show existence, first note that, if $\pi_s^1(N) \geq \pi_s^2(N-1)$, the claim is trivially true. If $\pi_s^1(1) < \pi_s^2(0)$ it is a timing-stage Nash equilibrium for everyone to vote in the second period. Otherwise, define $\tilde{N}^1$ be the largest N^1 such that $\pi_s^1(N^1) \geq \pi_s^2(N^1 - 1)$. It is trivial to see that $\tilde{N}^1$ voters choosing the first period is indeed a timing-stage Nash equilibrium since $\pi_s^1(\tilde{N}^1) \geq \pi_s^2(\tilde{N}^1 - 1)$ and $\pi_s^1(\tilde{N}^1 + 1) < \pi_s^2(\tilde{N}^1)$.

The second part of the proposition follows by repeating of the argument for $N^1 \geq X$.■

Proposition 1 implies that a sufficient condition for the existence of equilibria with sequential voting when voting-stage strategies are symmetric across voters is that, if $N^1 = n_J^1 = 1$, second-period voters vote for J with a probability higher than $\frac{1}{3}$. To provide a sharper characterization, we focus on the x-model for the case of large x.

A voting-stage strategy profile is *symmetric across candidates* if for any $\omega \in \Omega$, the strategy of each voter i maps (after the appropriate reordering) every permutation of the triple

$$(n_A^{t^i(\omega)-1}, v_A^i), (n_B^{t^i(\omega)-1}, v_B^i), (n_C^{t^i(\omega)-1}, v_C^i)$$

to an identical permutation of its image in Δ.[3] A voting-stage strategy profile is *simple* if it is pure and symmetric across voters and candidates.

In the x-model with x close to one, a sequential equilibrium with simple voting-stage strategies and sequential voting is easily obtained when voters vote (i) for the preferred candidate in the first period, and (ii) for the most preferred candidate among the (possibly unique) leading candidates in the second period when $N^1 < N - 1$. A single second-period voter's optimal strategy is derived easily by checking the possible histories. Obviously, given the above voting-stage profile, it cannot be an equilibrium that all voters vote in period 2 as the most preferred candidate of a unique first-period voter wins with certainty. Existence of sequential voting then follows by Proposition 1.[4] The next theorem shows that simultaneous voting in simple strategies is impossible for x close to one.

Theorem 2 *Consider the x-model. If x is sufficiently close to one, there does not exist a sequential equilibrium with simple voting-stage strategies in which all the voters vote in the same period.*

To prove the above theorem, we will make use of the following result. Let $r_c(x, y)$ be the probability of victory of C conditional upon the information that x voters have voted for A, y voters for B, and the remaining voters vote for candidate J with probability e_J. The next lemma states that if $e_A \geq e_B$, then – conditioning upon candidates having one vote each for A and B – changing the vote for B into a vote for A decreases the probability of victory of C.

Lemma 1 *Suppose that $e_A \geq e_B$. Then $r_c(2, 0) < r_c(1, 1)$.*

[3] Where $\left(n_J^{t^i(\omega)-1}\right)_{J\in\{A,B,C\}}$ is the null triple if i moves in period 1.

[4] Other, less shallow, equilibria with sequential voting also exist.

Proof. See Appendix. ■

Proof of Theorem 2: If all voters vote in the same period, simple voting-stage strategies vote for the highest valued candidate in equilibrium. Hence, there does not exist an equilibrium in which all voters vote in the first period since one voter would be better off moving to the second period. We need to show that if they all vote in the second period, one voter will move to the first period. Consider for simplicity the realization of votes $(n_A^1, n_B^1, n_C^1) = (1, 0, 0)$. There are four simple strategy profiles that might be an equilibrium in the second-period of the voting stage.

1. Voters may vote for either B or C and cease voting for A. For this type of strategies and x close to one, it is optimal for a voter to switch to the first period and vote for the lowest valued candidate as, for $N \geq 4$, the probability of victory of this candidate drops to zero.

2. Voters may choose persistent voting strategies. For x sufficiently close to one, a voter maximizes the expected payoff by minimizing the probability of victory of the candidate with the lowest valuation. Then, after an A vote in the first period, by Lemma 1, it is a strict best reply to vote for A when valued x (assuming everyone else votes for their favorite candidate). Hence persistent strategies are not an equilibrium.

3. Third, voters with preferences $(1, x, 0)$, $(x, 1, 0)$, $(1, 0, x)$, and $(x, 0, 1)$ vote for A and voters with the remaining preferences vote for the highest valued candidate. (It is clear that that by Lemma 1 this profile of strategies is an equilibrium for x close to one, but this is not needed for our conclusion.) In this case, the probability that the second-period voters vote for A after the realization of votes $(n_A^1, n_B^1, n_C^1) = (1, 0, 0)$ is higher than with $(n_A^1, n_B^1, n_C^1) = (0, 0, 0)$. Hence, a voter with preferences $(1, x, 0)$ prefers to move to the first period and vote for A.

4. All voters in the second period vote for A. Obviously, in this case a voter prefers to vote in the first period for the candidate whom he values most.■

The intuition behind this theorem is as follows. When only one voter votes in the first period, if x is close to one, there does not exist an equilibrium in which voters in the second period use persistent voting strategies. Thus the result from the preceding section does not apply. If the voters coordinate in equilibrium by ceasing to vote for the candidate receiving the first period vote, then it is optimal for one voter to move to the first period and vote for the option with zero value. If voters vote for the leading candidate with probability higher than $\frac{1}{3}$, it is optimal for one voter to move to the first period and vote for the option with the highest value. Thus at least one voter moves to the first period.

Arguments similar to those used in the proof of Lemma 1 establish that, when x is close to one, following certain vote outcomes in the first period it is not optimal to vote for the most preferred candidate in the second period. Thus, voting behavior is not equivalent to simultaneous voting.

6 Conclusions and Extensions

The main conclusion of this paper is that for sequential voting to arise endogenously, later voters must shift votes in favor of the second-best candidates when the probability of victory of their favorite candidate is not negligible. If voters desist from voting their favorite candidate only when her probability of victory is very small, voting early has a strategic disadvantage as it decreases the probability of victory of one's favorite candidate. Although our results were proved for the case of three candidates, the intuitions behind them carry over quite naturally to the case of more than three. While we have not verified the formal details, we expect the above results to hold in the latter case.

We now discuss some limitations and possible extensions of the model.

6.1 Asymmetric Voters

An important characteristic of US primary elections is that states differ in size and in the number of votes that they can cast. Our results raise some interesting questions for the timing of voting of states of different sizes. On the one hand, a large state alone could destabilize simultaneous voting by early voting whereas a small state alone may be unable to affect the behavior of later voters. On the other hand, a truly large state may have lower incentives than small states to vote early in the election. A large state could have a significant influence in determining the choices of later states but, as we have seen in Theorem 1, this can be detrimental to the probability of success of a large state's preferred candidate. To see this, suppose that there are three voters, and that voter i has i votes. Valuations are a random selection of permutations of $(1, x, y)$, where $x \in (0, 1)$, and, conditional on the valuations of two candidates being 1 and x respectively, $y = x - \varepsilon$ with probability $1 - \varepsilon$ and $y = 0$ with probability $\varepsilon > 0$, where $x + \varepsilon < 1$. Consider equilibria that are symmetric across candidates. In any equilibrium voter 3 always votes for her preferred candidate regardless of history and, when voting is simultaneous, every voter votes for the preferred candidate. If x is close to one and ε close to zero, it cannot be an equilibrium for all voters to vote in the second period. If voter 1 chooses to vote in the first period for her preferred candidate, voter 2 votes for this candidate if valued x when $y = 0$. Hence, the probability with which voter 1's preferred candidate wins increases.

We will now show that if x is close to one and ε is close to zero, then voter 3 never votes in the first period. Suppose first that voter 3 votes alone in the first period. Then, in equilibrium, either voters 1 and 2 vote for the candidate chosen by voter 3, or they vote for their preferred candidate. This is easily seen by observing that symmetry across candidates implies that a second-period voter votes for candidates receiving no votes with the same probability. If a second-period voter's most preferred candidate coincides with voter 3's vote, then he is indifferent between voting in the first or second period. However, if a second-period voter's most preferred candidate is not the same as voter 3's vote, this voter would be strictly better off moving to the first period as then only voter 3's candidate and her most preferred candidate will receive votes in the second period. Hence, voter 3 is not alone in the first period. If voter 3 votes in the first period with voter 1 (voter 2), then voter 2 (voter 1) votes for his best viable candidate regardless of the value of y. Note that in this case voter 3 is strictly worse off than in simultaneous voting (because of the same logic as in Lemma 1: the coordination over the candidate that 3 does not like is more harmful than the coordination on the candidate that 3 voted for). However, if voter 3 chooses instead to vote in the second period, the other second-period voter, when x is close to one, ε close to zero and $x + \varepsilon < 1$, votes for the best candidate unless $y = 0$, in which case he votes for the first period vote if valued x. Hence, with very high probability voting is equivalent in this case to simultaneous voting. The point, consistent with the other results in this paper, is that the presence of voter 3 in the first period causes second-period voters to switch to the second-best candidate more often than when voter 3 is absent. This effect is detrimental to voter 3's welfare.

6.2 Voting with Several Periods

Theorem 1 extends to the case of a finite but arbitrary number of periods. The proof follows from minor modifications of the arguments available in this paper and is not provided.

When voting has more than 2 periods, a strategic environment can be very complex and involve strategies that, at least at first glance, appear non-intuitive. For example, strategic voting for the second best candidate can occur when one's favorite is leading. Consider the x-model and suppose that voting takes place in 3 periods. Assume that $x = 1 - \varepsilon$ and that in period 1 the votes for A, B, and C are $n_A^1, n_A^1 - 1$, and $n_A^1 - 1$. Also suppose that one voter votes in period 2 and two voters vote in period 3 and that the preferences of the period 2 voter are $(1, 1 - \varepsilon, 0)$. If in period 2 the voter votes for A , the distribution of votes in period 3 is $n_A^1 + 1, n_A^1 - 1, n_A^1 - 1$ whereas voting for B yields $n_A^1, n_A^1, n_A^1 - 1$. In the latter case the subsequent voters will vote for their best between A and B to minimize their probability of getting 0, thereby giving the period 2 voter a payoff close to 1. If the distribution of votes in period 3 is $n_A^1 + 1, n_A^1 - 1, n_A^1 - 1$ and both period 3 voters have preferences $(0, 1 - \varepsilon, 1)$ (the probability of this event is $\frac{1}{36}$), the period 2 voter obtains a payoff equal to zero with probability $\frac{1}{2}$. Note the anti-herding feature in this example: an increase in the number of votes for B can lead a voter to vote for A instead.

Strategic voting can also induce voters to vote for their *least* preferred candidate. Consider again the x-model for x small and that voting takes 3 periods. The realization of votes in the first period is 1 vote for A, $N/2 - 1$ votes for C, and none for B. There is only one voter in period 2 and has preferences $(1, x, 0)$. If he votes for A or B, then B is still viable and, if x is small, it is an equilibrium for voters in period 3 to vote for the favorite candidate. Voting for C makes B not viable and hence voters will switch to the preferred candidate between A and C. This can increase the probability of A winning if N is large.

7 Appendix

Proof of Theorem 1: Clearly, it is an equilibrium for all voters to vote in the last period. Consider an equilibrium in which voting takes place in two periods and that is not equivalent to simultaneous voting. For simplicity in notation, we will assume that if two candidates cease to be viable in period 1, voters vote for their highest valued candidate in period 2. Of course, this assumption is inessential.

Consider a voter i who votes in period 1. Without loss of generality, suppose that A is the candidate having the highest valuation for i. We will first show that, for any N, i's payoff cannot decrease if he decides to vote for A in period 2. This will imply that he is at least as well off if he votes in period 2 for the best viable candidate. We will then show that, for $N \geq 6$, i's payoff must increase if he decides to vote for A in period 2.

Given a fixed number of period 1 voters, consider the events for which, if voter i deviates and votes in period 2, the voting of the other voters is affected. To define such pivotal events, let $\hat{n}_J^1$, $J = A, B, C$ denote the number of votes received by candidate J in period 1 excluding the vote of player i. Define the inequalities

$$N - \hat{n}_A^1 - \hat{n}_J^1 - 1 < \max\{\hat{n}_A^1 + 1, \hat{n}_J^1\} \tag{1}$$

$$N - \hat{n}_A^1 - \hat{n}_J^1 \geq \max\{\hat{n}_A^1, \hat{n}_J^1\} \tag{2}$$

where $J \in \{B, C\}$. If (1) and (2) hold for $J = B$ and both A and B are viable then period 2 voters whose most valuable candidate is C vote for their second most valuable candidate (A or B) when voter i votes for A in the first period (as then C is not viable by (1)), and vote for C when voter i votes for A in the second period (as then C is viable by (2)). A symmetric explanation holds for $J = C$.

Take $\gamma \leq \delta - 1$ such that, for $\hat{n}_A^1 = \gamma$ and $\hat{n}_B^1 = \delta$, (1) and (2) hold and both A and B are viable. Define the events

$$\begin{aligned} E_1(k) &= \{\hat{n}_A^1 = \gamma,\ \hat{n}_B^1 = \delta,\ k \text{ voters vote for } C \text{ in period 2}\} \\ E_2(k) &= \{\hat{n}_A^1 = \delta - 1,\ \hat{n}_B^1 = \gamma + 1,\ k \text{ voters vote for } C \text{ in period 2}\} \end{aligned}$$

and define $E(k) = E_1(k) \cup E_2(k)$. The following elementary fact is stated without proof.

Fact *If* (1) *and* (2) *hold and A and J are viable for $\hat{n}_A^1 = \gamma$ and $\hat{n}_J^1 = \delta$, where $\gamma \leq \delta - 1$, then* (1) *and* (2) *hold and A and J are viable for $\hat{n}_A^1 = \delta - 1$ and $\hat{n}_J^1 = \gamma + 1$.*

We will now show that the probability of A winning conditional on $E(k)$ does not decrease with k. To do so, we will replace one second-period vote for C with a vote for A or B with probability 0.5 each and show that the probability of A winning cannot increase. In particular, we will show that, conditional on $E_1(k)$, the effect of decreasing k is to weakly increase the probability that A wins. However, for the corresponding event $E_2(k)$ decreasing k weakly decreases the probability that A wins by an offsetting amount. Since $E_2(k)$ is at least as likely as $E_1(k)$, the overall effect conditional on $E(k)$ of decreasing k is to weakly decrease the probability that A wins. We will then show that for all remaining events with $\hat{n}_A^1 \geq \hat{n}_B^1$, the probability of A winning decreases as k decreases.

Let $\tilde{n}_A$ denote the total number of votes for candidate A in period 2 when the vote of voter i is not included, and N^1 the total number of voters in period 1 including voter i. First note that conditional on $E(k)$ and i voting for A in either the first or the second period, A wins with probability 1 if

$$\tilde{n}_A > \frac{N - 1 - k - \hat{n}_A^1 - (N^1 - \hat{n}_B^1)}{2} \tag{3}$$

and wins with probability 1/2 if

$$\tilde{n}_A = \frac{N - 1 - k - \hat{n}_A^1 - (N^1 - \hat{n}_B^1)}{2} \tag{4}$$

First, suppose that $N - 1 - k - \hat{n}_A^1 - (N^1 - \hat{n}_B^1)$ is even. Note that if it is even for one component event in $E(k)$, it is even for the other component as well. If k is decreased by one unit, the only N-tuples whose outcome is affected are those in (4). In this case, the probability of A winning is unchanged when k is decreased by one unit since the voter who ceases voting for C votes for A with probability 1/2 and B with probability 1/2.

Now consider the case of odd $N - 1 - k - \hat{n}_A^1 - (N^1 - \hat{n}_B^1)$ and suppose that one of the voters who vote for C in period 2 switches to A or B with probability 1/2 each. To evaluate the change in the probability of victory of A, in $E_1(k)$ we need to consider N-tuples for which

$$\tilde{n}_A = \frac{N - N^1 - k + (\delta - \gamma)}{2} \qquad (WL1)$$

$$\tilde{n}_A = \frac{N - N^1 - k + (\delta - \gamma)}{2} - 1 \qquad (LW1)$$

and, in $E_2(k)$,

$$\tilde{n}_A = \frac{N - N^1 - k - (\delta - \gamma)}{2} + 1 \qquad (WL2)$$

$$\tilde{n}_A = \frac{N - N^1 - k - (\delta - \gamma)}{2} \qquad (LW2)$$

For the events satisfying $WL1$ and $WL2$, if one of the k voters for C switches to voting A with probability $1/2$ and B with probability $1/2$, the probability of A winning decreases from 1 to $3/4$. For the events satisfying $LW1$ and $LW2$, the probability of A winning increases from 0 to $1/4$.

Since the distribution of $\tilde{n}_A$ conditional upon $E_1(k)$ or $E_2(k)$ is binomial and symmetric around $\dfrac{N - N^1 - k}{2}$, we have that

$$Prob(WL1 \mid E_1(k)) = Prob(LW2 \mid E_2(k)),$$

$$Prob(WL2 \mid E_2(k)) = Prob(LW1 \mid E_1(k)).$$

By the same token,

$$Prob(LW1 \mid E_1(k)) \geq Prob(WL1 \mid E_1(k)).$$

Since the symmetry of the binomial distribution also implies that the probability of $E_1(k)$ cannot exceed the probability $E_2(k)$, it can be easily verified that the probability of A winning cannot increase if one of the k second period C-voters switches to A or B with probability $1/2$ each.

In view of the Fact, to conclude that i's payoff never decreases when he votes for her highest valued candidate in period 2, we need to consider first period realizations for which $\hat{n}_A^1 \geq \hat{n}_B^1$. In particular, we want to show that, conditional upon the event $\{\hat{n}_A^1 \geq \hat{n}_B^1$, k voters vote for C in period 2$\}$, the probability of A winning the election decreases with k.

As before, we consider 2 cases. If $N - 1 - k - \hat{n}_A^1 - (N^1 - \hat{n}_B^1)$ is even and k decreases by one unit, the N-tuples whose outcome is affected are those in (4). As before, if a voter who ceases voting for C votes for A with probability $1/2$, the probability of A winning is unchanged.

If $N-1-k-\hat{n}_A^1-(N^1-\hat{n}_B^1)$ is odd, decreasing k by one unit changes the outcome of the election only when

$$\tilde{n}_A = \frac{N-k-\hat{n}_A^1-(N^1-\hat{n}_B^1)}{2}$$

or

$$\tilde{n}_A = \frac{N-k-\hat{n}_A^1-(N^1-\hat{n}_B^1)}{2} - 1.$$

Since $\hat{n}_A^1 \geq \hat{n}_B^1$, the event defined by the first equality is more likely. For an N-tuple satisfying the first equality, decreasing k decreases the probability of A winning from 1 to 3/4. For an N-tuple satisfying the second equality, decreasing k increases the probability of A winning from 0 to 1/4. Hence, decreasing k cannot increase the probability of A winning. Since conditional upon i voting for A, the probabilities of B or C winning the election are identical, voter i is not worse off voting for A in period 2.

We conclude the proof of theorem showing that, if $N \geq 6$, voter i is strictly better off voting in period 2 for the best viable candidate. First suppose that N is even. If $N^1 \geq \frac{N}{2}+2$, consider the event $\{\hat{n}_B^1 = \frac{N}{2}, \hat{n}_C^1 = N^1 - \frac{N}{2} - 1\}$. Conditional upon this event, if voter i votes for A in period 1, B wins the election with probability 1 whereas, if the second best is C, voting for C in period 2 gives C a positive probability of victory. If $N^1 = \frac{N}{2}+1$, (1) and (2) are satisfied for $\hat{n}_A^1 = 0$ and $\hat{n}_B^1 = \frac{N}{2}$. This corresponds to the case where $\gamma = 0$ and $\delta = \frac{N}{2}$ in the definition of $E(k)$. For $k=1$,

$$Prob(WL1 \mid E_1(1)) - Prob(LW1 \mid E_1(1)) = -\frac{1}{2} Prob\left(\tilde{n}_A = \frac{N}{2} - 2\right).$$

$N \geq 6$ implies that $Prob\left(\tilde{n}_A = \frac{N}{2} - 2\right) > 0$ and, since the probability of $E_1(1)$ is strictly smaller than the probability of $E_2(1)$, voter i is strictly better off voting in period 2.

Now suppose that N is odd. If $N^1 \geq \frac{N+5}{2}$, consider the event $\{\hat{n}_B^1 = \frac{N-1}{2}, \hat{n}_C^1 = N^1 - \frac{N-1}{2} - 1\}$. Conditional upon this event, A cannot win the election and, if voter i's second best is B, voting for B in period 2 gives

B the certainty of victory whereas voting for A in period 1 gives C a positive chance of winning the election. If $N^1 = \dfrac{N+3}{2}$, (1) and (2) are satisfied for $\hat{n}_A^1 = 1$ and $\hat{n}_B^1 = \dfrac{N-1}{2}$. This corresponds to the case where $\gamma = 1$ and $\delta = \dfrac{N-1}{2}$ in the definition of $E(k)$. For $k = 2$,

$$Prob(WL1 \mid E_1(k)) - Prob(LW1 \mid E_1(k)) = -\frac{1}{2} Prob\left(\tilde{n}_A = \frac{N-5}{2} - 1\right).$$

When $k = 2$, $Prob\left(\tilde{n}_A = \dfrac{N-5}{2} - 1\right) > 0$. But k can be equal to 2 if there are at least two voters (excluding i) in period 2, that is, $N \geq 7$. ■

Proof of Lemma 1: First, we show a preliminary lemma.

Lemma 2 *If $\mu_2 \geq \mu_1$, $\mu_2 + \mu_1 > \Lambda + 1$, and $\frac{1}{2} \leq e \leq 1$,*

$$\sum_{\mu=\mu_1}^{\mu_2} (1-e)^{\mu}\, e^{\Lambda-\mu} \binom{\Lambda}{\mu} < \sum_{\mu=\mu_1-1}^{\mu_2-1} (1-e)^{\mu}\, e^{\Lambda-\mu} \binom{\Lambda}{\mu}.$$

Proof of Lemma 2: The difference of the left and right hand sides simplifies to

$$-(1-e)^{\mu_1-1}\, e^{\Lambda-\mu_1+1} \tbinom{\Lambda}{\mu_1-1})) + (1-e)^{\mu_2}\, e^{\Lambda-\mu_2} \tbinom{\Lambda}{\mu_2})$$

Note that

$$\frac{(1-e)^{\mu_1-1}\, e^{\Lambda-\mu_1+1}}{(1-e)^{\mu_2}\, e^{\Lambda-\mu_2}} = \left(\frac{e}{1-e}\right)^{-\mu_1+1+\mu_2} \geq 1.$$

If $\mu_1 \geq \frac{\Lambda}{2}$, the claim obviously true as $\mu_2 \geq \mu_1$. If $\mu_1 < \frac{\Lambda}{2}$, then $\Lambda - (\mu_1 - 1) > \mu_1 - 1$. Moreover, $\mu_2 + \mu_1 > \Lambda + 1$ implies that $\Lambda - \mu_2 < \mu_2$. Hence, the claim follows since $\mu_1 - 1 > \Lambda - \mu_2$.■

To prove Lemma 1, define $q_c(v,k)$ to be the probability that C wins conditional upon exactly k voters voting for C, at least v for B, and at least

$2 - v$ for A, where $v \in \{0, 1\}$. Obviously, if $q_c(0, k) = 1$, then $q_c(1, k) = 1$. Now we will show that, if $0 < q_c(0, k) < 1$, then $q_c(1, k) > q_c(0, k)$.

Define $e = \frac{e_A}{e_A + e_B}$. First suppose that $k = \frac{N}{3}$. Then,

$$q_c(v, \frac{N}{3}) = \frac{1}{3} e^{\frac{N}{3} - 2 + v} (1 - e)^{\frac{N}{3} - v} \binom{\frac{2N}{3} - 2}{\frac{N}{3} - v}.$$

The above expression is increasing in v for $v \in \{0, 1\}$. If $k = \frac{N}{2}$, then $q_c(0, k) < 1$ and $q_c(1, k) = 1$. If $k = \frac{N-1}{2}$,

$$\begin{aligned} q_c(1, k) &= 1 - \frac{1}{2} e^{\frac{N-3}{2}} - \frac{1}{2} (1 - e)^{\frac{N-3}{2}} \\ q_c(0, k) &= 1 - e^{\frac{N-3}{2}} - \frac{1}{2} \cdot \frac{N-3}{2} e^{\frac{N-5}{2}} (1 - e) \end{aligned}$$

As $N \geq 4$, $q_c(0, k) < q_c(1, k)$. Suppose then that $\frac{N}{3} < k \leq \frac{N-2}{2}$. Then,

$$q_c(v, k) = \sum_{l=N-2k-v}^{k-v} \alpha_l \beta_l \binom{N-2-k}{l},$$

where $\beta_l = e^{N-2-k-l} (1 - e)^l$, $\alpha_l = \frac{1}{2}$ for $l = N - 2k - v$ or $l = k - v$, and $\alpha_l = 1$ for $N - 2k - v < l < k - v$.

Lemma 2 implies that

$$\sum_{l=N-2k}^{k} \beta_l \binom{N-2-k}{l} < \sum_{l=N-2k-1}^{k-1} \beta_l \binom{N-2-k}{l}$$

and

$$\sum_{l=N-2k+1}^{k-1} \beta_l \binom{N-2-k}{l} < \sum_{l=N-2k}^{k-2} \beta_l \binom{N-2-k}{l}$$

Since

$$2q_c(v, k) = \sum_{l=N-2k-v}^{k-v} \beta_l \binom{N-2-k}{l} + \sum_{l=N-2k-v+1}^{k-v-1} \beta_l \binom{N-2-k}{l}$$

it is easy to verify that the above inequalities imply that $q_c(0, k) - q_c(1, k) < 0$. The proof is then completed by taking expectations over k.■

References

[1] Ali S and N Kartik (2010) "Observational Learning with Collective Preferences", working paper, Columbia University.

[2] Bag PK, H Sabourian and E Winter "Multi-stage voting, sequential elimination and Condorcet consistency", *Journal of Economic Theory*, **144** 1278-1299

[3] Battaglini M (2005), "Sequential Voting with Abstention" *Games and Economic Behavior*, 445-463.

[4] Battaglini M, R Morton and T Palfrey (2007) "Efficiency, equity, and timing of voting mechanisms"*American Political Science Review*, **101** 409-424.

[5] Callander S (2007) "Bandwagons and momentum in sequential voting", *Review of Economic Studies*, **74** 653-684.

[6] Dekel E and M Piccione (2001), "On the Equivalence of Simultaneous and Sequential Binary Elections," *Journal of Political Economy*, **108** 34-55.

[7] Dekel E and M Piccione (2002), "Sequential Voting with Three Candidates", working paper.

[8] Deltas G, H. Herrera and M. Polborn (2010), "A Theory of Learning and Coordination in the Presidential Primary System", working paper.

[9] Fey M (1996): "Informational cascades, Sequential Elections, and Presidential Primaries," mimeo, Department of Politics, Princeton University.

[10] Knight B and N Schiff (2010) "Momentum and Social Learning in Presidential Primaries" *Journal of Political Economy*, **118** 1110-1150.

[11] Morton R and R Williams (1999): "Information Asymmetries and Simultaneous versus Sequential Voting," *American Political Science Review*, **93** 51-67.

[12] Sloth B (1993): "The Theory of Voting and Equilibria in non-Cooperative Games," *Games and Economic Behavior*, **5** (1) 152–169.

[13] Wit J (1997) "Herding Behavior in a Roll-Call Voting Game," mimeo, CREED, University of Amsterdam.