Attentiveness in Elections with Impressionable Voters

Costel Andonie and Daniel Diermeier
MAY 2019
Attentiveness in elections with impressionable voters

Costel Andonie* and Daniel Diermeier†

May 6, 2019

Abstract

We propose a model of attentiveness in elections with impressionable voters under various electoral rules. Voters’ behavior is determined by their attentiveness and impressions of candidates. We show that attentiveness is as important as voters’ preferences for the outcome of the election. Specifically, we show that candidates benefit from increased voter attention under all rules other than negative plurality. Second, less attentive voters have a larger impact on the election outcome under plurality and negative plurality rules, but not under approval voting. Third, under limited and endogenous attentiveness, a unanimously first-ranked candidate may not win under plurality and approval voting, but he always wins under negative plurality. We finally consider the case of tonality in news coverage and show that under plurality rule and approval voting candidates may benefit from frequent news coverage even if the news is negative.

JEL classification numbers: D03, D72, D83. Keywords: multi-candidate elections, adaptive behavioral voting, partial attentiveness.

1 Introduction


*Harris School of Public Policy, The University of Chicago, 1155 East 60th Street, Chicago, IL 60637, Email: candonie@uchicago.edu
†Office of the Provost, and Harris School of Public Policy, The University of Chicago, 5801 South Ellis Avenue, Chicago, IL 60637, ddiermeier@uchicago.edu and Canadian Institute for Advanced Research, CIFAR Program on Institutions, Organizations, and Growth, CIFAR, Toronto, ON M5G 1Z8.
Myatt (2007), Bouton & Castanheira (2012), Bouton (2013)). In contrast, the empirical research on elections has often argued that voters do not act rationally, but are subject to irrelevant features, such as the candidates’ performance in previous elections (Anagol & Fujiwara (2016)), their physical appearance (Bailenson, Iyengar, Yee & Collins (2008), Todorov, Mandisodza, Goren & Hall (2005)), or the performance of the world economy (Wolfers (2002)). Rather than carefully weighing issues, voters, it seems, are, disinterested, inattentive and ignorant of basic facts (Achen & Bartels (2004), Achen & Bartels (2016)). D. Kinder (Kinder (1998), p. 785) summarizes decades of research as follows:

Whether rational or not, the depth of ignorance demonstrated by modern mass publics can be quite breathtaking, inspiring such derogatory terminology as ”wretchedly-informed” (Converse, 1975), ”know-nothings” (Hyman & Sheatsley, 1947), and in modern parlance, ”clueless” (Morin, 1996).

In this paper, we focus on the implications of one such feature: voters lack of attention in politics. For this purpose, we develop a model of attentiveness in elections and then apply this model to three electoral rules: plurality, approval voting and negative plurality rules. The model has two building blocks. First, as in Andonie & Diermeier (2018), voters vote based on their ”impressions” of candidates (e.g., Bendor, Diermeier, Siegel & Ting (2011), Andonie & Diermeier (2017), Andonie & Diermeier (2019), Diermeier & Li (2017)). An impression of a candidate is the sum of two components. The first is a fixed (non-random) component that captures a voter’s like or dislike of a candidate, while the second is a random component that captures e.g., idiosyncratic events or irrelevant features. Impressions that are above a given cutoff (aspiration level) are experienced as ”positive”; while those under the cutoff are experienced as ”negative”. Voters are more likely to cast a ballot in favor of a candidate if they have a positive impression of the respective candidate, and more likely to cast a ballot against a candidate if they have a negative impressions of the candidate. Voters can also abstain; this will be the case if, e.g. under plurality rule, voters have negative impressions of all candidates.

The second building block is a model of attentiveness. Voters are assumed to only partially pay attention to campaigns and may be unaware of the full slate of candidates, their respective policy positions or standing in opinion polls, etc. We consider the case where voters’ attentiveness is event-driven, i.e., voters pay attention to candidates that attract attention e.g., candidates that do well in a debate, celebrities, or rank high in opinion polls. Voters form impressions of the candidates they pay attention to and update their propensities to vote or abstain based solely on those impressions.
The model induces a stochastic dynamic process, and we characterize its stationary distributions under the three electoral rules. The results we obtain are as follows. We first show the importance of voter attention in elections. Under plurality and approval voting, candidates that receive more attention are at an advantage compared to candidates that receive little attention. In contrast, under negative plurality, candidates benefit from flying under the radar. Consequently, e.g., a candidate which the electorate dislikes can win the election under plurality rule or approval voting, if voters pay much attention to that candidate at the expense of other candidates. Conversely, a candidate which the electorate dislikes can win under negative plurality, if voters pay little attention to that candidate and focus their attention on other candidates.

Second, we show that, under plurality and negative plurality rules, voters that focus their attention on a few candidates have a larger influence on the outcome than voters that pay attention to a large number of candidates. Under approval voting, in contrast, the vote measure a candidate obtains is not influenced by the amount of attention, or impressions voters have of other candidates. We then consider an example with two groups of voters, a minority and majority group, to demonstrate normative implications of inattentiveness. We show that the candidate that the minority group prefers can win under plurality or negative plurality rules, if the minority group focus their attention on the candidates they prefer, while the majority group of voters pays attention to all candidates in the election. In contrast, the candidate that the majority group favors (Condorcet winner) always wins under approval voting.

Third, we study two particular cases of attentiveness. Attentiveness can be endogenous, i.e., voters may pay attention to the two candidates leading in opinion polls, or exogenous, i.e., voters pay attention to a specific list of candidates because of prominence, prior performance, name recognition, or association with a party. That is, a candidate associated with a major party may receive more attention than a candidate from a minor party or an independent. In such cases, we can show the existence of momentum effects (e.g., Bartels (1988), Sides & Vavreck (2013)) where candidates capture public attention and experience rapidly growing electoral support, which, however will fade just as quickly, unless such candidates are preferred by a sufficient segment of the electorate. We then analyse the normative properties of endogenous attentiveness. If there is a candidate that all voters either prefer or all oppose, then, under full attentiveness, that candidate wins, or loses respectively under all three rules (Andonie & Diermeier (2018)). In contrast, we show that, if attentiveness is limited then the candidate that all voters prefer can lose under plurality rule, or approval voting, but wins under negative plurality. Conversely, the candidate that all voters oppose
cannot win under plurality rule or approval voting, but can win with positive probability under negative plurality.

Fourth, we consider the case of tonality in news coverage. That is, media coverage for a candidate may be neutral, good, or bad. While under plurality rule and approval voting good news is always beneficial, we show that a candidate may benefit from the increased attention due to frequent news coverage, even if the news is bad. Thus, notoriety, even if it is persistently negative, can benefit a candidate, especially in a crowded field with many candidates.

2 The model

The basic model follows Andonie & Diermeier (2018), to which we add the feature of attention. We consider an election, where there are \( n \) candidates \( (n \geq 2) \) and a continuum of voters. The set of candidates is \( N = \{1, 2, ..., n\} \). Voters are partitioned into groups. A given group \( \theta \) is characterized by vector \( u_\theta = \{v^1_\theta, ..., v^n_\theta\} \), and its share in the electorate is \( s_\theta \in [0,1] \). \( v^i_\theta \), for each candidate \( i \in N \), is the fixed factor that determines the impression of a voter in group \( \theta \) of candidate \( i \), defined as:

\[
\pi^i_\theta(i) := v^i_\theta + \alpha \epsilon^i_\theta(i)
\]

The impression consists of two components: the fixed component \( v^i_\theta \) and the random component \( \epsilon^i_\theta(i) \). The relative weight of the random component is \( \alpha \geq 0 \). The random component \( \epsilon^i_\theta(i) \) is iid across voters, candidates and time. We assume that \( \epsilon^i_\theta(i) \sim F(.) \), and its density function, denoted by \( f(.) \), is symmetric and has a single peak at 0, e.g., \( N(0,1) \).

We consider plurality, approval and negative plurality rules. The sets of ballots corresponding to these three rules are:

1. Plurality rule: \( B = \{b = (b_1, b_2, ..., b_n), \text{ where } b_j \in \{0,1\} \text{ for all } j = 1, 2, ..., n \text{ and } \sum_j b_j \in \{0,1\}\} \)

2. Approval voting: \( B = \{b = (b_1, b_2, ..., b_n), \text{ where } b_j \in \{0,1\} \text{ for all } j = 1, 2, ..., n\} \)

3. Negative plurality: \( B = \{b = (b_1, b_2, ..., b_n), \text{ where } b_j \in \{0,1\} \text{ for all } j = 1, 2, ..., n \text{ and } \sum_j b_j \in \{n-1, n\}\} \)

The measure of votes a candidate \( i \) receives among voters of group \( \theta \) under a given rule is \( S_\theta(i) \), while the total measure across all groups is \( S(i) = \sum_\theta s_\theta S_\theta(i) \). Under plurality rule,
\( S(i) \) represents the measure of voters that vote for candidate \( i \); under approval voting, it represents the measure of voters that approve candidate \( i \); and, under negative plurality, it represents the measure of voters that do not vote in opposition to candidate \( i \). The candidate with the largest measure of votes is the winner. If, e.g., \( S(i) > \max_{j \neq i} \{ S(j) \} \) then candidate \( i \) wins the election.

We analyse an election where voters update their propensities to vote throughout a campaign. \( t \) denotes time, which is discrete with \( t = 0, 1, 2, \ldots \). At each time \( t \), a voter of group \( \theta \) is characterized by the vector of propensities \( p_t^\theta = (p_t^\theta(1), \ldots, p_t^\theta(n)) \), a propensity to abstain \( p_t^\theta(A) \), and an aspiration \( a_t^\theta \). A propensity \( p_t^\theta(i) \) has the following meaning. Under plurality rule, \( p_t^\theta(i) \) is the probability that a voter of group \( \theta \) votes for candidate \( i \); under approval voting, \( p_t^\theta(i) \) is the probability that a voter of group \( \theta \) approves candidate \( i \); and, under negative plurality, \( p_t^\theta(i) \) is the probability that a voter of group \( \theta \) does not vote against candidate \( i \). \( p_t^\theta(A) \) is the probability that a voter abstains. Aspirations \( a_t^\theta \) divide impressions into "positive" and "negative" impressions. "Positive" impressions are those for which \( \pi_t^\theta(i) \geq a_t^\theta \), "negative" impressions are those for which \( \pi_t^\theta(i) < a_t^\theta \).

The adjustment of propensities and aspirations is based on voters’ impressions of candidates and their attentiveness. Impressions were defined above, while attentiveness is modeled as follows. For each candidate \( i \in N \), we let \( e_t^\theta(i) \in \{0, 1\} \) be the "event" variable associated with candidate \( i \). Specifically, \( e_t^\theta(i) = 1 \) means that an "event" associated with candidate \( i \) occurs in period \( t \), and voters of type \( \theta \) pay attention to candidate \( i \); while \( e_t^\theta(i) = 0 \) means that no "event" associated with candidate \( i \) occurs in period \( t \), and voters do not pay attention to the candidate. The word "event" refers to anything that may make voters pay attention to a candidate. It can correspond to an actual event, e.g., a speech, but it also may refer to e.g., being ranked high in a poll, being a celebrity, etc. The vector of all such \( n \) "event" variables is denoted by \( e_t^\theta = (e_t^\theta(1), \ldots, e_t^\theta(n)) \in \{0, 1\}^n \). We assume that \( e_t^\theta \) is a multi-dimensional random variable, and let \( \{ \Pr_t^\theta(e) \}_{e \in \{0, 1\}^n} \) denote its probability distribution. In our analysis, it may be more convenient to use the following alternative notation for the probability distribution of \( e_t^\theta \). For a given subset of candidates \( K \subseteq N \), we let \( \Pr_t^\theta(K) \) denote the (induced) probability that voters of type \( \theta \) pay attention only to candidates in the set \( K \), i.e.,

\[
\Pr_t^\theta(K) := \Pr(e_t^\theta(i) = 1 \text{ for all } i \in K, \text{ and } e_t^\theta(i) = 0 \text{ for all } i \notin K).
\]

\(^1\)See Section 4.1 for further discussion, and examples, of the factors that determine voters’ attentiveness.

\(^2\)The two ways to define the "event" probabilities \( \Pr_t^\theta(.) \) are equivalent, and each can be derived from the other.
The distributions of the individual "event" variables $e^t_\theta(i)$ can be correlated, e.g., all candidates participate in a debate, and also they can potentially vary across time.

Two particular cases of distribution of "events" $\{e^t_\theta(i)\}_{i \in \mathcal{N}}$ that we will analyze below are as follows. A first case is that of all voters paying attention to all candidates at all times $t \geq 0$, so that $\Pr^t_\theta(\mathcal{N}) = 1$, and $\Pr^t_\theta(K) = 0$ for all $K \subset \mathcal{N}$. This particular case, to which we refer as the "full attentiveness" case, corresponds to the impressionable voter model developed in Andonie & Diermeier (2018). A second particular case of interest is that of the "events" $\{e^t_\theta(i)\}_{i \in \mathcal{N}}$ being independent. In this case, we let $r^t_\theta(i) := \Pr(e^t_\theta(i) = 1)$ for each candidate $i \in \mathcal{N}$, i.e., $r^t_\theta(i)$ is the individual probability that a voter of type $\theta$ pays attention to candidate $i$ at each time $t$. $r^t_\theta(i)$ can be interpreted as the "amount of attention" a voter of type $\theta$ pays, on average, to candidate $i$. We note that, in this case, the probabilities $\Pr^t_\theta(K)$ defined above can be computed as: $\Pr^t_\theta(K) = \prod_{i \in K} r^t_\theta(i) \prod_{i \in \mathcal{N} \setminus K} (1 - r^t_\theta(i))$, for each subset $K \subseteq \mathcal{N}$.

At a given date $t$, and for a given realization of the vector $e^t_\theta$, suppose a voter of type $\theta$ pays attention to the candidates in the subset $K \subseteq \mathcal{N}$:

$$K := \{i \in \mathcal{N} : e^t_\theta(i) = 1\}.$$ 

We let $k = |K|$ denote the number of candidates in the set $K$, and call $k$ the "range of attentiveness". Given the set of candidates $K$ that a voter pays attention to at time $t$, the voter forms impressions of each candidate $i \in K$: $\{\pi^t_\theta(i)\}_{i \in K}$. The voter updates her aspirations $a^t_\theta$ and propensities $p^t_\theta$ based only on the impressions of the candidates in the set $K$. We let:

$$K^t_{pos} := \{i \in K : \pi^t_\theta(i) \geq a^t_\theta\},$$
$$K^t_{neg} := \{i \in K : \pi^t_\theta(i) < a^t_\theta\},$$
$$K^t_{inatt} := \mathcal{N} \setminus K.$$

In other words, $K^t_{pos}$ contains the candidates in the set $K$ of which the voter has positive impressions, $K^t_{neg}$ the candidates of which the voter has negative impressions, while $K^t_{inatt}$ contains the candidates to which the voter does not pay attention. We, also, let:

$$I^t_\theta(K) := |K^t_{pos}|,$$
$$J^t_\theta(K) := |K^t_{neg}|.$$

3Strictly speaking, we would have to index the set $K$ by subscript $\theta$ for the group of voters, and superscript $t$ for time. For simplicity of notation, however we drop these two indexes here and in analogous cases.
That is, $I^t_\theta(K)$ denotes the number of candidates voter pays attention to and has positive impression of, and $J^t_\theta(K)$ denotes the number of candidates voter pays attention to and has negative impression of. Finally, for each candidate $i \in K$, we let:

$$q^t_\theta(i) := \Pr(\pi^t_\theta(i) \geq a^t_\theta),$$

i.e., $q^t_\theta(i)$ is the probability of having a positive impression of candidate $i$.

The updating of voting propensities differs across the electoral rules we discuss. These are an extension of the updating rules from the full attentiveness model (Andonie & Diermeier (2018)), modified to take into account the fact that voters pay attention only to the candidates in the set $K$.

**Updating of propensities under plurality rule.** Under plurality rule, propensities are updated by a generalization of the Bush-Mosteller rule (Bush & Mosteller (1955)), which takes into account only the impressions of candidates a voter pays attention to.

1. For each candidate $i \in K^t_{pos}$, i.e., to which a voter pays attention and has a positive impression:

$$p^{t+1}_\theta(i) = (1 - \lambda_p)p^t_\theta(i) + \lambda_p \frac{1}{I^t_\theta(K)}$$

2. For each candidate $i \in K^t_{neg} \cup K^t_{inatt}$, i.e., to which a voter pays attention and has a negative impression, or to which a voter does not pay attention:

$$p^{t+1}_\theta(i) = (1 - \lambda_p)p^t_\theta(i)$$

The updating process implies that the propensity to abstain is updated as: $p^{t+1}_\theta(A) = (1 - \lambda_p)p^t_\theta(A) + \lambda_p I^t_\theta(K)$. The updating for candidates that voters pay attention to is analogous to those in the full attentiveness model, with the modification that they are based only on the impressions of candidates a voter pays attention to, i.e., in the set $K$. That is, a voter updates her propensities towards a distribution where she is uniformly likely to vote for each candidate to which she pays attention and has a positive impression, and she does not vote for candidates of which she has a negative impression. For the candidates that voters do not pay attention to, propensities are decreased by $(1 - \lambda_p)$. The basic principle behind the updating for candidates that do not attract attention is that voters, with their attention focused on candidates in the set $K$, will slowly "forget" about the candidates they do not pay attention to. We assume, for convenience, that the speed by which this process occurs is also given by $\lambda_p$, which can be interpreted as a "fading" rate.
Updating of propensities under approval voting. We now consider approval voting. At a given time \( t \), conditional on the set \( K \) of the candidates a voter of type \( \theta \) pays attention to, the updating of propensities is as follows.

1. For each candidate \( i \in K_{\text{pos}}^t \), i.e., to which a voter pays attention and has a positive impression:
   \[
   p_{\theta}^{t+1}(i) = (1 - \lambda_p)p_{\theta}^t(i) + \lambda_p
   \]
2. For each candidate \( i \in K_{\text{neg}}^t \cup K_{\text{inatt}}^t \), i.e., to which a voter pays attention and has a negative impression, or to which a voter does not pay attention:
   \[
   p_{\theta}^{t+1}(i) = (1 - \lambda_p)p_{\theta}^t(i)
   \]

As under plurality rule, the updating of abstention propensity is \( p_{\theta}^{t+1}(A) = (1 - \lambda_p)p_{\theta}^t(A) + \lambda_p \mathbf{1}_{\{\theta^t(K) = 0\}} \). Again, the updating of propensities of candidates voters pay attention to is analogous to the updating under full attentiveness under approval rule. Specifically, a voter updates her propensities towards a distribution where she approves each candidate she pays attention to and has a positive impression, and she does not approve candidates of which she has a negative impression. The propensities of candidates to which voters do not pay attention decrease at a rate of \( \lambda_p \). As under plurality rule, we can think of this as voters forgetting about candidates to which they do not pay attention, and so they decrease their voting propensities for such candidates. For convenience, we assume that the fading rate is also \( \lambda_p \).

Updating of propensities under negative plurality. Under negative plurality rule, propensities \( p_{\theta}^t \) are also updated by a generalized Bush-Mosteller rule, similar to that from plurality rule, and again, only takes into account the impressions of candidates in the set \( K \).

1. For each candidate \( i \in K_{\text{neg}}^t \), i.e., to which a voter pays attention and has a negative impression:
   \[
   p_{\theta}^{t+1}(i) = (1 - \lambda_p)p_{\theta}^t(i) + \lambda_p(1 - \frac{1}{J_{\theta}^t(K)})
   \]
2. For each candidate \( i \in K_{\text{pos}}^t \cup K_{\text{inatt}}^t \), i.e., to which a voter pays attention and has a positive impression, or to which a voter does not pay attention:
   \[
   p_{\theta}^{t+1}(i) = (1 - \lambda_p)p_{\theta}^t(i) + \lambda_p
   \]

The updating process implies that the propensity to abstain is updated as: \( p_{\theta}^{t+1}(A) = \ldots \)
\[(1 - \lambda_p) p^\theta_t(A) + \lambda_p 1_{\{J^\theta_t(K) = 0\}}.\] As in the other rules, the updating under negative plurality is similar to the case of full attentiveness, but defined relative to the set of candidates \(K\). Specifically, a voter updates her propensities towards a distribution where she is uniformly likely to vote in opposition to each candidate she pays attention to and has a negative impression of, while she does not vote in opposition to candidates she has positive impression of. Given the mirror nature of negative plurality, in contrast to plurality rule and approval voting, the propensities of candidates to which voters do not pay attention increase at a rate of \(\lambda_p\) towards 1. Again, we can think of this as candidates fading from memory, and so voters become less likely to cast a vote in opposition to such candidates. For convenience, the propensities in this case are adjusted at the same rate of \(\lambda_p\).

**Updating of aspirations.** The updating of aspirations follows the Cho-Matsui rule (Cho & Matsui (2005)), where aspirations in a certain period are the average of all previous impressions:

\[
a^{t+1}_\theta = \frac{1}{\sum_{\tau=1}^{t} |K^\tau_\theta|} \sum_{\tau=1}^{t} \sum_{j \in K^\tau_\theta} \pi^\tau_j(j)
\]

where \(K^\tau_\theta\), for each \(\tau = 1, ..., t\), denotes the set of candidates voters of type \(\theta\) pay attention to in period \(\tau\), and \(|K^\tau_\theta|\) is its cardinality.\(^4\) Intuitively, the aspirations in period \(t + 1\) are the average of all impressions of candidates (to which voters pay attention) from \(\tau = 1\) to \(t\).

### 3 Stationary distributions

We investigate the stationary distributions of the process. The proposition that follows characterizes the properties of the stationary distributions, when the distribution of the "event" vector \(e^t_\theta\) is stationary across time, i.e., the probabilities \(\{Pr^t_\theta(K)\}_{K \subseteq N}\) do not depend on time \(t\).\(^5\)

**Proposition 1** Consider the adjustment process under plurality, approval, and negative plurality rules previously defined where attentiveness is event-driven. Suppose that the distribution of the event vector \(e^t_\theta\) is stationary. Then in a stationary distribution of the process the distribution of votes within group \(\theta\) of voters is:

\(^4\)We note that \(K^\tau_\theta\) is a random variable, with the distribution \(\{Pr^t_\theta(K)\}_{K \subseteq N}\) as defined previously. We also note that if \(\sum_{\tau=1}^{t} |K^\tau_\theta| = 0\), i.e., voters do not pay attention to any candidate up to period \(t\), then the average of impressions from \(\tau = 1\) to \(t\) is not well-defined. In this situation, we can take, for example, \(a^{t+1}_\theta = a^0_\theta\) as an arbitrary value.

\(^5\)The proposition requires only the weaker assumption that the probabilities \(\{Pr^t_\theta(K)\}_{K \subseteq N}\) do not depend on time \(t\) in a stationary distribution of the adjustment process.
(a) under plurality rule: \( S_\theta(i) = q_\theta(i) \sum_{K \subseteq N \setminus \{i\}} \Pr_\theta(K \cup \{i\}) E[\frac{1}{1 + I_\theta(K)}] \) for each \( i \in N \),

(b) under approval voting: \( S_\theta(i) = q_\theta(i) \sum_{K \subseteq N \setminus \{i\}} \Pr_\theta(K \cup \{i\}) \) for each \( i \in N \),

(c) under negative plurality: \( S_\theta(i) = 1 - (1 - q_\theta(i)) \sum_{K \subseteq N \setminus \{i\}} \Pr_\theta(K \cup \{i\}) E[\frac{1}{1 + I_\theta(K)}(2 + I_\theta(K))] \) for each \( i \in N \).

Before proceeding, we make two observations. First, as in the full attentiveness case, the measure of votes \( S_\theta(i) \) increases with \( q_\theta(i) \) in all three rules, but decreases with \( I_\theta(K) \) under plurality, and negative plurality rules. \( S_\theta(i) \) is independent of \( I_\theta(K) \) under approval voting. Both the probability of positive impression \( q_\theta(i) \) and the count variable \( I_\theta(K) \) depend on preferences, i.e., they depend on the vectors of fixed components \( u_\theta \). Second, \( S_\theta(i) \) increases with the probability that voters pay attention to candidate \( i \), \( \Pr_\theta(K \cup \{i\}) \), under plurality rule and approval voting, but decreases with \( \Pr_\theta(K \cup \{i\}) \) under negative plurality. In other words, increased attentiveness is beneficial to a candidate under plurality rule and approval voting, but hurts the candidate under negative plurality rule. These two observations indicate that voter attentiveness is as important as preferences for the outcome of the election.

Let us consider now the difference in vote measures of two candidates \( i, j \in N \). To obtain sharper results, we assume that the events \( \{e_\theta(i)\}_{i \in N} \) are independent, though qualitatively the results will be similar in the general case where the events may be correlated. Then, we have the following proposition.

**Proposition 2** Consider the adjustment process under plurality, approval voting and negative plurality, and suppose that the distribution of events \( \{e_\theta(i)\}_{i \in N} \) is stationary and they are independent of each other. Then, for two distinct candidates \( i, j \in N \), the differences in vote measures of candidates \( i \) and \( j \) across the three rules in a stationary distribution are:

(a) under plurality rule:

\[
S_\theta(i) - S_\theta(j) = (r_\theta(i)q_\theta(i) - r_\theta(j)q_\theta(j)) \cdot \\
(\frac{1}{1 + I_\theta(K)} - r_\theta(l)q_\theta(l)E_{K \subseteq N \setminus \{i,j\}}[\frac{1}{1 + I_\theta(K)(2 + I_\theta(K))}]),
\]

(b) under approval voting:

\[
S_\theta(i) - S_\theta(j) = r_\theta(i)q_\theta(i) - r_\theta(j)q_\theta(j),
\]
The proposition has two consequences. First, under all three rules, the difference in vote measures $S_\theta(i) - S_\theta(j)$ is more likely to be positive when, on average, voters have a positive impression of candidate $i$ and a negative impression of candidate $j$, i.e., $q_\theta(i)$ is large, and $q_\theta(j)$ is small. On the other hand, the impact of "attention" probabilities differs across rules: under plurality and approval voting, the difference $S_\theta(i) - S_\theta(j)$ is more likely to be positive when, on average, voters pay more attention to candidate $i$ than $j$, i.e., $r_\theta(i)$ is large, and $r_\theta(j)$ is small, while under negative plurality, voters pay less attention to candidate $i$ than $j$, i.e., $r_\theta(i)$ is small, and $r_\theta(j)$ is large. This implies that under plurality rule and approval voting, a candidate that voters dislike, i.e., on average voters have negative impressions of the candidate, but that attracts attention, may defeat a candidate that voters like, but does not attract attention. Under negative plurality, in contrast, a candidate that voters dislike and to which voters do not pay attention may defeat a better liked candidate. We will further discuss these implications later in Sections 4.1 (in the context of exogenous attentiveness), and 4.2 (in the context of an example). Before moving to the second consequence, let us observe that the marginal effect of increased attentiveness is larger for candidates that voters prefer more under plurality and approval voting, and candidates that voters prefer less under negative plurality. In other words, under plurality and approval voting, candidates that voters like more will benefit more from higher attentiveness than the other candidates, while under negative plurality, candidates that voters like more will be hurt less from higher attentiveness than other candidates.

The second consequence concerns the effect of the other candidates $l \neq i, j$ on the difference in vote measures between candidates $i$ and $j$. We note that candidates $l \neq i, j$ impact the difference $S_\theta(i) - S_\theta(j)$ under plurality and negative plurality rules, but not under approval voting. The impact under plurality and negative plurality is as follows. First, the difference $S_\theta(i) - S_\theta(j)$ decreases with the attention probabilities $r_\theta(l)$, for each other candidate $l \in N \setminus \{i, j\}$ under both plurality and negative plurality rules. Intuitively, the difference in vote measures $S_\theta(i) - S_\theta(j)$ is larger (in absolute value) when, everything else equal, voters of group $\theta$ pay less attention to the other candidates $l \in N \setminus \{i, j\}$. In this case,

\[
S_\theta(i) - S_\theta(j) = (r_\theta(j)(1 - q_\theta(j)) - r_\theta(i)(1 - q_\theta(i))) \cdot (E_{K \subseteq N \setminus \{i, j\}}[1 + J_\theta(K)] - r_\theta(l)(1 - q_\theta(l)) E_{K \subseteq N \setminus \{i, j, l\}}[1 + J_\theta(K)][(1 + J_\theta(K))(2 + J_\theta(K))]).
\]
the impact of voters of group $\theta$ on the election outcome is larger as compared to the case where they pay more attention to the candidates $l \in N \setminus \{i, j\}$. We will study the normative implications of this observation in more detail later in Section 4.2. Second, the difference $S_\theta(i) - S_\theta(j)$ decreases with the probabilities of positive impression $q_\theta(l)$ under plurality rule, but increases with the probabilities $q_\theta(l)$ under negative plurality. This implies that, under plurality rule, once we control for the differences $r_\theta(i)q_\theta(i) - r_\theta(j)(q_\theta(j))$, differences in vote measures are larger at the top and smaller at the bottom of the distribution of votes. Under negative plurality in contrast, controlling for differences $r_\theta(j)(1-q_\theta(j)) - r_\theta(i)(1-q_\theta(i))$, differences in vote measures are smaller at the top and larger at the bottom of the distribution of votes.

Finally, we conclude this section with the following observation. In the particular case of full attentiveness, where voters pay attention, with probability one, to all candidates, i.e., $Pr_{t_\theta}(N) = 1$, while $Pr_{t_\theta}(K) = 0$ for all subsets $K \subset N$ and all $t \geq 0$, we obtain the measures of votes derived in the model of Andonie & Diermeier (2018):

(a) under plurality rule: $S_\theta(i) = q_\theta(i)E[\frac{1}{1+r_\theta(N \setminus \{i\})}]$ for each candidate $i \in N$,

(b) under approval voting: $S_\theta(i) = q_\theta(i)$ for each candidate $i \in N$,

(c) under negative plurality: $S_\theta(i) = 1 - (1-q_\theta(i))E[\frac{1}{1+r_\theta(N \setminus \{i\})}]$ for each candidate $i \in N$.

4 Attentiveness in multi-candidate elections

In this section, we first discuss two particular cases of attentiveness, endogenous and exogenous attentiveness, and their implications. This is followed by a normative analysis of limited attentiveness.

4.1 Exogenous and endogenous attentiveness

In this section, we discuss the factors that make voters pay attention to candidates, and analyse the implications. Voters may pay attention to a candidate because of factors external to the election e.g., because of name recognition (e.g., celebrities), party affiliation, or prior electoral success (e.g., Anagol & Fujiwara (2016)). We call this type of attentiveness exogenous. Alternatively, attention may shift during a campaign. Voters, for example, may pay attention to the top vote getters according to opinion polls. Media stories often feature horse-race coverage where heavily trailing candidates are usually ignored. In primary elections with many candidates, for example, there is a maximal number of candidates who can participate in a televised debate. We call this second type of attentiveness endogenous.
Formally, under exogenous attentiveness the distribution of the event vector $e_t^\theta$ is degenerate, with $\Pr_t^\theta(K) = 1$ for a given subset $K \subseteq N$, and $\Pr_t^\theta(K') = 0$ for all other subsets $K' \subseteq N$, $K' \neq K$. $K$ is a fixed set of candidates. Intuitively, the set $K$ contains the candidates to which voters pay attention for exogenous reasons, e.g., "celebrity" candidates, candidates associated with a well-known party, candidates that did well in prior elections, etc., while voters do not pay attention to the other candidates. The stationary distributions when attentiveness is exogenous are presented in the next corollary, which follows from Proposition 1.

**Corollary 1 (Exogenous Attentiveness.)** Consider the three voting rules, and suppose attentiveness is exogenous where all voters pay attention to the candidates in the subset $K_{exo}$ at each time $t \geq 0$. Then, under all three rules, the stochastic process has a unique stationary distribution where the distribution of measures of votes for each group $\theta$ of voters is:

(a) under plurality rule: $S_\theta(i) = q_\theta(i)E\left[\frac{1}{1+J_\theta(K_{exo}\setminus\{i\})}\right]$ for each $i \in K_{exo}$, and $S_\theta(i) = 0$ for each $i \notin K_{exo}$,

(b) under approval voting: $S_\theta(i) = q_\theta(i)$ for each $i \in K_{exo}$, and $S_\theta(i) = 0$ for each $i \notin K_{exo}$,

(c) under negative plurality rule: $S_\theta(i) = 1 - (1 - q_\theta(i))E\left[\frac{1}{1+J_\theta(K_{exo}\setminus\{i\})}\right]$ for each $i \in K_{exo}$, and $S_\theta(i) = 1$, for each $i \notin K_{exo}$.

The case of exogenous attentiveness illustrates clearly the effect of attentiveness on vote measures described in Proposition 1. In all rules but negative plurality, candidates that command attention for exogenous reasons are at an advantage compared to other candidates. In contrast, under negative plurality, candidates that attract attention, on average, obtain more negative votes than the other candidates. This has the perverse effect that candidates benefit from flying under the radar, avoiding attention if possible. This incentive to be invisible here, however, is limited to a specific rule (negative plurality).

Consider now endogenous attentiveness. For a given range of attentiveness $k \geq 1$, under endogenous attentiveness, the distribution of the event vector $e_t^\theta$ is degenerate: if $K$ is the set of the top $k$ candidates at time $t$, then $\Pr_t^\theta(K) = 1$, and $\Pr_t^\theta(K') = 0$ for all other $K' \subseteq N$, with $K' \neq K$. In other words, voters pay attention only to the candidates with the largest measures of votes, while they ignore the remaining candidates with the lowest measures of votes. In the next corollary (again, which follows from Proposition 1), we describe the stationary distributions under endogenous attentiveness for plurality rule, and approval voting. The stationary distributions under negative plurality are described in a separate proposition.

13
Corollary 2 (Endogenous Attentiveness.) Consider plurality rule and approval voting, and suppose attentiveness is endogenous where all voters pay attention to the top $k$ candidates at each time $t$. Then, for any set of $k$ candidates, $K = \{i_1, i_2, \ldots, i_k\} \subseteq N$, there exists a stationary distribution where the distribution of measures of votes for each group $\theta$ of voters is:

(a) under plurality rule: $S_\theta(i) = q_\theta(i)E\left[\frac{1}{1+r_\theta(K\setminus\{i\})}\right]$ for each $i \in K$, and $S_\theta(i) = 0$ for each $i \notin K$,

(b) under approval voting: $S_\theta(i) = q_\theta(i)$ for each $i \in K$, and $S_\theta(i) = 0$ for each $i \notin K$.

When the range of attentiveness is $k < n$, the process has more than one stationary distribution. As an illustrative case, if the range of attentiveness is $k = 2$, then there will be $\binom{n}{2} = \frac{n(n-1)}{2}$ distributions, where in a distribution, a set of two candidates share the electorate’s votes, while the other candidates obtain measures of zero votes. In this distribution, for say, the set of candidates $\{i, j\}$, candidates $i$ and $j$ measures of votes will be proportional to how well they are liked by the electorate, as specified by the fixed components $v_i^j$’s and $v_j^i$’s; while the other candidates $l \neq i, j$ will not obtain any votes, since such candidates will be out of the range of voters’ attentiveness.

Consider now negative plurality. Under endogenous attentiveness, and for a range of attentiveness $k < n$, the adjustment process does not have stationary distributions. Nevertheless, the process has distributions that are ”close” to a stationary distribution, when the adjustment parameter $\lambda_p$ of propensities is close to 0. For this reason, we will consider a weaker version of stationarity that requires that the distribution of measures of votes be in a ”small” neighbourhood around a fixed distribution of votes $\{S_\theta(i)\}_{\theta, i \in N}$ when $\lambda_p$ is close to 0. In this case, we call the distribution ”quasi-stationary”, which is defined as follows.

**Definition 1** A distribution of propensities and aspirations for each group $\theta$ of voters, $\{p_\theta, a_\theta\}_\theta$, is said to be quasi-stationary if the measures of votes the distribution generates

---

7In the case of full attentiveness, i.e., $k = n$, which corresponds to the model of Andonie & Diermeier (2018), under plurality rule and approval voting, the process has a unique stationary distribution.

8We note in passing that in the trivial case, where voters are minimally attentive and pay attention to a single candidate, i.e., $k = 1$, there are $n$ stationary distributions for each of the candidates. In each stationary distribution, one candidate $i$ will obtain $S(i) = 0.5$, while all others obtain $S(j) = 0$ (where $j \neq i$), and $S(A) = 0.5$. That is, half the electorate votes for the winning candidate, half the electorate abstains, and the remaining candidates receive zero votes.

9Under negative plurality, if attentiveness is endogenous then the distribution of the event vector $e_\theta^t$ can never be stationary, and so Proposition 1 does not apply. Moreover, it is easy to see that the process cannot have distributions where the vote measures are stationary.
is "close" to a fixed distribution of measures \( \{S_\theta(i)\}_{\theta,i \in N} \) in each period, i.e., it converges to the fixed distribution \( \{S_\theta(i)\}_{\theta,i \in N} \) as \( \lambda_p \to 0 \).

The characterization of the quasi-stationary distributions under negative plurality when attentiveness is "endogenous" for general \( n \) candidate elections is technically difficult. Nevertheless, we are able to fully describe the stationary and quasi-stationary distributions in elections with \( n = 3 \) candidates. A discussion of elections with \( n \) candidates can be found in the appendix (Section 7.4).

**Proposition 3** Consider an election with three candidates, i.e., \( N = \{1, 2, 3\} \) under the negative plurality rule, and suppose attentiveness is endogenous where all voters pay attention to the top \( k \) candidates at each time \( t \). Then, the quasi-stationary distributions of the process are as follows:

1. If \( k = 3 \) then the process has a unique stationary distribution where \( S(i) = \sum_\theta s_\theta (1 - (1 - q_\theta(i)) E[\frac{1}{1 + J_\theta(N \{i\})}]) \) for each candidate \( i = 1, 2, 3 \).

2. If \( k = 2 \) then: (a) if the average \( v^i_\theta \) (across all voters) for a given candidate \( i \) is substantially larger than the same average for the other two candidates, then the process has a quasi-stationary distribution where candidate \( i \) wins, and the other two candidates obtain approximately the same measures of votes, i.e., \( S(i) > \max_{j \neq i} S(j) \) and \( S(l) = \max_{j \neq i} S(j) \) for all \( l \neq i \) as \( \lambda_p \to 0 \); (b) if there is no candidate for which the average \( v^i_\theta \) (across all voters) is larger than the same average for the other candidates, then the process has a quasi-stationary distribution where all three candidates obtain approximately the same measures of votes, i.e., \( S(1) = S(2) = S(3) \) as \( \lambda_p \to 0 \).

3. If \( k = 1 \) then the process has a unique quasi-stationary distribution where the vote measures of all three candidates are the same, i.e., \( S(1) = S(2) = S(3) \) as \( \lambda_p \to 0 \).

The insights of Proposition 3 carry over to the general case with \( n \) candidates (see the discussion in the appendix, Section 7.4). If the range of attentiveness is less than \( n \) i.e., \( k < n \), the process has a quasi-stationary distribution where at least \( (n - k) + 1 \) candidates obtain equal measures of votes. These are the candidates for which their average of the fixed components \( v^i_\theta \)'s in the electorate is the lowest. The remaining candidates can obtain larger or the same measures of votes depending on their support measured in terms of the average \( v^i_\theta \)'s. Candidates with sufficiently large average \( v^i_\theta \)'s obtain measures of votes exceeding the level corresponding to the bottom \( (n - k) + 1 \) candidates. Candidates with an average \( v^i_\theta \)'s
not sufficiently large compared to that of the bottom \((n-k)+1\) candidates obtain the same measures of votes as those at the bottom level.

The previous results under endogenous and exogenous attentiveness have various consequences for our understanding of momentum. Consider the case of primary elections (under plurality rule). First, primaries can exhibit strong momentum effects. Candidates can come to the attention of voters for exogenous reasons (e.g., because of their celebrity status) or endogenous reasons (e.g., due to a combination of chance events that propels a "dark horse" in the spotlight). With the typical horse race coverage of the media, strong early showing by candidates in opinion polls further increases attention. For example, suppose a celebrity candidate \(i\) enters the race. Because of his celebrity status the media report on the event and voters pay attention (i.e., \(i \in K_{exo}\)). If \(K_{exo}\) is small, candidate \(i\) can improve in the polls very quickly. This improvement in the polls will be more noticeable when voters’ aspirations are low, potentially because of negative past impressions of the field of candidates or simply because of lack of familiarity.\(^{10}\) Such momentum, however, will only persist if candidate \(i\) has sufficient support in the electorate i.e., if the \(v_i\)'s for the candidate are large. Otherwise, candidate \(i\)'s momentum will quickly fade. More specifically, this will be the case if the candidate’s momentum is, largely, driven by lucky draws of \(\epsilon\), and not because of large fixed components \(v_i\)'s. In this case, voters’ aspirations will start to increase as a result of positive initial impressions. However, rising aspirations together with potential later negative draws of \(\epsilon\), imply that the initial momentum will quickly fade away. If momentum, however, is driven by large fixed components \(v_i\)'s, the momentum will continue, even though not at the initial peaks, again because of rising aspirations as a consequence of positive impressions. The model thus can account for the phenomenon of candidates that gather rapid support, peak, and fade away (Bartels (1988), Sides & Vavreck (2013)).\(^{11}\)

Second, a decreased attentiveness of voters can lead to a swift decrease in the number of candidates that are viable. For example, media coverage that only focuses on the top two

\(^{10}\)Recall that in our model aspirations are arithmetic means of past impressions. Therefore, if past impressions are negative, then aspirations will be low.

\(^{11}\)Sides & Vavreck (2013) have shown that these peak-shaped patterns of increasing and decreasing support in the 2012 presidential primary are correlated with similar dynamics in media interest which trigger attention. Sides and Vavrick suggest that emergence-peak-decline pattern can be explained by a shift in media coverage, where journalist first "discover" a candidate, but then shift to a more critical mode which finds problems or inconsistency in the candidate’s record. Our model provides an alternative explanation for the same pattern that does not depend on a change in the style of media coverage. Rather, candidates can gather support from a combination of sufficiently low aspiration and a series of attention catching events. This support, however, will fade if voters do not sufficiently like the candidate compared to others they pay attention to. This shift is entirely driven by the dynamics of attentiveness and does not require a change in media tone.
vote getters, or TV debate formats that only invite the leaders in opinion polls will quickly winnow down the number of viable candidates. That is, the model can imply outcomes as those suggested by Duverger’s law (Duverger (1954)). As an example, consider the case of an intermediate range of attentiveness, where \( k = 2 \). In this instance then, if attentiveness is endogenous, a voter will distribute her vote propensities among the top two candidates, with larger propensity on the candidate she likes more (i.e., with larger \( v^0_i \)). The voter will not vote for the other candidates. This behavior may appear indistinguishable from that of "strategic" voters to an observer, but it is driven by attentiveness rather than calculation of pivot probabilities as in rational voter models (Palfrey (1989), Myerson & Weber (1993), Myerson (2002), etc.). Note also that for higher ranges of attentiveness, i.e., \( k > 2 \) more than two candidates may receive positive vote measures, as suggested by Myatt (2007) and Myatt & Fisher (2002). But such candidates will only receive significant vote measures if they are sufficiently well liked by the electorate.

4.2 Normative analysis of limited attentiveness

In this section, we study the welfare implications of limited attentiveness in examples of multi-candidate elections. We will consider the following election, due to Myerson (2002). The set of candidates is \( N = \{1, 2, 3\} \). The electorate consists of two groups of voters, labeled 1 and 2. Their measures are \( s_1 = s \), and \( s_2 = 1 - s \), where we assume, without loss of generality, \( s \geq 0.5 \). The vectors of fixed components are \( u_1 = \{1, 0, v\} \), and \( u_2 = \{0, 1, v\} \) respectively, where \( v \) is a parameter. We will consider two cases. The first case is \( v < 0 \), the "one bad apple" election (Myerson (2002), example 2), where candidate 3 is ranked last by all voters. The second case is \( v > 1 \), the "above the fray" election (Myerson (2002), example 1), where candidate 3 is ranked first by all voters.

We illustrate the two consequences of Proposition 2. First, Proposition 2 implies that under plurality and approval voting, a candidate that voters dislike can win the election if voters pay sufficient attention to that candidate, at the expense of the candidates they prefer.\(^{12}\) Recall that in our model voters do not need to have well-defined preference orderings that they are aware of. Candidates only relate to voters through impressions and only if voters are aware of the candidates. Hence, voters may not know that some other feasible candidate may be preferable. In the "one bad apple" election, i.e., the case \( v < 0 \)

\(^{12}\) Conversely, under plurality and approval voting, a candidate that voters prefer can lose the election if voters do not pay attention to that candidate, but pay attention to other candidates they do not like. Under negative plurality, on the other hand, a candidate that voters prefer can lose the election if voters pay much attention to that candidate, and little attention to the other candidates. The analysis of this implication, i.e., the case of \( v > 1 \) in the election discussed by Myerson (2002), is analogous to the analysis below.
above, candidate 3 can win the election under plurality and approval voting, if the attention probabilities \( r_\theta(3) \) are sufficiently large relative to probabilities \( r_\theta(1) \), and \( r_\theta(2) \), for each group of voters \( \theta = 1, 2 \). Indeed, e.g., under plurality rule, the difference in vote measures of candidates 3 and 1, by Proposition 2 is:

\[
S(3) - S(1) = \sum_\theta s_\theta(r_\theta(3)q_\theta(3) - r_\theta(1)q_\theta(1))(1 - \frac{1}{2}r_\theta(2)q_\theta(2))
\]

For example, if \( r_\theta(1) < r_\theta(3)\frac{q_\theta(3)}{q_\theta(1)} \), for both groups of voters \( \theta = 1, 2 \), then we have \( S(3) - S(1) > 0 \), despite that fact that the probabilities of positive impression of candidate 3 are lower than the corresponding probabilities for candidate 1, i.e., \( q_\theta(3) < q_\theta(1) \) for each group \( \theta \). Similarly, if the attention probabilities of candidate 2 for the two groups of voters, \( r_\theta(2) \), are low enough, e.g., \( r_\theta(2) < r_\theta(3)\frac{q_\theta(3)}{q_\theta(2)} \) for each group \( \theta = 1, 2 \), then we have \( S(3) - S(2) > 0 \). Thus, in this case candidate 3 wins.

In contrast, under negative plurality, a candidate that voters dislike can win the election if voters pay little attention to that candidate, but pay more attention to other candidates they prefer. Again, consider the ”one bad apple” election. Then, under negative plurality, the difference in vote measures of candidates 3 and 1 is:

\[
S(3) - S(1) = \sum_\theta s_\theta(r_\theta(1)(1 - q_\theta(1)) - r_\theta(3)(1 - q_\theta(3)))(1 - \frac{1}{2}r_\theta(2)(1 - q_\theta(2)))
\]

If \( r_\theta(3) < r_\theta(1)\frac{1 - q_\theta(1)}{1 - q_\theta(3)} \) for each group of voters \( \theta = 1, 2 \), then \( S(3) - S(1) > 0 \), even though all voters have low probabilities of positive impression of candidate 3, relative to the probabilities for candidate 1, i.e., \( q_\theta(3) < q_\theta(1) \) for each group \( \theta \). Analogously, if \( r_\theta(3) < r_\theta(2)\frac{1 - q_\theta(2)}{1 - q_\theta(3)} \) for each group of voters \( \theta = 1, 2 \), then \( S(3) - S(2) > 0 \). In other words, candidate 3 has the largest measure of votes, and so he wins.

Second, Proposition 2 shows that under plurality and negative plurality rules, voters that focus their attention on a small subset of candidates will have a larger impact on the election outcome. Under approval voting, on the other hand, this effect is absent, as the vote measure of a given candidate is independent of the attention voters pay to or their impressions of other candidates. To illustrate the welfare implications, consider the election from above where \( v < 0 \) (i.e., ”one bad apple” election), so candidate 1 is the Condorcet winner. To simplify analysis and obtain clean results, we assume that the random component is uniformly distributed, i.e., \( \epsilon \sim U[-1, +1] \). Suppose first that all voters pay attention to all candidates with probability one, i.e., the case of full attentiveness. Then, under all three electoral rules, the outcome of this election is efficient (Andonie & Diermeier (2018)).
Proposition 4 (Andonie & Diermeier (2018)) Consider the "one bad apple" election. If all voters pay attention to all candidates with probability one, i.e., the case of full attentiveness, where $r^t_\theta(i) = 1$ for each candidate $i \in N$ and group of voters $\theta$, then under all three electoral rules, the process has a (unique) stationary distribution where candidate 1 wins.

Suppose now that voters of the second group are less attentive, and in particular they pay less attention to candidate 3, while they are still fully attentive to candidates 1 and 2. Then, the outcome of this election in this case is described in the proposition below.

Proposition 5 Consider the "one bad apple" election. Suppose voters of group 1 pay attention to all candidates with probability one, i.e., $r^t_1(i) = 1$ for each candidate $i \in N$, while voters of group 2 are fully attentive to candidates 1, and 2, but pay only partial attention to candidate 3, i.e., $r^t_2(i) = 1$ for $i \in \{1, 2\}$, and $r^t_2(3) = r^*_2 < 1$. Then, under all three electoral rules, the process has a (unique) stationary distribution where:

(1) Under plurality, and negative plurality rules, candidate 1 loses if the measure of voters of group 2 is sufficiently large, i.e., if $s \in (0.5, s^*)$, and, otherwise, candidate 1 wins.

(2) Under approval voting, candidate 1 wins for all $s > 0.5$.

Thus, if the measure of group 2 of voters is large enough i.e., $s$ is close to 0.5, then the Condorcet winner cannot win under plurality, and negative plurality rules. This is an illustration of the effect described above, where voters of group 2 pay less attention to candidate 3, which amplifies the impact of their votes on the election outcome. This effect, however, is absent under approval voting, where the Condorcet winner always wins for any $s > 0.5$.

Let us now discuss the stationary distributions of this election under endogenous attentiveness. Our goal is to study the qualitative consequences of limited attentiveness across the three rules. To simplify analysis, we consider the election at the beginning of this section and assume that $s_1 = s_2 = 50\%$, i.e., the election from Myerson (2002).

Proposition 6 Consider the election described above, and suppose that the measures of the two groups are the same, i.e., $s_1 = s_2 = 50\%$. If attentiveness is "endogenous" and the range of attentiveness is $k = 2$ for all voters, then:

(i) Under plurality and approval rules, there are three stationary distributions.
(a) A first distribution where \( S(1) = S(2) > S(3) = 0 \). Thus, in this distribution candidate 3 loses.

(b) A second distribution where \( S(1), S(3) > 0, \) and \( S(2) = 0 \). In this distribution, candidate 1 wins if \( v < \frac{1}{2} \), while candidate 3 wins if \( v > \frac{1}{2} \).

(c) A third distribution where \( S(2), S(3) > 0, \) and \( S(1) = 0 \). In this distribution, candidate 2 wins if \( v < \frac{1}{2} \), while candidate 3 wins if \( v > \frac{1}{2} \).

(ii) Under negative plurality, there is one (unique) quasi-stationary distribution. There exists a cut-off \( v^* > \frac{1}{2} \) such that, if \( v > v^* \) then in the (unique) quasi-stationary distribution \( S(3) > S(1) \approx S(2) \), so candidate 3 wins, and candidates 1 and 2 lose with (approximately) equal measures of votes. If \( v \leq v^* \), then in the (unique) quasi-stationary distribution \( S(1) \approx S(2) \approx S(3) \), i.e., all candidates obtain (approximately) equal measures of votes.\(^{13}\)

When \( v > 1 \), the candidate that all voters prefer (candidate 3) may lose under the plurality rule or approval voting. The reason for the existence of a stationary distribution where candidate 3 loses derives from the possibility that voters in the early states of a race may not pay sufficient attention to candidate 3, perhaps for exogenous reasons or due to a sequence of unlucky events that prevent candidate 3 from attracting attention. With \( k = 2 \), voters only pay attention to the top two candidates and candidate 3 will gradually fade from view, even though he was preferred by every voter to the eventual winner. Of course, voters may never know that such a candidate existed as their ranking over candidates is not common knowledge. Negative plurality avoids the vicious circles of plurality, or approval voting, where low-vote measure candidates do not attract attention, and vote measures of candidates that do not attract attention further decrease. In contrast, under negative plurality, candidates to which voters do not pay attention do not receive negative votes, and therefore their vote measures increase, making it more likely that voters pay attention to them. In the election from above where \( v > 1 \), all voters rank candidate 3 above candidates 1, and 2, in terms of their fixed components \( v_i^j \). The dynamics of the adjustment process under negative plurality implies that, in a quasi-stationary distribution, the vote measure of candidate 3 stabilizes at a level that is above the vote measures of candidates 1 and 2. That is, candidate 3 wins. However, as we have seen, negative plurality has other, problematic, properties. For example, it may select a candidate that all voters oppose (e.g., the case of \( v < 0 \) below).

\(^{13}\)The cutoff \( v^* \) is characterized by the implicit equation: 
\[
(3 - F(\frac{2v^* - 1}{4\alpha}))(8 - F(\frac{2v^* - 3}{4\alpha}) - F(\frac{1+2v^*}{4\alpha})) = 16.
\]
Second, when \( v < 0 \), under plurality rule and approval voting the candidate that all voters oppose (candidate 3) does not win in any stationary distribution. Intuitively, candidate 3 always loses, but in two different ways. In one case candidate 3 fails to attract the electorate’s attention and obtains a vote measure of \( S(3) = 0 \) (case \( i(a) \)). In the other two cases, candidate 3 is one of the two candidates that captures the electorate’s attention. But in that case candidate 3 loses against either candidate 1 or 2, but does capture a positive share of the vote. In contrast, under negative plurality there is a unique quasi-stationary distribution where candidate 3 has a positive probability of winning. Although all voters rank candidate 3 below candidates 1 and 2, there is no candidate that dominates the others in terms of the fixed components \( v_i \) across the electorate. The dynamics of the adjustment process under negative plurality implies that in the quasi-stationary distribution all candidates receive the same measures of votes.

To summarize, under endogenous and partial attentiveness, only plurality and approval rules are efficient when there is a candidate that all voters oppose. Conversely, only negative plurality is efficient when there is a candidate that all voters prefer. These efficiency results are not tied to the specifics of the electoral settings we considered. They apply to any election where there is a candidate that either all voters oppose, or all voters prefer. The analysis, and intuition are the same as above.

5 Good news versus bad news

As we have seen capturing the attention of the electorate, at least under plurality rule and approval voting, is very valuable for candidates. In many cases attention will be triggered by media coverage. This raises the question on what conditions “all news is good news”. That is, whether even negative media coverage can benefit a candidate, especially in a crowded field with many candidates, as was the case in the Republican primary during the 2016 U.S. Presidential election, when candidate Donald Trump captured disproportionate media attention through a sequence of controversial statements. While most of the media coverage was negative in tone, it kept then candidate Trump in the headlines and maintained the attention of the electorate.

To analyze this phenomenon, we develop an extension of the attentiveness model where voters’ attention is solely triggered by news coverage, and where news coverage of a candidate can be good, neutral or bad in tone. In the basic model, the distribution of impressions

---

\(^{14}\)In the quasi-stationary distribution under negative plurality, each of the three candidates can win, however the probability that candidate 3 wins is smaller than the probability that either candidate 1 or 2 wins.
across a group of voters was single-peaked and centered around the fixed component $v^i_\theta$ for a candidate. In other words, news on a candidate was always neutral. We can expand the model to capture the concepts of "bad" and "good" news by shifting the distribution of impressions to the left, or right respectively, of the fixed component $v^i_\theta$. Specifically, we assume that the impressions of voters of group $\theta$ of candidate $i$ have the following distribution:

$$\pi^i_\theta(i) := v^i_\theta + \mu^i + \alpha \epsilon^i_\theta(i),$$

where $\mu^i$ is a news parameter connected with candidate $i$ that reflects the average tone of news on candidate $i$. When $\mu^i < 0$, the distribution of impressions is shifted to the left of the fixed component $v^i_\theta$ for candidate $i$, and in this case we say that news is "bad". Conversely, when $\mu^i > 0$, the distribution of impressions is shifted to the right of the fixed component $v^i_\theta$ for candidate $i$, and in this case we say that news is "good". If $\mu^i = 0$ then news is neutral, which corresponds to the basic model. Furthermore, to focus the analysis, we assume that voters pay attention to candidates solely via news coverage. In other words, the attention probability $r^i_\theta(i)$ associated with a candidate $i$ corresponds to the average amount of media coverage of candidate $i$, which we call news frequency. For example, a high value of $r^i_\theta(i)$ corresponds to the case where candidate $i$ frequently appears in the news, and voters pay increased attention to the candidate. We note that the extended model is equivalent to the basic model, where the fixed component $v^i_\theta$ associated with each candidate $i$ shifts in the direction of the news parameter $\mu^i$ for that candidate, i.e., the vector of fixed components for each group $\theta$ is $u^\theta = \{w^1_\theta, ..., w^n_\theta\}$, where $w^i_\theta := v^i_\theta + \mu^i$ for each candidate $i \in N$.

Let us first analyze the impact of news tonality on aspirations and the probabilities of positive impressions in a stationary distribution. We note that, in a stationary distribution, the aspirations $a^\theta$ increase with $\mu^i$. Because the mean of the distribution of impressions of candidate $i$ increases with $\mu^i$, while the impressions of other candidate $j \neq i$ do not depend on $\mu^i$, we have that $g^i_\theta(i)$, the probability of positive impression of candidate $i$, increases with

---

15We can think of this specification as a simplified form of the following more detailed conception of news coverage. Suppose the distribution of impressions of candidate $i$ among voters of type $\theta$ is as follows:

$$\pi^i_\theta(i) := v^i_\theta + \chi^i + \xi^i + \alpha \epsilon^i_\theta(i),$$

where $\chi^i$ is a discrete random variable that can take values "$+\nu^i$" and "$-\nu^i$", with probabilities $\pi^i$, and $1 - \pi^i$ respectively; and $\xi^i$ is a continuous random variable with a symmetric, single-peaked, distribution around 0. The two additional random variables, $\chi^i$ and $\xi^i$, are to model the following two features. First, in some periods the news coverage of a candidate can be positive, i.e., when the realization of $\chi^i$ is $+\nu^i$, and in other periods news coverage can be negative, i.e., when the realization of $\chi^i$ is $-\nu^i$. Second, some voters may interpret some news as good, while others may interpret the same news as bad. This is captured by the second random variable $\xi^i$. This more detailed extension, however, is equivalent to the specification above, where $\mu^i$ corresponds to the expectation of $\chi^i$, and the uncertainty is combined into the random term $\epsilon^i_\theta(i)$. 

---

22
is particularly so when media increases coverage of candidate $i$ as the benefit of increased coverage may swamp the negative impact of bad news. This previously. A candidate $i$ is positive. In contrast, other candidates, $j \in K_{\text{exo}}$, obtain a vote measure of 0, even though $\mu^i$, while $q_\theta(j)$, the probability of positive impression of each candidate $j \neq i$, decreases with $\mu^i$.

Using the effect on impression probabilities described above, we can now study the effect on the distribution of votes $\{S_\theta(l)\}_{l \in N}$ Consider an election under plurality rule, and suppose, as before, that the events $\{e_\theta(l)\}_{l \in N}$ are independent. Then, analogously as in Proposition 2, the vote measure of a candidate $l \in N$ in the stationary distribution can be written as:

$$S_\theta(l) = r_\theta(l)q_\theta(l)(E_{K \subseteq N \setminus \{l,l'\}}[\frac{1}{1 + I_\theta(K)}] - r_\theta(l')q_\theta(l')E_{K \subseteq N \setminus \{l,l'\}}[\frac{1}{(1 + I_\theta(K))(2 + I_\theta(K))}]).$$

As $q_\theta(i)$ increases with $\mu^i$, while $q_\theta(j)$ decreases with $\mu^i$, better news increases the vote measure of candidate $i$, $S_\theta(i)$, and decreases, on average, the vote measure of each candidate $j \neq i$, $S_\theta(j)$ Conversely, worse news hurts candidate $i$, and benefits the other candidates $j \neq i$. Also, as we discussed previously, e.g., after Propositions 1 and 2, the vote measure of a candidate $i$, $S_\theta(i)$, increases with the frequency of media coverage of candidate $i$, $r_\theta(i)$, but decreases with the frequency of media coverage of other candidates $j \neq i$. Consequently, candidate $i$ may even benefit from (moderately) bad news, provided news coverage occurs at sufficiently high frequency, i.e., moderately low $\mu^i$, combined with large values of $r_\theta(i)$, as the benefit of increased coverage may swamp the negative impact of bad news. This is particularly so when media increases coverage of candidate $i$ at the expense of other candidates $j$. As an example, consider the case of exogenous attentiveness we discussed previously. A candidate $i$ benefits from being a candidate with exogenous attentiveness, $i \in K_{\text{exo}}$, even when news coverage is negative, as the vote measure the candidate obtains is positive. In contrast, other candidates, $i \notin K_{\text{exo}}$, obtain a vote measure of 0, even though $\mu^i$, while $q_\theta(j)$, the probability of positive impression of each candidate $j \neq i$, decreases with $\mu^i$.

---

16The effect of an increase of $\mu^i$ on the mean of impressions $\pi_\theta(i)$ is larger than the effect on the aspirations $a_\theta$, and so the probability of positive impression $q_\theta(i)$ increases with $\mu^i$.

17The discussion below of the effect of the news tonality on the distribution of votes is analogous to the discussion after Proposition 2 on the effect of attentiveness on vote measures.

18Because of its relevance in elections, e.g., in party primaries, we focus our analysis on plurality rule. The effect under the other two rules is similar, and we will describe it briefly afterwards.

19To see that, consider first the effect of an increase in the impression probability $q_\theta(i)$ on the distribution of vote measures. An increase of $q_\theta(i)$ (assuming $q_\theta(l)$, for $l \neq i$, are fixed) induces an increase in the vote measure of candidate $i$, $S_\theta(i)$, and a decrease in the vote measure of each candidate $j \neq i$, $S_\theta(j)$. Consider now in addition the effect of a decrease in the impression probabilities $q_\theta(l)$, for each candidate $l \neq i$. First, $S_\theta(i)$ further increases. Second, we note that the measure of abstentions $S_\theta(A)$ increases as well, as voters become less likely to have positive impressions of candidates $l \neq i$. This implies that, on average, the vote measure of a candidate $j \neq i$ further decreases. The vote measures $S_\theta(j)$, for $j \neq i$, decrease on average, however, the vote measure of some candidates $l \neq i$ may increase, depending on their ranks in the vector of fixed components $u_\theta$, and the shape of the distribution function of the random component $\epsilon$. 23
news coverage would be positive, were media to cover those candidates. Thus, notoriety, even if it is persistently negative, is valuable for candidates.

Consider finally the other two rules. First, under negative plurality, the analysis of good versus bad news is identical to that under plurality, though with the usual inverse direction, i.e., less attention is beneficial for candidates. Second, the analysis under approval voting is similar to the direct effect under plurality rule (i.e., the first effect described in footnote 19), where the change in news tonality or media coverage of a candidate $i$ impacts only the vote measure of candidate $i$, but not the vote measures of other candidates $j \neq i$.

6 Conclusion

We study a model of attentiveness under three electoral rules when voters vote based on impressions (Andonie & Diermeier (2018)). Positive impressions lead to an increased propensity to cast a ballot favoring a candidate, negative impressions decrease the propensity. Voters do not necessarily pay full attention to an electoral campaign and are only aware of a subset of the candidates. A candidate can receive attention exogenously, e.g., because of name recognition, or endogenously, e.g., when the media only pays attention to a subset of candidates that are leading in the polls. In any case, the support for a candidate can only increase if voters pay attention to the candidate.

We model this approach as a stochastic dynamic process and derive its stationary distributions on voting propensities for plurality rule, approval voting, and negative plurality rule for different levels of attentiveness and compare their properties. We show that, with less than full attentiveness, elections can fail to select a unanimously preferred candidate if that candidate fails to gather sufficient attention during the race. Electoral success thus depends on public support and the ability to capture the electorates’ attention. The voting rules vary in their propensity to fail to elect preferred candidates or to elect candidates that most of the electorate dislikes. Voting rules also vary in how attention matters. While under plurality and approval gaining attention is always desirable, candidates in elections under negative plurality rule prefer to “fly under the radar.”

Second, we show that the extent of inattentiveness influences the relative impact of voters on outcomes. Specifically, larger inattentiveness is associated with a larger impact on the election outcome under plurality and negative plurality rules, but not under approval voting. This implies that, under plurality or negative plurality rules, e.g., a candidate preferred by a minority group can win if the minority group pays less attention to candidates than the

---

20For technical reasons in the case of negative plurality we derive “quasi-stationary” distributions.
opposing majority group.

Under limited attentiveness stationary distributions exhibit momentum effects (Bartels (1988), Sides & Vavreck (2013)). Of particular interest is the case where some candidates have an exogenous advantage in capturing public attention, e.g., because of name recognition (Anagol & Fujiwara (2016)). Such candidates can experience rapidly growing support that is further enhanced by "horse-race" coverage in the media. However, unless such candidates have sufficient public support, their electoral fortunes will peak and they will fade into irrelevance. Exogenous attentiveness thus only provides a material advantage if there is sufficient electoral support for the candidate. "Sufficient support”, however, does not mean that the candidate commands a majority. In multi-candidate races under plurality rule and approval voting even unanimously preferred candidates may lose. These insights suggest an alternative role for political parties: they may serve as mechanisms to ensure attention to their official candidates.

Finally, we consider the case where media coverage of candidate can be favorable or unfavorable. We show that, under plurality rule and approval voting, a candidate may benefit from even bad news, as the benefit of increased attention due to frequent news coverage may swamp the negative impact of bad news.

Overall, elections with impressionable voters can perform well, especially when attentiveness is high. But the effect of attentiveness on outcomes depends on the electoral rule. These insights point to the crucial role of the media in influencing public attention, both social and traditional. Another interesting venue to explore pertains to the role of campaign finances. In our setting more financial resources can be used to secure attention for a candidate. This impact can be substantial, especially in the context of a primary with multiple candidates. Future work is needed to fully incorporate the effect of the mass media or competitive campaign advertising into models of impressionable voting.
Appendix

7.1 Proof of Proposition 1

**Plurality rule.** Given a candidate $i \in N$ and a subset of candidates $K \subseteq N \setminus \{i\}$, with probability $\Pr_{\theta}^t(K \cup \{i\})$ a voter of group $\theta$ pays attention only to the candidates in the subset $K \cup \{i\}$. In this case, the equation for the voting propensity $p_{\theta}^t(i)$ is:

$$p_{\theta}^{t+1}(i) = (1 - \lambda_p)p_{\theta}^t(i) + \lambda_p \frac{1}{1 + I_{\theta}^t(K)} \mathbb{1}(\pi_{\theta}^t(i) \geq a_{\theta}^t).$$

With probability $\Pr_{\theta}^t(K)$ a voter of group $\theta$ pays attention only to the candidates in the subset $K$, and, in this case, the voting probability $p_{\theta}^t(i)$ is updated as:

$$p_{\theta}^{t+1}(i) = (1 - \lambda_p)p_{\theta}^t(i)$$

Averaging over all subsets $K \subseteq N \setminus \{i\}$, we can compute the expectation of propensities $p_{\theta}^{t+1}(i)$ as:

$$E[p_{\theta}^{t+1}(i)] = (1 - \lambda_p)E[p_{\theta}^t(i)] + \lambda_p \sum_{K \subseteq N \setminus \{i\}} q_{\theta}(i) \Pr_{\theta}^t(K \cup \{i\}) E\left[\frac{1}{1 + I_{\theta}^t(K)} | \pi_{\theta}^t(i) \geq a_{\theta}^t\right]$$

In a stationary distribution, both distributions of $p_{\theta}^t$ and $a_{\theta}^t$ are time independent, which implies $E[p_{\theta}^{t+1}(i)] = E[p_{\theta}^t(i)]$. Using this condition in the previous equation and omitting the time $t$, we have that in a stationary distribution the average propensity of voting for candidate $i$ among voters of group $\theta$ is:

$$E[p_{\theta}(i)] = q_{\theta}(i) \sum_{K \subseteq N \setminus \{i\}} \Pr_{\theta}(K \cup \{i\}) E\left[\frac{1}{1 + I_{\theta}(K)}\right]$$

As we assume a continuum of voters, the average voting propensity coincides with the measure of voters of group $\theta$ that vote for candidate $i$, and therefore we can write:

$$S_{\theta}(i) = q_{\theta}(i) \sum_{K \subseteq N \setminus \{i\}} \Pr_{\theta}(K \cup \{i\}) E\left[\frac{1}{1 + I_{\theta}(K)}\right]$$

**Approval voting.** The proof is similar to that for plurality rule. Given a candidate $i \in N$ and a subset of candidates $K \subseteq N \setminus \{i\}$, with probability $\Pr_{\theta}^t(K \cup \{i\})$ a voter of group $\theta$ pays attention only to the candidates in the subset $K \cup \{i\}$. In this case, the equation for the voting propensity $p_{\theta}^t(i)$ is:

$$p_{\theta}^{t+1}(i) = (1 - \lambda_p)p_{\theta}^t(i) + \lambda_p \frac{1}{1 + I_{\theta}(K)} \mathbb{1}(\pi_{\theta}^t(i) \geq a_{\theta}^t).$$

With probability $\Pr_{\theta}^t(K)$ a voter of group $\theta$ pays attention only to the candidates in the subset $K$, and, in this case, the voting probability $p_{\theta}^t(i)$ is updated as:

$$p_{\theta}^{t+1}(i) = (1 - \lambda_p)p_{\theta}^t(i)$$

Averaging over all subsets $K \subseteq N \setminus \{i\}$, we can compute the expectation of propensities $p_{\theta}^{t+1}(i)$ as:

$$E[p_{\theta}^{t+1}(i)] = (1 - \lambda_p)E[p_{\theta}^t(i)] + \lambda_p \sum_{K \subseteq N \setminus \{i\}} q_{\theta}(i) \Pr_{\theta}^t(K \cup \{i\}) E\left[\frac{1}{1 + I_{\theta}(K)} | \pi_{\theta}^t(i) \geq a_{\theta}^t\right]$$

In a stationary distribution, both distributions of $p_{\theta}^t$ and $a_{\theta}^t$ are time independent, which implies $E[p_{\theta}^{t+1}(i)] = E[p_{\theta}^t(i)]$. Using this condition in the previous equation and omitting the time $t$, we have that in a stationary distribution the average propensity of voting for candidate $i$ among voters of group $\theta$ is:

$$E[p_{\theta}(i)] = q_{\theta}(i) \sum_{K \subseteq N \setminus \{i\}} \Pr_{\theta}(K \cup \{i\}) E\left[\frac{1}{1 + I_{\theta}(K)}\right]$$

As we assume a continuum of voters, the average voting propensity coincides with the measure of voters of group $\theta$ that vote for candidate $i$, and therefore we can write:

$$S_{\theta}(i) = q_{\theta}(i) \sum_{K \subseteq N \setminus \{i\}} \Pr_{\theta}(K \cup \{i\}) E\left[\frac{1}{1 + I_{\theta}(K)}\right]$$

21 When $t \to +\infty$, the aspirations $a_{\theta}$ converge to the average of the fixed components $\{v_i^\theta\}_{i \in N}$, weighted by the attention probabilities $\{\Pr_{\theta}(K)\}_{K \subseteq N}$. Because the random variables $\frac{1}{1 + I_{\theta}(K)}$ and $\pi_{\theta}^t(i) \geq a_{\theta}^t$ are independent, we can omit the condition $\pi_{\theta}^t(i) \geq a_{\theta}^t$ in the expectation. (see Andonie & Diermeier (2019)).

26
\(i \in N\) and a subset of candidates \(K \subseteq N \setminus \{i\}\), a voter from group \(\theta\) pays attention only to candidates in the subset \(K \cup \{i\}\) with probability \(\Pr_{\theta}^{t}(K \cup \{i\})\) and the equation for the updating of \(p_{\theta}^{t}(i)\) is:

\[
p_{\theta}^{t+1}(i) = (1 - \lambda_{p})p_{\theta}^{t}(i) + \lambda_{p}1_{\{\pi_{\theta}^{t}(i) \geq a_{\theta}^{t}\}}
\]

On the other hand, a voter from group \(\theta\) pays attention only to candidates in the set \(K\) with probability \(\Pr_{\theta}^{t}(K)\) and, in this case, the equation for the updating of \(p_{\theta}^{t}(i)\) is:

\[
p_{\theta}^{t+1}(i) = (1 - \lambda_{p})p_{\theta}^{t}(i)
\]

If we average across all subsets \(K \subseteq N \setminus \{i\}\), we obtain:

\[
E[p_{\theta}^{t+1}(i)] = (1 - \lambda_{p})E[p_{\theta}^{t}(i)] + \lambda_{p}q_{\theta}(i) \sum_{K \subseteq N \setminus \{i\}} \Pr_{\theta}^{t}(K \cup \{i\})
\]

As we are interested in a stationary distribution, the distributions of \(p_{\theta}^{t}\) and \(a_{\theta}^{t}\) are time independent, and so \(E[p_{\theta}^{t+1}(i)] = E[p_{\theta}^{t}(i)]\). Similarly to the case of plurality rule, we then obtain:

\[
E[p_{\theta}(i)] = q_{\theta}(i) \sum_{K \subseteq N \setminus \{i\}} \Pr_{\theta}^{t}(K \cup \{i\})
\]

and so the measure of voters from group \(\theta\) that approve candidate \(i\) is:

\[
S_{\theta}(i) = q_{\theta}(i) \sum_{K \subseteq N \setminus \{i\}} \Pr_{\theta}^{t}(K \cup \{i\})
\]

**Negative plurality.** The proof of the stationary distribution for negative plurality follows similar steps to those for plurality and approval voting. For an arbitrary candidate \(i \in N\) and a subset of candidates \(K \subseteq N \setminus \{i\}\), if a voter of group \(\theta\) pays attention only to the candidates in the subset \(K \cup \{i\}\), which occurs with probability \(\Pr_{\theta}^{t}(K \cup \{i\})\), then:

\[
p_{\theta}^{t+1}(i) = (1 - \lambda_{p})p_{\theta}^{t}(i) + \lambda_{p}1\left(\frac{1}{1 + J_{\theta}^{t}(K)}1_{\{\pi_{\theta}^{t}(i) < a_{\theta}^{t}\}}\right)
\]

If a voter of group \(\theta\) pays attention only to the candidates in the subset \(K\), which occurs with probability \(\Pr_{\theta}^{t}(K)\), then:

\[
p_{\theta}^{t+1}(i) = (1 - \lambda_{p})p_{\theta}^{t}(i) + \lambda_{p}
\]
As before, we average across all subsets \( K \subseteq N \setminus \{i\} \) and compute the expectation of \( p^t_{\theta^i} \) as:

\[
E[p^t_{\theta^i}] = (1 - \lambda_p)E[p^t_{\theta}] + \lambda_p(1 - \sum_{K \subseteq N \setminus \{i\}} (1 - q_\theta(i)) \Pr_\theta(K \cup \{i\})E\left[\frac{1}{1 + J_\theta(K)}|\pi_\theta(i) < a^\prime_\theta\right]
\]

In a stationary distribution, we have \( E[p^t_{\theta^i}] = E[p^t_{\theta}] \), which yields:

\[
E[p_\theta(i)] = 1 - (1 - q_\theta(i)) \sum_{K \subseteq N \setminus \{i\}} \Pr_\theta(K \cup \{i\})E\left[\frac{1}{1 + J_\theta(K)}\right]
\]

Finally, the fact there is a continuum of voters implies that the vote measure of candidate \( i \) is:

\[
S_\theta(i) = 1 - (1 - q_\theta(i)) \sum_{K \subseteq N \setminus \{i\}} \Pr_\theta(K \cup \{i\})E\left[\frac{1}{1 + J_\theta(K)}\right]
\]

### 7.2 Proof of Proposition 2

Consider first plurality rule. Given two distinct candidates \( i,j \in N \), using Proposition 1, the measure of votes candidate \( i \) obtains among voters of group \( \theta \) in a stationary distribution can be written as:

\[
S_\theta(i) = \sum_{K \subseteq N \setminus \{i\}} \Pr_\theta(K \cup \{i\})E\left[\frac{1}{1 + I_\theta(K)}\right]
\]

\[
= \sum_{K \subseteq N \setminus \{i,j\}} \{\Pr_\theta(K \cup \{i\})E\left[\frac{1}{1 + I_\theta(K)}\right] + \Pr_\theta(K \cup \{i,j\})E\left[\frac{1}{1 + I_\theta(K \cup \{j\})}\right]\}
\]

\[
= \sum_{K \subseteq N \setminus \{i,j\}} \{\Pr_\theta(K \cup \{i\}) + \Pr_\theta(K \cup \{i,j\})\}E\left[\frac{1}{1 + I_\theta(K)}\right]
\]

\[
- \sum_{K \subseteq N \setminus \{i,j\}} \Pr_\theta(K \cup \{i,j\})E\left[\frac{1}{1 + I_\theta(K)(2 + I_\theta(K))}\right]
\]

Therefore, the difference in measure of votes \( S_\theta(i) - S_\theta(j) \) is:

\[
S_\theta(i) - S_\theta(j) = \sum_{K \subseteq N \setminus \{i,j\}} \{\Pr_\theta(K \cup \{i\}) + \Pr_\theta(K \cup \{i,j\})\}E\left[\frac{1}{1 + I_\theta(K)}\right]
\]

\[
- \sum_{K \subseteq N \setminus \{i,j\}} \Pr_\theta(K \cup \{j\})E\left[\frac{1}{1 + I_\theta(K)}\right]
\]
When the attention events are independent, we have:

\[
\Pr_\theta(K \cup \{i\}) + \Pr_\theta(K \cup \{i, j\}) = r_\theta(i) \prod_{l \in K} r_\theta(l') \prod_{l' \notin K \cup \{i, j\}} (1 - r_\theta(l'))
\]

\[
\Pr_\theta(K \cup \{j\}) + \Pr_\theta(K \cup \{i, j\}) = r_\theta(j) \prod_{l \in K} r_\theta(l') \prod_{l' \notin K \cup \{i, j\}} (1 - r_\theta(l'))
\]

Therefore, the difference in measures of votes becomes:

\[
S_\theta(i) - S_\theta(j) = (r_\theta(i)q_\theta(i) - r_\theta(j)q_\theta(j)) \sum_{K \subseteq N \setminus \{i, j\}} \prod_{l \in K} r_\theta(l') \prod_{l' \notin K \cup \{i, j\}} (1 - r_\theta(l'))E\left[\frac{1}{1 + I_\theta(K)}\right]
\]

For a fixed candidate \(l \in N \setminus \{i, j\}\), the weight on the difference \((r_\theta(i)q_\theta(i) - r_\theta(j)q_\theta(j))\) can be written as follows:

\[
= \sum_{K \subseteq N \setminus \{i, j\}} \prod_{l \in K} r_\theta(l') \prod_{l' \notin K \cup \{i, j\}} (1 - r_\theta(l'))E\left[\frac{1}{1 + I_\theta(K)}\right]
\]

\[
+ \sum_{K \subseteq N \setminus \{i, j\}} \prod_{l \in K} r_\theta(l') \prod_{l' \notin K \cup \{i, j\}} (1 - r_\theta(l'))E\left[\frac{1}{1 + I_\theta(K \cup \{l\})}\right]
\]

\[
- r_\theta(l)q_\theta(l) \sum_{K \subseteq N \setminus \{i, j, l\}} \prod_{l \in K} r_\theta(l') \prod_{l' \notin K \cup \{i, j, l\}} (1 - r_\theta(l'))E\left[\frac{1}{(1 + I_\theta(K))(2 + I_\theta(K))}\right]
\]

\[
= E_{K \subseteq N \setminus \{i, j, l\}}\left[\frac{1}{1 + I_\theta(K)}\right] - r_\theta(l)q_\theta(l)E_{K \subseteq N \setminus \{i, j, l\}}\left[\frac{1}{(1 + I_\theta(K))(2 + I_\theta(K))}\right]
\]

This proves part (a) of the proposition.

The proof in the case of approval voting is trivial. So, finally, let us consider negative plurality rule. Similarly to plurality rule, for two distinct candidates \(i, j \in N\), using Proposition [1] we write the vote measure that candidate \(i\) obtains among voters of group \(\theta\) in a
stationary distribution as follows:

\[ S_\theta(i) = 1 - (1 - q_\theta(i)) \sum_{K \subseteq N \setminus \{i\}} \Pr_\theta(K \cup \{i\}) E\left[ \frac{1}{1 + J_\theta(K)} \right] \]
\[ = 1 - (1 - q_\theta(i)) \sum_{K \subseteq N \setminus \{i,j\}} \{ \Pr_\theta(K \cup \{i\}) E\left[ \frac{1}{1 + J_\theta(K)} \right] + \Pr_\theta(K \cup \{i,j\}) E\left[ \frac{1}{1 + J_\theta(K \cup \{j\})} \right] \} \]
\[ = 1 - (1 - q_\theta(i)) \sum_{K \subseteq N \setminus \{i,j\}} \{ \Pr_\theta(K \cup \{i\}) + \Pr_\theta(K \cup \{i,j\}) \} E\left[ \frac{1}{1 + J_\theta(K)} \right] \]
\[ + (1 - q_\theta(i))(1 - q_\theta(j)) \sum_{K \subseteq N \setminus \{i,j\}} \Pr_\theta(K \cup \{i,j\}) E\left[ \frac{1}{(1 + J_\theta(K))(2 + J_\theta(K))} \right] \]

This implies that the difference \( S_\theta(i) - S_\theta(j) \) is:

\[ S_\theta(i) - S_\theta(j) = -(1 - q_\theta(i)) \sum_{K \subseteq N \setminus \{i,j\}} \{ \Pr_\theta(K \cup \{i\}) + \Pr_\theta(K \cup \{i,j\}) \} E\left[ \frac{1}{1 + J_\theta(K)} \right] \]
\[ + (1 - q_\theta(j)) \sum_{K \subseteq N \setminus \{i,j\}} \{ \Pr_\theta(K \cup \{j\}) + \Pr_\theta(K \cup \{i,j\}) \} E\left[ \frac{1}{1 + J_\theta(K)} \right] \]

If the attention events are independent, then the difference in vote measures \( S_\theta(i) - S_\theta(j) \) becomes:

\[ S_\theta(i) - S_\theta(j) = \left( r_\theta(j)(1 - q_\theta(j)) - r_\theta(i)(1 - q_\theta(i)) \right) \cdot \sum_{K \subseteq N \setminus \{i,j\}} \prod_{l' \in K \setminus \{i,j\}} r_\theta(l') \prod_{l \notin K \cup \{i,j\}} (1 - r_\theta(l')) E\left[ \frac{1}{1 + J_\theta(K)} \right] \]

For a fixed candidate \( l \in N \setminus \{i,j\} \), we can write the weight on the difference \( (r_\theta(j)(1 -
\[ q_\theta(j) - r_\theta(i)(1 - q_\theta(i)) \] as follows:

\[
\sum_{K \subseteq N \setminus \{i,j\}} \prod_{l' \in K \cup \{i\}} r_\theta(l') \prod_{l' \notin K \cup \{i,j\}} (1 - r_\theta(l')) E\left[ \frac{1}{1 + J_\theta(K)} \right]
\]

\[
= \sum_{K \subseteq N \setminus \{i,j,l\}} \prod_{l' \in K} r_\theta(l') \prod_{l' \notin K \cup \{i,j\}} (1 - r_\theta(l')) E\left[ \frac{1}{1 + J_\theta(K \cup \{l\})} \right]
\]

\[
+ \sum_{K \subseteq N \setminus \{i,j,l\}} \prod_{l' \in K} r_\theta(l') \prod_{l' \notin K \cup \{i,j\}} (1 - r_\theta(l')) E\left[ \frac{1}{1 + J_\theta(K \cup \{l\})} \right]
\]

\[
- r_\theta(l)(1 - q_\theta(l)) \sum_{K \subseteq N \setminus \{i,j,l\}} \prod_{l' \in K} r_\theta(l') \prod_{l' \notin K \cup \{i,j,l\}} (1 - r_\theta(l')) E\left[ \frac{1}{(1 + J_\theta(K))(2 + J_\theta(K))} \right]
\]

\[
= E_{K \subseteq N \setminus \{i,j\}} \left[ \frac{1}{1 + J_\theta(K)} \right] - r_\theta(l)(1 - q_\theta(l)) E_{K \subseteq N \setminus \{i,j,l\}} \left[ \frac{1}{(1 + J_\theta(K))(2 + J_\theta(K))} \right]
\]

This proves part (c) of the proposition.

### 7.3 Proof of Proposition 3

(1) The result for the range of attentiveness \( k = 3 \) follows directly from the case of full attentiveness for \( n = k = 3 \).

(2) The proof for the range of attentiveness \( k = 2 \) relies on the following result. Consider the following dynamic (deterministic) adjustment process. The parameters of the process are: \( \lambda_p \in (0, 1) \), which we assume to be close to 0, together with the fixed constants \( S_{12}(1), S_{12}(2), S_{13}(1), S_{13}(3), S_{23}(2) \) and \( S_{23}(3) \in (0, 1) \). The variables \( S^t(1), S^t(2), \) and \( S^t(3) \) are adjusted according to the following equations.

(1) If \( S^t(1) > S^t(3) \), and \( S^t(2) > S^t(3) \) then:

\[
S^{t+1}(1) = (1 - \lambda_p)S^t(1) + \lambda_p S_{12}(1)
\]

\[
S^{t+1}(2) = (1 - \lambda_p)S^t(2) + \lambda_p S_{12}(2)
\]

\[
S^{t+1}(3) = (1 - \lambda_p)S^t(3) + \lambda_p
\]
(2) If \( S^t(1) > S^t(2) \), and \( S^t(3) > S^t(2) \) then:

\[
\begin{align*}
S^{t+1}(1) &= (1 - \lambda_p)S^t(1) + \lambda_p S_{13}(1) \\
S^{t+1}(2) &= (1 - \lambda_p)S^t(2) + \lambda_p S_{23}(2) \\
S^{t+1}(3) &= (1 - \lambda_p)S^t(3) + \lambda_p S_{23}(3)
\end{align*}
\]

(3) If \( S^t(2) > S^t(1) \), and \( S^t(3) > S^t(1) \) then:

\[
\begin{align*}
S^{t+1}(1) &= (1 - \lambda_p)S^t(1) + \lambda_p S_{23}(2) \\
S^{t+1}(2) &= (1 - \lambda_p)S^t(2) + \lambda_p S_{23}(3) \\
S^{t+1}(3) &= (1 - \lambda_p)S^t(3) + \lambda_p S_{23}(3)
\end{align*}
\]

Before we analyze the properties of the process, consider the following three conditions:

(C1) \( \frac{1 - S_{12}(1)}{1 - S_{12}(2)} + \frac{1 - S_{13}(1)}{1 - S_{13}(3)} < 1 \)

(C2) \( \frac{1 - S_{12}(2)}{1 - S_{12}(1)} + \frac{1 - S_{23}(2)}{1 - S_{23}(3)} < 1 \)

(C3) \( \frac{1 - S_{13}(3)}{1 - S_{13}(1)} + \frac{1 - S_{23}(3)}{1 - S_{23}(2)} < 1 \)

In regard to these three conditions, we can show the following.

**Lemma 1** Either only (C1) holds, or only (C2) holds, or only (C3) holds, or neither (C1),(C2) and (C3) hold.

**Proof.** Suppose, for example, that (C1) holds. Then this implies that both \( \frac{1 - S_{12}(1)}{1 - S_{12}(2)} < 1 \) and \( \frac{1 - S_{13}(1)}{1 - S_{13}(3)} < 1 \). Therefore we have \( S_{12}(1) > S_{12}(2) \) and \( S_{13}(1) > S_{13}(3) \), and moreover the amount by which \( S_{12}(1) \) exceeds \( S_{12}(2) \), and \( S_{13}(1) \) exceeds \( S_{13}(3) \) must be large enough. But this implies that, for example, condition (C2) cannot hold. If (C2) were to hold as well, then a similar argument as above would imply that \( S_{12}(2) > S_{12}(1) \) and \( S_{23}(2) > S_{23}(3) \), which is a contradiction. Thus if (C1) holds, for example, then the other two conditions cannot hold at the same time. Finally, it is possible that none of the three conditions holds. This is the case if the six constants \( S_{12}(1), S_{12}(2), S_{13}(1), S_{13}(3), S_{23}(2) \) and \( S_{23}(3) \) are relatively close to each other. As an example when they all have the same value, neither (C1), (C2), or (C3) hold. ■

Let us now return to the adjustment process. We claim the process possibly has two types of quasi-stationary distributions, which we describe in the lemma that follows.
Lemma 2 If condition (C1) holds, i.e. if \( \frac{1 - S_{12}(1)}{1 - S_{12}(2)} + \frac{1 - S_{13}(1)}{1 - S_{13}(2)} < 1 \) then the process has a unique quasi-stationary distribution where \( S(1) > S(2) = S(3) \) as \( \lambda_p \to 0 \). Similarly, for the cases where (C2) or (C3) hold. Finally, if none of the three conditions holds, then the process has a unique quasi-stationary distribution where \( S(1) = S(2) = S(3) \) as \( \lambda_p \to 0 \).

Proof. Let us call (S12) the state of the process where \( S'(1) > S'(3) \), and \( S'(2) > S'(3) \); and similarly for the other two states (S13) and (S23). Because all six parameters \( S_{12}(1), S_{12}(2), S_{13}(1), S_{13}(3), S_{23}(2) \) and \( S_{23}(3) \) are between 0 and 1, the process cannot be trapped in one single state, e.g. it cannot be trapped in state (S12) for example. Thus, the process must either be cycling between two of the three states, or between all three states ad infinitum. Let us consider the case where, for example, the process cycles between states (S12) and (S13) only, ad infinitum. We will show that this is possible if only if condition (C1) from above holds and \( \lambda_p \) is sufficiently close to 0, in which case the process has a quasi-stationary distribution where \( S(1) > S(2) = S(3) \) as \( \lambda_p \to 0 \). So let us start by assuming that the process cycles between states (S12) and (S13) only. Then, we note that the vector of the three variables \( (S'(1), S'(2), S'(3)) \) moves linearly towards either vector \( (S_{12}(1), S_{12}(2), 1) \) or vector \( (S_{13}(1), 1, S_{13}(3)) \). Therefore, if \( t \) is sufficiently large, independent of the initial starting point, the vector of variables \( (S'(1), S'(2), S'(3)) \) will end up on the line between points \( (S_{12}(1), S_{12}(2), 1) \) and \( (S_{13}(1), 1, S_{13}(3)) \), i.e. in the set of all convex combinations between \( (S_{12}(1), S_{12}(2), 1) \) and \( (S_{13}(1), 1, S_{13}(3)) \). Formally, this is:

\[
T = \{(S(1), S(2), S(3)) = \beta(S_{12}(1), S_{12}(2), 1) + (1 - \beta)(S_{13}(1), 1, S_{13}(3)),
\text{where } 0 \leq \beta \leq 1\}.
\]

Let us denote by \( (\overline{S}(1), \overline{S}(2), \overline{S}(3)) \) the point in set \( T \), where the variables \( S(2) \) and \( S(3) \) have the same value: \( \overline{S}(2) = \overline{S}(3) \). We can show, after a simple calculation, that:

\[
(\overline{S}(1), \overline{S}(2), \overline{S}(3)) = \overline{\beta}(S_{12}(1), S_{12}(2), 1) + (1 - \overline{\beta})(S_{13}(1), 1, S_{13}(3)),
\]

where \( \overline{\beta} = \frac{1 - S_{13}(3)}{1 - S_{12}(3) + 1 - S_{12}(2)} \). We note that if condition (C1) holds, then we will have \( \overline{S}(1) > \overline{S}(2) = \overline{S}(3) \). Now, for any vector of variables \( (S(1), S(2), S(3)) \in T \) between \( (\overline{S}(1), \overline{S}(2), \overline{S}(3)) \) and \( (S_{13}(1), 1, S_{13}(3)) \), or, equivalently, for \( \beta < \overline{\beta} \), we have \( S(3) < S(2) \), and so the process is in state (S12). In this case, the vector of variables \( (S(1), S(2), S(3)) \) moves linearly towards \( (S_{12}(1), S_{12}(2), 1) \). Conversely, if \( (S(1), S(2), S(3)) \) is between \( (\overline{S}(1), \overline{S}(2), \overline{S}(3)) \) and \( (S_{12}(1), S_{12}(2), 1) \), or, equivalently, for \( \beta > \overline{\beta} \), then we have \( S(3) > S(2) \), and so the process is in state (S13). In this case, the vector of variables \( (S(1), S(2), S(3)) \) moves linearly towards \( (S_{13}(1), 1, S_{13}(3)) \). Therefore, it is clear that as \( t \) gets large, the vector of
variables \((S(1), S(2), S(3))\) will settle on the line connecting vectors \((S_{12}(1), S_{12}(2), 1)\) and \((S_{13}(1), 1, S_{13}(3))\), and at a distance no larger than \(\lambda_p M\) (where \(M\) is a constant that depends on the parameters\(^{22}\)) from \((\vec{S}(1), \vec{S}(2), \vec{S}(3))\). Because \(\vec{S}(1) > \vec{S}(2) = \vec{S}(3)\) (which, as we noted above, is equivalent to condition (C1)), if \(\lambda_p\) is close enough to 0, then for all vectors \((S(1), S(2), S(3))\) at a distance no larger than \(\lambda_p M\) from \((\vec{S}(1), \vec{S}(2), \vec{S}(3))\), we will have that \(S(1) > S(2)\), and \(S(1) > S(3)\). Thus condition (C1) ensures that, for example, whenever the vector of the three variables \((S(1), S(2), S(3))\) is between \((\vec{S}(1), \vec{S}(2), \vec{S}(3))\) and \((S_{13}(1), 1, S_{13}(3))\), we have both \(S(1) > S(3)\) and \(S(2) > S(3)\), so the process is in state \((S_{12})\) as claimed above. Similarly, for the case where \((S(1), S(2), S(3))\) is between \((\vec{S}(1), \vec{S}(2), \vec{S}(3))\) and \((S_{12}(1), S_{12}(2), 1)\). Therefore, as \(\lambda_p \to 0\), the vector of the three variables \((S(1), S(2), S(3))\) gets arbitrarily close to \((\vec{S}(1), \vec{S}(2), \vec{S}(3))\) and in the limit it will coincide with \((\vec{S}(1), \vec{S}(2), \vec{S}(3))\). Thus, in summary, if condition (C1) holds then the process has a unique quasi-stationary distribution where \(S(1) > S(2) = S(3)\) as \(\lambda_p \to 0\).

A similar reasoning shows that if condition (C2) holds then the process has a unique quasi-stationary distribution where \(S(2) > S(1) = S(3)\) as \(\lambda_p \to 0\), and if condition (C3) holds then the process has a unique quasi-stationary distribution where \(S(3) > S(1) = S(2)\) as \(\lambda_p \to 0\). Finally, if neither (C1), (C2) or (C3) hold, then the process will cycle through all three states \((S_{12}), (S_{13})\) and \((S_{23})\), in which case we will have \(S(1) = S(2) = S(3)\) as \(\lambda_p \to 0\). When the process cycles among all three states \((S_{12}), (S_{13})\) and \((S_{23})\), the vector of variables \((S(1), S(2), S(3))\) moves linearly towards either \((S_{12}(1), S_{12}(2), 1), (S_{13}(1), 1, S_{13}(3))\) or \((1, S_{23}(2), S_{23}(3))\). Therefore, if \(t\) is sufficiently large, the process will end up in the convex hull of these three points, i.e. the set:

\[ T = \{(S(1), S(2), S(3)) = \beta_{12}(S_{12}(1), S_{12}(2), 1) + \beta_{13}(S_{13}(1), 1, S_{13}(3)) + \beta_{23}(1, S_{23}(2), S_{23}(3)), \text{ where } \beta_{12}, \beta_{13}, \beta_{23} \geq 0 \text{ and } \beta_{12} + \beta_{13} + \beta_{23} = 1\}. \]

The rest of the proof then follows an argument similar to that for proving Lemma 3 below. We will explain later how to use the argument of the proof to Lemma 3 to show that if neither (C1), (C2), or (C3) hold, then the process has a unique quasi-stationary distribution where \(S(1) = S(2) = S(3)\) as \(\lambda_p \to 0\). ■

We will now use Lemma 2 to prove the claims for the range of attentiveness \(k = 2\). We are looking for a quasi-stationary distribution when \(\lambda_p \to 0\), as given in Definition 1. Let us denote by \(q_0(i)\), without a time index, the probability that, in a given quasi-stationary

\(^{22}\)The precise value of \(M\) is: \(M = \max\{\vec{S}, 1 - \vec{S}\}\)(\(S_{12}(1), S_{12}(2), 1\) - \(S_{13}(1), 1, S_{13}(3)\)), where \(|x - y|| denotes the distance between vectors \(x\) and \(y\).
distribution, a voter of type $\theta$ has a positive impression of candidate $i$ (assuming the voter pays attention to candidate $i$).

Suppose, for example, that at time $t$ we have $S^t(1) > S^t(3)$ and $S^t(2) > S^t(3)$. That is, candidates 1 and 2 have the two largest measures of votes. Because the attentiveness is endogenous, and the range of attentiveness is $k = 2$, we have $K^t = \{1, 2\}$ for all voters. Using the updating equations for propensities $p^t_\theta(1)$, we have:

$$E[p^{t+1}_\theta(1)] = E[p^{t+1}_\theta(1)|1 \in K^t_\text{pos}] \Pr(1 \in K^t_\text{pos}) + E[p^{t+1}_\theta(1)|1 \in K^t_\text{neg}] \Pr(1 \in K^t_\text{neg})$$

$$= (1 - \lambda_p)E[p^t_\theta(1)] + \lambda_p \Pr(1 \in K^t_\text{pos}) + \lambda_p E[J^t \theta - 1 \bigg| 1 \in K^t_\text{neg}] \Pr(1 \in K^t_\text{neg})$$

$$= (1 - \lambda_p)E[p^t_\theta(1)] + \lambda_p \left(1 - \frac{1}{2}(1 - q^t_\theta(1))(1 + q^t_\theta(2))\right)$$

Alternatively, we can write the equation we obtained as: $S^{t+1}_\theta(1) = (1 - \lambda_p)S^t_\theta(1) + \lambda_p \left(1 - \frac{1}{2}(1 - q^t_\theta(1))(1 + q^t_\theta(2))\right)$. Summing over all types $\theta$ of voters, we then obtain:

$$S^{t+1}(1) = (1 - \lambda_p)S^t(1) + \lambda_p \left(1 - \frac{1}{2} \sum_\theta s_\theta(1 - q^t_\theta(1))(1 + q^t_\theta(2))\right).$$

An analogous calculation for candidate 2 shows that:

$$S^{t+1}(2) = (1 - \lambda_p)S^t(2) + \lambda_p \left(1 - \frac{1}{2} \sum_\theta s_\theta(1 - q^t_\theta(2))(1 + q^t_\theta(1))\right).$$

Finally, since candidate 3 $\notin K^t$ the equation for updating the propensity associated with candidate 3 is $p^{t+1}_\theta(3) = (1 - \lambda_p)p^t_\theta(3) + \lambda_p$. Taking expectation, we have $E[p^{t+1}_\theta(3)] = (1 - \lambda_p)E[p^t_\theta(3)] + \lambda_p$ or, alternatively, $S^{t+1}_\theta(3) = (1 - \lambda_p)S^t_\theta(3) + \lambda_p$. Finally, summing over all types $\theta$ of voters, we have:

$$S^{t+1}(3) = (1 - \lambda_p)S^t(3) + \lambda_p.$$

The other cases of $S^t(1) > S^t(2)$ and $S^t(3) > S^t(2)$; and $S^t(2) > S^t(1)$ and $S^t(3) > S^t(1)$ are similar. For example, when $S^t(1) > S^t(2)$ and $S^t(3) > S^t(2)$, the updating equations for
\( S^t(1), S^t(2) \) and \( S^t(3) \) are:

\[
S^{t+1}(1) = (1 - \lambda_p)S^t(1) + \lambda_p(1 - \frac{1}{2} \sum_\theta s_\theta(1 - q_\theta(1))(1 + q_\theta(3))) \\
S^{t+1}(3) = (1 - \lambda_p)S^t(3) + \lambda_p(1 - \frac{1}{2} \sum_\theta s_\theta(1 - q_\theta(3))(1 + q_\theta(1))) \\
S^{t+1}(2) = (1 - \lambda_p)S^t(2) + \lambda_p.
\]

Therefore, the equations characterizing the evolution of the three variables \( S^t(1), S^t(2) \) and \( S^t(3) \) follow the equations of the adjustment process that we previously studied. The constants are \( S_{ij}(i) = 1 - \frac{1}{2} \sum_\theta s_\theta(1 - q_\theta(i))(1 + q_\theta(j)) \) for each pair of candidates \( i \) and \( j \). Therefore, we can use Lemma 2 to conclude that the process can have one of two possible types of quasi-stationary distributions. In the first type, if, for example, condition (C1) holds then the process has a quasi-stationary distribution where \( S(1) > S(2) = S(3) \) as \( \lambda_p \to 0 \).

Condition (C1) in the present context is:

\[
(C1) : \frac{\sum_\theta s_\theta(1 - q_\theta(1))(1 + q_\theta(2))}{\sum_\theta s_\theta(1 - q_\theta(2))(1 + q_\theta(1))} + \frac{\sum_\theta s_\theta(1 - q_\theta(3))(1 + q_\theta(1))}{\sum_\theta s_\theta(1 - q_\theta(3))(1 + q_\theta(1))} < 1
\]

In our context, (C1) can hold only if candidate 1 has sufficiently large support in the electorate relative to the support of candidates 2 and 3.

To see that, we note that a necessary condition for (C1) to hold is that both:

\[
\frac{\sum_\theta s_\theta(1 - q_\theta(1))(1 + q_\theta(2))}{\sum_\theta s_\theta(1 - q_\theta(2))(1 + q_\theta(1))} < 1 \quad \text{and} \quad \frac{\sum_\theta s_\theta(1 - q_\theta(3))(1 + q_\theta(1))}{\sum_\theta s_\theta(1 - q_\theta(3))(1 + q_\theta(1))} < 1.
\]

These two necessary sub-conditions are equivalent to \( \sum_\theta s_\theta q_\theta(2) < \sum_\theta s_\theta q_\theta(1) \) and \( \sum_\theta s_\theta q_\theta(3) < \sum_\theta s_\theta q_\theta(1) \). For example, when the random component \( \epsilon \sim U[-1, +1] \) the two sub-conditions come to \( \sum_\theta s_\theta v_\theta^2 < \sum_\theta s_\theta v_\theta^1 \) and \( \sum_\theta s_\theta v_\theta^3 < \sum_\theta s_\theta v_\theta^1 \). Thus, in this case, (C1) can possibly hold only if the weighted average of fixed components \( v_\theta^1 \) for candidate 1 exceeds the corresponding weighted averages for candidates 2 and 3, and moreover \( \sum_\theta s_\theta v_\theta^2 \) exceeds both \( \sum_\theta s_\theta v_\theta^3 \) and \( \sum_\theta s_\theta v_\theta^3 \) by a sufficiently "large" amount. In other words, (C1) holds only if candidate 1 has sufficiently large support in the electorate relative to candidates 2 and 3, support which is measured in terms of the average of the fixed components \( v_\theta^1 \)’s. In this case, then we have that \( S(1) > S(2) = S(3) \) as \( \lambda_p \to 0 \). That is, if \( \lambda_p \) is sufficiently small, then candidate 1 wins the election for certain, while candidates 2 and 3 will have approximately equal measures of votes.

Finally, in the second type of quasi-stationary distributions, if none of the conditions (C1),
(C2) or (C3) hold, then the process has a quasi-stationary distribution where \( S(1) = S(2) = S(3) \), as \( \lambda_p \to 0 \). That is, if none of the three candidates has “sufficiently” large support in the electorate relative to the other two candidates, then the process has a quasi-stationary distribution where all three candidates have the same measures of votes as \( \lambda_p \to 0 \).

(3) The proof for the range of attentiveness \( k = 1 \) relies on the following result. Consider the following dynamic (and deterministic) adjustment process. The parameters of the process are: \( \lambda_p \in (0, 1) \), which we assume to be close to 0, together with the fixed constants \( Q(1) \), \( Q(2) \) and \( Q(3) \in (0, 1) \). The variables \( S^t(1) \), \( S^t(2) \), and \( S^t(3) \) are adjusted according to the following equations.

(1) If \( S^t(1) > S^t(2) \), and \( S^t(1) > S^t(3) \) then:

\[
\begin{align*}
S^{t+1}(1) &= (1 - \lambda_p)S^t(1) + \lambda_p Q(1) \\
S^{t+1}(2) &= (1 - \lambda_p)S^t(2) + \lambda_p \\
S^{t+1}(3) &= (1 - \lambda_p)S^t(3) + \lambda_p
\end{align*}
\]

(2) If \( S^t(2) > S^t(1) \), and \( S^t(2) > S^t(3) \) then:

\[
\begin{align*}
S^{t+1}(1) &= (1 - \lambda_p)S^t(1) + \lambda_p \\
S^{t+1}(2) &= (1 - \lambda_p)S^t(2) + \lambda_p Q(2) \\
S^{t+1}(3) &= (1 - \lambda_p)S^t(3) + \lambda_p
\end{align*}
\]

(3) If \( S^t(3) > S^t(1) \), and \( S^t(3) > S^t(2) \) then:

\[
\begin{align*}
S^{t+1}(1) &= (1 - \lambda_p)S^t(1) + \lambda_p \\
S^{t+1}(2) &= (1 - \lambda_p)S^t(2) + \lambda_p \\
S^{t+1}(3) &= (1 - \lambda_p)S^t(3) + \lambda_p Q(3)
\end{align*}
\]

Then we have:

**Lemma 3** The process has a unique quasi-stationary distribution where \( S(1) = S(2) = S(3) \) as \( \lambda_p \to 0 \).

**Proof.** We observe first that, as all three parameters \( Q(1) \), \( Q(2) \) and \( Q(3) \) are between 0 and 1, the process will cycle across all three possible states ad infinitum. We will show that as \( \lambda_p \) gets close to zero, the variables \( S^t(1) \), \( S^t(2) \) and \( S^t(3) \) will all be in a small neighborhood
of $Q$, where $\overline{Q}$ is defined by: \( \frac{1}{1-\overline{Q}} = \frac{1}{1-Q(1)} + \frac{1}{1-Q(2)} + \frac{1}{1-Q(3)} \). Moreover, as $\lambda_p \to 0$, the neighbourhood gets arbitrarily small and so the variables $S'(1)$, $S'(2)$ and $S'(3)$ all converge to $\overline{Q}$. We will provide a graphical argument for this claim. \(^{23}\)

First, we note that the vector of the three variables $(S(1), S(2), S(3))$ always moves linearly, towards either $(Q(1), 1, 1)$, $(1, Q(2), 1)$ or $(1, 1, Q(3))$. Therefore, independent of the initial values, the vector of the three variables $(S(1), S(2), S(3))$ must eventually lie in the convex hull (or, triangle) formed by the three points $\{(Q(1), 1, 1), (1, Q(2), 1), (1, 1, Q(3))\}$. This is the set:

\[
T = \{(S(1), S(2), S(3)) = \beta_1(Q(1), 1, 1) + \beta_2(1, Q(2), 1) + \beta_3(1, 1, Q(3)),
\]

where $\beta_1, \beta_2, \beta_3 \geq 0$ and $\beta_1 + \beta_2 + \beta_3 = 1$.

Let us partition the set $T$ into the following three regions:

\[
R_1 = \{(S(1), S(2), S(3)) \in T : S(1) \geq S(2), \text{ and } S(1) \geq S(3)\}
\]

\[
R_2 = \{(S(1), S(2), S(3)) \in T : S(2) \geq S(1), \text{ and } S(2) \geq S(3)\}
\]

\[
R_3 = \{(S(1), S(2), S(3)) \in T : S(3) \geq S(1), \text{ and } S(3) \geq S(2)\}.
\]

That is, in region $R_1$ $S(1)$ has the largest value, and similarly for regions $R_2$ and $R_3$. The intersection of the three regions, denoted by point $C$ in Figure \[1\], corresponds to the case where $S(1) = S(2) = S(3) = \overline{Q}$, where $\overline{Q}$ is characterized by \( \frac{1}{1-\overline{Q}} = \frac{1}{1-Q(1)} + \frac{1}{1-Q(2)} + \frac{1}{1-Q(3)} \).

For a graphical representation of the triangle $T$, together with the three-region partition, see Figure \[1\].

Given the adjustment process from above, if the vector of the three variable $(S(1), S(2), S(3))$ is in region $R_1$, then $(S(1), S(2), S(3))$ will move linearly towards the opposite vertex $(Q(1), 1, 1)$. At each step, it will get closer to $(Q(1), 1, 1)$ at a rate of $\lambda_p$ percent the current distance to $(Q(1), 1, 1)$. Similarly, when $(S(1), S(2), S(3))$ is in region $R_2$, it will move linearly towards $(1, Q(2), 1)$. Finally, when the vector $(S(1), S(2), S(3))$ is in region $R_3$, it will move linearly towards $(1, 1, Q(3))$, again at a rate of $\lambda_p$ percent the current distance to $(1, 1, Q(3))$. A simple graphical analysis shows that, starting from any point inside triangle $T$, the process will be attracted into a small neighbourhood around $C$, at a distance no larger than $\lambda_p M$ (where $M$ is a constant that depends on the parameters $Q(1)$, $Q(2)$, and $Q(3)$) from $C$, and it will stay there forever. Moreover, as $\lambda_p \to 0$, the neighbourhood containing the vector $(S(1), S(2), S(3))$ will get arbitrarily small, and it will collapse into point $C$ in the limit. In

\(^{23}\)An analytical proof is available on request.
other words, the process has a unique quasi-stationary distribution where \( S(1) = S(2) = S(3) \) as \( \lambda_p \to 0 \).

Finally, let us go back to Lemma 2 and show how to use the above argument to prove the last part of that Lemma, i.e. if neither of the conditions (C1), (C2), or (C3) from Lemma 2 holds, then the process has a unique quasi-stationary distribution where \( S(1) = S(2) = S(3) \) as \( \lambda_p \to 0 \). We noted above that if neither (C1), (C2), or (C3) holds, then the process cycles across all three states \((S12), (S13), \) and \((S23),\) and it will end up in the set:

\[
T = \{(S(1), S(2), S(3)) = \beta_{12}(S_{12}(1), S_{12}(2), 1) + \beta_{13}(S_{13}(1), 1, S_{13}(3)) + \beta_{23}(1, S_{23}(2), S_{23}(3)),
\]

where \( \beta_{12}, \beta_{13}, \beta_{23} \geq 0 \) and \( \beta_{12} + \beta_{13} + \beta_{23} = 1 \).

Analogously as above, we partition the set \( T \) into the following three regions:

\[
R_{12} = \{(S(1), S(2), S(3)) \in T : S(1) \geq S(3), \text{ and } S(2) \geq S(3)\}
\]

\[
R_{13} = \{(S(1), S(2), S(3)) \in T : S(1) \geq S(2), \text{ and } S(3) \geq S(2)\}
\]

\[
R_{23} = \{(S(1), S(2), S(3)) \in T : S(2) \geq S(1), \text{ and } S(3) \geq S(1)\}.
\]

The intersection of the three regions corresponds to the case where \( S(1) = S(2) = S(3) = S \),
and let us denote the intersection point by \( C = (\mathcal{S}, \mathcal{S}, \mathcal{S}) \). Given the adjustment process specified in Lemma 2, if the vector of the three variables \((S(1), S(2), S(3)) \in R_{12}\) then \((S(1), S(2), S(3))\) will move linearly towards \((S_{12}(1), S_{12}(2), 1)\), at a rate of \(\lambda_p\) per cent the current distance from \((S_{12}(1), S_{12}(2), 1)\). Similarly, for the other two cases where \((S(1), S(2), S(3)) \in R_{13}\) and \((S(1), S(2), S(3)) \in R_{23}\). As above, a graphical analysis shows that, starting from any point inside the set \( T \), the vector of the three variables \((S(1), S(2), S(3))\) will settle in a small neighbourhood of \((\mathcal{S}, \mathcal{S}, \mathcal{S})\), with a radius no larger than \(\lambda_p M\), where \( M \) is a constant that depends on the parameters \( S_{12}(1), S_{12}(2), S_{13}(1), S_{13}(3), S_{23}(2) \) and \( S_{23}(3)\). Thus, as \(\lambda_p \to 0\), the neighbourhood will become arbitrarily small, and, therefore, in the limit the vector \((S(1), S(2), S(3))\) will converge to \((\mathcal{S}, \mathcal{S}, \mathcal{S})\). In other words, the process has a unique quasi-stationary distribution where \( S(1) = S(2) = S(3) \) as \(\lambda_p \to 0\). \(\blacksquare\)

We can now use Lemma 3 to prove the claim for the range of attentiveness \( k = 1 \). We are looking for quasi-stationary distributions as \( \lambda_p \to 0 \), in the sense of the definition from the text (Definition 1). Let us denote by \( q_\theta(i) \), without a time index, the probability that, in a quasi-stationary distribution, a voter of type \( \theta \) has a positive impression of candidate \( i \), assuming the voter pays attention to candidate \( i \).

Suppose that at time \( t \) candidate \( i \) has the largest measure of votes, ie. \( S^t(i) > S^t(j) \) for all \( j \neq i \). Since we assume endogenous attentiveness and \( k = 1 \), we have \( K^t = \{i\} \) for all voters. For each type of voters \( \theta \), using the updating equations for \( p_\theta^t(i) \), we have:

\[
E[p_\theta^{t+1}(i)] = E[p_\theta^{t+1}(i)|i \in K^t_{pos}] \Pr(i \in K^t_{pos}) + E[p_\theta^{t+1}(i)|i \in K^t_{neg}] \Pr(i \in K^t_{neg})
\]

\[= (1 - \lambda_p)E[p_\theta^t(i)] + \lambda_p \Pr(i \in K^t_{pos}) + \lambda_p E[p_\theta^t(i) - 1\] \(|i \in K^t_{neg}) \Pr(i \in K^t_{neg})
\]

\[= (1 - \lambda_p)E[p_\theta^t(i)] + \lambda_p q_\theta(i)\]

Alternatively, we can write this as: \( S^{t+1}_\theta(i) = (1 - \lambda_p)S^t_\theta(i) + \lambda_p q_\theta(i) \). Summing over all types \( \theta \) of voters, we obtain \( S^{t+1}(i) = (1 - \lambda_p)S^t(i) + \lambda_p \sum_\theta s_\theta q_\theta(i) \).

---

\(^{24}\)The intersection point is:

\[(\mathcal{S}, \mathcal{S}, \mathcal{S}) = \mathcal{B}_{12}(S_{12}(1), S_{12}(2), 1) + \mathcal{B}_{13}(S_{13}(1), 1, S_{13}(3)) + \mathcal{B}_{23}(1, S_{23}(2), S_{23}(3))\]

where the weights are: \( \mathcal{B}_{12} = \frac{\Delta_1}{\Delta_1 + \Delta_2 + \Delta_3} \), \( \mathcal{B}_{13} = \frac{\Delta_1}{\Delta_1 + \Delta_2 + \Delta_3} \), and \( \mathcal{B}_{23} = \frac{\Delta_3}{\Delta_1 + \Delta_2 + \Delta_3} \), and where:

\[
\begin{align*}
\Delta_1 &= (1 - S_{13}(3))(1 - S_{23}(2)) + (1 - S_{13}(1))(1 - S_{23}(3)) - (1 - S_{13}(1))(1 - S_{23}(2)), \\
\Delta_2 &= (1 - S_{12}(2))(1 - S_{23}(3)) + (1 - S_{12}(1))(1 - S_{23}(2)) - (1 - S_{12}(1))(1 - S_{23}(3)), \\
\Delta_3 &= (1 - S_{12}(1))(1 - S_{13}(3)) + (1 - S_{12}(2))(1 - S_{13}(1)) - (1 - S_{12}(2))(1 - S_{13}(3)).
\end{align*}
\]

Because conditions (C1), (C2), and (C3) do not hold, this implies that \( \Delta_1 \geq 0 \), \( \Delta_2 \geq 0 \) and \( \Delta_3 \geq 0 \).
Finally, for each candidate \( j \neq i \) (and so \( j \notin K^t \)), the equation for the updating of voting propensities is: \( p^t_\theta(j) = (1 - \lambda_p)p^t_\theta(j) + \lambda_p \). Taking expectation, we obtain: \( E[p^t_{\theta}^{t+1}(j)] = (1 - \lambda_p)E[p^t_{\theta}(j)] + \lambda_p \) or, equivalently, \( S^t_{\theta}^{t+1}(j) = (1 - \lambda_p)S^t_{\theta}(j) + \lambda_p \). Therefore, summing over all types \( \theta \) of voters, we obtain \( S^{t+1}(j) = (1 - \lambda_p)S^t(j) + \lambda_p \).

Thus, the updating equations for the measures of votes \( S^t(1), S^t(2) \) and \( S^t(3) \) follow the model described above, where the constants are \( Q(i) = \sum_{\theta} s_{\theta q_{\theta}(i)} \) for each \( i = 1, 2, 3 \). Therefore, we can use Lemma 3 to conclude that the process has a unique quasi-stationary distribution where, as \( \lambda_p \to 0 \), the measures of votes of the three candidates are \( S(1) = S(2) = S(3) \).

### 7.4 Quasi-stationary distributions in elections with \( n \) candidates under negative plurality rule and endogenous attentiveness

In this section, we discuss the quasi-stationary distributions under endogenous attentiveness in the negative rule, when the number of candidates is \( n \), with \( n \geq 3 \). As we will see below, the qualitative results are similar to those of the \( n = 3 \) candidates case we analyzed above.

To start, suppose the set of candidates is \( N = \{1, 2, \ldots, n\} \), where \( n \geq 3 \). Attentiveness is endogenous, i.e. voters pay attention to the top candidates only, and suppose that the range of attentiveness is \( k \), with \( 1 \leq k \leq n \), for all voters. The case of \( k = n \) corresponds to full attentiveness (Andonie & Diermeier (2018)) so we consider here only the cases where \( k < n \).

The basic idea is similar to that from the case of \( n = 3 \) candidates. Let us first consider the following dynamic (deterministic) adjustment process. The parameters of the process are: \( \lambda_p \in (0, 1) \), which we assume to be close to 0, together with the sets of fixed constants \( \{S_{i_1i_2\ldots i_k}(i)\}_{i \in N} \), for each subset of \( k \) candidates \( \{i_1, i_2, \ldots i_k\} \subset N \). We assume:

\[
S_{i_1i_2\ldots i_k}(i) = \begin{cases} 
   \in (0, 1), & \text{if } i \in \{i_1, i_2, \ldots i_k\} \\
   1, & \text{if } i \notin \{i_1, i_2, \ldots i_k\}
\end{cases}
\]

The variables of the process are \( S^t(1), S^t(2), \ldots S^t(n) \), which are adjusted as follows.

Suppose at time \( t \), where \( t \geq 1 \), we have: \( \min\{S^t(i_1), S^t(i_2), \ldots S^t(i_k)\} > \max\{S^t(i_{k+1}), \ldots S^t(i_n)\} \)

i.e. candidates \( i_1, i_2, \ldots i_k \) are the top \( k \) candidates. Then:

\[
S^{t+1}(i) = (1 - \lambda_p)S^t(i) + \lambda_p S_{i_1i_2\ldots i_k}(i) \text{ for each } i = 1, 2, \ldots n.
\]

We want to characterize the quasi-stationary distributions of the process, when \( k < n \). Let us denote by \( S(i_1i_2\ldots i_k) \) the state where candidates \( i_1, i_2, \ldots i_k \) are the top \( k \) candidates,
i.e. if $\min\{S'(i_1), S'(i_2), \ldots S'(i_k)\} > \max\{S'(i_{k+1}), \ldots S'(i_n)\}$ we call the state $S(i_1i_2 \ldots i_k)$. Because $S_{i_1i_2 \ldots i_k}(i) \in (0, 1)$ for $i \in \{i_1, i_2, \ldots i_k\}$, and $S_{i_1i_2 \ldots i_k}(i) = 1$ for $i \notin \{i_1, i_2, \ldots i_k\}$, and given the adjustment equations described above, we note the following.

1. The process cannot be trapped in one single state. Thus, the process will cycle across various (at least two) states.

2. For each candidate $i \in N$, the process cannot cycle across a set of states where $i$ is never among the top $k$ candidates. That is, the process cannot cycle across a set of states $\{S(i_1i_2 \ldots i_k)\}_{i_1, i_2, \ldots i_k \in N}$ where $i \notin \{i_1, i_2, \ldots i_k\}$ for all states $S(i_1i_2 \ldots i_k)$.

3. The adjustment process allows for the possibility that for all states across which the process cycles, some candidate $i$ may always be among the top $k$ candidates. That is, given the set of states $\{S(i_1i_2 \ldots i_k)\}_{i_1, i_2, \ldots i_k \in N}$ across which the process cycles, we have $i \in \{i_1, i_2, \ldots i_k\}$ for all states $S(i_1i_2 \ldots i_k)$.

The maximum number of candidates that can have the property described in item (3) above is $k - 1$, while minimum is 0. Depending on the number of candidates that have this property we distinguish among various cases.

Let us consider the case where this number is $k - 1$, and suppose, for convenience, these are candidates $1, 2, \ldots k - 1$. That is, the states across which the process cycles are: $S(12 \ldots k - 1k), S(12 \ldots k - 1k + 1), \ldots S(12 \ldots k - 1n)$. Because the process cycles only among states $S(12 \ldots k - 1k), S(12 \ldots k - 1k + 1), \ldots S(12 \ldots k - 1n)$, then the vector of the $n$ variables $(S'(1), S'(2), \ldots S'(n))$ will eventually end up in the convex hull:

$$T = \{(S(1), S(2), \ldots S(n)) = \beta_k(S_{12 \ldots k - 1k}(1), \ldots S_{12 \ldots k - 1k}(k), 1, \ldots 1)$$
$$+ \beta_{k+1}(S_{12 \ldots k - 1k + 1}(1), \ldots S_{12 \ldots k - 1k + 1}(k + 1), 1, \ldots 1) + \ldots$$
$$+ \beta_n(S_{12 \ldots k - 1n}(1), \ldots S_{12 \ldots k - 1n}(k - 1), 1, \ldots 1, S_{12 \ldots k - 1n}(n)), \text{ where } \beta_k, \ldots \beta_n \geq 0 \text{ and}$$
$$\sum_{j=k}^{n} \beta_j = 1\}.$$

Lemma 4 below shows that if $\sum_{j=k}^{n} \frac{1 - S_{12 \ldots k - 1j}(i)}{1 - S_{12 \ldots k - 1j}(i)} < 1$ for each candidate $i = 1, 2, \ldots k - 1$, then the process has a quasi-stationary distribution where:

$$(S(1), S(2), \ldots S(k - 1), S(k), \ldots S(n)) \rightarrow (S^*(1), S^*(2), \ldots S^*(k - 1), \overline{S}, \ldots \overline{S}),$$

with $\min\{S^*(1), S^*(2), \ldots S^*(k - 1)\} > \overline{S}$, as $\lambda_p \rightarrow 0$.  

42
If the above conditions do not hold for any subset of \((k - 1)\) candidates, then the process does not have a quasi-stationary distribution where the same \((k - 1)\) candidates are always among the top \(k\) candidates in the set of states across which the process cycles. In this case, the process may have quasi-stationary distributions where \(l - 1\), with \(1 \leq l \leq k\) candidates are always among the top \(k\) candidates in the set of states across which the process cycles. The basic idea is similar to that for the case where there are \(k - 1\) among the top candidates, and it is described briefly below.

Let us consider the possibility of a quasi-stationary distribution where the process cycles only among the states \(S(1, 2, \ldots, \l - 1, i_1, \ldots, i_k)\), where \(\{i_1, \ldots, i_k\}\) is a subset of \(k - l - 1\) candidates of the set \(\{l, l + 1, \ldots, n\}\). Because the process cycles only among these states, the vector of the \(n\) variables \((S^t(1), S^t(2), \ldots, S^t(n))\) will end up in the convex hull:

\[
T = \{(S(1), S(2), \ldots, S(n)) = \sum_{\{i_l, \ldots, i_k\} \subset \{l, l + 1, \ldots, n\}} \beta(i_1, \ldots, i_k)(S_{12, \ldots, l-1, j_1, \ldots, j_k}(1), \ldots, S_{12, \ldots, l-1, j_1, \ldots, j_k}(n)),
\]

where \(\beta(i_l, \ldots, i_k) \geq 0\) for each \(\{i_l, \ldots, i_k\} \subset \{l, l + 1, \ldots, n\}\) and

\[
\sum_{\{i_l, \ldots, i_k\} \subset \{l, l + 1, \ldots, n\}} \beta(i_l, \ldots, i_k) = 1. \]

Therefore, if \((S(1), S(2), \ldots, S(n)) \in T\) then there exists a set of weights \(\{\beta(i_l, \ldots, i_k)\}_{\{i_l, \ldots, i_k\} \subset \{l, l + 1, \ldots, n\}}\), with \(\beta(i_l, \ldots, i_k) \geq 0\) for each subset \(\{i_l, \ldots, i_k\}\), and \(\sum_{\{i_l, \ldots, i_k\} \subset \{l, l + 1, \ldots, n\}} \beta(i_l, \ldots, i_k) = 1\), such that:

\[
S(i) = \sum_{\{i_l, \ldots, i_k\} \subset \{l, l + 1, \ldots, n\}} \beta(i_1, \ldots, i_k) S_{12, \ldots, l-1, j_1, \ldots, j_k}(i) \text{ for each } i = 1, 2, \ldots, l - 1
\]

\[
S(j) = \sum_{\{j, i_1, \ldots, i_k\} \subset \{l, l + 1, \ldots, n\}} \beta(j, i_1, \ldots, i_k) S_{12, \ldots, l-1, j_1, \ldots, j_k}(j) + \sum_{\{i_l, i_1, \ldots, i_k\} \subset \{l, l + 1, \ldots, n\} \setminus j} \beta(i_l, \ldots, i_k)
\text{ for each } j = l, l + 1, \ldots, n
\]

Let \(\{\beta^*(i_l, \ldots, i_k)\}_{\{i_l, \ldots, i_k\} \subset \{l, l + 1, \ldots, n\}}\) be a set of weights such that \(S^*(l) = S^*(l + 1) = \ldots = S^*(n) = S\). For each \(i = 1, 2, \ldots, l - 1\), let us define:

\[
S^*(i) = \sum_{\{i_l, \ldots, i_k\} \subset \{l, l + 1, \ldots, n\}} \beta^*(i_l, \ldots, i_k) S_{12, \ldots, l-1, j_1, \ldots, j_k}(i)
\]

The conditions required for the existence a quasi-stationary distribution where candidates \(1, 2, \ldots, l - 1\) are always among the top \(k\) candidates is \(S^*(i) > S\) for each \(i = 1, 2, \ldots, l - 1\). If these conditions are met, then the process has a quasi-stationary distribution where:

\[
(S(1), S(2), \ldots, S(l-1), S(l), \ldots, S(n)) \rightarrow (S^*(1), S^*(2), \ldots, S^*(l-1), S, \ldots, S),
\]
with \( \min\{S^*(1), S^*(2), ... S^*(l - 1)\} > \mathcal{S} \), as \( \lambda_p \to 0 \).

We now finalize the discussion of the quasi-stationary distributions of the process with the following lemma. The lemma characterizes the exact conditions under which the process has a quasi-stationary distribution where candidates \( 1, 2, ... k - 1 \) are always among the top \( k \) candidates.

**Lemma 4** If \( \sum_{j=k}^{n} \frac{1 - S_{12...k-1j}(i)}{1 - S_{12...k-1j}(j)} < 1 \) for each candidate \( i = 1, 2, ... k - 1 \), then the process has a quasi-stationary distribution where:

\[
(S(1), S(2), ... S(k - 1), S(k), ... S(n)) \to (S^*(1), S^*(2), ... S^*(k - 1), \bar{\mathcal{S}}, ... \bar{\mathcal{S}}),
\]

with \( \min\{S^*(1), S^*(2), ... S^*(k - 1)\} > \mathcal{S} \), as \( \lambda_p \to 0 \).

**Proof.** Let us define \( \mathcal{S} \) as:

\[
\frac{1}{1 - \mathcal{S}} = \sum_{j=k}^{n} \frac{1}{1 - S_{12...k-1j}(j)}.
\]

Because \( S_{12...k-1j}(j) \in (0, 1) \) for each \( j = k, k + 1, ... n \), it follows that also \( \mathcal{S} \in (0, 1) \), and moreover \( \mathcal{S} > S_{12...k-1j}(j) \) for each \( j = k, k + 1, ... n \). If we let \( m = \min_{j=k,k+1,...n}\{S_{12...k-1j}(j)\} \), then we also have \( \mathcal{S} > m \). Let us also define:

\[
\beta^*_j = \frac{1 - \mathcal{S}}{1 - S_{12...k-1j}(j)} \quad \text{for each } j = k, k + 1, ... n,
\]

and

\[
S^*(i) = \sum_{j=k}^{n} \beta^*_j S_{12...k-1j}(i) \quad \text{for each } i = 1, 2, ... k - 1.
\]

Using this notation, the condition from lemma is equivalent to \( S^*(i) > \mathcal{S} \) for each \( i = 1, 2, ... k - 1 \). We will show that there exists a quasi-stationary distribution of the process, where \( S(i) \to S^*(i) \) for each \( i = 1, 2, ... k - 1 \), and \( S(j) \to \mathcal{S} \) for each \( j = k, k + 1, ... n \) as \( \lambda_p \to 0 \).

Consider a state of the process \((S(1), S(2), ... S(n))\) where:

\[
\mathcal{S} - \lambda_p(\mathcal{S} - m) \leq S(j) \leq \mathcal{S} + \lambda_p\left(\frac{(\mathcal{S} - m)^2}{1 - \mathcal{S}}\right)
\]

for each \( j = k, k + 1, ... n \). Because \((S(1), S(2), ... S(n)) \in T\), then there exists a set of
constants \( \{\beta_j\}_{j=k,k+1,...n} \) with \( \beta_j \geq 0 \) and \( \sum_{j=k}^{n} \beta_j = 1 \) such that:

\[
S(i) = \sum_{j=k}^{n} \beta_j S_{12...k-1j}(i) \quad \text{for each } i = 1, 2, ... k - 1
\]

\[
S(j) = \beta_j S_{12...k-1j}(j) + 1 - \beta_j \quad \text{for each } j = k, k + 1, ... n.
\]

Therefore \( \beta_j = \frac{1 - S(j)}{1 - S_{12...k-1j}(j)} \) for each \( j = k, k + 1, ... n \), and using the above bounds for \( S(j) \) we obtain:

\[
\frac{1 - S}{1 - S_{12...k-1j}(j)} - \lambda_p \frac{(\overline{S} - m)^2}{(1 - S)(1 - S_{12...k-1j}(j))} \leq \beta_j \leq \frac{1 - S}{1 - S_{12...k-1j}(j)} + \lambda_p \frac{\overline{S} - m}{1 - S_{12...k-1j}(j)}
\]

or, equivalently,

\[
\beta^*_j - \lambda_p \frac{(\overline{S} - m)^2}{(1 - S)(1 - S_{12...k-1j}(j))} \leq \beta_j \leq \beta^*_j + \lambda_p \frac{\overline{S} - m}{1 - S_{12...k-1j}(j)}
\]

Multiplying by \( S_{12...k-1j}(i) \) and summing over all \( j = k, k + 1, ... n \) we obtain:

\[
S^*(i) - \lambda_p \sum_{j=k}^{n} \frac{S_{12...k-1j}(i)(\overline{S} - m)^2}{(1 - S)(1 - S_{12...k-1j}(j))} \leq S(i) \leq S^*(i) + \lambda_p \sum_{j=k}^{n} \frac{S_{12...k-1j}(i)(\overline{S} - m)}{1 - S_{12...k-1j}(j)}
\]

for each \( i = 1, 2, ... k - 1 \). Therefore, if \( \lambda_p \) is close to zero, then each \( S(i) \), for \( i = 1, 2, ... k - 1 \), is in a close neighbourhood of \( S^*(i) \). At the same time, each \( S(j) \), for \( j = k, k + 1, ... n \) is in a close neighbourhood of \( \overline{S} \). Because \( S^*(i) > \overline{S} \) for each \( i = 1, 2, ... k - 1 \) by assumption, if \( \lambda_p \) is sufficiently small, then we also have \( S(i) > S(j) \) for all \( i = 1, 2, ... k - 1 \) and \( j = k, k + 1, ... n \).

We will now show that after variables \( (S(1), S(2), ..., S(n)) \) are adjusted, the new variables \( (S'(1), S'(2), ..., S'(n)) \) satisfy the same set of inequalities as \( (S(1), S(2), ..., S(n)) \). That is, for each \( i = 1, 2, ... k - 1 \), we have:

\[
S^*(i) - \lambda_p \sum_{j=k}^{n} \frac{S_{12...k-1j}(i)(\overline{S} - m)^2}{(1 - S)(1 - S_{12...k-1j}(j))} \leq S'(i) \leq S^*(i) + \lambda_p \sum_{j=k}^{n} \frac{S_{12...k-1j}(i)(\overline{S} - m)}{1 - S_{12...k-1j}(j)}
\]

and for each \( j = k, k + 1, ... n \), we have:

\[
\overline{S} - \lambda_p(\overline{S} - m) \leq S(j) \leq \overline{S} + \lambda_p \frac{(\overline{S} - m)^2}{1 - \overline{S}}
\]

In other words, the new adjusted variables \( (S'(1), S'(2), ..., S'(k - 1), S'(k), ... S'(n)) \) remain in the same neighbourhood of \( (S^*(1), S^*(2), ..., S^*(k - 1), \overline{S}, ... \overline{S}) \), which implies that the
process has a quasi-stationary distribution where \((S(1), S(2), ...S(k-1), S(k), ...S(n)) \rightarrow (S^*(1), S^*(2), ...S^*(k-1), \overline{S}, ...\overline{S})\) as \(\lambda_p \rightarrow 0\).

To prove this claim, we observe first that \(\min_{j=k,k+1,...,n} S(j) < \overline{S} < \max_{j=k,k+1,...,n} S(j)\). To see that, suppose, for example, that \(S(j) \leq \overline{S}\) for all \(j = k, k+1, ...n\), and where at least one inequality is strict. This implies \(\frac{1-S(j)}{1-S_{12...k-1}^*(j)} \geq \frac{1-S}{1-S_{12...k-1}^*(j)}\), or, equivalently, \(\beta_j \geq \beta_j^*\) for all \(j = k, k+1, ...n\), with at least one inequality strict. Thus \(\sum_{j=k}^n \beta_j > \sum_{j=k}^n \beta_j^*\), which is impossible as \(\sum_{j=k}^n \beta_j = 1\) and \(\sum_{j=k}^n \beta_j^* = 1\). Similarly, \(S(j) \geq \overline{S}\) for all \(j = k, k+1, ...n\) leads to a contradiction. Therefore, we must have \(\min_{j=k,k+1,...,n} S(j) < \overline{S} < \max_{j=k,k+1,...,n} S(j)\).

Let us denote by \(l\), the candidate for which \(S(l) = \max_{j=k,k+1,...,n} S(j)\).

Given that \(S(i) > S(j)\) for all \(i = 1, 2, ...k-1\) and \(j = k, k+1, ...n\), and \(S(l) > S(j)\) for all \(j = k, k+1, ...n\) with \(j \neq l\), then candidates \(1, 2, ...k-1, l\) are the top candidates. Therefore, the adjustment equations for candidates \(j = k, k+1, ...n\) are:

\[
\begin{align*}
S'(l) &= (1 - \lambda_p)S(l) + \lambda_p S_{12...k-1}^*(l) \\
S'(j) &= (1 - \lambda_p)S(j) + \lambda_p \text{ for all } j \neq l.
\end{align*}
\]

As \(S(l) > \overline{S}\) and \(S_{12...k-1}^*(l) \geq m\), we can write:

\[
S'(l) \geq (1 - \lambda_p)\overline{S} + \lambda_p m = \overline{S} - \lambda_p (\overline{S} - m)
\]

For the other candidates \(j \neq l\), we have \(S(j) \geq \overline{S} - \lambda_p (\overline{S} - m)\), and so we can write:

\[
\begin{align*}
S'(j) &\geq (1 - \lambda_p)(\overline{S} - \lambda_p (\overline{S} - m)) + \lambda_p = \overline{S} - \lambda_p (\overline{S} - m) + \lambda_p (1 - (\overline{S} - \lambda_p (\overline{S} - m))) \\
&\geq \overline{S} - \lambda_p (\overline{S} - m)
\end{align*}
\]

As the adjusted vector of variables \((S'(1), S'(2), ...S'(n)) \in T\), there exists a set of constants \(\{\beta'_j\}_{j=k,k+1,...n}\) with \(\beta'_j \geq 0\) and \(\sum_{j=k}^n \beta'_j = 1\) such that:

\[
\begin{align*}
S'(i) &= \sum_{j=k}^n \beta'_j S_{12...k-1}^*(i) \text{ for each } i = 1, 2, ...k-1 \\
S'(j) &= \beta'_j S_{12...k-1}^*(j) + 1 - \beta'_j \text{ for each } j = k, k+1, ...n.
\end{align*}
\]

The last equations imply \(\beta'_j = \frac{1-S'(j)}{1-S_{12...k-1}^*(j)}\) for each \(j = k, k+1, ...n\). As \(\sum_{j=k}^n \beta'_j = 1\), this
further implies \( \sum_{j=k}^{n} \frac{1-S'(j)}{1-S_{12\ldots k-1} \cdot (j)} = 1 \). Therefore, for each \( j = k, k+1, \ldots, n \), we can write:

\[ S'(j) = S_{12\ldots k-1} \cdot (j) + (1 - S_{12\ldots k-1} \cdot (j)) \sum_{j'=k,j' \neq j}^{n} \frac{1 - S'(j')}{1 - S_{12\ldots k-1} \cdot j'(j')} \leq S_{12\ldots k-1} \cdot (j) + (1 - S + \lambda_p(S - m)) \sum_{j'=k,j' \neq j}^{n} \frac{1}{1 - S_{12\ldots k-1} \cdot j'(j')} \]

\[ = \frac{S + \lambda_p(S - m)}{1 - S - S_{12\ldots k-1} \cdot (j)} \leq S + \lambda_p(S - m)^2 \]

As above, the fact that

\[ S - \lambda_p(S - m) \leq S'(j) \leq S + \lambda_p(S - m)^2 \]

for each \( j = k, k+1, \ldots, n \) implies:

\[ S^*(i) - \lambda_p \sum_{j=k}^{n} \frac{S_{12\ldots k-1} \cdot (i)(S - m)^2}{(1 - S)(1 - S_{12\ldots k-1} \cdot (j))} \leq S'(i) \leq S^*(i) + \lambda_p \sum_{j=k}^{n} \frac{S_{12\ldots k-1} \cdot (i)(S - m)}{1 - S_{12\ldots k-1} \cdot (j)} \]

for each \( i = 1, 2, \ldots, k-1 \). Thus, the adjusted variables \((S'(1), S'(2), \ldots, S'(k-1), S'(k), \ldots, S'(n))\) remain in the same neighbourhood of \((S^*(1), S^*(2), \ldots, S^*(k-1), S, \ldots, S)\), and therefore the process has a quasi-stationary distribution where

\[ (S(1), S(2), \ldots, S(k-1), S(k), \ldots, S(n)) \to (S^*(1), S^*(2), \ldots, S^*(k-1), S, \ldots, S) \]

as \( \lambda_p \to 0 \). □

We can now use this results to discuss the possible quasi-stationary distributions of the dynamic process under negative plurality rule and endogenous attentiveness. We denote by \( q_\theta(i) \), without a time index, the probability that, in a given quasi-stationary distribution, a voter of type \( \theta \) has a positive impression of candidate \( i \) (assuming the voter pays attention to candidate \( i \)).

Suppose, for example, that at time \( t: \min\{S'(i_1), \ldots, S'(i_k)\} > \max\{S'(i_{k+1}), \ldots, S'(i_n)\} \}, so the set of candidates to which voters pay attention is \( K' = \{i_1, i_2, \ldots, i_k\} \). Consider a candidate
Using the equations for updating propensities $p^\theta_t(i)$, we can write:

\[
E[p^t_{\theta+1}(i)] = E[p^t_{\theta}(i) | i \in K^t_{pos}] \Pr(i \in K^t_{pos}) + E[p^t_{\theta}(i) | i \in K^t_{neg}] \Pr(i \in K^t_{neg})
\]

\[
= (1 - \lambda_p)E[p^t_{\theta}(i)] + \lambda_p \Pr(i \in K^t_{pos}) + \lambda_p E\left[\frac{1}{J^K_{\theta} \cdot \left(\frac{1}{J^K_{\theta}} - 1\right)} | i \in K^t_{neg}\right] \Pr(i \in K^t_{neg})
\]

Multiplying by $s_{\theta}$ and summing over all types $\theta$, we obtain:

\[
S^{t+1}(i) = (1 - \lambda_p)S^t(i) + \lambda_p \sum_{\theta} s_{\theta}(1 - (1 - q_{\theta}(i))E\left[\frac{1}{1 + J^K_{\theta} (-i)}\right]).
\]

Consider now a candidate $i \notin K^t$. Then using the equations for updating propensities $p^\theta_t(i)$, we can write:

\[
E[p^t{_{\theta+1}}(i)] = (1 - \lambda_p)E[p^t_{\theta}(i)] + \lambda_p
\]

Multiplying by $s_{\theta}$ and summing over all types $\theta$, we obtain:

\[
S^{t+1}(i) = (1 - \lambda_p)S^t(i) + \lambda_p
\]

Therefore, the equations characterizing the evolution of measures of votes $\{S^t(i)\}_i \in N$ follow the equations of the adjustment process from above. The parameters of the process are:

\[
S_{i_1i_2...i_k}(i) = \begin{cases} 
\sum_{\theta} s_{\theta}(1 - (1 - q_{\theta}(i))E\left[\frac{1}{1 + J^K_{\theta} (-i)}\right]), & \text{if } i \in \{i_1, i_2, ...i_k\} \\
1, & \text{if } i \notin \{i_1, i_2, ...i_k\}
\end{cases}
\]

for each set of $k$ candidates $\{i_1, i_2, ...i_k\}$. Therefore, we can use the results from our discussion to determine the quasi-stationary distributions of the process.

7.5 Proof of Proposition 5

In the stationary distribution, the aspirations of the groups $\theta = 1, 2$ of voters are $a_1 = \frac{1 + v}{3}$, and $a_2 = r_2(3) \frac{1 + v}{3} + (1 - r_2(3)) \frac{1}{2}$ respectively. We note that if $r_2(3) = 1$ then $a_2 = a_1$, while for $r_2(3) < 1$ the aspirations of group 2, $a_2$, decrease with $r_2(3)$. The probabilities of positive
impression of each candidate for the two groups of voters, in the stationary distribution, are:

\[ q_1(1) = \Pr(1 + \alpha \epsilon_1(1) > a_1) = 1 - F\left(\frac{a_1 - 1}{\alpha}\right) \]
\[ q_1(2) = \Pr(\alpha \epsilon_2(2) > a_1) = 1 - F\left(\frac{a_1}{\alpha}\right) \]
\[ q_1(3) = \Pr(v + \alpha \epsilon_1(3) > a_1) = 1 - F\left(\frac{a_1 - v}{\alpha}\right) \]
\[ q_2(1) = \Pr(\alpha \epsilon_2(1) > a_2) = 1 - F\left(\frac{a_2}{\alpha}\right) \]
\[ q_2(2) = \Pr(1 + \alpha \epsilon_2(2) > a_2) = 1 - F\left(\frac{a_2 - 1}{\alpha}\right) \]
\[ q_2(3) = \Pr(v + \alpha \epsilon_2(3) > a_2) = 1 - F\left(\frac{a_2 - v}{\alpha}\right) \]

Consider first plurality rule. By Proposition 2, the difference in vote measures of candidates 1 and 2 under plurality rule is:

\[ S(1) - S(2) = s(q_1(1) - q_1(2))(1 - \frac{1}{2} q_1(3)) + (1 - s)(q_2(1) - q_2(2))(1 - \frac{1}{2} r_2(3) q_2(3)) \]
\[ = s(F\left(\frac{a_1}{\alpha}\right) - F\left(\frac{a_1 - 1}{\alpha}\right))(1 - \frac{1}{2}(1 - F\left(\frac{a_1 - v}{\alpha}\right))) \]
\[ - (1 - s)(F\left(\frac{a_2}{\alpha}\right) - F\left(\frac{a_2 - 1}{\alpha}\right))(1 - \frac{1}{2} r_2(3)(1 - F\left(\frac{a_2 - v}{\alpha}\right))) \]

If \( r_2(3) = 1 \), then \( a_2 = a_1 \) and therefore in this case \( S(1) - S(2) > 0 \) for all \( s > \frac{1}{2} \). Consider now the case where \( r_2(3) < 1 \). First, because \( F(.) \sim U[-1, +1] \), then the difference \( F\left(\frac{a_2}{\alpha}\right) - F\left(\frac{a_2 - 1}{\alpha}\right) \) does not depend on \( a_2 \). Second, as the attention probability \( r_2(3) \) decreases, the aspirations \( a_2 \) of group 2 increase. These two observations imply that \( \left(F\left(\frac{a_2}{\alpha}\right) - F\left(\frac{a_2 - 1}{\alpha}\right)\right)(1 - \frac{1}{2} r_2(3)(1 - F\left(\frac{a_2 - v}{\alpha}\right))) \) increases as the attention probability \( r_2(3) \) decreases. Therefore, in the case of \( r_2(3) < 1 \), there exists a threshold \( s^*_p \) such that \( S(1) - S(2) < 0 \) if \( s \in (\frac{1}{2}, s^*_p) \), while \( S(1) - S(2) > 0 \) if \( s > s^*_p \). As the measure of votes candidate 3 obtains is always lower than that of either candidates 1 or 2, then candidate 2 wins if \( s \in (\frac{1}{2}, s^*_p) \), and candidate 1 wins if \( s > s^*_p \).

Second, consider negative plurality rule. Again, we use Proposition 2 to compute the difference in vote measures of candidates 1 and 2:

\[ S(1) - S(2) = s(q_1(1) - q_1(2))(1 - \frac{1}{2} (1 - q_1(3))) + (1 - s)(q_2(1) - q_2(2))(1 - \frac{1}{2} r_2(3)(1 - q_2(3))) \]
\[ = s(F\left(\frac{a_1}{\alpha}\right) - F\left(\frac{a_1 - 1}{\alpha}\right))(1 - \frac{1}{2} F\left(\frac{a_1 - v}{\alpha}\right)) \]
\[ - (1 - s)(F\left(\frac{a_2}{\alpha}\right) - F\left(\frac{a_2 - 1}{\alpha}\right))(1 - \frac{1}{2} r_2(3) F\left(\frac{a_2 - v}{\alpha}\right)) \]

49
If \( r_2(3) = 1 \) then \( a_2 = a_1 \), and so \( S(1) - S(2) > 0 \) for all \( s > \frac{1}{2} \). Suppose now that \( r_2(3) < 1 \). It can be shown, after differentiating the expression \( r_2(3)F(\frac{a_2-v}{\alpha}) \) and additional analysis, that \( r_2(3)F(\frac{a_2-v}{\alpha}) \) decreases as the attention probability \( r_2(3) \) decreases. As the difference \( (F(\frac{a_2}{\alpha}) - F(\frac{a_2-1}{\alpha})) \) does not depend on \( a_2 \), then \( (F(\frac{a_2}{\alpha}) - F(\frac{a_2-1}{\alpha}))(1 - \frac{1}{2}r_2(3)F(\frac{a_2-v}{\alpha})) \) increases as probability \( r_2(3) \) decreases. Therefore, when \( r_2(3) < 1 \), there exists a threshold \( s_n^* > \frac{1}{2} \) such that \( S(1) - S(2) < 0 \) for all \( s \in (\frac{1}{2}, s_n^*) \), and \( S(1) - S(2) > 0 \) for \( s > s_n^* \). On the other hand, the difference between the vote measures of candidates 1 and 3 is:

\[
S(1) - S(3) = s(q_1(1) - q_1(3))(1 - \frac{1}{2}(1 - q_2(2)))
+ (1 - s)(r_2(3)(1 - q_2(3)) - (1 - q_2(1))(1 - \frac{1}{2}(1 - q_2(2)))
+ s(F(\frac{a_1-v}{\alpha}) - F(\frac{a_1-1}{\alpha}))(1 - \frac{1}{2}F(\frac{a_1}{\alpha}))
+ (1 - s)(r_2(3)F(\frac{a_2-v}{\alpha}) - F(\frac{a_2}{\alpha}))(1 - \frac{1}{2}F(\frac{a_2-1}{\alpha}))
\]

When \( s \) is large the difference \( S(1) - S(3) > 0 \) because the first term above, which is positive, dominates the second term. Combining the observations on the differences \( S(1) - S(2) \) and \( S(1) - S(3) \), we obtain that, under negative plurality, if \( s < s_n^* \) then candidate 1 cannot win the election, while if \( s \) is large, i.e., \( s \) close to one, then candidate 1 wins.

Finally, consider approval voting. The difference in vote measures of candidates 1 and 2, by Proposition 2, is:

\[
S(1) - S(2) = s(q_1(1) - q_1(2)) + (1 - s)(q_2(1) - q_2(2))
= s(F(\frac{a_1}{\alpha}) - F(\frac{a_1-1}{\alpha})) - (1 - s)(F(\frac{a_2}{\alpha}) - F(\frac{a_2-1}{\alpha}))
\]

If \( F(.) \) follows the uniform distribution, i.e., the function \( F(.) \) is linear, then the difference \( F(\frac{a_1}{\alpha}) - F(\frac{a_1-1}{\alpha}) \) does not depend on \( a \), and therefore \( F(\frac{a_1}{\alpha}) - F(\frac{a_1-1}{\alpha}) = F(\frac{a_2}{\alpha}) - F(\frac{a_2-1}{\alpha}) \). In this case then, \( S(1) - S(2) > 0 \) for all \( s > \frac{1}{2} \). As in the case of plurality rule, the vote measure of candidate 3 is lower than that of either candidates 1 or 2, and therefore candidate 1 wins for all \( s > \frac{1}{2} \).

### 7.6 Proof of Proposition 6

Because the number of candidates is \( n = 3 \), the range of attentiveness is \( k = 2 \), and attentiveness is endogenous, by Corollary 2, under both plurality and approval rules there will be three stationary distributions, one distribution for each pair of candidates in the set \( N = \{1, 2, 3\} \).
In the stationary distribution corresponding to candidates 1 and 2, we have $S(1), S(2) > 0$ and $S(3) = 0$. As $s_1 = s_2 = 50\%$, and the vectors of fixed components $u_1$ and $u_2$ are symmetric with respect to candidates 1 and 2, in this distribution we have $S(1) = S(2)$, for both the plurality and approval rules.

In the second stationary distribution corresponding to candidates 1 and 3, we have $S(1), S(3) > 0$ and $S(2) = 0$. This distribution corresponds to the stationary distribution of a two-candidate election with candidates 1 and 3, and where all voters have a range of attentiveness of $k = 2$. Using the results from the two-candidate elections, in both the plurality and approval rules, candidate 1 wins if $\sum_\theta s_\theta F\left(\frac{v_3 - v_2}{2\alpha}\right) > \frac{1}{2}$. In the context of the election we analyze here, this condition is $\frac{1}{2} F\left(\frac{1-v}{2\alpha}\right) + \frac{1}{2} F\left(\frac{-v}{2\alpha}\right) > \frac{1}{2}$, which is equivalent to $v < \frac{1}{2}$. Therefore, in both the plurality and approval rules, if $v < \frac{1}{2}$ then candidate 1 wins in this distribution, and if $v > \frac{1}{2}$ then candidate 3 wins.

Finally, the analysis of the third stationary distribution corresponding to candidates 2 and 3 is similar to that of the stationary distribution from above, corresponding to candidates 1 and 3. In this distribution, we have $S(2), S(3) > 0$ and $S(1) = 0$. In both the plurality and approval rules, if $v < \frac{1}{2}$ then candidate 2 wins in this distribution, and if $v > \frac{1}{2}$ then candidate 3 wins.

Let us now analyze the quasi-stationary distributions in the negative plurality rule. By Proposition 3, when the range of attentiveness is $k = 2$ there are two possible types of quasi-stationary distributions. In the first type, one candidate wins, and the other two candidates obtain approximately equal measures of votes. This distribution can occur if the average $v_i^j$ for one candidate is substantially larger than the same average of each of the other candidates. In the second type of distribution, all three candidates obtain approximately equal measures of votes. This distribution occurs when no candidate has an average $v_i^j$ substantially larger than the same average for each of the other candidates. In our context, because of the symmetry between candidates 1 and 2, there cannot be a quasi-stationary distribution where candidate 1 or candidate 2 wins. Therefore, the process can either have a quasi-stationary distribution where $S(3) > S(1) = S(2)$, or a quasi-stationary distribution where $S(1) = S(2) = S(3)$.

Let us analyze the conditions under which we have a quasi-stationary distribution where candidate 3 wins, ie. $S(3) > S(1) = S(2)$ as $\lambda_p \to 0$. Intuitively, this requires that $v$ be large enough. Using the proof of Proposition 3, the quasi-stationary distribution where $S(3) > S(1) = S(2)$ requires condition (C3) to hold. That is, we need to have:

$$(C3) : \frac{\sum_\theta s_\theta(1 - q_\theta(3))(1 + q_\theta(1))}{\sum_\theta s_\theta(1 - q_\theta(1))(1 + q_\theta(3))} + \frac{\sum_\theta s_\theta(1 - q_\theta(3))(1 + q_\theta(2))}{\sum_\theta s_\theta(1 - q_\theta(2))(1 + q_\theta(3))} < 1$$
In our present context, this condition is:

\[
\frac{1}{2}(1 - q_1(3))(1 + q_1(1)) + \frac{1}{2}(1 - q_2(3))(1 + q_2(1)) \\
\frac{1}{2}(1 - q_1(1))(1 + q_1(3)) + \frac{1}{2}(1 - q_2(1))(1 + q_2(3)) \\
\frac{1}{2}(1 - q_1(3))(1 + q_1(2)) + \frac{1}{2}(1 - q_2(3))(1 + q_2(2)) \\
\frac{1}{2}(1 - q_1(2))(1 + q_1(3)) + \frac{1}{2}(1 - q_2(2))(1 + q_2(3)) < 1
\]

To find the range of \( v \) for which the condition holds, let us first compute the aspirations for the two types of voters in this stationary distribution. Because of the symmetry between candidates 1 and 2, in this stationary distribution the process will cycle between two states where either candidates 1 and 3 (state S13), or candidates 2 and 3 (state S23) are the top two candidates, with equal frequency. Therefore, the stationary aspirations are:

\[
a_1 = \frac{1}{2} + \frac{v}{4} \quad \text{and} \quad a_2 = \frac{1}{2} + \frac{v}{4} = \frac{1 + 2v}{4}.
\]

Given these aspirations, we can now compute the probabilities of positive impression for each type of voters, and for each candidate as follows:

\[
q_1(1) = \Pr(\pi_1(1) > a_1) = \Pr(1 + \alpha \epsilon > \frac{1 + 2v}{4}) = 1 - F\left(\frac{2v - 3}{4\alpha}\right)
\]

\[
q_1(2) = \Pr(\pi_1(2) > a_1) = \Pr(\alpha \epsilon > \frac{1 + 2v}{4}) = 1 - F\left(\frac{1 + 2v}{4\alpha}\right)
\]

\[
q_1(3) = \Pr(\pi_1(3) > a_1) = \Pr(v + \alpha \epsilon > \frac{1 + 2v}{4}) = F\left(\frac{2v - 1}{4\alpha}\right)
\]

\[
q_2(1) = \Pr(\pi_2(1) > a_2) = \Pr(\alpha \epsilon > \frac{1 + 2v}{4}) = 1 - F\left(\frac{1 + 2v}{4\alpha}\right)
\]

\[
q_2(2) = \Pr(\pi_2(2) > a_2) = \Pr(1 + \alpha \epsilon > \frac{1 + 2v}{4}) = 1 - F\left(\frac{2v - 3}{4\alpha}\right)
\]

\[
q_2(3) = \Pr(\pi_2(3) > a_2) = \Pr(v + \alpha \epsilon > \frac{1 + 2v}{4}) = F\left(\frac{2v - 1}{4\alpha}\right)
\]

Using these probabilities in the above condition (C3), and upon some manipulations, condition (C3) is equivalent to:

\[
(3 - F\left(\frac{2v - 1}{4\alpha}\right))(8 - F\left(\frac{2v - 3}{4\alpha}\right) - F\left(\frac{1 + 2v}{4\alpha}\right)) < 16
\]

We note that the left side of the inequality is a decreasing function of \( v \), it takes a value of 17.5 when \( v = \frac{1}{2} \) and a value of 12 when \( v = +\infty \). Therefore, there exists a unique threshold \( v^* > \frac{1}{2} \) at which:

\[
(3 - F\left(\frac{2v^* - 1}{4\alpha}\right))(8 - F\left(\frac{2v^* - 3}{4\alpha}\right) - F\left(\frac{1 + 2v^*}{4\alpha}\right)) = 16
\]
In addition, if \( v > v^* \) then condition (C3) holds, and if \( v < v^* \) then (C3) does not hold.

Therefore, we can conclude that if \( v > v^* \) then the process has a quasi-stationary distribution where \( S(3) > S(1) = S(2) \) as \( \lambda_p \to 0 \). If, however, \( v \leq v^* \) then the process has a quasi-stationary distribution where \( S(1) = S(2) = S(3) \) as \( \lambda_p \to 0 \).
References


