Games and Economic Behavior 74 (2012) 257–268

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

The fragility of information aggregation in large elections

Michael Mandler

Department of Economics, Royal Holloway College, University of London, Egham, Surrey, TW20 0EX, United Kingdom

ARTICLE INFO

Article history:
Received 3 September 2009
Available online 7 May 2011

JEL classification:
C72
D72

Keywords:
Information aggregation
Elections
Common values
Exchangeability

ABSTRACT

In a common-values election where voters receive a signal about which candidate is superior, suppose there is a small amount of uncertainty about the conditional likelihood of the signal's outcome, given the correct candidate. Once this uncertainty is resolved, the signal is i.i.d. across agents. Information can then fail to aggregate. The candidate less likely to be correct given agents' signals can be elected with probability near 1 in a large electorate even if the distribution of signal likelihoods is arbitrarily near to a classical model where agents are certain that a particular likelihood obtains given that a specific candidate is correct.

1. Introduction

The Condorcet jury theorem shows that agents with independent information who vote for the candidate or proposal that their signal indicates is superior will be virtually certain to elect the correct candidate when the population is large (see, e.g., Miller, 1986; Young, 1988). The subsequent game-theoretic literature, where voters take into account when they are likely to be pivotal, has given information aggregation strategic foundations. As long as the identity of the correct candidate has some impact on the signal probabilities, then in the Nash equilibria where agents optimally use their available information the likelihood that the correct candidate is elected will again approach 1 as the population size increases (see Feddersen and Pesendorfer, 1996, 1997; Myerson 1998, 2000; Wit, 1998; following Austen-Smith and Banks, 1996).

In both the Condorcet and game theory approaches, agents have common values (the same preferences over candidates, given the same information) and receive a signal of candidate quality that is i.i.d. once we fix the correct candidate. We make one adjustment: agents have a little uncertainty about the probability of the signal's outcome, again fixing the correct candidate. Conditional on the resolution of this uncertainty, the signal is i.i.d. This change to the environment can lead information to fail to aggregate: the candidate that usually has the smaller conditional probability of being correct given the voters' signals can be elected with near certainty.

Labeling the candidates 0 and 1 and the signal realizations α and β, the traditional approach supposes that the signal's probability of coming up α, which we call the *signal bias*, is certain to equal some number q_0 when 0 is correct and is certain to equal some other number q_1 when 1 correct. We assume instead that agents have two densities over the signal bias, one when 0 is correct and the other when 1 is correct. But once the uncertainty about the signal bias is resolved, each voter faces the same i.i.d. signal. This set-up allows agents to be virtually sure that the bias is near q_i when i is correct, $i = 0, 1$. Since signals are conditionally i.i.d., the signal realizations are *exchangeable*.[1] Exchangeability means that all finite permutations of a sequence of realizations are equiprobable; e.g., for any three tosses of a coin, HHT, HTH, and THH, are

E-mail address: m.mandler@rhul.ac.uk.

[1] By de Finetti's theorem (1975), the exchangeability of a random variable is in fact equivalent to the random variable being conditionally i.i.d.

0899-8256/$ – see front matter
doi:10.1016/j.geb.2011.03.004

equally likely, which is weaker and more plausible than assuming that tosses are independent. The Bayesian tradition has argued forcefully that exchangeability is the only coherent and defensible reformulation of the independence assumptions of classical probability theory (Kreps, 1988, is persuasive and well known in economics). We therefore take the view that, just as with coin tosses, an exchangeability approach is considerably more plausible than an i.i.d. model. See Mandler (2010) for more on the Bayesian view of large elections.

When agents are uncertain about the signal bias, the symmetric Nash equilibria of the resulting voting game are determined by the local characteristics of the two densities that govern the signal bias. It turns out that it is only the slopes of the densities at specific critical values that matter for the candidates' probabilities of victory. Exploiting this fact, we show that the candidate that is less likely to be correct given the agents' signals can be elected with high probability: for any $\varepsilon > 0$ there is a model where, in a large population, the probability is greater than $1 - \varepsilon$ that the candidate that is less likely to be correct given agents' signals is elected. Since only the slopes of the densities are relevant, this conclusion holds even if the distributions that govern the signal bias are arbitrarily near to a classical model where agents are certain that a specific bias obtains (conditional on a specific candidate being correct). The naive behavior of the older Condorcet jury theorem, where agents throw strategic caution to the wind and just vote their signals,[2] can therefore perform much better. In fact, even an electorate that ignored its signals can outperform a Nash electorate.

A small amount of uncertainty about the signal bias can have such dramatic effects due to the role of *critical signal biases* at which the expected vote shares of the two candidates are equal. As long as a neighborhood of a critical signal bias has positive probability, agents in a large population will infer that it is highly likely that the true signal bias will lie in this neighborhood when they condition on being pivotal (the election being tied). In contrast, the conditional probability that the signal bias will lie outside of such a neighborhood, given that there is a tie, will converge to 0 with exponential speed as the population size increases. The global facts about the signals then become irrelevant; agents pay attention only to the neighborhood of the critical signal bias.

Although critical signal biases will draw all of the attention of optimizing voters, election models that rule out such biases are problematic. If there is no critical signal bias, the probability of a tie will diminish will exponential speed as the population size increases. Hence a perturbation of the model that introduces a ε probability of a critical signal bias will dominate every voter's calculation of how to act when pivotal: if the population is large then, conditional on a tie, agents will be practically sure the signal bias lies near the critical value introduced by the perturbation.

This methodological point is the broader agenda of this paper. Any model of elections with three ingredients – (1) a large number of agents who take independent actions, (2) expected vote shares for candidates that are unequal, (3) agents who condition on being pivotal – is vulnerable to perturbation. If we introduce new parameters where expected vote shares are equal those parameters will be the focus of all of the voters, even if they have only small probability. The new parameters therefore govern the equilibrium properties of the model. And inserting the new parameters typically does not require major surgery; they will appear when the independent actions are remodeled as an exchangeable process.

An important precedent, Ladha (1993), analyzes information aggregation with exchangeable signals and shows that aggregation obtains when agents naively vote their signals. It is therefore the combination of strategic voting and exchangeable signals that can lead to poor aggregation performance.

We will assume that there is an odd number of agents, none of whom abstain, which puts the focus on the simplest way a single voter can be pivotal, namely where the votes of the remainder of the electorate are tied. None of our results hinge on this modeling decision – see the working paper version of this article, Mandler (2011), for the more detailed model where an arbitrary number of voters all have the option to abstain.

There are other ways to adapt the classical model of Feddersen and Pesendorfer (1997) that will also undermine information aggregation. Feddersen and Pesendorfer (1997, Section 6), Kim and Fey (2007), and Bhattacharya (2007) show that the heterogeneity of voters' preferences and uncertainty about voters' preferences can block information aggregation. The present paper, in contrast, sticks to a pure common-values setting and our results hold for *arbitrarily small* (ε) perturbations of a classical information aggregation model. In a different literature, the argument in Acemoglu et al. (2009) resembles our main point: in a model where different individuals observe a sequence of state-dependent signals, they show that a small amount of uncertainty about the distribution of signals can lead to asymptotic differences in the individuals' beliefs.

2. Elections with uncertain signal bias

There are two candidates/proposals, 0 and 1. The prior probabilities that 0 and 1 are correct equal $\pi_0 > 0$ and $\pi_1 > 0$, respectively, where $\pi_0 + \pi_1 = 1$. When j denotes a specific candidate then $\{-j\} = \{0, 1\} \backslash \{j\}$.

Voters receive one of two signals, α or β. The probability of receiving a particular signal is affected by which candidate is correct and signals therefore convey information that can influence voting. The *signal bias* q will be the probability that any given agent receives the α signal and $1 - q$ is the probability that an agent receives the β signal. We assume that the signals that agents receive are conditionally independent, given q. But the signal bias q itself is uncertain even given that

[2] That is, where agents vote for the candidate that is more likely to be correct, conditioning on their information but not on being pivotal.

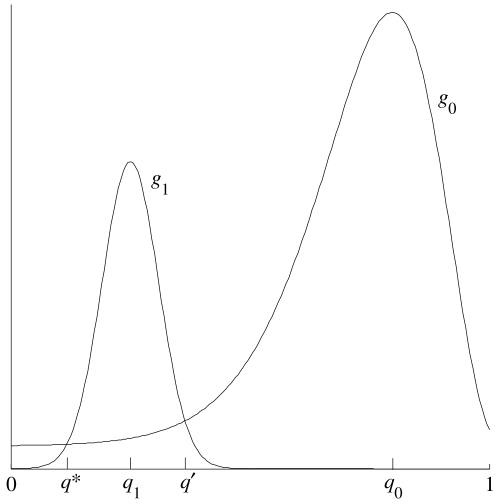

Fig. 1. Signal densities that lead to a failure of information aggregation.

one of the candidates is correct: q has the continuous conditional density $\frac{1}{\pi_0} g_0(q)$ when 0 is correct and $\frac{1}{\pi_1} g_1(q)$ when 1 is correct.[3] The timeline of events is as follows.

Timeline:

(1) nature chooses a correct candidate i and a signal bias q,
(2) each agent receives an i.i.d. signal that equals α with probability q,
(3) the agents vote.

In stage (2), all agents face the *same* probabilities q and $1 - q$ of the α and β signals. If instead each agent were to face an i.i.d. draw from a distribution of possible values for q and then, once q is drawn, a further i.i.d. draw that comes up α with probability q, we could collapse stages (1) and (2) into a single lottery where each agent receives an i.i.d. signal. We would then return to the classical jury model.

The present density model can however be arbitrarily near to a classical model with i.i.d. signals. In particular, when a signal in a classical model indicates that a particular candidate is more likely to be correct – ignoring any complications from conditioning on being pivotal – then the same candidate will remain more likely to be correct following the same signal in any nearby density model. For example, the densities in Fig. 1 lead to distributions that are near to a classical model in which α is an i.i.d. signal that occurs with probability q_0 when 0 is correct and probability q_1 when 1 is correct. Since $q_0 > q_1$, receiving the α signal in the classical model will raise one's estimate that 0 is correct while receiving β will lower that estimate. The signals retain this interpretation in any nearby density model. Nevertheless we will see that with the densities in Fig. 1 information can fail to aggregate once agents condition on being pivotal.

We say that an agent who receives signal $i = \alpha, \beta$, is an agent of *type i*.

Call a $q^* \in [0, 1]$ such that $g_0(q^*) = g_1(q^*)$ an *intersection point*. To avoid some knife-edge cases, we assume throughout that, for any intersection point q^*, the functions g_0 and g_1 cross at q^* rather than being tangent. That is, we assume that if q^* is an intersection point then there is an open interval $I \subset (0, 1)$ that contains q^* such that, for any $p, q \in I$ with $q < q^*$ and $p > q^*$, $(g_0(q) - g_1(q))(g_0(p) - g_1(p)) < 0$ and where $(g_0(r) - g_1(r))/(r - q^*)$ is bounded away from 0 for all $r \in I$.

The information aggregation properties of equilibria will turn on whether g_0 cuts g_1 from above or below. We say that an intersection q^* *is a* $I_{0,1}$ *intersection* if, for all q in the interval I given in the previous paragraph, $(q < q^* \Rightarrow g_0(q) > g_1(q))$ and $(q > q^* \Rightarrow g_1(q) > g_0(q))$, and *a* $I_{1,0}$ *intersection* if, for all q in I, $(q < q^* \Rightarrow g_1(q) > g_0(q))$ and $(q > q^* \Rightarrow g_0(q) > g_1(q))$. In Fig. 1, q^* is a $I_{0,1}$ intersection and q' is a $I_{1,0}$ intersection. (Keep in mind that each g_i is an unconditional density and hence integrates to π_i, not 1.)

The population size equals $n + 1$, where n is an even integer. Each agent therefore thinks that the election excluding his or her vote could in principle be tied but that one vote leads are impossible.

If we fix the signal bias q, then the candidate with the (weakly) higher conditional probability of being correct, given q, is any $i = 0, 1$ such that $g_i(q) \geqslant g_{-i}(q)$. Since agents taken together receive no pertinent information beyond q, the signal bias can be understood as the state variable for the model. Although reference to which candidate is 'correct' given q will facilitate interpretation, what is important is that, conditional on q, *every* agent prefers the candidate i with the larger value of $g_i(q)$ over the other candidate. We could in fact take these conditional preferences as primitive and make no mention

[3] We could allow the distribution of q to have atoms, which would be a natural modeling step when seeking an approximation of a classical model, where q is certain.

of which candidate is correct or incorrect. Such a 'preferences' description would match the Feddersen and Pesendorfer (1997) modeling strategy, except that we specialize by going to the extreme common values assumption that the agents have unanimous preferences given q.

3. Equilibrium

An agent of type $i=\alpha,\beta$, either votes for 0 or 1 and the probabilities that i casts these votes are given by $v_i^0 \geqslant 0$ and $v_i^1 \geqslant 0$, respectively, where $v_i^0+v_i^1=1$. We consider only symmetric equilibria, that is, where all agents of the same type adopt the same voting probabilities.

Let $P(E)$ denote the probability of an event E. Given (v_i^0, v_i^1) for $i=\alpha,\beta$, $P(T|q)$ and $P(j \text{ leads by } y \text{ votes}|q)$ will indicate respectively the probability of a tie and the probability that j leads by y votes, conditional on the signal bias q and given that the population consists of n agents (not $n+1$). The binomial formula gives explicit expressions for these probabilities. When we need to emphasize that probabilities and events are a function of n, we use n as a subscript, e.g., $P_n(T|q)$. Although the probabilities of election outcomes are partly determined by $(v_i^0, v_i^1)_{i=\alpha,\beta}$, we suppress this dependence in the notation.

We assume that each agent takes the other agents' voting probabilities as fixed and acts to maximize the conditional probability that the correct candidate is elected given the signal the agent observes. Suppose that an arbitrary agent votes for j and that the remaining agents' actions are governed by $(v_i^0, v_i^1)_{i=\alpha,\beta}$. We then use $C^j(i)$ to denote the (unconditional) probability that the correct candidate is elected and that the arbitrary agent under consideration receives the i signal.

To illustrate how $C^j(i)$ is calculated, consider as an example $C^0(\alpha)$ and observe that

$$C^0(\alpha) = P(0 \text{ is correct, } 0 \text{ is elected, } \alpha \text{ is observed}) + P(1 \text{ is correct, } 1 \text{ is elected, } \alpha \text{ is observed}),$$

where the probabilities above are calculated assuming that the agent under consideration votes for 0 and the remaining votes are determined by $(v_i^0, v_i^1)_{i=\alpha,\beta}$. Let's begin with the first term, $P(0 \text{ is correct, } 0 \text{ is elected, } \alpha \text{ is observed})$. Since the action taken is a vote for 0, the conditional probability that 0 wins and α is observed given q is simply $qP(0 \text{ leads by} \geqslant 0 \text{ votes}|q)$, here exploiting the fact that we have ruled out a one vote loss for 0 among n voters. Since we are dealing with the case where 0 is correct, when we step back to take account of the uncertainty regarding q, we must integrate using the density for q that holds when 0 is correct:

$$P(0 \text{ wins, } 0 \text{ is correct, } \alpha \text{ is observed}) = \int_{[0,1]} qP(0 \text{ leads by} \geqslant 0 \text{ votes}|q) g_0(q)\, dq.$$

The second term is similar. Since we are considering a vote for 0, the conditional probability that 1 wins and α is observed given q is $qP(1 \text{ leads by } \geqslant 2 \text{ votes}|q)$ and so

$$P(1 \text{ wins, } 1 \text{ is correct, } \alpha \text{ is observed}) = \int_{[0,1]} qP(1 \text{ leads by} \geqslant 2 \text{ votes}|q) g_1(q)\, dq.$$

Hence

$$C^0(\alpha) = \int_{[0,1]} q\big[P(0 \text{ leads by} \geqslant 0 \text{ votes}|q) g_0(q) + P(1 \text{ leads by} \geqslant 2 \text{ votes}|q) g_1(q)\big]\, dq.$$

The expression $C^j(i)$ gives the unconditional probability of the event that the correct candidate is elected and the agent under consideration observes the i signal, on the assumption that the agent votes for j. The agent chooses j to maximize the conditional probability that the correct candidate is elected given that the i signal is observed, that is,

$$\frac{C^j(i)}{P(i \text{ is observed})}.$$

Since the denominator $P(i \text{ is observed})$ remains unchanged regardless of which candidate type i votes for, the sign of $C^0(i) - C^1(i)$ by itself determines whether voting for 0 or 1 is optimal when signal i is observed.

We therefore say that type i *weakly* (*resp. strictly*) *prefers voting for* j *to voting for* k if $C^j(i) - C^k(i) \geqslant 0$ (resp. > 0).[4]

Definition 1. An *equilibrium* is a $(v_i^0, v_i^1)_{i=\alpha,\beta}$ such that, for $i=\alpha,\beta$, and $j=0,1$, $v_i^j > 0 \Rightarrow i$ weakly prefers voting for j to voting for $-j$.

[4] Preference can instead be viewed in expected utility terms. If utility equals 1 when the correct candidate is elected and 0 if not, then $C^j(i)$ is the probability of the event that utility equals 1 and signal i is observed, assuming the agent votes for j. Division by $P(i \text{ is observed})$ then gives the conditional expectation of the agent's utility given that i is observed and assuming the agent votes for j. Hence voting for j is weakly preferred to voting for k if and only if the agent's conditional expected utility is weakly larger when voting for j.

It will be useful to have explicit expressions for the $C^0(i) - C^1(i)$. We have

$$C^0(\alpha) = \int_{[0,1]} q\big[P(0 \text{ leads by} \geqslant 0 \text{ votes}|q)g_0(q) + P(1 \text{ leads by} \geqslant 2 \text{ vote}|q)g_1(q)\big]\,dq,$$

$$C^1(\alpha) = \int_{[0,1]} q\big[P(0 \text{ leads by} \geqslant 2 \text{ votes}|q)g_0(q) + P(1 \text{ leads by} \geqslant 0 \text{ vote}|q)g_1(q)\big]\,dq.$$

Since $P(j \text{ leads by } \geqslant 0 \text{ vote}|q) = P(j \text{ leads by } \geqslant 2 \text{ vote}|q) + P(T|q)$,

$$C^0(\alpha) - C^1(\alpha) = \int_{[0,1]} qP(T|q)\big(g_0(q) - g_1(q)\big)\,dq. \tag{3.1}$$

Similarly,

$$C^0(\beta) - C^1(\beta) = \int_{[0,1]} (1-q)P(T|q)\big(g_0(q) - g_1(q)\big)\,dq. \tag{3.2}$$

4. Information aggregation

We first set down a benchmark for information aggregation. Since the distribution of the signals is determined by q, q provides all of the payoff-relevant information of the model and is strictly more informative than the particular vector of signals that agents happen to receive; as we earlier pointed out, q can therefore be interpreted as a state variable. We define a candidate i to be (weakly) *more likely to be correct* given q if $g_i(q) \geqslant g_{-i}(q)$. Our common values assumption implies that if q were known then all agents would agree to vote for a candidate i that is more likely to be correct; and, outside of deviations on a measure 0 set of q's, selecting such a candidate is the only decision rule that will maximize the ex ante probability of selecting a correct candidate. Consequently, given a sequence of equilibria indexed by the population size n, we define information aggregation to obtain if for every q a candidate more likely to be correct given q is asymptotically sure to be elected.

Definition 2. Information aggregation obtains for the sequence $\langle (v_i^0(n), v_i^1(n))_{i=\alpha,\beta}\rangle$, if and only if, for all realizations of q, the probability that the elected candidate is more likely to be correct converges to 1 as $n \to \infty$.

Definition 2 closely follows Feddersen and Pesendorfer (1997), but in FP the preference of the median voter determines which candidate must win, given q, if information aggregation is to obtain, whereas in our setting all voters agree on which candidate should win, given q.

Since the state variable q can be accurately estimated in a large electorate from the sample proportion of agents who receive the α signal, the Definition 2 benchmark is in principle achievable. In fact, as long as the two types adopt distinct strategies, $v_\alpha^0 \neq v_\beta^0$, the sample proportion who receive α can be inferred from an election's equilibrium returns: the proportion of votes that 0 receives, say x, will converge to $v_\alpha^0 q + v_\beta^0(1-q)$ as n increases and so we can back out q from $x = v_\alpha^0 q + v_\beta^0(1-q)$ in a large electorate. But although the information the population needs to determine the optimal candidate can be deduced from its aggregate behavior, the optimal candidate need not be the likeliest to win.[5] As we will see, the performance of elections misses the Definition 2 benchmark by a wide margin; far from electing the more-likely-to-be-correct candidate with high probability for every q when n is large, the electorate can select the less-likely-to-be-correct candidate for *most* values of q.

The main principle of our analysis will be that if the population is sufficiently large then ties are most likely to arise at those signal biases near the critical value where the expected vote shares of the candidates are equal. Since agents care about their votes only when they are pivotal, they can ignore all other signal biases. As we will see, this principle implies that in an equilibrium for a large electorate a critical signal bias can occur only at an intersection point q^* of the signal densities – otherwise both types would vote for the same candidate and there would be no critical signal bias at all. The classification of intersections given in Section 2 then determines which type of agent is more likely to vote for each candidate. Having witnessed a α signal, the α types will assign higher probability than the β types to the larger q's in a neighborhood of a critical signal bias. Suppose for instance that a critical signal bias q^* is a $I_{0,1}$ intersection. Then $g_1(q) > g_0(q)$ for the q's slightly greater than q^*, and so the α type will be more likely than the β type to vote for 1. Since the expected proportion of α types in the population equals q and since the expected vote shares for the two candidates are equal at q^*, candidate 1 will be the expected victor when $q > q^*$ and 0 will be the expected victor when $q < q^*$. The

[5] Since the state variable can be inferred from the votes, the failure of information aggregation in the present model differs from the failure in Theorem 4 of Feddersen and Pesendorfer (1997).

$I_{1,0}$ case leads to the opposite result. These conclusions are based only on local facts – the intersection classification – and not on which candidate is globally more likely to be correct; as a consequence, equilibria can fail to aggregate information.

Definition 3. The strategies $(v_i^0, v_i^1)_{i=\alpha,\beta}$ *lead to a potentially close election at* q if and only if

$$v_\alpha^0 q + v_\beta^0(1-q) = v_\alpha^1 q + v_\beta^1(1-q).$$

Since we have ruled out abstention, the left- and right-hand sides above will both equal $\frac{1}{2}$. We use the label 'potentially close' because the expected number of votes for 0 and 1 are equal when Definition 3 applies.

Definition 4. A sequence of equilibria $\langle (v_i^0(n), v_i^1(n))_{i=\alpha,\beta}\rangle$ *has the critical signal bias* $\hat{q}$ if and only if the sequence converges to a $(\bar{v}_i^0, \bar{v}_i^1)_{i=\alpha,\beta}$ that leads to a potentially close election at $\hat{q}$ and $\bar{v}_\alpha^0 \neq \bar{v}_\beta^0$.

The condition $\bar{v}_\alpha^0 \neq \bar{v}_\beta^0$ (or equivalently $\bar{v}_\alpha^1 \neq \bar{v}_\beta^1$) ensures that the sequence of equilibria is potentially close at one q rather than at every q.

We can then restate our main principle as: if n is large enough agents will conclude that ties are most likely to occur in a neighborhood of a critical signal bias. To understand this principle, suppose a sequence of equilibria has the critical signal bias $\hat{q}$, let $\widehat{Q}$ be an interval of signal biases with $\hat{q}$ in its interior, and let $\widehat{Q}^c$ be the complement of $\widehat{Q}$ in $[0,1]$. Then $\frac{P_n(T|\widehat{Q}^c)}{P_n(T|\widehat{Q})}$, the ratio of a probability of a tie conditional on q being in $\widehat{Q}^c$ relative to the probability of a tie conditional on q being in $\widehat{Q}$, will converge to 0 as $n \to \infty$. The law of large numbers helps explain this fact. If q is the true signal bias, the law of large numbers implies that for large n it is a virtual certainty that $1/(n+1)$ times the difference between the number of votes for 0 and 1 is near

$$\big(\bar{v}_\alpha^0 q + \bar{v}_\beta^0(1-q)\big) - \big(\bar{v}_\alpha^1 q + \bar{v}_\beta^1(1-q)\big),$$

where $(\bar{v}_i^0, \bar{v}_i^1)_{i=\alpha,\beta}$ is the limit of the sequence of equilibria. Since this quantity equals 0 when $q = \hat{q}$ and is bounded away from 0 for $q \in \widehat{Q}^c$, ties are much more likely when n is large if we condition on $\widehat{Q}$ rather than $\widehat{Q}^c$.[6]

The first important implication of the principle that, when n is large, ties are more likely near critical signal biases is that in equilibrium critical signal biases must be intersection points. If some sequence of equilibria had a critical signal bias $\hat{q}$ that is not an intersection point then $g_j(\hat{q}) > g_{-j}(\hat{q})$ for one of the candidates j and each agent would conclude, when n is large, that if he is pivotal then j is more likely to be correct. Agents of both types would therefore adopt the pure strategy of voting for j when n is large and hence the sequence could not in fact have a critical signal bias at $\hat{q}$ (or any other q). To make the same argument more formally, suppose the critical signal bias $\hat{q}$ is not an intersection point and set the interval $\widehat{Q}$ to contain $\hat{q}$ and be small enough that $\widehat{Q}$ contains no intersection points; the sign of $g_0(q) - g_1(q)$ is then constant on $\widehat{Q}$. Since $\frac{P_n(T|\widehat{Q}^c)}{P_n(T|\widehat{Q})} \to 0$, the $C^j(i)$ are determined only by the q's in $\widehat{Q}$: (3.1) and (3.2) therefore imply

$$C_n^0(\alpha) - C_n^1(\alpha) \approx \int_{\widehat{Q}} q P_n(T|q)\big(g_0(q) - g_1(q)\big)\, dq,$$

$$C_n^0(\beta) - C_n^1(\beta) \approx \int_{\widehat{Q}} (1-q) P_n(T|q)\big(g_0(q) - g_1(q)\big)\, dq,$$

when n is large. Both these expressions have the same sign, the sign of $g_0(q) - g_1(q)$ on $\widehat{Q}$. Hence, for large enough n, $v_\alpha^0(n)$ and $v_\beta^0(n)$ must either both equal 1 or both equal 0, which implies that the limit $(\bar{v}_i^0, \bar{v}_i^1)_{i=\alpha,\beta}$ cannot lead to a potentially close election at $\hat{q}$. We therefore rule out sequences of equilibria with a critical signal bias that is not an intersection point. Conversely,

Proposition 1. *For any intersection point q^* there exists a sequence of equilibria that has a critical signal bias equal to q^*.*

[6] This argument does not lead to a proof since the law of large numbers does not give the probability of an exact tie. Consider, however, two signal biases $\hat{q}$ and q', where some $(v_i^0, v_i^1)_{i=\alpha,\beta}$, leads to a potentially close election at $\hat{q}$ but not at q'. It is well known that $P_n(T|\hat{q})$ converges to 0 at rate $\frac{1}{\sqrt{n}}$ while $P_n(T|q')$ converges to 0 at rate e^{-cn}, where $c > 0$. So $\frac{P_n(T|q')}{P_n(T|\hat{q})} \to 0$. Given that the likelihood of a signal bias is governed by a density, the probability that the bias assumes any particular value is zero. But a similar conclusion holds for sets of biases. If an interval $\widehat{Q} \subset [0,1]$ contains $\hat{q}$ in its interior and if the closure of $Q' \subset [0,1]$ contains no q at which $(v_i^0, v_i^1)_{i=\alpha,\beta}$ leads to a potentially close election, then $P_n(T|\widehat{Q})$ converges to 0 at rate $\frac{1}{n}$ while $P_n(T|Q')$ converges to 0 at rate e^{-cn}, $c > 0$. Thus $\frac{P_n(T|Q')}{P_n(T|\widehat{Q})} \to 0$. See Good and Mayer (1975), Chamberlain and Rothschild (1981), and Mandler (2010) for these rate of convergence results.

Proposition 1 is our main technical result and is proved in Appendix A.

We now argue that for the sequences of equilibria given by Proposition 1 the global information aggregation properties are determined by the local fact of whether the intersection point q^* is of class $I_{0,1}$ or $I_{1,0}$. Given that agents will ignore any signal bias not in a neighborhood of q^*, it is no surprise that it is only the properties of the signal densities near q^* that matter.

Consider a sequence of equilibria with limit $(\bar{v}_i^0, \bar{v}_i^1)_{i=\alpha,\beta}$ that has a critical signal bias at the intersection point q^*. To illustrate, assume q^* is a $I_{0,1}$ intersection and that we are in the typical case where $q^* \neq \frac{1}{2}$ (as in Fig. 1). Then, since expected vote shares are equal in the limit when the signal bias is q^*, at least one of the types must be mixing between voting for 0 and voting for 1 when n is sufficiently large.[7] If, say, type β mixes then (3.2) implies that

$$C_n^0(\beta) - C_n^1(\beta) = \int_{[0,1]} (1-q) P_n(T|q)\big(g_0(q) - g_1(q)\big)\, dq = 0, \tag{4.1}$$

for all large n. In deciding between 0 and 1, the type α agent, using (3.1), examines the sign of

$$C_n^0(\alpha) - C_n^1(\alpha) = \int_{[0,1]} q P_n(T|q)\big(g_0(q) - g_1(q)\big)\, dq. \tag{4.2}$$

Our main principle implies that we may ignore all q except those near q^*. Now the α type, having witnessed the α signal, will assign higher posterior probability to the larger values of q in a neighborhood of q^* – specifically those greater than q^* – than the β type assigns. In the above integrals, this difference appears as the weight q in (4.2) versus the weight $1-q$ in (4.1). Consequently, given (4.1) and that q^* is a $I_{0,1}$ intersection, it must be that the α type strictly prefers to vote for 1 when n is large: for q's that are slightly larger than q^*, 1 is more likely to be correct and the α type views these q's as more likely. Similarly, if we assume it is the α type who is mixing, then the β type, who believes that the smaller values of q near q^* are more likely, will strictly prefer voting for 0 to voting for 1 when n is large. Whoever is mixing, therefore, our assumption that q^* is a $I_{0,1}$ intersection implies that, for large n, the α type has a higher probability of voting for 1 (or lower probability of voting for 0) than the β type.

Now if the true q turns out to be greater than q^* then the expected proportion of type α voters will be larger at q than at q^*, while if the true q is less than q^* then the expected proportion of type α voters will be smaller at q than at q^*. Since for large n expected vote shares are equal when the true signal bias is q^*, our conclusion that the α type is more likely than the β type to vote for 1 when q^* is a $I_{0,1}$ intersection means that if the true q is greater than q^* then candidate 1 will have the larger expected vote share while if the true q is less than q^* then it is 0 that will have the larger expected vote share. So if $q > q^*$ the law of large numbers reports that as $n \to \infty$ the probability that 1 is elected converges to 1 while if $q < q^*$ then the probability that 0 is elected converges to 1.

The analysis of $I_{1,0}$ intersections is entirely parallel, but the conclusions of the last two paragraphs are reversed: when n is large the α type will have a higher probability of voting for 0 than the β type, and so when the true q is greater than q^* the probability that 0 is elected converges to 1 and when the true q is less than q^* the probability that 1 is elected converges to 1.

Proposition 2. *Let a sequence of equilibria have the critical signal bias q^*. If q^* is a $I_{0,1}$ intersection then for $q > q^*$ the probability that* 1 *wins approaches* 1 *as $n \to \infty$ while for $q < q^*$ the probability that* 0 *wins approaches* 1 *as $n \to \infty$. If q^* is a $I_{1,0}$ intersection then for $q > q^*$ the probability that* 0 *wins approaches* 1 *as $n \to \infty$ while for $q < q^*$ the probability that* 1 *wins approaches* 1 *as $n \to \infty$.*

Proposition 2 implies that information aggregation can fail to obtain. For example, if an intersection q^* is near 0 or 1 then there will be a sequence of equilibria with the critical signal bias q^* such that for most values of q the same candidate is elected (with probability near 1 when n is large). Yet, because the behavior of the g_i near q^* is unrelated to their behavior away from q^*, it could be that the candidate who is rarely elected is more likely to be correct for most values of q. See Fig. 1 where q^* is a $I_{0,1}$ intersection near 0, leading 1 to be elected at most values of q, but where nevertheless $g_0(q) > g_1(q)$ for most values of q. Thus not only can information aggregation fail – there are some q's where the less-likely-to-be-correct candidate, conditional on q, is elected – but it could be that for most values of q the candidate less likely to be correct is elected. If agents pooled their information they could estimate the true signal bias with near perfect accuracy when n is large and could therefore usually select the candidate more likely to be correct given q. In fact, even if agents cannot pool their information they could well be better off if they simply all voted for one of the candidates regardless of their signals. If, as in Fig. 1, a $I_{0,1}$ intersection q^* is near 0 and $\pi_0 > \pi_1$, then when n is large the agents would have higher expected utility if they all voted for 0 compared to the equilibrium with the same n that is part of a sequence that has q^* as a critical signal bias. (Recall that π_i is the prior probability that i is correct.)[8]

---

[7] If in the limit both types play pure strategies and expected vote shares are equal then $(\bar{v}_\alpha^0 - \bar{v}_\alpha^1)q^* + (\bar{v}_\beta^0 - \bar{v}_\beta^1)(1-q^*) = 0$ must be solved at a $(\bar{v}_\alpha^0 - \bar{v}_\alpha^1, \bar{v}_\beta^0 - \bar{v}_\beta^1)$ equal to either $(1,-1)$ or $(-1,1)$, which is possible only when $q^* = \frac{1}{2}$.

[8] Notice the characteristic failure of the g_i in Fig. 1 to satisfy a monotone likelihood ratio property. We discuss this further below.

Poor information aggregation performance does not depend on signal biases that show significant dispersion. To establish this, we compare the classical model, where a signal has a fixed probability of flashing one of its two values given that 0 or 1 is correct, with a nearby model that endows the signal bias with a density. We fix the probabilities that 0 and 1 are correct, π_0 and π_1. A 'classical model' is then simply a pair $(q_0, q_1) \in [0,1] \times [0,1]$ where $q_0 \neq q_1$: the signal is sure to have the bias q_0 if 0 is correct and is sure to have the bias q_1 if 1 is correct. A classical model therefore assumes that q has a distribution defined by

$$P(q \leqslant \tilde{q} | i \text{ is correct}) = \begin{cases} 0 & \text{if } \tilde{q} < q_i, \\ 1 & \text{if } \tilde{q} \geqslant q_i, \end{cases}$$

for $i = 0, 1$. The 'density model' of this paper, in contrast, is defined by a pair of densities (g_0, g_1) where each $\int_{[0,1]} g_i(q)\,dq = \pi_i$. A density model therefore approximates a classical model if the distributions of q given that 0 or 1 is correct are near to the distributions of the classical model, as in Fig. 1. To define 'near', let q_i^t denote the random variable on [0, 1] with density $\frac{1}{\pi_i} g_i^t$.

Definition 5. A sequence of density models $\langle (g_0^t, g_1^t) \rangle$ converges to (q_0, q_1) if and only if, for $i = 0, 1$, q_i^t converges in probability to q_i.[9]

In a classical model, the symmetric Nash equilibria of a voting game will lead the probability that the correct candidate wins to converge to 1 as $n \to \infty$, assuming that we refine away equilibria where agents have no chance of being decisive; each candidate i will therefore be elected with probability π_i when n is large. Since the candidate that is more likely to be correct given the signal bias will be elected with probability that converges to 1 as n increases, information aggregation obtains.

When a density model is near to a classical model then information aggregation will require similar outcomes in the two models: each candidate i must be elected with probability near π_i when n is large. Equilibrium behavior in the density model can be quite different however. Since a density model (g_0^t, g_1^t) near to (q_0, q_1) can have arbitrary values or slopes for the g_i^t in any small interval, we may construct a sequence of density models $\langle (g_0^t, g_1^t) \rangle$ that converges to (q_0, q_1) such that each (g_0^t, g_1^t) has a $I_{0,1}$ intersection q^{*t} and where q^{*t} converges to 1. So, for any large t, there will be a sequence of equilibria where the probability that 0 is elected will converge to a value near 1 as $n \to \infty$. We can build the sequence of models so that at the same time each (g_0^t, g_1^t) has another $I_{0,1}$ intersection q^{**t} where q^{**t} converges to 0 as $t \to \infty$, leading there to be, for any large t, a second sequence of equilibria where the probability 1 is elected will also converge to a value near 1 as $n \to \infty$. Hence, letting P_n^t denote probabilities for the model (g_0^t, g_1^t), we have

Theorem 1. *For any classical model of signals (q_0, q_1), there is a sequence of density models $\langle (g_0^t, g_1^t) \rangle$ that converges to (q_0, q_1) where for candidate $i = 0$ or $i = 1$, any $\varepsilon > 0$, and for all t sufficiently large, there is a sequence of equilibria such that*

$$\lim_{n \to \infty} P_n^t(i \text{ wins}) > 1 - \varepsilon.$$

We omit a formal proof which would add little to the explanation we have already given; see Mandler (2011) for an explicit construction.

We may in fact choose the (g_0^t, g_1^t) so that each g_i^t is single-peaked and has full support, $g_i^t(q) > 0$ for all $q \in [0, 1]$. Theorem 1 holds for any (π_0, π_1). So for example π_1 could be near 1 but still 0 can be elected with high probability when t and n are large, the polar opposite of what information aggregation requires.

We can resurrect a version of information aggregation if we suppose that (g_0, g_1) has exactly one intersection point q^*. Recall from Proposition 2 that if a sequence of equilibria has a critical signal bias at a $I_{0,1}$ intersection point q^* then $q > q^* \Rightarrow$ (the probability that 1 wins converges to 1 (as $n \to \infty$)) and $q < q^* \Rightarrow$ (the probability that 0 wins converges to 1). Moreover, if q^* is a $I_{0,1}$ intersection *and* is the sole intersection point, and if the realization q is greater than q^*, then 1 is more likely to be the correct candidate (since $g_1(q) > g_0(q)$ for any $q > q^*$). Similarly, if the realization q is less than q^* then 0 is more likely to be the correct candidate. The $I_{1,0}$ case is of course parallel, and the more-likely-to-be-correct candidate given q will again be elected with probability near 1 in a large electorate. Thus we have,

Proposition 3. *If (g_0, g_1) has exactly one intersection point q^* and a sequence of equilibria has a critical signal bias at q^*, then information aggregation obtains: for all realizations of q the probability that a candidate that is more likely to be correct is elected converges to 1 as $n \to \infty$.*

Proposition 3 gives little reassurance however: it is hard to see why a (g_0, g_1) should have just one intersection point. Even if each g_i is single-peaked and leads to a distribution function that approximates the step function of a classical model

9 That is, that, for any $\varepsilon > 0$, $\lim_{t \to \infty} \int_{|q - q_i| < \varepsilon} (1/\pi_i) g_i^t(q)\,dq = 1$. We could allow the sequence of approximating models to have distributions of q that have atoms at q_0 and q_1 without sacrificing the conclusion of Theorem 1.

of signals, (g_0, g_1) can have arbitrarily many intersection points. So, while a (g_0, g_1) that satisfies a monotone likelihood ratio property cannot have multiple intersections, a (g_0, g_1) can be arbitrarily near a classical model and yet fail to satisfy the MLRP.

Theorem 1 and Proposition 3 taken together deliver the message that information aggregation does not hinge on (g_0, g_1) being near to a classical model: Theorem 1 says that if (g_0, g_1) is near a classical model then information aggregation can fail to obtain while Proposition 3 shows that information aggregation does not require proximity to a classical model. Since a (g_0, g_1) with a single intersection need not satisfy the MLRP, Proposition 3 also shows that information aggregation does not depend on likelihood ratios being monotone.

Proposition 3 always applies locally in the sense that, for any intersection q^* – even if it is not the sole intersection – there is a neighborhood of q^* such that for any sequence of equilibria that has a critical signal bias at q^* and any realization of q in this neighborhood, the more-likely-to-be-correct candidate will be elected with high probability when n is large. When we add a global assumption that (g_0, g_1) has only one intersection, this local optimality extends globally.

5. Conclusion

In those classical models of elections where voters take independent actions and expected vote shares are unequal, ties become extraordinary unlikely when the population is sufficiently large. As a consequence, if we introduce parameters where expected vote shares *are* equal when the population size is large, the new parameters will rule the roost: no matter how small the likelihood of the new parameters, a voter who conditions on being pivotal will consider the new parameters to be much more likely than the original parameters when the population is large.

The Bayesian-exchangeability approach to large elections weakens independence by allowing agents' actions to be conditionally independent, given the outcome of some random variable. One structural advantage of this approach is that voters, when they care about their vote only when they are pivotal, do not have to consider mind-bogglingly unlikely events. In a society of 100,000 i.i.d. voters, where the difference between the probabilities of voting for 0 and 1 is a mere 0.05, the odds of a tie are less than a decillion (10^{33}) to one. Such extreme odds are a consequence of the fact that, unless the difference in voting probabilities is exactly 0, the probability of tie in an electorate of independent voters will converge to 0 at an exponential rate as the population size increases. In contrast, in the equilibria given in Proposition 1 where voters' actions are only conditionally i.i.d., the probability of a tie will converge to 0 at the more sedate rate of $1/n$.

Acknowledgments

Let me thank Sophie Bade, Roger Myerson, and two referees for helpful advice and Stephen Morris for suggesting this topic.

Appendix A

Throughout the appendix (as in the text), $P_n(T|Q)$ is defined relative to a specification of strategies, $(v_i^0(n), v_i^1(n))_{i=\alpha,\beta}$, and gives the probability of a tie when n (not $n+1$) agents vote.

Definition 6. Given a subsequence $\langle (v_i^0(n), v_i^1(n))_{i=\alpha,\beta} \rangle$ of (possibly nonequilibrium) strategies, candidates j and k, and signal i, we write $j \succsim_i k$ (resp. $j \succ_i k$) if and only if there exists a $\bar{n}$ such that, for all n in the subsequence with $n > \bar{n}$, the type i agent weakly (resp. strictly) prefers voting for j to voting for k when the probabilities of ties are determined by $(v_i^0(n), v_i^1(n))_{i=\alpha,\beta}$.

Lemma 1. *Suppose $\langle (v_i^0(n), v_i^1(n))_{i=\alpha,\beta} \rangle$ is a convergent subsequence with $\lim_{n\to\infty} v_\alpha^0(n) \neq \lim_{n\to\infty} v_\beta^0(n)$ that leads to a potentially close election at some intersection point q^*. If q^* is a $I_{0,1}$ intersection then*

$$0 \succsim_\alpha 1 \quad \Rightarrow \quad 0 \succ_\beta 1, \qquad 1 \succsim_\beta 0 \quad \Rightarrow \quad 1 \succ_\alpha 0,$$

while if q^* is a $I_{1,0}$ intersection then

$$1 \succsim_\alpha 0 \quad \Rightarrow \quad 1 \succ_\beta 0, \qquad 0 \succsim_\beta 1 \quad \Rightarrow \quad 0 \succ_\alpha 1,$$

Proof. All four implications have essentially the same proof; we prove only the first. So assume q^* is of a $I_{0,1}$ intersection and $0 \succsim_\alpha 1$. Then, for all n in the subsequence sufficiently large,

$$\int\limits_{[0,1]} q P_n(T|q)\big(g_0(q) - g_1(q)\big)\, dq \geqslant 0. \tag{A.1}$$

The fact that the subsequence leads to a potentially close election at q^* implies there is a $\hat{n}$ such that, for all q and $n > \hat{n}$, $P_n(T|q) > 0$. (We can have $P_n(T|q) = 0$ early in the subsequence since both types might then vote for the same candidate with certainty.) Let I denote an interval containing q^* such that, for all $q \in I$, $(q < q^* \Rightarrow g_0(q) > g_1(q))$ and $(q > q^* \Rightarrow g_1(q) > g_0(q))$.

We begin with some preliminary facts. Fix $n > \hat{n}$. If $q \in I$ and $q > q^*$ then $\frac{q(1-q^*)}{q^*} > 1 - q$ and, by assumption, $g_1(q) > g_0(q)$. Hence

$$\int_{\{q \in I:\, q > q^*\}} q\frac{(1-q^*)}{q^*} P_n(T|q)\big(g_1(q) - g_0(q)\big)\, dq > \int_{\{q \in I:\, q > q^*\}} (1-q) P_n(T|q)\big(g_1(q) - g_0(q)\big)\, dq.$$

Similarly,

$$\int_{\{q \in I:\, q < q^*\}} q\frac{(1-q^*)}{q^*} P_n(T|q)\big(g_0(q) - g_1(q)\big)\, dq < \int_{\{q \in I:\, q < q^*\}} (1-q) P_n(T|q)\big(g_0(q) - g_1(q)\big)\, dq.$$

Hence

$$\int_I q\frac{(1-q^*)}{q^*} P_n(T|q)\big(g_0(q) - g_1(q)\big)\, dq < \int_I (1-q) P_n(T|q)\big(g_0(q) - g_1(q)\big)\, dq. \tag{A.2}$$

We can write potential closeness at q^* as

$$\big(v_\alpha^0(n)q^* - v_\beta^0(n)q^*\big) - \big(v_\alpha^1(n)q^* - v_\beta^1(n)q^*\big) + \big(v_\beta^0(n) - v_\beta^1(n)\big)\frac{q^*}{q} \to \big(v_\beta^0(n) - v_\beta^1(n)\big)\frac{q^*}{q} - \big(v_\beta^0(n) - v_\beta^1(n)\big),$$

and multiplying by $\frac{q}{q^*}$ gives

$$\big(v_\alpha^0(n)q + v_\beta^0(n)(1-q)\big) - \big(v_\alpha^1(n)q + v_\beta^1(n)(1-q)\big) \to \big(v_\beta^0(n) - v_\beta^1(n)\big)\left(1 - \frac{q}{q^*}\right). \tag{A.3}$$

Since if $\lim_{n\to\infty}(v_\beta^0(n) - v_\beta^1(n)) = 0$ were to obtain then $\lim_{n\to\infty} v_\beta^0(n) = \lim_{n\to\infty} v_\beta^1(n) = \frac{1}{2}$, potential closeness would imply $\lim_{n\to\infty} v_\alpha^0(n) = \lim_{n\to\infty} v_\alpha^1(n) = \frac{1}{2}$, violating the assumption that $\lim_{n\to\infty} v_\alpha^0(n) \neq \lim_{n\to\infty} v_\beta^0(n)$. Hence

$$q \neq q^* \implies \lim_{n\to\infty}\big(v_\beta^0(n) - v_\beta^1(n)\big)\left(1 - \frac{q}{q^*}\right) \neq 0. \tag{A.4}$$

We now use these preliminaries to show that there is a $c > 0$ such that

$$e^{cn}\int_I P_n(T|q)\, dq \to \infty \quad \text{and} \quad e^{cn}\int_{[0,1]\setminus I} P_n(T|q)\, dq \to 0. \tag{A.5}$$

To address the first part of (A.5), let $P_{n,v(n')}(T|Q)$ indicate the conditional probability of a tie in a population of n agents when agents play $v(n') \equiv (v_i^0(n'), v_i^1(n'))_{i=\alpha,\beta}$, given that $q \in Q$. Since the left-hand side of (A.3) converges to 0 when $q = q^*$, Mandler (2010, proof of Theorem 7) shows that if Q is an interval with $q^* \in \operatorname{int} Q$ and we fix n' at a sufficiently large value then $nP_{n,v(n')}(T|Q)$ converges to a strictly positive constant as $n \to \infty$. We must set n' to be large enough that Q contains the q such that $v(n')$ leads to a potentially close election at q, which is possible since $\lim_{n\to\infty} v_\alpha^0(n) \neq \lim_{n\to\infty} v_\beta^0(n)$ implies there is a unique such q when n' is large and that this q converges to q^* as $n' \to \infty$. Since $v(n')$ converges as $n' \to \infty$ and since the convergence of $nP_{n,v(n')}(T|Q)$ as $n \to \infty$ is uniform across all sufficiently large n', $\lim_{n\to\infty} nP_n(T|Q)$ is also well defined and strictly positive for any interval Q with $q^* \in \operatorname{int} Q$. Hence, for any $c > 0$, $e^{cn}P_n(T|I) = e^{cn}\int_I P_n(T|q)\, dq \to \infty$.

For the second part of (A.5), define $Y_1, \ldots, Y_n$ to be i.i.d. random variables that equal 1 with probability p and -1 with probability $1 - p$, define $S_n \equiv Y_1 + \cdots + Y_n$, and let $\Pr(\cdot)$ denote the resulting probabilities. The Hoeffding (1963) inequality states that for any $t > 0$ there is a $k(t) > 0$ such that, for any $p \in [0, 1]$,

$$\Pr\left(\left|\frac{S_n}{n} - EY_i\right| > t\right) \leqslant e^{-k(t)n},$$

and, in addition, $\lim_{t\to 0} k(t) = 0$. Now let $\tilde{q}$ be an element of the closure of $[0,1]\setminus I$, which we denote by $\operatorname{cl} I^c$. Letting $\bar{v}_j^i$ denote $\lim_{n\to\infty} v_j^i(n)$, we have

$$v_\alpha^i(n)\tilde{q} + v_\beta^i(n)(1-\tilde{q}) \to \bar{v}_\alpha^i\tilde{q} + \bar{v}_\beta^i(1-\tilde{q})$$

for $i = 0, 1$ and convergence is uniform across $\tilde{q} \in \operatorname{cl} I^c$. Given (A.4), $(\bar{v}_\beta^0(n) - \bar{v}_\beta^1(n))(1 - \frac{\tilde{q}}{q^*}) \neq 0$ for $\tilde{q} \in \operatorname{cl} I^c$. So (A.3) then implies that there is a $\bar{n}$ and a $\varepsilon > 0$ such that, for all $n > \bar{n}$ and $\tilde{q} \in \operatorname{cl} I^c$,

$$\left|\left(v_\alpha^0(n)\tilde{q} + v_\beta^0(n)(1-\tilde{q})\right) - \left(v_\alpha^1(n)\tilde{q} + v_\beta^1(n)(1-\tilde{q})\right)\right| > \varepsilon.$$

Hence there is a $\bar{t}$ such that, for all $n > \bar{n}$ and $\tilde{q} \in \operatorname{cl} I^c$, and defining the Y_i using $p = v_\alpha^0(n)\tilde{q} + v_\beta^0(n)(1-\tilde{q})$ and $1 - p = v_\alpha^1(n)\tilde{q} + v_\beta^1(n)(1-\tilde{q})$, we have $0 < \bar{t} < |EY_i|$. Then, for the same Y_i, $\Pr(S_n = 0) \leqslant \Pr(|\frac{S_n}{n} - EY_i| > \bar{t})$. Since $P_n(T|\tilde{q}) = \Pr(S_n = 0)$, Hoeffding gives $P_n(T|\tilde{q}) \leqslant e^{-k(\bar{t})n}$ for $n > \bar{n}$ and $\tilde{q} \in \operatorname{cl} I^c$. Setting $c \in (0, k(\bar{t}))$, we then have both parts of (A.5).

Given (A.5), (A.2), our no-tangency assumption on intersection points, and the slower convergence of $P_n(T|q)$ to 0 for q near q^* than for $q \not\equiv I$, we have

$$e^{cn}\left[\int\limits_{[0,1]} (1-q)P_n(T|q)\big(g_0(q) - g_1(q)\big)\,dq - \int\limits_{[0,1]} q\frac{(1-q^*)}{q^*}P_n(T|q)\big(g_0(q) - g_1(q)\big)\,dq\right] \to \infty.$$

Since (A.1) implies $e^{cn}\int_{[0,1]} q\frac{(1-q^*)}{q^*}P_n(T|q)(g_0(q) - g_1(q))\,dq \geqslant 0$ for all large n,

$$\int\limits_{[0,1]} (1-q)P_n(T|q)\big(g_0(q) - g_1(q)\big)\,dq > 0$$

for all n sufficiently large, i.e., $0 \succ_\beta 1$. □

Proof of Proposition 1. Since agents can play mixed strategies, it is routine to show that the model has an equilibrium for any n. So we need only show that there are equilibria $\langle (v_i^0(n), v_i^1(n))_{i=\alpha,\beta}\rangle$ for all n sufficiently large that converge to a $(\bar{v}_i^0, \bar{v}_i^1)_{i=\alpha,\beta}$ that leads to a potentially close election at q^* and with $\bar{v}_\alpha^0 \neq \bar{v}_\beta^0$.

There are six cases depending on whether $q^* < \frac{1}{2}$, $q^* = \frac{1}{2}$, $q^* > \frac{1}{2}$, and on whether q^* is a $I_{0,1}$ or $I_{1,0}$ intersection. Since the proofs for the six cases are quite similar, we cover only the case where $q^* > \frac{1}{2}$ and q^* is a $I_{0,1}$ intersection.

Define the variables $\theta_\alpha \equiv v_\alpha^0 - v_\alpha^1$, $\theta_\beta \equiv v_\beta^0 - v_\beta^1$, and the constant $\theta_\alpha^* \equiv \frac{(q^*-1)}{q^*}$ and note that, since $q^* > \frac{1}{2}$, $\theta_\alpha^* \in (-1, 0)$. If the signal bias equals q then the two candidates receive equal vote shares in expectation if

$$q\theta_\alpha + (1-q)\theta_\beta = 0. \tag{A.6}$$

For θ_α in a sufficiently small open neighborhood of θ_α^* and for θ_β in a sufficiently small open neighborhood of 1, let $\tilde{q}(\theta_\alpha, \theta_\beta)$ denote the unique q that satisfies (A.6). Specifically, $\tilde{q}(\theta_\alpha^*, 1) = q^*$. Differentiation shows that, for θ_α near θ_α^* and for θ_β near 1, the function $\tilde{q}$ is strictly increasing in both θ_α and θ_β. We may therefore pick $\delta > 0$ and $\varepsilon > 0$ such that (i) for any $\theta_\beta \in [1-\delta, 1]$, $\tilde{q}(\theta_\alpha^* - \varepsilon, \theta_\beta) < q^*$ and $\tilde{q}(\theta_\alpha^* + \varepsilon, \theta_\beta) > q^*$, (ii) if $q \in \{\tilde{q}(\theta_\alpha, \theta_\beta)\colon\ \theta_\alpha \in [\theta_\alpha^* - \varepsilon,\ \theta_\alpha^* + \varepsilon], \theta_\beta \in [1-\delta, 1]\}$ and $q \neq q^*$ then $g_0(q) \neq g_1(q)$, and (iii) $[\theta_\alpha^* - \varepsilon, \theta_\alpha^* + \varepsilon] \subset (-1, 0)$. Observe that if a $\langle (v_i^0(n), v_i^1(n))_{i=\alpha,\beta}\rangle$ has $v_\alpha^0(n) - v_\alpha^1(n) \in [\theta_\alpha^* - \varepsilon, \theta_\alpha^* + \varepsilon]$ and $v_\beta^0(n) - v_\beta^1(n) \in [1-\delta, 1]$ for all n sufficiently large then, for small enough δ and ε, $v_\alpha^0(n) \neq v_\beta^0(n)$ for the same values of n.

Let $\Delta \equiv \{y \in \mathbb{R}_+^2\colon\ y_1 + y_2 = 1\}$ and define the best response correspondence for a population of size $n+1$, $B^n : \Delta \times \Delta \to \Delta \times \Delta$, by

$$B_i^n\big((\tilde{v}_l^0, \tilde{v}_l^1)_{l=\alpha,\beta}\big) = \big\{\big(v_i^0, v_i^1\big) \in \Delta\colon\ v_i^j > 0 \Rightarrow C^j(i) \geqslant C^{-j}(i),\ j = 0, 1\big\},$$

for $i = \alpha, \beta$, where the $P_n(\cdot|q)$ that define the $C^j(i)$ are calculated using $(\tilde{v}_l^0, \tilde{v}_l^1)_{l=\alpha,\beta}$.

Next, define a correspondence ξ^n with domain and range equal to the convex, compact set $\Delta^\alpha \times \Delta^\beta \equiv \{(v_\alpha^0, v_\alpha^1) \in \Delta\colon\ \theta_\alpha \in [\theta_\alpha^* - \varepsilon, \theta_\alpha^* + \varepsilon]\} \times \{(v_\beta^0, v_\beta^1) \in \Delta\colon\ \theta_\beta \in [1-\delta, 1]\}$. Let $\xi_i^n((v_l^0, v_l^1)_{l=\alpha,\beta})$, $i = \alpha, \beta$, equal $B_i^n((v_l^0, v_l^1)_{l=\alpha,\beta})$ if $B_i^n((v_l^0, v_l^1)_{l=\alpha,\beta}) \cap \Delta^i \neq \varnothing$ and otherwise equal the point in Δ^i nearest in Euclidean distance to $B_i^n((v_l^0, v_l^1)_{l=\alpha,\beta})$. Since ξ^n is convex-valued and upper hemicontinuous, ξ^n has a fixed point $(v_i^0(n), v_i^1(n))_{i=\alpha,\beta}$.

Given the compactness of $\Delta^\alpha \times \Delta^\beta$, $\langle (v_i^0(n), v_i^1(n))_{i=\alpha,\beta}\rangle$ has a convergent subsequence. To see that $\langle (v_i^0(n), v_i^1(n))_{i=\alpha,\beta}\rangle$ has only one accumulation point and that, for any large n, $(v_i^0(n), v_i^1(n))_{i=\alpha,\beta}$ is an equilibrium, define a $(v_i^0, v_i^1)_{i=\alpha,\beta}$ to be *interior* if it satisfies $v_\alpha^0 - v_\alpha^1 \in (\theta_\alpha^* - \varepsilon, \theta_\alpha^* + \varepsilon)$ and $v_\beta^0 - v_\beta^1 \in (1-\delta, 1]$. Since any interior $(v_i^0(n), v_i^1(n))_{i=\alpha,\beta}$ is an equilibrium, it is sufficient to show that if $(\bar{v}_i^0, \bar{v}_i^1)_{i=\alpha,\beta}$ is an accumulation point of $\langle (v_i^0(n), v_i^1(n))_{i=\alpha,\beta}\rangle$ then it is interior and satisfies $\bar{v}_\alpha^0 q^* + \bar{v}_\beta^0(1-q^*) = \bar{v}_\alpha^1 q^* + \bar{v}_\beta^1(1-q^*)$ and $\bar{v}_\beta^0 - \bar{v}_\beta^1 = 1$, in which case $\langle (v_i^0(n), v_i^1(n))_{i=\alpha,\beta}\rangle$ will have just one accumulation point. Observe first that if there is a subsequence that converges to an interior $(\bar{v}_i^0, \bar{v}_i^1)_{i=\alpha,\beta}$ then $\bar{v}_\beta^0 - \bar{v}_\beta^1 = 1$: if $\bar{v}_\beta^0 - \bar{v}_\beta^1 < 1$ then $1 \succsim_\beta 0$ for the subsequence and hence by the Lemma $\bar{v}_\alpha^0 = 0$, violating interiority. Next suppose that $(\bar{v}_i^0, \bar{v}_i^1)_{i=\alpha,\beta}$ is interior and satisfies $\bar{v}_\alpha^0 q^* + \bar{v}_\beta^0(1-q^*) \neq \bar{v}_\alpha^1 q^* + \bar{v}_\beta^1(1-q^*)$. Let $\bar{q}$ denote $\tilde{q}(\bar{v}_\alpha^0 - \bar{v}_\alpha^1, \bar{v}_\beta^0 - \bar{v}_\beta^1)$ and let $N(\bar{q})$ be an arbitrary open neighborhood of $\bar{q}$ such that $N(\bar{q}) \subset \{\tilde{q}(\theta_\alpha, \theta_\beta)\colon\ \theta_\alpha \in [\theta_\alpha^* - \varepsilon, \theta_\alpha^* + \varepsilon],\ \theta_\beta \in [1-\delta, 1]\}$. Then, for any subsequence that converges to $(\bar{v}_i^0, \bar{v}_i^1)_{i=\alpha,\beta}$,

$$\liminf_{n\to\infty} n\int\limits_{N(\bar{q})} P_n(T|q)\,dq > 0 \quad \text{and} \quad n\int\limits_{[0,1]\setminus N(\bar{q})} P_n(T|q)\,dq \to 0 \tag{A.7}$$

(see Mandler, 2010, proof of Theorem 7, and the proof of the Lemma). Hence, for large n, $B^n_\alpha((v^0_i(n), v^1_i(n))_{i=\alpha,\beta}) = \{(0,1)\}$ if $\bar{q} > q^*$ (since then $g_1(\bar{q}) > g_0(\bar{q})$) or $B^n_\alpha((v^0_i(n), v^1_i(n))_{i=\alpha,\beta}) = \{(1,0)\}$ if $\bar{q} < q^*$ (since then $g_0(\bar{q}) > g_1(\bar{q})$). In either case, we have a violation of interiority. To finish, we show that $\langle (v^0_i(n), v^1_i(n))_{i=\alpha,\beta}\rangle$ cannot have an accumulation point $(\bar{v}^0_i, \bar{v}^1_i)_{i=\alpha,\beta}$ that is not interior. First if $(\bar{v}^0_i, \bar{v}^1_i)_{i=\alpha,\beta}$ were to satisfy $\bar{v}^0_\alpha - \bar{v}^1_\alpha = \theta^*_\alpha - \varepsilon$ then, letting $\bar{q}$ denote $\tilde{q}(\bar{v}^0_\alpha - \bar{v}^1_\alpha, \bar{v}^0_\beta - \bar{v}^1_\beta)$, property (i) implies that $\bar{q} < q^*$. But then, since (A.7) again obtains and since $g_0(\bar{q}) > g_1(\bar{q})$, $B^n_\alpha((v^0_i(n), v^1_i(n))_{i=\alpha,\beta}) = \{(1,0)\}$ for n sufficiently large, contradicting the assumption that $\bar{v}^0_\alpha - \bar{v}^1_\alpha = \theta^*_\alpha - \varepsilon$. The case where $\bar{v}^0_\alpha - \bar{v}^1_\alpha = \theta^*_\alpha + \varepsilon$ proceeds similarly. Finally if $\bar{v}^0_\alpha - \bar{v}^1_\alpha \in (\theta^*_\alpha - \varepsilon, \theta^*_\alpha + \varepsilon)$ and $\bar{v}^0_\beta - \bar{v}^1_\beta = 1 - \delta$, then $1 \succsim_\beta 0$ for the relevant subsequence. Hence, by the Lemma, $1 \succ_\alpha 0$, violating the assumption that $\bar{v}^0_\alpha - \bar{v}^1_\alpha \in (\theta^*_\alpha - \varepsilon, \theta^*_\alpha + \varepsilon)$. □

Proof of Proposition 2. Let $\langle (v^0_i(n), v^1_i(n))_{i=\alpha,\beta}\rangle$ be the assumed sequence.

(I) We first show that (i) if q^* is a $I_{0,1}$ intersection then $v^1_\alpha(n) > v^1_\beta(n)$ and $v^0_\beta(n) > v^0_\alpha(n)$ for all large n and (ii) if q^* is a $I_{1,0}$ intersection then $v^0_\alpha(n) > v^0_\beta(n)$ and $v^1_\beta(n) > v^1_\alpha(n)$ for all large n. We prove (i) and leave (ii) to the reader.

Partition the sequence into subsequences for which $1 \succsim_\beta 0$ obtains or for which $0 \succ_\beta 1$ obtains. In the first case then, by the Lemma, $1 \succ_\alpha 0$, and so $v^1_\alpha(n) = 1$ for all large n. But since q^* is a critical signal bias for the sequence, $\lim_{n\to\infty} v^1_\alpha(n) \neq \lim_{n\to\infty} v^1_\beta(n)$, and so $v^1_\beta(n) < 1$ for large n. So, for large n, $v^1_\beta(n) < v^1_\alpha(n)$ and hence $v^0_\beta(n) > v^0_\alpha(n)$. If $0 \succ_\beta 1$ obtains then $v^1_\beta(n) = 0$ for large n and, since $\lim_{n\to\infty} v^1_\alpha(n) \neq \lim_{n\to\infty} v^1_\beta(n)$, we have $v^1_\alpha(n) > 0$ for large n. Thus $v^1_\beta(n) < v^1_\alpha(n)$ and hence $v^0_\beta(n) > v^0_\alpha(n)$ for large n.

(II) As shown in the proof of the Lemma, the fact that $v^0_\alpha(n)q^* + v^0_\beta(n)(1-q^*) \to v^1_\alpha(n)q^* + v^1_\beta(n)(1-q^*)$ implies

$$\big(v^0_\alpha(n)q + v^0_\beta(n)(1-q)\big) - \big(v^1_\alpha(n)q + v^1_\beta(n)(1-q)\big) \to \big(v^0_\beta(n) - v^1_\beta(n)\big)\left(1 - \frac{q}{q^*}\right).$$

So, given q and $\varepsilon > 0$, the law of large numbers implies that the probability that the difference between the proportion of the population that votes for 0 and the proportion that votes for 1 is within ε of $\lim_{n\to\infty}(v^0_\beta(n) - v^1_\beta(n))(1 - \frac{q}{q^*})$ converges to 1 as n increases. With a q^* that is a $I_{0,1}$ intersection, (I) showed that $v^1_\beta(n) < v^1_\alpha(n)$ and $v^0_\beta(n) > v^0_\alpha(n)$ for all large n. So if $\lim_{n\to\infty}(v^1_\beta(n) - v^0_\beta(n)) \geqslant 0$ then $\lim_{n\to\infty}(v^1_\alpha(n) - v^0_\alpha(n)) \geqslant 0$; since if either statement holds with strict inequality we have a violation of potential closeness at q^* and if equality holds for both then $\lim_{n\to\infty} v^0_\alpha(n) = \lim_{n\to\infty} v^0_\beta(n)$ implying that q^* would not be a critical signal bias, it must be that $\lim_{n\to\infty}(v^0_\beta(n) - v^1_\beta(n)) > 0$. Since therefore $\lim_{n\to\infty}(v^0_\beta(n) - v^1_\beta(n)) \times (1 - \frac{q}{q^*})$ is decreasing in q and equals 0 only when $q = q^*$, the probability that 1 wins approaches 1 when $q > q^*$ and the probability that 0 wins approaches 1 when $q < q^*$. Similarly if q^* is a $I_{1,0}$ intersection then $\lim_{n\to\infty}(v^0_\beta(n) - v^1_\beta(n))(1 - \frac{q}{q^*})$ is increasing in q. And so the probability 0 wins approaches 1 when $q > q^*$ and the probability 1 wins approaches 1 when $q < q^*$. □

References

Acemoglu, D., Chernozhukov, V., Yildiz, M., 2009. Fragility of asymptotic agreement under Bayesian learning. Mimeo, MIT.

Austen-Smith, D., Banks, J., 1996. Information aggregation, rationality and the Condorcet jury theorem. Amer. Polit. Sci. Rev. 90, 34–45.

Bhattacharya, S., 2007. Preference monotonicity and information aggregation in elections. Mimeo, University of Pittsburgh.

Chamberlain, G., Rothschild, M., 1981. The probability of casting a decisive vote. J. Econ. Theory 25, 152–162.

de Finetti, B., 1975. Theory of Probability, vol. 2. Wiley, New York.

Feddersen, T., Pesendorfer, W., 1996. The swing voter's curse. Amer. Econ. Rev. 86, 408–424.

Feddersen, T., Pesendorfer, W., 1997. Voting behavior and information aggregation in large elections with private information. Econometrica 65, 1029–1058.

Good, I., Mayer, L., 1975. Estimating the efficacy of a vote. Behav. Sci. 20, 25–33.

Hoeffding, W., 1963. Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58, 13–30.

Kim, J., Fey, M., 2007. The swing voter's curse with adversarial preferences. J. Econ. Theory 135, 236–252.

Kreps, D., 1988. Notes on the Theory of Choice. Westview, Boulder.

Ladha, K., 1993. Condorcet's jury theorem in light of de Finetti's theorem. Soc. Choice Welfare 10, 69–85.

Mandler, M., 2010. How to win a large election. Mimeo, Royal Holloway College, University of London.

Mandler, M., 2011. The fragility of information aggregation in large elections. Mimeo, Royal Holloway College, University of London.

Miller, N., 1986. Information, electorates, and democracy: some extensions and interpretations of the Condorcet jury theorem. In: Grofman, B., Owen, G. (Eds.), Information Pooling and Group Decision Making. JAI Press, Greenwich, pp. 173–192.

Myerson, R., 1998. Extended Poisson games and the Condorcet jury theorem. Games Econ. Behav. 25, 111–131.

Myerson, R., 2000. Large Poisson games. J. Econ. Theory 94, 7–45.

Wit, J., 1998. Rational choice and the Condorcet jury theorem. Games Econ. Behav. 22, 364–376.

Young, P., 1988. Condorcet's theory of voting. Amer. Polit. Sci. Rev. 82, 1231–1244.