Condorcet's Jury Theorem without Symmetry *

Carl Heese[†] Stephan Lauermann[‡]

December 12, 2024

Classic results on the Condorcet jury theorem assume two forms of symmetry to guarantee information aggregation in large electorates: equilibria are symmetric, and voters draw their signals and preferences from identical distributions, making them ex-ante identical (Feddersen and Pesendorfer, 1997, 1998; Bhattacharya, 2013). This paper relaxes these restrictions by allowing heterogeneous preference and signal distributions and permitting asymmetric equilibria. We provide several examples showing that information aggregation may fail without the symmetry restrictions.

However, our main result demonstrates that information aggregates in all non-trivial asymmetric equilibria when voters' preference types satisfy a richness condition. This finding generalizes the Condorcet jury theorem beyond symmetric domains and highlights a constructive role for preference diversity.

The "modern" Condorcet Jury Theorem addresses a collective choice problem where voters must decide on the best course of action under uncertainty. Voters form judgments based on their noisy and private information, resulting in diverging opinions. A classic example is a jury deliberating on a defendant's uncertain guilt; another is shareholders voting on a proposed merger with uncertain profitability. The literature identifies conditions under which, for any equilibrium of a large majority election game, the majority choice is asymptotically equivalent to the choice with full information. These results establish conditions under which elections are effective

* *Preliminary and incomplete. Comments are welcome.* We are grateful for helpful discussions with Nageeb Ali, Francesc Dilme, Mehmet Ekmekci, Andy Kleiner, Wolfgang Pesendorfer, Justus Preusser, and, especially, Emir Kamenica. This work was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation) under Germany's Excellence Strategy EXC 2126/1-390838866 and through the CRC TR 224 (Project B04).

† The University of Hong Kong, Faculty of Business and Economics, Pokfulam Road, Hong Kong, heese@hku.hk

‡ University of Bonn, Department of Economics, s.lauermann@uni-bonn.de.

mechanisms to aggregate voters' private information (Bhattacharya, 2013; Feddersen and Pesendorfer, 1997, 1998).

However, these results critically rely on symmetry assumptions. First, the environment is symmetric: Voters are assumed to be ex-ante identical, drawing preferences and signals from the same distributions. In addition to the symmetric environment, voters are assumed to behave symmetrically, following the same strategy. Consequently, any two voters are indistinguishable and, in equilibrium, submit votes with the same distribution.

These assumptions are a severe restriction. In many natural applications, electorates are heterogeneous, and voters exhibit predictable differences in behavior: voters may receive information of differing precision due to factors like local access to news (Strömberg, 2004; DellaVigna and Kaplan, 2007), micro-targeted campaign messages (Nickerson and Rogers, 2014), or tailored presentations for key shareholders (Yermack, 2010). Voters may also hold distinct preferences based on their identities, locations, and jobs. For example, in distributive politics in the context of trade reforms, some groups of voters may stand to benefit while others may lose depending on what industry they work in (Fernandez and Rodrik, 1991; Kim and Fey, 2007; Acharya, 2016; Ali, Mihm, and Siga, 2024 (forthcoming). Furthermore, shared customs or public news may asymmetrically coordinate voter behavior, even in symmetric environments. This raises the question of whether the Condorcet jury theorem is robust to such asymmetries.

Our analysis provides a nuanced answer. We present examples (Section 1) of asymmetric equilibria where information fails to aggregate. This includes an example demonstrating that such equilibria may even arise in "fully symmetric" settings, such as the pure-common values framework of Feddersen and Pesendorfer (1998). Analogous equilibria arise naturally in related settings with ex-ante asymmetric voters (heterogeneous signal precisions and preferences): They often lead to higher payoffs than all symmetric equilibria.

Our main result shows that such failures of information aggregation cannot happen in environments in which voters' preference types are sufficiently diverse.

Specifically, we study the setting of Feddersen and Pesendorfer (1997) and Bhattacharya (2013) where $2n+1$ voters decide between two possible alternatives, A and B, in a simple majority election with an uncertain state $\omega \in \{\alpha, \beta\}$. Depending on their realized preference type, some voters may prefer A in state α and B in β, with heterogeneous "thresholds of doubt" when uncertain about the state. Others may be "partisans," who prefer one alternative regardless of the state. The preference types are drawn independently across voters and are private information. In addition, all voters privately receive information about the state in the form of a conditionally independent, noisy signal. We allow ex-ante asymmetries, meaning each voter's private signal and preference type may be drawn from distinct distributions.

We analyze the outcomes of large elections as $n \to \infty$. As shown in Section 1, information aggregation may fail in some equilibria when voters share a common threshold of doubt. Therefore, we require that the distribution of the voters' preference types is sufficiently "rich:" (i) voters have a positive chance of being partisan for either alternative; (ii) the voters' threshold of doubt satisfies a full-support condition. The latter implies that any belief change has a strictly positive chance of altering a voter's ranking of the alternatives.

Under these richness conditions, Theorem 1 establishes that the outcome of all undominated equilibria—including asymmetric ones—converges to the majority outcome under full information (i.e., when the state is known).

A central technical difficulty is that asymmetries persist in large elections and do not necessarily wash out. With ex-ante symmetric voters and symmetric strategies, all voters draw the same inference from being pivotal. However, when voters behave asymmetrically, even with an arbitrarily large electorate, two voters may draw markedly different inferences from being "pivotal" for the election outcome. As a result, voting incentives can differ substantially between voters, rationalizing asymmetric behavior by voters with identical preferences and signals even in large elections. The observation is also related to the mechanism driving the "swing voters' curse" (Feddersen and Pesendorfer, 1996; McMurray, 2013), which also remains operational in large elections.[1]

When voters make different inferences from being pivotal, standard arguments by contradiction to rule out failures of information aggregation no longer apply. The example in Section 1 illustrates this. The key ingredient in extending the Condorcet Jury Theorem (CJT) is therefore Lemma 3, which shows that the richness conditions,

[1] The "swing voters' curse" originates from the differing inferences drawn when being pivotal with a vote for A versus B, reflecting the difference between A being behind and ahead by one vote, respectively. Feddersen and Pesendorfer (1996) cite anecdotal evidence for the "swing-voters curse" and connect it to "roll-off"-voting, where voters may vote on some items on a ballot—such as the governor—but abstain from others, such as a referendum on a constitutional change.

in particular the chance of voters being partisan, implies a bound on how different the pivotal inference can be. The bound is violated in the example.

Another problem is the characterization of the likelihood that a voter is pivotal (there is a tie among the other voters). For the symmetric case, it is relatively straightforward to characterize the tying probability. It is equivalent to the probability of exactly n successes among $2n$ independent of Bernoulli trials with identical success probabilities. Asymmetric elections correspond to sequences of Bernoulli random variables that are independent but not identical, and the probability of a tie follows a Poisson binomial distribution (Darroch, 1964), for which no simple characterization is available in the voting literature. For our proof, we, therefore, import results from statistics on the properties of Poisson binomial distributions and theories of large deviations (Dembo, 2009). In particular, we present a characterization of the probability of an exact tie for a sequence of independent, non-identically distributed Bernoulli random variables, tailored to the voting setting (Theorem 1). We also establish comparative static results for this probability (Lemma 1 and 2), using properties of the Poisson binomial distribution (Darroch, 1964) as well as stochastic dominance tools. These statistical results may be of independent interest to other voting theorists interested in large elections with asymmetric voters.

Our results concern the role of symmetry in the basic setting of the Condorcet jury theorem. A large literature uses variations of this setting to analyze the performance of elections with uncertainty about decision-relevant facts, showing the effects of various complications for information aggregation. For example, outcomes do not satisfy full-information equivalence when there is aggregate uncertainty with respect to the distribution of preferences (Feddersen and Pesendorfer, 1997, Section 6) or signals (Mandler, 2012). Bhattacharya (2013, 2018) notes the necessity of monotone aggregate preferences. Other limitations are identified in the literature, for example, by Razin (2003), Ali, Mihm, and Siga (2024 (forthcoming), Ekmekci and Lauermann (2020), Barelli, Bhattacharya, and Siga (2022), and Kosterina (2023). It could be interesting to study the role of symmetry restrictions in these variations, too.

Section 1 presents the counterexample. Section 2 presents the statistical results. Section 3 introduces the voting model based on (Feddersen and Pesendorfer, 1997; Bhattacharya, 2013), states Theorem 1 (the "Condorcet jury theorem without symmetry"), characterizes the best response of voters as a function of the probability of a tie among $2n$ Bernoulli variables, and derives the bound on the posterior differences across voters. Section 4 proves Theorem 1. Section 5 provides some extensions and applies them to a persuasion problem.

1 Failure of the CJT with Asymmetric Equilibria

We present a counterexample in a simple symmetric environment with homogeneous voters. There are $2n+1$ voters who choose between two policies, A and B. Voters have pure common values and aim for the policy to match an uncertain state $\omega \in \{\alpha, \beta\}$, with each voter obtaining a payoff of 1 when the state matches the outcome and 0 otherwise. The states are equally likely ex-ante and each voter $i \in \{1, 2, ..., 2n+1\}$ obtains a binary signal $s_i \in \{a, b\}$, drawn independently and identically across voters, with precision

$$\Pr(s_i = a|\alpha) = \Pr(s_i = b|\beta) = r \text{ for all } i, \text{ and } \frac{1}{2} < r < 1.$$

In the common jury interpretation, the state corresponds to whether a defendant is guilty (state α), and the jurors have diverging opinions on whether to convict the defendant (choose A) based on their realized private signals (some observe a while others observe b).

The voting game is as follows. The voters simultaneously vote for A or B based on their private signals; then, the majority choice is implemented. We consider Bayesian Nash equilibria in which each voter i chooses a voting strategy $\sigma_i : \{a, b\} \to [0, 1]$ that is a best response to σ_{-i}, with $\sigma_i(s_i)$ the probability that voter i votes for A with signal s_i. A voting profile $\sigma = (\sigma_i)_{i=1}^{2n+1}$ is "trivial" if there is a voter who is never pivotal for the decision. For example, voting for A with probability 1 by all voters is a trivial equilibrium.

Condorcet Jury Theorem. *(Feddersen and Pesendorfer (1998)) For each n large enough, there is a unique nontrivial symmetric equilibrium σ^* with*

$$\sigma_i^*(a) = 1 - \sigma_i^*(b) = 1, \text{ and } \lim_{n\to\infty} \Pr(A \text{ is elected}|\alpha) = \lim_{n\to\infty} \Pr(B \text{ is elected }|\beta) = 1.$$

The unique equilibrium is given by "sincere" voting where voters vote A after an a-signal and B after a b-signal. The equilibrium "aggregates information", meaning that the elected outcome is the one preferred by the majority of the voters under full information about the state, as the electorate grows large.

However, in the current setting with pure common values, the Condorcet Jury theorem fails without the restriction to symmetric equilibrium:[2]

Example. *For every $n > 0$, there exists a non-trivial, undominated, and asymmetric equilibria for which*

$$\Pr(A \textit{ is elected}|\alpha) = \Pr(B \textit{ is elected } |\beta) = r < 1.$$

Split the voters into one expert (voter $i = 1$) and $2n$ non-experts (voters $i = 2, 3, ..., 2n + 1$). There is an asymmetric equilibrium in which the expert votes sincerely, but the non-experts behave as follows: Every even-numbered non-expert $i \in \{2, 4, 6, ..., 2n\}$ votes A and every odd-numbered non-expert $i \in \{3, 5, 7, ..., 2n + 1\}$ votes for B.

The equilibrium logic is simple. The votes of the non-experts cancel each other. So, the expert is always the one who is pivotal for the election outcome, and it is optimal for her to vote according to her signal. For the non-experts, this is different. Consider voter $i = 2$, who is supposed to vote for A. From her perspective, among the other non-experts, there is one more vote for B than for A. Hence, she is pivotal for the election outcome if and only the expert votes A. Since this implies that the expert's signal is a, she has a strict incentive to vote A if her own signal is a. She also has a weak incentive to vote A if her own signal is b because she knows that the expert's signal is a, implying that her posterior conditional on being pivotal and having a b signal equals her prior. Thus, it is a (weak) best response for all even-numbered voters to vote A. Analogously, it is a (weak) best response for all odd-numbered voters to vote B. Intuitively, the non-experts face a "bias" among the other non-experts that their vote can undo.

The described behavior is an asymmetric equilibrium for every n. It does not aggregate information: The probability of the correct outcome remains equal to r, even as $n \to \infty$; so, the Condorcet jury theorem fails to extend to asymmetric equilibria.

A key feature of the equilibrium is that the difference in the inference for odd- versus even-numbered voters is large enough to overturn the difference in inference from the a versus b signals.

The asymmetric equilibrium involves best responses for the non-experts that are only weak (since they are indifferent when their signal contradicts the expert's inferred

[2] The example is from Justus Preusser (unpublished bachelor's thesis).

signal). However, analogous equilibria also arise when voters are ex-ante asymmetric due to heterogeneous signal precisions or preferences, and these equilibria are strict.

Suppose that voter $i = 1$ truly is an "expert" whose signal has higher precision than the others. Then, when non-expert voter $i = 2$ receives signal b and infers the expert's signal is a, she strictly prefers to vote for A, because the expert's more precise signal outweighs her own. Moreover, when the expert's signal is sufficiently precise, the described asymmetric equilibrium is not only strict but also the best possible equilibrium for the voters.[3]

Similarly, the asymmetric voting behavior becomes a strict best response if the even-numbered voters have some small preference bias toward A and the odd-numbered voters have a small bias toward B.[4]

Finally, voting rules that differ from the simple majority rule can introduce asymmetries, too. For example, if the voting rule specifies that A wins whenever more than $n + d$ votes are for A, then for any $d > 0$, the voter's best equilibrium is asymmetric (Ladha, Miller, and Oppenheimer, 1996).[5]

In short, since asymmetric settings naturally imply asymmetric behavior, the symmetry assumptions of the existing Condorcet jury theorem are a substantial restriction. In the following, we systematically study the election outcomes without the symmetry restriction.

2 Large Deviation Theory for Voting Applications

We present several technical results about large deviation probabilities for sequences of independent but not identically distributed Bernoulli random variables. The results are later used to prove the "Condorcet jury theorem without symmetry" and may be

[3]There is also a symmetric equilibrium where all non-experts vote sincerely. In this case, being pivotal does not convey information about the expert's signal. However, when the expert's signal is sufficiently precise, this equilibrium yields a lower probability of the correct outcome because the noise from the non-experts dilutes the single expert's signal. The inferiority of the symmetric equilibrium is especially evident when the expert's signal perfectly reveals the state while the non-experts obtain no information.

[4]In the context of our model below, this would correspond to the former group having a threshold of doubt just below 1/2 and the latter group having a threshold just above 1/2.

[5]The voter's best asymmetric equilibrium features a deterministic set of voters who vote for B independently of their signal, which counteracts the bias introduced by the voting rule. Ladha, Miller, and Oppenheimer (1996) provides experimental evidence that participants utilize asymmetric strategies and achieve payoffs above the theoretical maximum of those in symmetric equilibria.

interesting to voting theorists more generally.

Typical textbook results about large deviation theory of sequences of non-identical random variables—most prominently the Gaertner-Ellis theorem—are in terms of probabilities of non-point events such as of intervals, $\Pr\left(\sum_{i=1}^{2n} X_i \leq \lfloor 2n\gamma \rfloor\right)$, and express these in terms of a minimizing "Fenchel-Legendre transform". Section 2.1 states a variant (Theorem 1) that is specifically tailored to voting applications: First, by considering point probabilities instead of interval probabilities; second, by giving a formulation in terms of the minimizing expected Kullback-Leibler divergence. We provide an entirely elementary proof in the main text and explain the formal relation of Theorem 1 to the Gaertner-Ellis theorem in the Appendix.

Section 2.2 provides comparative statics of the point probabilities when comparing different sequences of non-identically distributed Bernoulli random variables. Lemma 1 states a monotonicity property that we derive utilizing results by Darroch (1964) about the Poisson binomial distribution. Lemma 2 shows how the formulation in terms of the expected Kullback-Leibler divergence gives a gateway to characterizing behavior aggregated across the sequence. This is key in voting scenarios where outcomes are determined by the aggregate behavior of the voters, such as the number of votes for a given alternative.

2.1 The Point Probabilities for Independent but not Identical Bernoulli Random Variables

Consider a sequence of independent Bernoulli random variables $(X_i)_{i=1}^{\infty}$ with $X_i \in \{0, 1\}$. For any n, let $F^n(q) = \frac{1}{2n} | \{i : q_i \leq q \text{ and } i \leq 2n\}$ the cumulative distribution function of the success probabilities of the first $2n$ trials. We assume that F^n converges pointwise almost everywhere to some F and

$$\mathbb{E}[q] := \int_0^1 q dF(q) \geq \frac{1}{2}.$$

We allow for general c.d.f's F, including those admitting atoms.

We want to characterize the probability that exactly γn out of $2n$ trials are a success, with $\gamma \in (0, 1)$, that is,

$$\Pr\left(\sum_{i=1}^{2n} X_i = \gamma n\right)$$

This would be simple with *identical* success probabilities $\Pr(X_i = 1) = q$ for all i. In this case, given any $\gamma \in (0,1)$ with $\gamma 2n \in \mathbb{N}$, the probability of exactly $\gamma 2n$ successes out of $2n$ trials is well-known to be[6]

$$\Pr\left(\sum_{i=1}^{2n} X_i = \gamma 2n\right) = \exp\left[-2n\mathrm{KL}(\gamma, q) + o(n)\right] \tag{1}$$

with the Kullback-Leibler divergence,

$$\mathrm{KL}(\gamma, q) = \gamma \log\left(\frac{\gamma}{q}\right) + (1-\gamma)\log\left(\frac{1-\gamma}{1-q}\right);$$

The proof idea is due to Cramér (1938): The idea is to perform a change of measure, the "Escher transform" (Escher, 1932). Considering the binomial under which the event is not rare but rather typical, $Z_n \sim \mathcal{B}(2n, \gamma)$, (1) follows from observing that[7]

$$\frac{\Pr\left(\sum_{i=1}^{2n} X_i = \gamma 2n\right)}{\Pr(Z_n = \gamma 2n)} = \exp\left[-2n\mathrm{KL}(\gamma, q)\right] \tag{2}$$

and that[8]

$$\Pr(Z_n = \gamma 2n) = \exp\left[o(n)\right]. \tag{3}$$

For a sequence of independent Bernoulli variables with distinct success probabilities q_i, we now show that its rate function minimizes an analogous *expected* Kullback-Leibler divergence. For the statement of the result, let $B(\gamma)$ denote the set of functions $a : [0,1] \to [0,1]$ that are integrable with respect to the measure implied by F and have mean γ,

$$B(\gamma) = \{a : [0,1] \to [0,1] : \int_0^1 a(q)dF(q) = \gamma\}.$$

Theorem 1 *Consider a sequence of independent Bernoulli random variables* $(X_i)_{i=1}^{\infty}$ *with* $\Pr(X_i = 1) =: q_i \in [\varepsilon, 1-\varepsilon]$ *for some* $\varepsilon > 0$. *Let* F^n *denote the cumulative distribution function of the first* n *success probabilities* q_i. *If there is some c.d.f.* F

[6]Recall that a function f is in $o(n)$ if $\frac{f(n)}{n}$ vanishes to 0 for $n \to \infty$. In a standard abuse of notation, we write $f(n) = o(n)$. So, $g(n) = \exp[o(n)]$ means that there is some function $f(n)$ such that $g(n) = \exp[f(n)]$ and $f(n) \in o(n)$; that is, the exponent grows more slowly than linearly in n.

[7]Note that $\frac{\Pr(Z_n = \gamma 2n)}{\Pr\left(\sum_{i=1}^{2n} X_i = \gamma 2n\right)} = \frac{\gamma}{q}^{2n\gamma}\frac{1-\gamma}{1-q}^{2n(1-\gamma)} = \exp\left[\ln\left(\frac{\gamma}{q}^{2n\gamma}\frac{1-\gamma}{1-q}^{2n(1-\gamma)}\right)\right] = \exp\left[2n\left[\gamma\ln(\frac{\gamma}{q}) + (1-\gamma)\ln(\frac{1-\gamma}{1-q})\right]\right]$.

[8]This is because the p.d.f of the binomial peaks at its mean, implying $\Pr(Z_n = \gamma 2n) \in [\frac{1}{2n}, 1]$. But for any sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in [\frac{1}{2n}, 1]$, it holds $x_n = \exp\left[\ln(x_n)\right] = \exp\left[o(n)\right]$.

such that

$$F^n \text{ converges pointwise almost everywhere to } F,$$

then, for any $\gamma \in (0,1)$,

$$\Pr\left(\sum_{i=1}^{2n} X_i = \lfloor 2n\gamma \rfloor\right) = \exp\left[-2nc^{KL} + o(n)\right]$$

for

$$c^{\mathrm{KL}}(\gamma) = \inf_{a \in B(\gamma)} \int_q \mathrm{KL}(a(q), q) dF(q). \tag{4}$$

Proof. In the i.i.d. scenario, the proof via the Escher transform shows that the ratio (2) between two p.d.f.'s is the relevant term in order to measure point probabilities. This makes it intuitive why a distance measure between two distributions arises, the Kullback-Leibler divergence.

When F^n is a step-function, i.e., the corresponding distribution has finite support, $\{p_1^n, ..., p_D^n\}$, the proof can be extended naturally and is provided here. The proof highlights how the independence of the X_i implies that the *expected* Kullback-Leibler divergence provides the accurate generalization of the i.i.d. result. (It is without loss to assume that D does not depend on n).

We prepare the proof with some notation: Denote by $f_d^n \in (0,1)$ the likelihood of p_d^n. Let us consider only the first $2n$ random variables, $(X_i)_{i=1}^{2n}$. Let $n_d = \sum_{i=1}^{2n} 1_{q_i = p_d^n}$ be the number of i for which $q_i = p_d^n$ and let their share be $\eta_d = \frac{n_d}{2n}$. Given any realized $x \in \{0,1\}^{2n}$, let $m_d(x) = \sum_{i=1}^{2n} 1_{q_i = p_d^n \textit{ and } x_{i=1}}$ be the number of successes among the X_i with $q_i = p_d^n$ and let its (empirical) share be $a_d(x) = \frac{m_d(x)}{n_d}$.

To evaluate

$$\Pr\left(\sum_{i=1}^{2n} X_i = \lfloor 2n\gamma \rfloor\right) = \sum_{m' : \sum_{d=1}^{D} m'_d = \lfloor 2n\gamma \rfloor} \Pr\left(m(X) = m'\right), \tag{5}$$

the key insight is that the independence of the X_i implies that the likelihood of any vector of successes $m' \in \prod_{d=1}^{D} \{0, ..., n_d\}$ is the product of the component-wise success probabilities;

$$\Pr\left(m(X) = m'\right) = \prod_{d=1}^{D} \Pr\left(m_d(X) = m'_d\right). \tag{6}$$

We apply the result (1) of the Esscher transform to each component. For any $m'_d \in \{0, \ldots, n_d\}$ and $a'_d = \frac{m'_d}{n_d}$,

$$\Pr\left(m_d\left(X\right) = m'_d\right) = \exp\left[-n_d \mathrm{KL}(a'_d, p^n_d) + o(n_d)\right]. \tag{7}$$

A lower bound for (5) is the maximal probability of a success vector. Since the product of the exponentials in (7) translates into sums of their exponents, this lower bound is in terms of the sum of the KL-divergence, weighted by the empirical frequencies n_d of the success probabilities,

$$\Pr\left(\sum_{i=1}^{2n} X_i = \lfloor 2n\gamma \rfloor\right) \geq \max_{m' : \sum m'_d = \lfloor 2n\gamma \rfloor} \Pr\left(m\left(X\right) = m'\right)) = \exp\left[-2nc^{\mathrm{KL-D}} + o\left(n\right)\right] \tag{8}$$

with $c^{\mathrm{KL-D}}\left(\gamma\right) = \min_{m' : \sum_{d=1}^{D} m'_d = \lfloor 2n\gamma \rfloor} \sum_{d=1}^{D} \eta_d \mathrm{KL}_B\left(\frac{m'_d}{n_d}, p^n_d\right)$. An upper bound for (5) is the maximal probability of a success vector times the number of possible success vectors, $\#\{m' \in \prod_{d=1}^{D}\{1, \ldots, n_d\}\} \leq (2n)^D$. Since $(2n)^D = e^{D\ln(2n)} = e^{o(n)}$, the upper bound equals the lower bound,

$$\Pr\left(\sum_{i=1}^{2n} X_i = \lfloor 2n\gamma \rfloor\right) \leq (2n)^D \max_{m' : \sum m'_d = \lfloor 2n\gamma \rfloor} \Pr\left(m\left(X\right) = m'\right) = \exp\left[-2nc^{KL-D}\left(\gamma\right) + o\left(n\right)\right]. \tag{9}$$

The lower and the upper bound (8) and (9) imply the claim when the support of F is discrete.[9]

In the Appendix, we prove the general case by approximating sequences $(X_i)_{i=1}^{\infty}$ with general F from above and below by sequences where F is a step-function. The monotonicity property recorded in the next Lemma 1 will imply that the point probabilities are then also approximated from above and below; this way, an application of the squeeze lemma will finally imply the theorem's formula for general F. ■

The condition that the q_i's are bounded away from 0 and 1 can be weakened. For example, the condition that $F(0) < \gamma < 1 - F(1^-)$ is sufficient for the conclusion (we also need to require $q_i \in (0, 1)$ since the proof requires dividing by q_i and $1 - q_i$).

[9] Note that the minimizing vector $c^{\mathrm{KL-D}}\left(\gamma\right)$ on the finite grid $\{m' \in \prod_{d=1}^{D}\{1, \ldots, n_d\} : \sum_{d=1}^{D} m'_d = \lfloor 2n\gamma \rfloor\} \cong \{a' \in \prod_{d=1}^{D}\{\frac{1}{n_d}, \ldots, \frac{n_d}{n_d}\} : \sum_{d=1}^{D} \eta_d a'_d = \frac{\lfloor 2n\gamma \rfloor}{2n} \approx \gamma\}$ converges to the infimum $c^{\mathrm{KL}}\left(\gamma\right)$ across all vectors of possible realizations from the continuous set $[0, 1]^D$, as $n \to \infty$. This is because $n \to \infty$ implies $n_d \to \infty$ for all $d = 1, \ldots, D$, so the finite grid becomes an arbitrarily fine approximation of $[0, 1]^D$.

The condition rules out degenerate cases; for example, cases in which $q_i \in \{0,1\}$ for all i, so that the number of successes equals $\lfloor 2n\gamma \rfloor$ with a probability exactly 0 or 1, respectively.

2.2 Comparative Statics

Lemma 1 *Consider two vectors $(X_i)_{i=1}^{2n}$ and $(X_i')_{i=1}^{2n}$ of independent Bernoulli random variables with*

$$\Pr(X_i{}' = 1) \geq \Pr(X_i = 1) \text{ for all } i$$

(strictly for some $i \leq 2n$). Then, for $S_n = \sum_{i=1}^{2n} X_i$ and $S_n' = \sum_{i=1}^{2n} X_i'$,

$$\Pr[S_n = k] \quad > \quad \Pr[S_n' = k] \;\; \text{for } k \in \{1, 2, ..., \lfloor \mathbb{E}(S_n) - 1 \rfloor\} \tag{10}$$

$$\Pr[S_n = k] \quad < \quad \Pr[S_n' = k] \;\; \text{for } k \in \{\lfloor \mathbb{E}(S_n') + 1 \rfloor, ..., 2n\}. \tag{11}$$

The proof is established in the Appendix by using a property of Poisson binomial distributions like S_n: The p.d.f. of S_n is "bell-shaped."[10] It either has a unique mode or two consecutive modes, with the mode(s) differing from the mean $\mathbb{E}[S_n]$ by at most 1 (Darroch, 1964). Thus,

$$\Pr[S_n = k-1] < \qquad \Pr[S_n = k] \;\text{ for } k \in \{1, ..., \lfloor \mathbb{E}[S_n] - 1 \rfloor\} \tag{12}$$

$$\Pr[S_n = k] > \quad \Pr[S_n' = k+1] \;\text{ for } k \in \{\lfloor \mathbb{E}[S_n] + 1 \rfloor, ..., 2n-1\}. \tag{13}$$

Given the bell-shape, Lemma 1 makes an intuitive statement: On the left of both the mode(s) of S_n and S_n', the density of the distribution with the lower mode(s) is strictly higher; on the right of all modes, the density of the distribution with the higher mode(s) is strictly higher.

The monotonicity property of the point probabilities given by (10) and (11) will play an important role for establishing the "Condorcet Jury Theorem without symmetry". We will apply it to the point event when a voter is "pivotal".

The other tool will be Theorem 1 in connection with the following Lemma 2. The lemma presents the connection between the asymptotic expected success probability

[10] So, the p.d.f. is convex-concave-convex, and, in particular, the p.d.f. is strictly increasing below the mode(s) and strictly decreasing above.

$\mathbb{E}_F(q)$ and the minimizing expected Kullback-Leibler divergence

$$c_F^{\mathrm{KL}}(\gamma) = \inf_{\mu \in B(\gamma)} \int_q \mathrm{KL}(\mu(q), q) dF(q).$$

This relation will be key in our voting setting. It will allow us to characterize $\mathbb{E}_F(q)$, which will correspond to the aggregate behavior of the voters in equilibrium and, consequently, outcomes, which are determined by the voters' aggregate behavior.

Lemma 2 *Let F and $\tilde{F}$ be the cumulative distribution functions of two random variables ordered by first-order stochastic dominance, $\tilde{F}(q) \leq F(q)$ for all $q \in [0,1]$ (with a strict inequality for some q). For all $\gamma \in (0,1)$:*

$$\mathbb{E}_F(q) < \mathbb{E}_{\tilde{F}}(q) \leq \gamma \Rightarrow c_F^{\mathrm{KL}}(\gamma) > c_{\tilde{F}}^{\mathrm{KL}}(\gamma). \tag{14}$$

$$\mathbb{E}_{\tilde{F}}(q) > \mathbb{E}_F(q) \geq \gamma \Rightarrow c_F^{\mathrm{KL}}(\gamma) < c_{\tilde{F}}^{\mathrm{KL}}(\gamma). \tag{15}$$

The proof of Lemma 2 uses that first-order stochastic dominance implies that there is a "monotone coupling" between the distributions of F and $\tilde{F}$; this is an instance of Strassen's theorem.[11] A coupling v is a joint measure that preserves the marginals: For any Lebesgue measurable $U, \tilde{U} \subseteq [0,1]$,[12]

$$v(U \times [0,1]) = \Pr_F(U), \tag{16}$$

$$v([0,1] \times \tilde{U}) = \Pr_F(\tilde{U}). \tag{17}$$

It is "monotone" if

$$v(\{(q,\tilde{q}) : q \leq \tilde{q}\}) = 1 \text{ and } v(\{(q,\tilde{q}) : q > \tilde{q}\}) > 0. \tag{18}$$

The monotone coupling gives us an explicit way to relate the two distributions. Figure 1 shows an example where the distribution of F has a singleton support and that of $\tilde{F}$ is binary, illustrating that each $q \in \mathrm{supp}(F)$ may not be associated deterministically to some $\tilde{q} \in \mathrm{supp}(F)$.

This requires us to consider an enlarged minimization program allowing randomization. For any $\gamma \in (0,1)$, consider the following pairs (v, a) of couplings v and

[11] Strassen's theorem asserts that there is a coupling with $v(\{(q,\tilde{q}) : q \leq \tilde{q}\}) = 1$; see Theorem 17.59 in Klenke (2020) and the discussion before it. Our definition of monotonicity parallels first-order stochastic dominance, which likewise includes strict differences. In fact, first-order stochastic dominance is equivalent to the existence of a monotone coupling defined via (18).

[12] We denote by $\Pr_F(U)$ the likelihood of $q \in U$ given F.

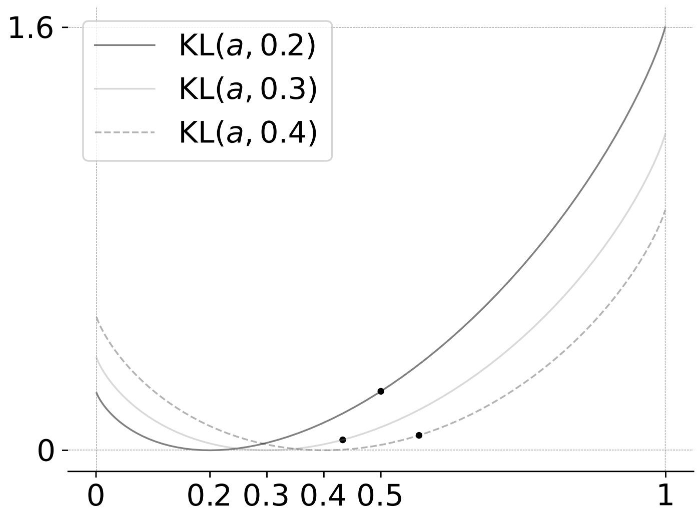

Figure 1: An example where $\gamma = \frac{1}{2}$, $\text{supp}(F) = \{0.2\}$, and $\tilde{F}$ is distributed uniformly on its support $\text{supp}(\tilde{F}) = \{0.3, 0.4\}$. The minimizer of (23) for F is given by $a_F(0.2) = \frac{1}{2}$ and the minimizer for $\tilde{F}$ is given by $a_{\tilde{F}}(0.3) = 1 - a_{\tilde{F}}(0.4) = 0.43303$. We indicate the Kullback-Leibler divergence of the minimizers with the three dots.

measurable functions a,

$$R(\gamma) = \{(a, v) : \int_{(q,\tilde{q})} a(q, \tilde{q}) dv(q, \tilde{q}) = \gamma\}, \tag{19}$$

and the minimization problem

$$\inf_{(a,v)\in R(\gamma)} \int_{(q,\tilde{q})} \text{KL}(a(q, \tilde{q}), \tilde{q}) dv(q, \tilde{q}). \tag{20}$$

The enlarged problem has the same solution as the original one since the strict convexity of the Kullback-Leibler divergence implies that any minimizer features a deterministic function $a(q) = a(q, \tilde{q})$ for all $\tilde{q} \in \text{supp}(\tilde{F})$;

$$\inf_{a\in B(\gamma)} \int_{q} \text{KL}(a(\tilde{q}), \tilde{q}) d\tilde{F}(q) = \inf_{(a,v)\in R(\gamma)} \int_{(q,\tilde{q})} \text{KL}(a(q, \tilde{q}), \tilde{q}) dv(q, \tilde{q}). \tag{21}$$

The rest of the proof is in the Appendix. We define an explicit coupling m—the "quantile coupling" or "Frechet-Hoeffding coupling" (see, e.g., Rachev and Rüschendorf, 2006)— show that it is monotone and finally construct a randomization $\tilde{a}$ so that

$$(\tilde{a}, m) \in R(\gamma)$$

and which yields a point-wise improvement relative to the minimizer a^* of (23) for $\tilde{F}$; for all $(q, \tilde{q}) \in \text{supp}(m)$,

$$\text{KL}(a^*(q), q) \leq \text{KL}(\tilde{a}(q, \tilde{q}), \tilde{q}), \tag{22}$$

and the inequality is strict with a positive m-measure. The details are in the Appendix.

2.3 Triangular Arrays

Theorem 1 extends to triangular arrays $\left((X_{i,n})_{i \leq n} \right)_{n=1}^{\infty}$ of independent Bernoulli variables. The proof is as before, just involving an additional index.

Theorem 2 *Consider a triangular array of independent Bernoulli random variables* $\left((X_{i,n})_{i \leq n} \right)_{n=1}^{\infty}$ *with* $\Pr(X_{i,n} = 1) =: q_{i,n} \in [\varepsilon, 1 - \varepsilon]$ *for some* $\varepsilon > 0$. *Let* F^n *denote the cumulative distribution function of the* n *success probabilities* $q_{i,n}$. *If there is some c.d.f.* F *such that*

$$F^n \text{ converges pointwise almost everywhere to } F,$$

then, for any $\gamma \in (0, 1)$,

$$\Pr\left(\sum_{i=1}^{2n} X_{i,2n} = \lfloor 2n\gamma \rfloor \right) = \exp\left[-2nc^{KL} + o(n) \right]$$

for

$$c^{\text{KL}}(\gamma) = \inf_{a \in B(\gamma)} \int_q \text{KL}(a(q), q) dF(q). \tag{23}$$

3 Condorcet's Jury Theorem without Symmetry

3.1 Voting Model and Equilibrium

As in the example, there are $2n + 1$ voters $i \in \{1, ..., 2n + 2\}$ (she) who choose between A and B with a simple majority vote. There are two states $\omega \in \{\alpha, \beta\}$, with $\Pr(\alpha) = p_0 \in (0, 1)$. Other than in the example, voters do not share a common type, but we allow for private preference types. Moreover, voters can be ex-ante heterogeneous, drawing their preferences and signals from different distributions.

Each voter has a private signal $s_i \in S_i$ from a finite signal set S_i and a private preference type given by a "threshold of doubt" $y_i \in [0,1]$. A voter with a threshold of doubt y prefers A over B if she believes the probability of α is above y. Voters with thresholds of doubt $y \in \{0,1\}$ are "partisans" who prefer A and B, respectively, no matter their beliefs.[13] The signal's distribution is given by $\left(\Pr_i\left(s_i = s|\omega\right)_{\omega\in\{\alpha,\beta\}}\right)_{s\in S_i}$, has c.d.f $\Psi_{i,\omega}$ in ω, and, conditional on the state, is independent of the preference type's distribution, which has a continuous and strictly increasing c.d.f. Φ_i. Types are drawn independently across voters and signals independently conditional on the state.

To consider a large election with asymmetric voters, we fix a sequence of preference type and signal distributions, varying the number of voters n. We impose the following uniformity conditions on n. First, there exists some $\varepsilon > 0$ such that the expected share of A- and B-partisans is bounded away from $\frac{1}{2}$,

$$\frac{1}{2n+1}\sum_{i=1}^{2n+1}\Phi_i\left(0\right) < \frac{1}{2} - \varepsilon \text{ and } \frac{1}{2} + \varepsilon < \liminf_{n\to\infty}\frac{1}{2n+1}\sum_{i=1}^{2n+1}\Phi_i\left(1^-\right), \tag{24}$$

and from 0 and 1,

$$\varepsilon \leq \Phi_i\left(0\right) \text{ and } \Phi_i\left(1^-\right) \leq 1-\varepsilon \text{ for all } i \in \{1,2,...\}. \tag{25}$$

Given (24), under full information about the state, the realized majority preference is A in α and B in β with probability going to 1, as $n \to \infty$; this is a consequence of Kolmogorov's strong law of large numbers for non-identically distributed sequences (see, e.g., Theorem 5.8 in McDonald and Weiss, 2004).

[13] Here is a simple formulation in terms of payoff types $\hat{y} \in \mathbb{R}$: For a voter with type $\hat{y}$, the payoff from A is $1-\hat{y}$ in α and $-\hat{y}$ in β and the payoff from B is 0 in both states. With this specification, a voter prefers A whenever she believes the probability of α to be above $\hat{y}$. Types with $\hat{y} \leq 0$ and $\hat{y} \geq 1$ are "partisans" who prefer A and B, respectively, independently of their beliefs. An atomless distribution of $\hat{y}$ with $\mathbb{R}$ as its support induces a distribution of thresholds of doubt $y \in [0,1]$ (with atoms at 0 and 1) via $y = \max\{\min\{\hat{y},1\},0\}$.

Second, the signals remain boundedly informative, for all $i \in \{1, 2, ...\}$,

$$\varepsilon \leq \Pr_i (s_i = s|\omega) \leq 1 - \varepsilon \quad \text{for all } s \in S_i, \text{ and } \omega \in \{\alpha, \beta\} \tag{26}$$

$$\inf_{s \in S_i} \frac{\Pr_i (s_i = s|\alpha)}{\Pr_i (s_i = s|\beta)} < 1 - \varepsilon < 1 + \varepsilon < \sup_{s \in S_i} \frac{\Pr_i (s_i = s|\alpha)}{\Pr_i (s_i = s|\beta)} \tag{27}$$

Third, there is a uniform Lipschitz bound $L > 0$ for all Φ_i,

$$\frac{1}{L}(x - y) \geq \Phi_i(x) - \Phi_i(y) \geq L(x - y) \text{ for all } i \in \{1, 2, ...\} \text{ and } x > y. \tag{28}$$

A strategy of voter i is a mapping $\sigma_{i,n} : S \times [0,1] \to [0,1]$, where $\sigma_{i,n}(s_i, y_i)$ is the probability that voter i votes A with signal s_i and threshold of doubt y_i. Let $\sigma_n = (\sigma_{i,n})_{i=1}^{2n+1}$ denote the strategy vector. A strategy profile σ_n is "undominated" if $\sigma_{i,n}(s, 1) = 0$ and $\sigma_{i,n}(s, 0) = 1$ for all i, meaning that partisans vote for their alternative. All undominated strategy profiles are nontrivial; there is a positive chance of a tie among any $2n$ voters. Henceforth, we only consider undominated strategies. An " (undominated) equilibrium" is some (undominated) profile σ_n^* with $\sigma_{i,n}^* : S \times [0,1] \to [0,1]$ such that $\sigma_{i,n}^*(s, y)$ is best response to $\sigma_{-i,n}^*$ for all $i \in \{1, \ldots, 2n+1\}$, $s \in S_i$, and $y \in (0,1)$.

When $\Phi_i = \Phi_j$ and $\left(\Pr_i (s_i = s|\omega)_{\omega \in \{\alpha,\beta\}}\right)_{s \in S_i} = \left(\Pr_j (s_j = s|\omega)_{\omega \in \{\alpha,\beta\}}\right)_{s \in S_j}$ for any two voters i, j, the environment is "(ex-ante) symmetric." A strategy profile σ_n^* is symmetric if $\sigma_i^* = \sigma_j^*$ for any two voters.

3.2 Main Result

Bhattacharya (2013) has shown that, in a symmetric environment, for any sequence of symmetric (undominated) equilibria, the probability that A wins in α and B in β converges to 1. Our main result generalizes this.

Theorem 3 *Given a sequence of preferences distributions $(\Phi_i)_{i=1}^{\infty}$ and a sequence of signal distributions $\left(\left(\Pr_i (s_i = s|\omega)_{\omega \in \{\alpha,\beta\}}\right)_{s \in S_i}\right)_{i=1}^{\infty}$ that satisfy the uniform bounds (24) - (28), and any sequence of undominated equilibria $(\sigma_n^*)_{n=1}^{\infty}$,*

$$\lim_{n \to \infty} \Pr (A \textit{ is elected} \mid \alpha; \sigma_n^*, n) = 1 \textit{ and } \lim_{n \to \infty} \Pr (B \textit{ is elected} \mid \beta; \sigma_n^*, n) = 1.$$

3.3 Characterization of Best Responses

This section explains the relation of the voting model to our results on point probabilities of sequences of Bernoulli variables. We show that the point probabilities related to the "pivotal" voting events fully determine the voters' best response. Let

$$q_i(\omega;\sigma_{i,n}) = \int_{s_i,y_i} \sigma(s_i,y_i)d\Phi_i(y_i)d\Psi_{i,\omega}(s_i) \tag{29}$$

be the probability that agent i votes A in state ω. Thus, in state α, a strategy profile σ implicitly defines $2n+1$ independent but not identically distributed Bernoulli random variables $X_{i,n}(\alpha)$, with

$$\Pr\left(X_{i,n}(\alpha)=1\right) = q_i(\alpha;\sigma_{i,n}),$$

and similarly in state β,

$$\Pr\left(X_{i,n}(\beta)=1\right) = q_i(\beta;\sigma_{i,n})$$

The probability that voter i is pivotal is

$$\Pr\left(\text{piv}_i|\omega;\sigma_{-i,n}\right) = \Pr\left(\sum_{j=1}^{2n} X_{j,n}(\omega) = n\right).$$

The posterior probability of α conditional on voter i being pivotal is denoted $\Pr\left(\alpha|s,\text{piv}_i;\sigma_{-i}\right)$ and the posterior likelihood ratio satisfies

$$\frac{\Pr\left(\alpha|\text{piv}_i;\sigma_{-i,n},n\right)}{\Pr\left(\beta|\text{piv}_i;\sigma_{-i,n},n\right)} = \frac{p_0}{1-p_0}\frac{\Pr\left(\sum_{j=1}^{2n} X_{j,n}(\alpha)=n\right)}{\Pr\left(\sum_{j=1}^{2n} X_{j,n}(\beta)=n\right)}.$$

Conditional on being pivotal and signal $s\in S_i$, it is

$$\frac{\Pr\left(\alpha|s,\text{piv}_i;\sigma_{-i,n},n\right)}{\Pr\left(\beta|s,\text{piv}_i;\sigma_{-i,n},n\right)} = \frac{p_0}{1-p_0}\frac{\Pr_i\left(s_i=s|\alpha\right)}{\Pr_i\left(s_i=s|\beta\right)}\frac{\Pr\left(\sum_{j=1}^{2n} X_{j,n}(\alpha)=n\right)}{\Pr\left(\sum_{j=1}^{2n} X_{j,n}(\beta)=n\right)}.$$

Given the strategy profile $\sigma_{-i,n}$ of the other voters, the strategy $\sigma_{i,n}$ is a best response for voter i if $\sigma_i(s,y)=1$ if $\Pr\left(\alpha|s,\text{piv}_i;\sigma_{-i,n},n\right)>y$ and $\sigma_i(s,y)=0$ if $\Pr\left(\alpha|s,\text{piv}_i;\sigma_{-i,n},n\right)<y$,

3.4 Representation and Existence of Equilibrium

We follow an idea from Bhattacharya (2013) to represent equilibrium as a fixed point in beliefs: Consider voter i and suppose her belief conditional on being pivotal is $p_{i,n} = \Pr(\alpha|\text{piv}_i; \sigma_{-i,n})$. Then, her posterior conditional on signal $s \in S_i$ is

$$\frac{p_{i,n} \Pr_i(s_i = s|\alpha)}{p_{i,n} \Pr_i(s_i = s|\alpha) + (1 - p_{i,n}) \Pr_i(s_i = s|\beta)}. \tag{30}$$

Hence, the probability that i votes A when playing a best response given a belief $p_{i,n} \in (0,1)$ is

$$\hat{q}_i(\omega; p_{i,n}) = \sum_{s \in S_i} \Phi_i \left(\frac{p_{i,n} \Pr_i(s_i = s|\alpha)}{p_{i,n} \Pr_i(s_i = s|\alpha) + (1 - p_{i,n}) \Pr_i(s_i = s|\beta)} \right) \Pr_i(s_i = s|\omega) \tag{31}$$

and

$$\hat{q}_i(\omega; 1) = \Phi_i\left(1^-\right).$$

Given the role of $p_{i,n}$ in (30), Bhattacharya (2013) terms it the induced prior. From (31), the induced prior is sufficient to determine a voter's best response. In turn, given a vector $p_n = (p_{1,n}, ..., p_{2n+1,n})$ of induced priors and the best response voting probabilities $\hat{q}_i(\omega, p_i)$ from (31), we find a new posterior, denoted $\Pr(\alpha|\text{piv}_i; p_n)$. Given this discussion, for any equilibrium σ_n^*, the induced priors $p_{i,n}^* = \Pr\left(\alpha|\text{piv}_i; \sigma_{-i,n}^*\right)$ must satisfy

$$p_{i,n}^* = \Pr\left(\alpha|\text{piv}_i; p_{-i,n}^*, n\right) \text{ for all } i \in \{1, 2 \ldots, 2n+1\}. \tag{32}$$

Conversely, any profile of induced priors satisfying (32) induces an equilibrium. We use the fixed-point property (32) to prove the existence and, later, the main result.

Theorem 4 *For every n, there exists an undominated equilibrium σ_n^*.*

Proof. Fix n. For all $p_{i,n} \in [0,1]$, the vote share $\hat{q}_i(\omega; p_{i,n})$ is uniformly bounded away from 0 and 1 across i given (25), and continuous in p_i. It follows that $\Pr(\alpha|\text{piv}_i; p_{-i})$ is uniformly bounded away from 0 and 1 across i by some distance $\delta > 0$ and continous in $p_n \in [0,1]^{2n+1}$. Application of Kakutani's fixed point theorem establishes the existence of some $p_n^* \in [\delta, 1-\delta]^{2n+1}$ that solves (32). Picking, for each voter i, the best-response given the induced prior $p_{i,n}^*$ yields an undominated equilibrium profile σ_n^*. ■

3.5 A Bound on Updating Differences implied by Partisans

The difference in the induced prior ratios $\left(\frac{\Pr(\text{piv}_i|\alpha;p_n)}{\Pr(\text{piv}_i|\beta;p_n)}\right)$ and $\left(\frac{\Pr\left(\text{piv}_j|\alpha;p_n\right)}{\Pr\left(\text{piv}_j|\beta;p_n\right)}\right)$ of any two voters $i \neq j$ is bounded as follows:

Lemma 3 *For any i and j, and any induced prior vector $p_n \in (0,1)^{2n+1}$,*

$$\prod_{i'=i,j} \min_{s \in S_{i'}} \left(\frac{x_{i'}(\alpha)}{x_{i'}(\beta)}, \frac{1-x_{i'}(\alpha)}{1-x_{i'}(\beta)}, \frac{1-x_{i'}(\alpha)}{x_{i'}(\beta)}, \frac{x_{i'}(\alpha)}{1-x_{i'}(\beta)} \right)$$
$$\leq \left(\frac{\Pr(\text{piv}_i|\alpha;p_n)}{\Pr(\text{piv}_i|\beta;p_n)} \right) / \left(\frac{\Pr\left(\text{piv}_j|\alpha;p_n\right)}{\Pr\left(\text{piv}_j|\beta;p_n\right)} \right)$$
$$\leq \prod_{i'=i,j} \max_{s \in S_{i'}} \left(\frac{x_{i'}(\alpha)}{x_{i'}(\beta)}, \frac{1-x_{i'}(\alpha)}{1-x_{i'}(\beta)}, \frac{1-x_{i'}(\alpha)}{x_{i'}(\beta)}, \frac{x_{i'}(\alpha)}{1-x_{i'}(\beta)} \right),$$

for $x_{i'}(\omega) = \Pr(s_{i'} = s|\omega)$ for $i' \in \{i,j\}$ and any $\omega \in \{\alpha,\beta\}$, with strict inequality if the maximum (or, minimum) is taken at $\max_{s \in S_{i'}}\left(\frac{x_{i'}(\alpha)}{x_{i'}(\beta)}\right)$ (or, $\min_{s \in S_{i'}}\left(\frac{x_{i'}(\alpha)}{x_{i'}(\beta)}\right)$) for either $i' = i$ or $i' = j$.

The proof is in the Appendix. The lemma implies that the difference between the induced priors of any two voters is uniformly bounded; given (26)

$$\frac{\varepsilon}{1-\varepsilon} \leq \left(\frac{\Pr(\text{piv}_i|\alpha;p)}{\Pr(\text{piv}_i|\beta;p)} \right) / \left(\frac{\Pr\left(\text{piv}_j|\alpha;p\right)}{\Pr\left(\text{piv}_j|\beta;p\right)} \right) \leq \frac{1-\varepsilon}{\varepsilon} \quad \text{for all } i,j \in \{1,2,\ldots,2n+1\}. \tag{33}$$

For the pure common-values setting of Section 1, the lemma states that the difference in the induced prior ratios is *strictly* bounded by the inference from any signal realizations of the two voters,[14]

$$\left(\frac{1-r}{r}\right)^2 < \left(\frac{\Pr(\text{piv}_i|\alpha;p_n)}{\Pr(\text{piv}_i|\beta;p_n)} \right) < \left(\frac{r}{1-r}\right)^2 \tag{34}$$

for $r = \Pr(s_i = a|\alpha) = \Pr(s_i = b|\beta)$ for all i.

Without partisans, the inequalities of the lemma would not be strict. To see this, recall the sequence of asymmetric (and undominated) equilibria for which information

[14] The proof of the lemma is based on elementary arithmetic calculations; this means it also applies to the example setting.

aggregation fails. In the constructed equilibria, the fourth voter votes for A deterministically, while the fifth votes for B deterministically, and this is supported by the induced priors

$$p_{4,n} = \frac{r}{1-r} \text{ and } p_{5,n} = \frac{1-r}{r}, \tag{35}$$

which imply that $i = 4$ weakly prefers A even after a b-signal and $i = 5$ weakly prefers B even after an a-signal. The lemma rules out the possibility of induced prior pairs such as in (35) *whenever* each voter is a partisan for A and B with arbitrarily small probability $\varepsilon > 0$. In other words, the example equilibrium breaks down whenever there are some partisans.

4 Proof: Equilibria Aggregate Information

We now prove Theorem 3. The first two steps prepare the other two. The third step uses Lemma 1 and 3 to show that, in equilibrium, all converging sequences of equilibrium induced priors $(p_n)_{n=1}^{\infty} = (p_{i,n})_{n=1,\ldots,\infty}$ have an interior limit. The fourth step shows this implies that the vote shares for A are strictly ordered across the states asymptotically, as $n \to \infty$. Finally, we prove the theorem by applying our large deviation tools, Lemma 2 and Theorem 2.

Consider any sequence of equilibria $(\sigma_n^*)_{n=1}^{\infty}$. Let $(p_n^*)_{n=1}^{\infty}$ be the corresponding sequence of equilibrium induced prior vectors; for each i,

$$p_{i,n}^* = \Pr\left(\alpha|\text{piv}_i; \sigma_{-i,n}^*\right).$$

Equilibrium induced priors must satisfy the fixed point equation (32),

$$p_{i,n}^* = \Pr\left(\alpha|\text{piv}_i; p_{-i,n}^*, n\right). \tag{36}$$

Step 1 *For any all $p_n = (p_{i,n})_{i=1}^{2n+1}$ with $p_{i,n} \in (0,1)$ for all i,*

$$\hat{q}_i(\beta; p_{i,n}) \leq \hat{q}_i(\alpha, p_{i,n}) \text{ for all } i.$$

Given any i and $p_{i,n} \in (0,1)$, the distribution of the private signals of i induces a distribution of posteriors (30) in states α and β. Given the bounds on the informativeness of the private signals, (26) and (27), the distribution in α first-order stochastically dominates that in β. Finally, the claim follows from (31) and since Φ_i is strictly increasing.

Step 2 *For any $\delta > 0$, there is $d > 0$ so that for all $n \in \mathbb{N}$ and all $p_n = (p_{i,n})_{i=1}^{2n+1}$ with $p_{i,n} \in (\delta, 1-\delta)$ for all i,*

$$\hat{q}_i(\beta; p_{i,n}) < \hat{q}_i(\alpha, p_{i,n}) + d \;\; \textit{for all } i.$$

Take any i and $p_{i,n} \in (\delta, 1-\delta)$. Given the uniform bounds on the informativeness of the private signals, (26) and (27), there is $\delta' > 0$ so that the support of the distributions of posteriors (30) in the states α and β is in $[\delta', 1-\delta']$ for all i. Finally, given the uniform bounds on the likelihood ratio of the signals, (27), and the uniform Lipschitz bound for Φ_i, (28), we see that the probability to vote A is larger in α than in β by at least some margin d uniformly.

Step 3 *Voters cannot become certain of the state conditional on being pivotal; that is, the inference from the pivotal event must remain bounded:*

$$0 < \liminf_{n\to\infty} p^*_{i,n} \;\textit{and}\; \limsup_{n\to\infty} p^*_{i,n} < 1 \;\textit{for all}\; i \in \{1, 2, \ldots, 2n+1\}. \tag{37}$$

We prove the claim by contradiction. Suppose $\lim_{n\to\infty} p^*_{1,n} = 1$.[15] By Lemma 3 (and the implied uniform bound (33)), this implies that the induced priors converge to 1 uniformly, that is, for every $\delta < 1$, we have $p^*_{i,n} \geq \delta$ for all n large enough and all i. For large enough δ, then $n < (\frac{1}{2} + \varepsilon)2n < \sum_{i=2}^{2n+1} q_i(\beta; p_{i,n})$, given (24). This together with Step 1 means that we can apply Lemma 1 to $X_i = X_{i,n}(\beta)$ and $X'_i = X_{i,n}(\alpha)$ for $i \in \{1, \ldots, 2n+1\}$ (these are the Bernoulli variables given by the voting probabilities $q_i(\omega; p_{i,n})$; compare to Section 3.3), and $k = n$. This yields that being pivotal is indicative of β. Consequently, the posterior conditional on being pivotal is below the prior,

$$\limsup_{n\to\infty} \Pr\left(\text{piv}_1 | \sigma^*_n; p^*_{1,n}, n\right) \leq \Pr(\alpha);$$

this is a contradiction to the starting hypothesis $\lim_{n\to\infty} p^*_{1,n} = 1$.

We prepare the statement of the next step: Helly's selection theorem (Helly, 1912)

[15] It is sufficient to show the contradiction for any converging subsequence, given that the values of the sequence are in the compact set $[0, 1]$. We identify the subsequence with the original sequence to omit the subsequence notation.

implies that there is a subsequence $(b(n))_{n\in\mathbb{N}}$ for which[16]

$$F_{\omega}^{b(n)}(q) = \frac{1}{2b(n)+1} | \left\{ i : \hat{q}_i(\omega, p_{i,n}^*) \le q \text{ and } i \le 2b(n)+1 \right\} \tag{38}$$

converges pointwise to some c.d.f F_ω for all states $\omega \in \{\alpha, \beta\}$. We identify the subsequence with the original sequence to omit the subsequence notation in the following.

Step 4 *F_α first order-stochastically dominates F_β.*

Suppose $\lim_{n\to\infty} p_{1,n}^* = \bar{p}_1$ (see Footnote 15). From Step 3, $\bar{p}_1 \in (0,1)$. Moreover, Lemma 3 implies $\delta > 0$ and $\bar{n} \in \mathbb{N}$ so that $\delta < p_{i,n}^* < 1-\delta$ for all i and $n \ge \bar{n}$. Step 2 then yields some $d > 0$ such that $\hat{q}_{i,n}\left(\alpha; p_{i,n}^*\right) - \hat{q}_{i,n}\left(\beta; p_{i,n}^*\right) \ge d$ for all i and $n \ge \bar{n}$. This implies the step's claim.

Note that convergence in distribution implies that the expected vote share for A in ω converges also,

$$\frac{1}{2n+1}\sum_{i=1}^{2n+1} \hat{q}_i(\omega, p_i^*) = \int_q q dF_\omega^n(q) \overset{n\to\infty}{\rightarrow} \int_q q dF_\omega(q) = \mathbb{E}_{F_\omega}(q). \tag{39}$$

Given Step 4, the expected vote share of A is strictly larger in α than in β for n large enough,

$$\mathbb{E}_{F_\alpha}(q) > \mathbb{E}_{F_\beta}(q). \tag{40}$$

We conclude the proof by leveraging our results from Section 2 on the point probabilities of sequences of Bernoulli random variables: We claim that

$$\mathbb{E}_{F_\alpha}(q) > \frac{1}{2} > \mathbb{E}_{F_\beta}(q). \tag{41}$$

Suppose, for example that $\mathbb{E}_{F_\beta}(q) \ge \frac{1}{2}$. Step 4 and (40) imply that the conditions of Lemma 2 are satisfied for $F = F_\alpha$ and $\tilde{F} = F_\beta$. Lemma 2 and Theorem 2 then imply

$$\lim_{n\to\infty} \frac{\Pr\left(\text{piv}_i|\alpha; p_n^*\right)}{\Pr\left(\text{piv}_i|\beta; p_n^*\right)} = 0 \text{ for all } i;$$

this contradicts Step 3. Thus, $\mathbb{E}_{F_\beta}(q) < \frac{1}{2}$. The analogous argument proves $\mathbb{E}_{F_\alpha}(q) > \frac{1}{2}$. Finally, an application of Kolmogorov's strong law of large numbers for non-identically distributed sequences implies that the realized share of votes for A con-

[16] A reference in English is p. 220 in Natanson (1961).

verges almost surely to the expected share $\mathbb{E}_{F_\omega}(q)$ of the votes for A in each state. This together with (41) implies that the full-information outcome is elected,

$$\lim_{n\to\infty} \Pr\left(A \text{ is elected} \mid \alpha; \sigma_n^*, n\right) = 1 \text{ and } \lim_{n\to\infty} \Pr\left(B \text{ is elected} \mid \beta; \sigma_n^*, n\right) = 1,$$

which was the claim of the theorem.

5 An Extension and Application to Voter Persuasion

The conclusion of our Theorem 3 that information aggregates with probability 1 extends to a more general setting that allows for

(i) prior beliefs and each voter i's signal distribution to vary with the size $2n+1$ of the electorate.

(ii) any signal sets S_i (e.g., a continuum of signals is permitted),

(iii) arbitrarily informative private signals.

Formally, for each $n \in \mathbb{N}$, we drop the requirement that the signal set S_i is finite and allow for arbitrary sets S_i; we consider signal distributions $\Psi_{i,n}$ for all $i \leq n$ (as before, these are distributions over S_i for each state). These define the game of $2n+1$ voters. Varying the electorate size n, the signal distributions form a triangular array. We drop condition (26) for all signal distributions $\Psi_{i,n}$ to allow for arbitrarily informative private signals. Condition (27) is maintained; it requires bounds on the supremum and infimum of the likelihood ratio of signal realizations and slightly abuses notation in the case of uncountable signal sets.

Theorem 5 *Given any sequence of prior beliefs* $(\Pr(\alpha|n))_{n\in\mathbb{N}}$, *any sequence of preference distributions* $(\Phi_i)_{i=1}^\infty$, *any sequence of signal sets* $(S_i)_{i=1}^\infty$, *and any triangular array of signal distributions* $\Psi_{i,n}$ *that satisfy the uniform bounds* (24), (25), (27), *and* (28) *across all* i *and* n, *and any sequence of (undominated) equilibria* $(\sigma_n^*)_{n=1}^\infty$,

$$\lim_{n\to\infty} \Pr(\alpha|n) \Pr\left(A \text{ is elected} \mid \alpha; \sigma_n^*, n\right) + \Pr(\beta|n) \Pr\left(B \text{ is elected} \mid \beta; \sigma_n^*, n\right) = 1.$$

Some Impossibility Results for Persuasion. This generalization has an immediate implication in terms of "Bayesian persuasion" (Kamenica and Gentzkow, 2011) of voters, studied in Hese and Lauermann (2024). Consider the voting model from Section 3, and modify it by introducing a sender: Suppose a sender can provide each voter $i \leq 2n+1$ with additional information about the state. He can specify a set of additional signals M_i and a distribution of the additional signals conditional on each state.

Theorem 5 applies to the scenario where the sender's additional signals are not allowed to be targeted to the voters' private information (preference types and private signals). They are independent of it, conditional on the state. However, they can be tailored to the voters' identities since distributions may differ across i.

Conditionally Independent Signals. Theorem 5 implies that, given any sender's signals that are independent across voters conditional on the state, information aggregates in all sequences of asymmetric equilibria. In other words, as $n \to \infty$, the sender's signals are ineffective and do not alter election outcomes.

Public Signals. The theorem also implies that, given any sender's signals that are perfectly correlated across voters conditional on the state ("public signals"), information aggregates in all sequences of asymmetric equilibria. This is because the theorem implies that, given any possible sequence of realized public signals $(s_n)_{n\in\mathbb{N}}$ and their common posteriors $(\Pr(\alpha|s_n))_{n\in\mathbb{N}}$, in any sequence of continuation equilibria information aggregates with probability 1.

Sketch of the Proof of Theorem 5. The full proof is in the Appendix.

The proof is by contradiction. It starts by assuming that information aggregation would fail in a state—say in β—for which the prior probability does not vanish as $n \to \infty$, and then derives two implications: First, the expected number of votes for A strictly exceeds n in both states when n is large. Therefore, by the logic of Lemma 1, the inference from being pivotal is in the direction of β for all i and n, $\Pr(\alpha|\mathrm{piv}_{i,n}) \leq \Pr(\alpha)$. Second, as $n \to \infty$, for a share of voters i arbitrarily close to 1, the likelihood to vote A must be arbitrarily close in both states. Otherwise, the logic of the previous proof would imply information aggregation; see the argument after Step 4 in Theorem 5's proof.

We then use these two observations to argue that information aggregation does not fail in β, yielding a contradiction: The second observation implies that the likelihood to vote A must become arbitrarily close to the minimum or maximum bound for most voters, so arbitrarily close to either $\Phi_i(0) < \frac{1}{2}$ or to $\frac{1}{2} < \Phi_i(1^-)$), in both states. Otherwise, the voters' private signals would imply a uniform lower bound on their difference in the two states. But then, the first observation implies they cannot be arbitrarily close to $\Phi_i(1^-)$ since this would require arbitrarily strong signals for α with a higher likelihood in β than possible given Bayes consistency. So, the expected vote share for A must be smaller than $\frac{1}{2}$ in β, in contradiction to the initial hypothesis.

6 Conclusion

In the classic Condorcet setting, agents choose between two alternatives with a simple majority vote. The "modern" Condorcet jury theorem states that the outcome of strategic voting converges to that under full information about the state as the number of voters grows large (Bhattacharya, 2013; Feddersen and Pesendorfer, 1997, 1998), provided that all voters are ex-ante symmetric and attention is restricted to symmetric equilibria. Thus, in these models, the equilibrium distribution of votes is ex-ante identical for all voters.

To address the pervasive ex-ante heterogeneity of voters, we have revisited this setting to relax the conventional symmetry assumptions, considering asymmetric equilibria and allowing for the possibility that voters draw their types (signals and preferences) from non-identical distributions.

We presented examples demonstrating that the Condorcet jury theorem does not extend immediately to asymmetric settings. Even in "fully symmetric" settings, such as the pure-common values framework of Feddersen and Pesendorfer (1998), there are asymmetric equilibria where information fails to aggregate. Analogous equilibria arise naturally in related settings with ex-ante asymmetric voters (heterogeneous signal precisions and preference distributions): They often lead to higher payoffs than all symmetric equilibria.

Our main result, the "Condorcet jury theorem without symmetry," showed that sufficient heterogeneity of the voters' (ex-post) preference types is a sufficient condition for information aggregation, even without assumptions of ex-ante asymmetry or the restriction to symmetric equilibria. Thus, our result identifies a positive role for diversity in collective decision-making. Our proof draws on statistical results about

Poisson binomial distributions and their large sample properties. These may be of independent interest to voting theorists.

7 Appendix

7.1 Proof of Lemma 1

Take any $k \in \{1, \ldots, \lfloor \mathbb{E}(S_n) - 1 \rfloor\}$ and $i \leq 2n$ for which $\Pr(X_i' = 1) > \Pr(X_i = 1)$. Denoting $S_{n \setminus i} = \sum_{j \in \{1,\ldots,2n\} \setminus \{i\}} X_j$,

$$\Pr(S_n = k) = \Pr(X_i = 1) \Pr\left(S_{n\setminus i} = k - 1\right) + \Pr(X_i = 0) \Pr\left(S_{n\setminus i} = k\right). \tag{42}$$

The sum $S_{n\setminus i}$ is distributed according to a Poisson binomial distribution. We use that the p.d.f of $S_{n\setminus i}$ is "bell-shaped"(Darroch, 1964).[17] It either has a unique mode or two consecutive modes, with the mode(s) differing from the mean $\mathbb{E}\left[S_{n\setminus i}\right]$ by at most 1; thus,

$$\begin{aligned} \Pr\left[S_{n\setminus i} = k-1\right] &< \Pr\left[S_{n\setminus i} = k\right] \text{ for } k \in \left\{1, ..., \lfloor \mathbb{E}(S_{n\setminus i}) - 1 \rfloor\right\}, & (43)\\ \Pr\left[S_{n\setminus i} = k\right] &> \Pr\left[S_{n\setminus i} = k+1\right] \text{ for } k \in \left\{\lfloor \mathbb{E}(S_{n\setminus i}) + 1 \rfloor, ..., 2n-1\right\} & (44) \end{aligned}$$

Together, (42), (43), and (44) show that replacing the variable X_i with X_i' strictly decreases $\Pr(S_n = k)$ given that $\Pr(X_i' = 1) > \Pr(X_i = 1)$. Replacing X_j with $X_{j'}$ for $j \in \{1, \ldots, 2n\} \setminus \{i\}$ iteratively, we can repeat the argument (obtaining a weak inequality each time) and finally conclude that (10) holds. The argument for (11) is analogous.

7.2 Proof of Theorem 1

We use the monotonicity property (10) and (11) to show that the rate function for any sequence $(X_i)_{i=1}^{\infty}$ can be "sandwiched" by the rate functions of sequences for which the success probabilities q_i only take finitely many values. In particular, given any

[17] So, the p.d.f. is convex-concave-convex, and, in particular, the p.d.f. is strictly increasing below the mode(s) and strictly decreasing above.

integer D, define $X_D^+ = \left(X_{i,D}^+\right)_{i=1}^\infty$ by

$$\Pr\left(X_{i,D}^+ = 1\right) = \frac{\lceil q_i D \rceil}{D}.$$

and define $\left\{X_{i,D}^-\right\}$ by

$$\Pr\left(X_{i,D}^- = 1\right) = \frac{\lfloor q_i D \rfloor}{D}.$$

This way,

$$\Pr\left(X_{i,D}^- = 1\right) \leq \Pr\left(X_i = 1\right) \leq \Pr\left(X_{i,D}^+ = 1\right) \text{ for all } i.$$

Consider any $\gamma \in (0,1)$ for which $\mathbb{E}_F(q) = \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{2n}\sum_{i=1}^{2n} X_i\right] > \gamma$. Then, there is some $\bar{n}$ such for all $n \geq \bar{n}$, it holds $\mathbb{E}\left[\sum_{i=1}^{2n} X_i\right] > \lfloor 2n\gamma \rfloor$. Hence, $\mathbb{E}\left[\sum_{i=1}^{2n} X_{i,D}^+\right] > \lfloor 2n\gamma \rfloor$ and $\mathbb{E}\left[\sum_{i=1}^{2n} X_{i,D}^-\right] > \lfloor 2n\gamma \rfloor$ for any D large enough. Now, we apply the monotonicity property (10) and (11), setting $k = \lfloor 2n\gamma \rfloor$ and $X_i' = X_{i,D}^+$ or $X_i' = X_{i,D}^-$ respectively and obtain

$$\Pr(\sum_{i=1}^{2n} X_i^- = \lfloor 2n\gamma \rfloor) \geq \Pr \sum_{i=1}^{2n} X_i = \lfloor 2n\gamma \rfloor) \geq \Pr(\sum_{i=1}^{2n} X_i^+ = \lfloor 2n\gamma \rfloor). \tag{45}$$

Our characterization in the main text derives the rate functions $c_+^{\mathrm{KL-D}}$ and $c_-^{\mathrm{KL-D}}$ of the sequences X_D^+ and X_D^+, as functions of the limits F_D^+ and F_D^- of the distribution of success probabilities. When $D \to \infty$, then F_D^+ and F_D^- converge pointwise almost everywhere to F.[18] Finally, the continuity of the expected Kullback-Leibler divergence $\int_q \mathrm{KL}(a(q), q) dF(q)$ in the measure given by a c.d.f F implies

$$\lim_{D\to\infty} c_D^{\mathrm{KL}+} = \lim_{D\to\infty} c_D^{\mathrm{KL}-}.$$

Given (45), an application of the squeeze lemma yields that the rate function of $(X_i)_{i=1}^\infty$ is

$$c^{\mathrm{KL}} = \lim_{D\to\infty} c_D^{\mathrm{KL}+}. \tag{46}$$

This finishes the proof of the theorem in this case. The proof in the case where $\mathbb{E}_F(q) = \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{2n}\sum_{i=1}^{2n} X_i\right] < \gamma$ is anologous. When $\mathbb{E}_F(q) = \gamma$, then the identity map $\mu(q) = q$ is in $B(\gamma)$, implying $c^{\mathrm{KL}} = 0$. In this case, the claim of the theorem follows since the density of the Poisson binomial $S_n = \sum_{i=1}^{2n} X_i$ is uniformly bounded

[18] Specifically, they converge at any q that is not an atom of the distribution corresponding to F.

above by $\frac{1}{2n} = e^{\ln(\frac{1}{2n})} = e^{o(n)}$ for n large.[19]

7.3 Proof of Lemma 2

We consider the case when

$$\mathbb{E}_F(q) < \mathbb{E}_{\tilde{F}}(q) \leq \gamma. \tag{47}$$

The proof in the other case is analogous.

We start by relating the distributions of F and $\tilde{F}$ with a "monotone coupling". We use the "quantile coupling", also known as "Frechet-Hoeffding coupling". For any closed intervals $U = [q_1, q_2]$ and $\tilde{U} = [\tilde{q}_1, \tilde{q}_2]$, it is given by

$$m(U \times \tilde{U}) = \lambda([F(q_1^-), F(q_2)] \cap [F(\tilde{q}_1^-), F(\tilde{q}_2)]), \tag{48}$$

where λ is the Lebesgue-Borel measure. It "matches" $q \in U$ to $\tilde{q} \in \tilde{U}$ with a likelihood proportional to the overlap in the quantiles of U and $\tilde{U}$. Since $\tilde{F}$ first-order stochastically dominates F, m is monotone. To see why, take any $\tilde{q} < q$ and any closed intervals $U = [q_1, q_2]$ and $\tilde{U} = [\tilde{q}_1, \tilde{q}_2]$ with $q \in U$ and $\tilde{q} \in \tilde{U}$ and $\tilde{q}_2 < q_1$. Since $\tilde{q}_2 < q_1$ implies $\tilde{F}(\tilde{q}_2) \leq \tilde{F}(q_1^-) \leq F(q_1^-)$, it must be that $m(U \times \tilde{U}) = 0$. Since we picked arbitrary $\tilde{q} < q$, the argument implies $m(\{(q, \tilde{q}) : \tilde{q} < q\}) = 0$. The other monotonicity condition $m(\{(q, \tilde{q}) : q < \tilde{q}\}) > 0$ follows since first-order stochastic dominance implies a closed interval $[q_1, q_2] \subseteq [0, 1]$ with $q_1 < q_2$ so that $\tilde{F}(\tilde{q}) < F(q)$ for all $q, \tilde{q} \in [q_1, q_2]$; this way, $m([q_1, q_2] \times [q_2, 1]) > 0$.

The proof now constructs a randomization $\tilde{a}$ so that

$$(\tilde{a}, m) \in R(\gamma) \text{ and } \int_{(q,\tilde{q})} \mathrm{KL}(\tilde{a}(q, \tilde{q}), \tilde{q}) dm(q, \tilde{q}) < \inf_{a \in B(\gamma)} \int_q \mathrm{KL}(a(q), q) dF(q). \tag{49}$$

For this, pick a minimizer $a^* \in \arg\inf_{a \in B(\gamma)} \int_q \mathrm{KL}(a(q), q) dF(q)$ and define

$$a_1(q, \tilde{q}) = \max(a^*(q), \tilde{q}),$$

which satisfies

[19] This can be seen as follows: Since the density of the Poisson binomial can be approximated by the density $\phi(\frac{x - \mathbb{E}(S_n)}{\mathrm{Var}(S_n)^{\frac{1}{2}}})$ where ϕ is the density of the standard normal, with the approximation error bounded by $\frac{C}{\sqrt{n}}$ for some universal $C > 0$; see Theorem 3.5 in Tang and Tang (2023) and Platonov (1980) for the primary reference.

$$\int_{(q,\tilde{q})} a_1(q,\tilde{q})dm(q,\tilde{q}) \geq \int_{(q,\tilde{q})} a^*(q)dm(q,\tilde{q}) = \int_q a^*(q)dF(q) = \gamma.$$

Since m is a coupling, $\int_{(q,\tilde{q})} \tilde{q}dm(q,\tilde{q})) = \mathrm{E}_{\tilde{F}}(\tilde{q}) < \gamma$. Thus, there is $\tilde{a}$ so that

$$(\tilde{a}, m) \in R(\gamma) \text{ and } \tilde{q} \leq \tilde{a}(q,\tilde{q}) \leq a_1(q,\tilde{q}) \text{ for all } (q,\tilde{q}). \tag{50}$$

The rest of the proof establishes an ordering of $q, a^*(q), \tilde{q}$, and $\tilde{a}(q)$ on the support of m, and translates this to a pointwise ordering of the Kullback-Leibler divergence. First, we claim that F- almost everywhere

$$q < a^*(q). \tag{51}$$

To see why, note that $\frac{\partial \mathrm{KL}(a^*(q),q)}{\partial a} = \lambda$ holds F-almost everywhere, where $\lambda \in \mathbb{R}$ is the Lagrange multiplier of the minimization problem (23). Second, $\mathrm{KL}(x,y)$ is strictly convex with $\frac{\partial \mathrm{KL}(x,y)}{\partial x} = 0$ at $x = y$. Therefore, $\mathrm{E}_F(q) < \mathrm{E}_F(a^*(q))$ implies $\lambda > 0$ and $a^*(q) > q$ almost everywhere.

Second, since either $\tilde{q} \leq a^*(q)$ or $a^*(q) < \tilde{q}$, the strict part of the monotonicity of m together with (50) and (51) implies that with strictly positive m-measure, either

$$q < \tilde{q} \leq a(q,\tilde{q}) \leq a^*(q), \text{ or} \tag{52}$$

$$q < a^*(q) < \tilde{q} = a(q,\tilde{q}). \tag{53}$$

The monotonicity of m further implies that, either (52) holds with the strict inequality $q < \tilde{q}$ replaced by the weak inequality $q \leq \tilde{q}$ m-almost everywhere or (53) holds m-almost everywhere. Hence,

$$\mathrm{KL}(a^*(q), q) < \mathrm{KL}(\tilde{a}(q,\tilde{q}), \tilde{q}) \tag{54}$$

with strictly positive m-measure and a weak inequality m-almost everyhwere. Given that the enlarged minimization program has the same solution as (23) (recall (21)), we obtain the ordering (15) claimed by the lemma for the case $\mathbb{E}_F(q) < \mathbb{E}_{\tilde{F}}(q) \leq \gamma$ considered here.

7.4 Proof of Lemma 3

Fix ω. Given the vector of induced priors $p_{(-i,-,j),n}$ of the $2n-1$ voters other than i and j, let $P(n,n-1)$ denote the probability that precisely n others are voting A and $n-1$ others are voting B; likewise, $P(n-1,n)$ is the probability that precisely $n-1$ others are voting A and n others are voting B. Then,

$$\Pr\left(\text{piv}_i|\omega;p_n\right)=P\left(n-1,n\right)q_j\left(\omega;p_{j,n}\right)+P\left(n,n-1\right)\left(1-q_j\left(\omega;p_{j,n}\right)\right) \tag{55}$$

voter i is pivotal if either j votes B and precisely n others are voting A and $n-1$ others are voting B or if j votes A and precisely $n-1$ others are voting A and n others are voting B. The analogous formula holds for the pivotal likelihood of voter j.

In what follows, we repeatedly use that for any four positive numbers a,b,c,d and $\gamma\in[0,1]$, the inequality $\max\left(\frac{a}{b},\frac{c}{d}\right)\geq\frac{(1-\gamma)a+\gamma c}{(1-\gamma)b+\gamma d}\geq\min\left(\frac{a}{b},\frac{c}{d}\right)$ holds, with strict inequalities if the maximum and the minimum do not coincide and $\gamma\notin\{0,1\}$.

Combining this fact with (55) yields

$$\begin{aligned}
&\left(\frac{\Pr\left(\text{piv}_i|\alpha;p_n\right)}{\Pr\left(\text{piv}_j|\alpha;p_n\right)}\right)\Big/\left(\frac{\Pr\left(\text{piv}_i|\beta;p_n\right)}{\Pr\left(\text{piv}_j|\beta;p_n\right)}\right)\\
\leq\quad&\max\left\{\frac{1-\hat{q}_i\left(\alpha;p_{i,n}\right)}{1-\hat{q}_j\left(\alpha;p_{j,n}\right)},\frac{\hat{q}_i\left(\alpha;p_{i,n}\right)}{\hat{q}_j\left(\alpha;p_{j,n}\right)}\right\}\Big/\min\left\{\frac{1-\hat{q}_i\left(\beta;p_{i,n}\right)}{1-\hat{q}_j\left(\beta;p_{j,n}\right)},\frac{\hat{q}_i\left(\beta;p_{i,n}\right)}{\hat{q}_j\left(\beta;p_{j,n}\right)}\right\}.
\end{aligned} \tag{56}$$

Another application of the fact yields

$$\frac{\hat{q}_{i'}\left(\alpha;p_{i',n}\right)}{\hat{q}_{i'}\left(\beta;p_{i',n}\right)}=\frac{\sum_{s\in S_{i'}}\Pr(s_{i'}=s|\alpha)\sigma_i(s)}{\sum_{s\in S_{i'}}\Pr(s_{i'}=s|\beta)\sigma_i(s)}<\max_{s\in S_i}\left(\frac{\Pr(s_{i'}=s|\alpha)}{\Pr(s_{i'}=s|\beta)}\right) \tag{57}$$

and likewise

$$\frac{\hat{q}_{i'}\left(\alpha;p_{i',n}\right)}{1-\hat{q}_{i'}\left(\beta;p_{i',n}\right)}<\max_{s\in S_i}\left(\frac{\Pr(s_{i'}=s|\alpha)}{1-\Pr(s_{i'}=s|\beta)}\right), \tag{58}$$

$$\frac{1-\hat{q}_{i'}\left(\alpha;p_{i',n}\right)}{\hat{q}_{i'}\left(\beta;p_{i',n}\right)}<\max_{s\in S_i}\left(\frac{1-\Pr(s_{i'}=s|\alpha)}{\Pr(s_{i'}=s|\beta)}\right), \tag{59}$$

for any $i'=i,j$ and $\sigma_i(s)=\Phi_i\left(\frac{p_{i,n}\Pr_i(s_i=s|\alpha)}{p_{i,n}\Pr_i(s_i=s|\alpha)+(1-p_{i,n})\Pr_i(s_i=s|\beta)}\right)$. The strictness of the inequality (57) here stems from the uniform bound on the likelihood ratio of the signals, (27), and the lower bound on the share of the partisans for each alternative, (25). The upper bounds (56) - (59) jointly prove the upper bound claimed by the lemma. A parallel argument establishes the lower bound.

7.5 Proof of Theorem 5

Consider any sequence of equilibria $(\sigma_n^*)_{n=1}^{\infty}$. Let $(p_n^*)_{n=1}^{\infty}$ be the corresponding sequence of equilibrium induced prior vectors; for each i and n,

$$p_{i,n}^* = \Pr\left(\alpha|\text{piv}_i; \sigma_{-i,n}^*, n\right).$$

Suppose that information aggregation fails with positive probability given some subsequence (which we identify with the original subsequence for simplicity). Denoting by $X_i(\omega)$ the Bernoulli variable with success probability $q_i(\omega; p_{i,n}^*)$ and by $S_{2n+1}(\omega) = \sum_{i=1}^{2n+1} X_i(\omega)$ their sum across i, this means,

$$\lim_{n\to\infty} \Pr(\alpha|n) \Pr(S_{2n+1}(\alpha) < n) + \Pr(\beta|n) \Pr(S_{2n+1}(\beta) > n) > 0. \tag{60}$$

Without loss, consider the case where

$$\lim_{n\to\infty} \Pr(\beta|n) > 0 \tag{61}$$

and

$$\lim_{n\to\infty} \Pr(S_{2n+1}(\beta) > n) > 0. \tag{62}$$

We now derive a contradiction to (60) in six steps.

Step 1 $\mathrm{E}(\sum_{i=1}^{2n} q_i(\omega; p_{i,n}^*)) > n$ *for n sufficiently large and both states ω.*

Kolmogorov's law of large numbers implies that the realized share of votes converges in distribution to the expected share of votes.[20] Thus, for $\omega = \beta$,

[20] See Theorem 5.8 in McDonald and Weiss (2004).

$$\begin{aligned}\lim_{n\to\infty} \Pr(\frac{\mathbb{E}(S_{2n+1}(\beta))}{2n+1} > \frac{1}{2}+\varepsilon) = & \quad 1-\lim_{n\to\infty} \Pr(\frac{\mathbb{E}(S_{2n+1}(\beta))}{2n+1} \leq \frac{1}{2}+\varepsilon) \\ = & \quad 1-\lim_{n\to\infty} \Pr(\frac{S_{2n+1}(\beta)}{2n+1} \leq \frac{1}{2}+\varepsilon) \\ = & \quad \lim_{n\to\infty} \Pr(\frac{S_{2n+1}(\beta)}{2n+1} > \frac{1}{2}+\varepsilon) > 0,\end{aligned}$$

where the last inequality is implied by (62) for all $\varepsilon > 0$ for which $(2n+1)(\frac{1}{2}+\varepsilon) < n+1$. Since $\mathbb{E}(S_{2n+1}(\beta))$ is deterministic, this implies that for all n sufficiently large,

$$\frac{\mathbb{E}(S_{2n+1}(\beta))}{2n+1} > \frac{1}{2}+\varepsilon \text{ for all } \varepsilon \text{ with } (2n+1)(\frac{1}{2}+\varepsilon) < n+1$$

and hence

$$\mathbb{E}(S_{2n+1}(\beta)) \geq n+\varepsilon' \text{ for all } \frac{1}{2} < \varepsilon' < 1$$

Since $q_{2n+1}(\omega, p^*_{i,n}) \leq \Phi_i(1^-) < 1$,

$$\mathbb{E}(S_{2n}(\beta)) > n.$$

The claim for $\omega = \alpha$ follows since

$$\hat{q}_i(\beta; p^*_{i,n}) \leq \hat{q}_i(\alpha, p^*_{i,n}) \text{ for all } i \text{ and all } n; \tag{63}$$

compare to Step 1 in Section 4.[21]

Step 2 *There is $\bar{n} \in \mathbb{N}$ so that*

$$\Pr(\alpha|\text{piv}_{i,n}, n) \leq \Pr(\alpha|n) \text{ for all } i \text{ and } n \geq \bar{n}.$$

Given Step 1 and (63), the proof of Lemma 1 yields the step's claim.

Step 3 *For all $d > 0$,*

$$\lim_{n\to\infty} \frac{\left|\left\{i \leq 2n+1 : |\hat{q}_i(\alpha, p^*_{i,n}) - \hat{q}_i(\beta, p^*_{i,n})| < d\right\}\right|}{2n+1} = 1$$

[21] This observation does not require that signals are boundedly informative but only that Φ_i are strictly increasing.

Otherwise, there are $d > 0$ and $x > 0$ and a subsequence $a(n)$ so that

$$\frac{\left|\{i \leq 2n+1 : |\hat{q}_i(\alpha, p^*_{i,n}) - \hat{q}_i(\beta, p^*_{i,n})| \geq d\right\}|}{2n+1} > x \tag{64}$$

for all $a(n)$. We claim this implies a subsequence $b(n)$ of $a(n)$ for which information aggregates in β, i.e.,

$$\lim_{n\to\infty} \Pr(S_{2b(n)+1}(\beta) > b(n)) = 0, \tag{65}$$

which implies a contradiction to the initial assumption that it does not, (60).

Applying Helly's selection theorem, we pick a subsequence $b(n)$ along which F_ω converges pointwise for both states ω to some c.d.f. This implies convergence of the means to $\mathbb{E}_{F_\omega}(q)$. Then, (63), (64), and (39) together imply

$$\mathbb{E}_{F_\alpha}(q) > \mathbb{E}_{F_\beta}(q). \tag{66}$$

Finally, similar to the proof of Step 4 in Theorem 3's proof, we show that (66) necessitates information aggregation in β: Suppose that $\mathbb{E}_{F_\beta}(q) \geq \frac{1}{2}$. The observation (66) together with (63) implies that F_α first-order stochastically dominates F_β. We see that the conditions of Lemma 2 are satisfied for $F = F_\alpha$ and $\tilde{F} = F_\beta$. Lemma 2 and Theorem 2 then imply

$$\lim_{n\to\infty} \frac{\Pr\left(\text{piv}_i|\alpha; p^*_{i,n}\right)}{\Pr\left(\text{piv}_i|\beta; p^*_{i,n}\right)} = 0 \text{ for all } i;$$

This, however implies that, for all i, $\hat{q}_i(\beta; p^*_{i,n}) \to \Phi_i(0)$ as $n \to \infty$ and thus $E_{F_\beta}(q) \leq \frac{1}{2} - \varepsilon$, given (24), which yields a contradiction to the initial assumption. We conclude that $\mathbb{E}_{F_\beta}(q) < \frac{1}{2}$. But then applying Kolgomorov's strong law of large numbers shows that information aggregates in β.

Step 4 *For all $d > 0$ and $\omega \in \{\alpha, \beta\}$,*

$$\lim_{n\to\infty} \frac{\left|\{i \leq 2n+1 : \min\left(\hat{q}_i(\omega, p^*_{i,n}) - \Phi_i(0), \Phi_i(1^-) - \hat{q}_i(\omega, p^*_{i,n})\right) \leq d\right\}|}{2n+1} = 1 \tag{67}$$

Suppose not. Then, there are $\omega \in \{\alpha, \beta\}$, $d > 0$, $x > 0$, and a subsequence $a(n)$ so that for all n, there is a share x of agents $i \leq 2a(n) + 1$ with $q(\omega; p^*_{i,a(n)}) \in (\Phi_i(0) + d, \Phi_i(1^-) - d)$.

Given Step 3, there are $0 < d' < d$ and $0 < x' < x$ so that, for all n, there is a share x of agents $i \leq 2a(n) + 1$ with

$$q(\omega; p^*_{i,a(n)}) \in (\Phi_i(0) + d, \Phi_i(1^-) - d) \text{ for } \textit{all } \omega \in \{\alpha, \beta\}.$$

Now we claim that this implies $d_0 > 0$ so that, uniformly, for all i and n with $q(\omega; p^*_{i,a(n)})(\Phi_i(0) + d, \Phi_i(1^-) - d)$ for $\omega \in \{\alpha, \beta\}$, it holds

$$\Pr(\alpha|\text{piv}_{i,a(n)}, n) \in (d_0, 1 - d_0). \tag{68}$$

Note that, for any $d_1 > 0$ sufficiently small,

$$\Pr(\alpha|\text{piv}_{i,a(n)}, n) > 1 - d_1 \tag{69}$$

implies numbers $d_2 > 0, d_3 > 0$ so that $\Pr(\alpha|s_i, \text{piv}_{i,a(n)}, n) > 1 - d_2$ with probability larger than $1 - d_3$ in α. This follows from Bayesian updating, independently of i and n: Only arbitrarily strong signals for β could balance out the arbitrarily strong posterior conditional on being pivotal, (69), but these can only be sent with vanishing probability in α. These numbers can be chosen to satisfy $d_2, d_3 \to 1$ as $d_1 \to 1$, implying that, for d_1 exceeding some sufficiently large upper bounder $\bar{d}_1$, the property $q(\omega; p^*_{i,a(n)}) \leq \Phi_i(1^-) - d$ would be violated, given (31).

Note that this upper bound $\bar{d}_1$ for d_1 can be chosen uniformly across i and n given the uniform upper Lipschitz bound of Φ_i, (28). The analogous argument yields a uniform lower bound $\Pr(\alpha|\text{piv}_{i,a(n)}, n) \geq \bar{d}_2$. We set $d_0 = \min(\bar{d}_1, \bar{d}_2)$ to establish (68).

Finally, given the uniform bounds on the likelihood ratio of the signals, (27), and the uniform lower Lipschitz bound for Φ_i, (28), we see that there is $d'' > 0$ so that for all n, there is a share $x' > 0$ of agents $i \leq 2a(n) + 1$ with $|q(\alpha; p^*_{i,a(n)}) - q(\beta; p^*_{i,a(n)})| > d''$. This yields a contradiction to Step 3, finishing the proof by contradiction of this step's claim (67).

Step 5 *For all $d > 0$ and $\omega \in \{\alpha, \beta\}$,*

$$\lim_{n\to\infty} \frac{\left|\{i \leq 2n + 1 : \hat{q}_i(\beta, p^*_{i,n}) - \Phi_i(0) \leq d\}\right|}{2n + 1} = 1$$

Step 2 states that the pivotal inference is in favor of β for n large enough and all i. Since the prior likelihood of β is bounded above 0, (61), only after an arbitrarily strong signal for α the voting probability can be arbitrarily close to $\Phi_i(1^-)$, but these can only be received with vanishing probability in β. So, averaging over all signals, the voting probability $\hat{q}_i(\beta, p^*_{i,n})$ cannot be close to $\Phi_i(1^-)$ for any i when n is sufficiently large. Step 4 then implies the step's claim.

Step 6 $\lim_{n\to\infty} \Pr(S_{2n+1}(\beta) > n) = 0$

Step 5 implies that the limit expected vote share in β is $\Phi_i(0)$ for each i, so that $\mathrm{E}_{F_\beta}(q) = \lim_{n\to\infty} \frac{1}{2n+1} \sum_{i=1}^{2n+1} \Phi_i(0)$. Since $\lim_{n\to\infty} \frac{1}{2n+1} \sum_{i=1}^{2n+1} \Phi_i(0) < \frac{1}{2}$, given (24), an application of Kolmorogov's strong law of large numbers shows the step's claim.

Step 6 contradicts the initial assumption that A is chosen in β, (62). An analogous argument derives a contradiction if B is chosen in α. This finishes the proof of Theorem 5.

7.6 Relation to the Gärtner-Ellis Theorem: Interval vs Point Probabilities

We start by stating a version of the Gärtner-Ellis theorem for sequences of real-valued random variables (Dembo, 2009).

Gärtner-Ellis Theorem. *(Gärtner (1977); Ellis (1984)) Suppose that $(Y_n)_{n\in\mathbb{N}}$ is a sequence of real-valued random variables such that*

$$\Lambda(t) = \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}\left[e^{tnY_n}\right] \tag{70}$$

exists for all $t \in \mathbb{R}$. If Λ is differentiable, then, for all $\gamma \in (0,1)$, it holds that

$$\lim_{n\to\infty} \Pr(Y_n \leq \frac{\lfloor \gamma 2n \rfloor}{n}) = e^{-nc^{\mathrm{FL}}(\gamma)+o(n)}$$

for

$$c^{\mathrm{FL}}(\gamma) = \inf_{x\in[0,\gamma]} \Lambda^*(x) \quad \text{and } \Lambda^*(x) = \sup_{t\in\mathbb{R}} (xt - \Lambda(t)).$$

The function Λ is called the "cumulant generating function" and Λ^* is called the "Fenchel-Legendre transform" of Λ.

Now, we apply the theorem to our setting. Consider a sequence of independent Bernoulli random variables $\{X_i\}_{i=1}^{\infty}$ with $\Pr(X_i = 1) =: q_i \in [\varepsilon, 1-\varepsilon]$ for some $\varepsilon >$

0. As before, suppose that the cumulative distribution function F_n of the first $2n$ success probabilities q_i converges almost surely to a c.d.f. F. Let $S_n = \sum_{i=1}^{2n} X_i$ and $Y_{2n} = \frac{1}{2n} S_n$. Given the convergence of F_n to F, the limit Λ exists for $Y_{2n} = \frac{1}{2n} S_n$ and is differentiable in t. Application of the theorem to $(Y_{2n})_{n \in \mathbb{N}}$ yields

$$\Pr\left(S_n \leq \lfloor \gamma 2n \rfloor\right) = e^{-2nc^{FL}(\gamma)+o(n)}. \tag{71}$$

In what follows, we connect the point probability $\Pr\left(S_n = \lfloor \gamma 2n \rfloor\right)$ to the interval probability $\Pr\left(S_n \leq \lfloor \gamma 2n \rfloor\right)$, by using the properties of the Poisson Binomial distribution. For example, consider the case where

$$\lim_{n \to \infty} \mathbb{E}\left[\frac{S_n}{2n}\right] > \gamma.$$

Then, Darroch (1964)'s result about the p.d.f. of the Poisson Binomial, (12), implies

$$\Pr\left(S_n = \lfloor x 2n \rfloor\right) \leq \Pr\left(S_n = \lfloor \gamma 2n \rfloor\right) \text{ for all } x \in [0, \gamma],$$

and thus

$$\begin{aligned}\Pr\left(S_n \leq \lfloor \gamma 2n \rfloor\right) = \sum_{r=1}^{\lfloor \gamma 2n \rfloor} \Pr\left(S_n = r\right) \leq \sum_{r=1}^{\lfloor \gamma 2n \rfloor} \Pr\left(S_n = \lfloor \gamma 2n \rfloor\right) = 2n \Pr\left(S_n = \lfloor \gamma 2n \rfloor\right) \\ \Rightarrow \frac{1}{2n} \Pr\left(S_n \leq \lfloor \gamma 2n \rfloor\right) \leq \Pr\left(S_n = \lfloor \gamma 2n \rfloor\right)\end{aligned}$$

Finally, since $2n = e^{\ln(2n)} = e^{o(n)}$, the interval probability formula (71) implies $\Pr\left(S_n = \lfloor \gamma 2n \rfloor\right) \geq e^{-2nc^{FL}(\gamma)+o(n)}$ and therefore also the equality

$$\Pr\left(S_n = \lfloor \gamma 2n \rfloor\right) = e^{-2nc^{FL}(\gamma)+o(n)}.$$

Comparison with our Theorem 1 implies an identity of the minimizing expected Fenchel-Legendre transform c^{FL} and the minimizing expected Kullback-Leibler divergence c^{KL}

$$c^{\mathrm{FL}}(\gamma) = c^{\mathrm{KL}}(\gamma) \text{ for all } \gamma \in (0,1). \tag{72}$$

This identity is discussed in more generality in Lemma 6.2.13 in Dembo (2009).

References

Acharya, A. (2016): "Information aggregation failure in a model of social mobility," *Games and Economic Behavior*, 100, 257–272.

Ali, S. N., M. Mihm, and L. Siga (2024 (forthcoming)): "The Political Economy of Zero-Sum Thinking," *Econometrica*.

Barelli, P., S. Bhattacharya, and L. Siga (2022): "Full information equivalence in large elections," *Econometrica*, 90(5), 2161–2185.

Bhattacharya, S. (2013): "Preference monotonicity and information aggregation in elections," *Econometrica*, 81(3), 1229–1247.

——— (2018): "Condorcet Jury theorem in a spatial model of elections," Working paper.

Cramér, H. (1938): "On a new limit theorem in probability theory (Translation of 'Sur un nouveau théoreme-limite de la théorie des probabilités')," *arXiv preprint arXiv:1802.05988*.

Darroch, J. N. (1964): "On the distribution of the number of successes in independent trials," *The Annals of Mathematical Statistics*, 35(3), 1317–1321.

DellaVigna, S., and E. Kaplan (2007): "The Fox News effect: Media bias and voting," *The Quarterly Journal of Economics*, 122(3), 1187–1234.

Dembo, A. (2009): *Large deviations techniques and applications*. Springer.

Ekmekci, M., and S. Lauermann (2020): "Manipulated electorates and information aggregation," *The Review of Economic Studies*, 87(2), 997–1033.

Ellis, R. S. (1984): "Large deviations for a general class of random vectors," *The Annals of Probability*, 12(1), 1–12.

Escher, F. (1932): "On the probability function in the collective theory of risk," *Skand. Aktuarie Tidskr.*, 15, 175–195.

Feddersen, T., and W. Pesendorfer (1997): "Voting behavior and information aggregation in elections with private information," *Econometrica*, pp. 1029–1058.

——— (1998): “Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting,” *American Political science review*, 92(1), 23–35.

Feddersen, T. J., and W. Pesendorfer (1996): “The swing voter’s curse,” *The American Economic Review*, pp. 408–424.

Fernandez, R., and D. Rodrik (1991): “Resistance to reform: Status quo bias in the presence of individual-specific uncertainty,” *The American Economic Review*, pp. 1146–1155.

Gärtner, J. (1977): “On large deviations from the invariant measure,” *Theory of Probability & Its Applications*, 22(1), 24–39.

Helly, E. (1912): *Ueber lineare Funktionaloperationen*. Kaiserlich-koniglichen Hof und Staatsdruckerei.

Hese, C., and S. Lauermann (2024): “Persuasion in Elections when Voting is Costly,” working paper.

Kamenica, E., and M. Gentzkow (2011): “Bayesian persuasion,” *The American Economic Review*, 101(6), 2590–2615.

Kim, J., and M. Fey (2007): “The swing voter’s curse with adversarial preferences,” *Journal of Economic Theory*, 135(1), 236–252.

Klenke, A. (2020): *Probability theory: a comprehensive course*. Springer Science & Business Media, 3 edn.

Kosterina, S. (2023): “Information structures and information aggregation in threshold equilibria in elections,” *Economic Theory*, 75, 493–522.

Ladha, K., G. Miller, and J. Oppenheimer (1996): “Information aggregation by majority rule: Theory and experiments,” Working paper.

Mandler, M. (2012): “The fragility of information aggregation in large elections,” *Games and Economic Behavior*, 74(1), 257–268.

McDonald, J. N., and N. A. Weiss (2004): “A course in real analysis,” .

McMurray, J. C. (2013): "Aggregating information by voting: The wisdom of the experts versus the wisdom of the masses," *Review of Economic Studies*, 80(1), 277–312.

Natanson, I. (1961): *Theory of functions of a real variable, Volume I.* Dover Publications, Inc. USA.

Nickerson, D. W., and T. Rogers (2014): "Political campaigns and big data," *Journal of Economic Perspectives*, 28(2), 51–74.

Platonov, M. (1980): "Combinatorial numbers of a class of mappings and their applications," *Nauka, Moscow [Russian].*

Rachev, S. T., and L. Rüschendorf (2006): *Mass Transportation Problems: Volume 1: Theory.* Springer Science & Business Media.

Razin, R. (2003): "Signaling and election motivations in a voting model with common values and responsive candidates," *Econometrica*, 71(4), 1083–1119.

Strömberg, D. (2004): "Radio's impact on public spending," *The Quarterly Journal of Economics*, 119(1), 189–221.

Tang, W., and F. Tang (2023): "The Poisson binomial distribution—Old & new," *Statistical Science*, 38(1), 108–119.

Yermack, D. (2010): "Shareholder voting and corporate governance," *Annu. Rev. Financ. Econ.*, 2(1), 103–125.