# Condorcet's Jury Theorem without Symmetry

Carl Heese Stephan Lauermann

October 4, 2024

In the classic Condorcet setting, agents choose between two alternatives $A$ or $B$ with a simple majority vote. Voters have potentially heterogeneous private preferences that depend on an unknown state, $\alpha$ or $\beta$, in a monotone way: A random voter type is more likely to prefer $A$ when he considers $\alpha$ more likely; see Bhattacharya (2013). Each voter holds a private signal about the state. The Condorcet jury theorem states that the outcome of strategic voting is the same as under full information about the state as the number of voters grows large ("information aggregation"), provided that the private preferences are i.i.d across voters, the private signals i.i.d. conditional on the state, and when considering any sequence of non-trivial symmetric equilibria. We show that the conclusion does not require signals and preferences to be identically distributed and also not the restriction to symmetric equilibrium.

The "modern" Condorcet jury theorem rests on symmetry (Bhattacharya, 2013; Feddersen and Pesendorfer, 1997, 1998). First, the environment is symmetric: Voters are assumed to be ex-ante identical, drawing preferences and signals from the same distributions independently conditional on the state. In addition to the symmetric environment, voters are assumed to behave symmetrically, following the same strategy.

Ex-ante *asymmetry* of voter preferences and information is natural in many situations. It captures the typical feature that the voting body consists of multiple "interest groups". Preferences and also information differ in distribution across interest groups. Often, two groups hold opposed preferences, such as in distributive politics, e.g., in the context of education or trade reforms; see Fernandez and Rodrik (1991); Kim and Fey (2007); Acharya (2016); Ali, Mihm, and Siga (2018). Similarly, public news may coordinate voter behavior in an asymmetric way; in particular, the literature on "public persuasion" of voters has studied how to exploit such asymmetric coordination; see, e.g., Alonso and Câmara (2016).

The symmetry assumptions are not innocent. In the simple symmetric environment of Feddersen and Pesendorfer (1998) in which voters have the same preference type, there are asymmetric equilibria of the majority voting game for which information aggregation fails. We provide an example of such an equilibrium in the next section.

Our main result is to show that such failures of information aggregation cannot happen in any non-trivial and possibly asymmetric equilibrium when each voter may be a partisan voter for either alternative, even if just with minimal probability.

We prove such "Condorcet jury theorem without symmetry" (Theorem 1 and 4) within the setting of next section's example modified with partisans, and we extend the result to the canonical setting by Bhattacharya (2013) with heterogeneous preferences modified to include ex-ante asymmetry:

As in Bhattacharya (2013)'s setting, a majority election decides over two possible alternatives—$A$ and $B$. Voters' preferences over alternatives are private, heterogeneous, and depend on an unknown state, $\alpha$ or $\beta$. Some voters may prefer $A$ in state $\alpha$ and $B$ in $\beta$, with heterogeneous "thresholds of doubt" when they are uncertain about the state, while others may be "partisans" who prefer one alternative or the other, independently of the state. All voters privately receive a private noisy signal about the state. Preferences are independent across voters and signals independent conditional on the state.[1] There is a finite set of $2n + 1$ voters, and, here, we allow them to be ex-ante asymmetric: Each voter has a distribution from which his private signal and his private preference type are drawn. Distributions can be non-identical across voters.

A central technical difficulty of extending the CJT to asymmetric settings is that asymmetries do not wash out in large elections. When voters behave asymmetrically, then even with an arbitrarily large number of voters, two different voters can make substantially different inferences from being "pivotal" for the election outcome. Thus, the voting incentives can also differ substantially for any two voters.

The underlying cause is related to the so-called "swing voters' curse" (Feddersen

[1] Heese and Lauermann (2019) show that lifting the standard independence assumption for the signal distributions is not innocuous, so we maintain it, and also the independence assumption for the preference distributions.

and Pesendorfer, 1996). The "swing voters' curse" emerges from the fact that when the vote count for an alternative increases even by just one vote, the posterior probability of that alternative increases strictly, and this effect does not vanish when the number of voters grows large, $n \to \infty$.[2]

We explain how the failure of information aggregation in our example relies on the voter's different posteriors. Critically, when there are partisans, they imply a bound on the difference (Lemma 3) violated in the example equilibrium. This way, the logic of the equilibrium in which aggregation fails will break down. This bound will be a key ingredient in proving the "Condorcet jury theorem without symmetry" in general.

Our proof relies on a set of technical results. To establish these, we leverage statistical tools, particularly from the theories of large deviations and stochastic dominance. We present a characterization of the probability of a tie for a sequence of independent but not identically distributed Bernoulli random variables tailored to the voting setting (Theorem 1). Then, we establish comparative static results for this probability (Lemma 1 and 2), using an instance of Strassen's theorem on stochastic dominance as well as properties of the Poisson binomial distribution (Darroch, 1964). The technical results may be of independent interest to voting theorists.

The remaining paper is structured as follows: Section 1 presents the example. Section 2 presents the technical results. Section 3 starts with the model based on Bhattacharya (2013), characterizes the best response of voters in terms 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 the "Condorcet jury theorem without symmetry." Section 5 concludes.

# 1 Failure of the CJT with Asymmetric Equilibria

Consider the following symmetric environment from Feddersen and Pesendorfer (1998): There are $2n + 1$ voters who choose between two alternatives, $A$ and $B$. There are two states, $\omega \in \{\alpha, \beta\}$, that are equally likely ex-ante. Voters have pure common values: Each voter obtains a payoff of 1 when the state matches the outcome and 0

[2] 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 not on others—such as a referendum on a constitutional change.

otherwise. Finally, 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

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

The voting game is as follows. The voters simultaneously vote for $A$ or $B$; 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 $\sigma_{-i}$, 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 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 $\sigma^*$ with*

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

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.

We now give a counter-example showing that the Condorcet Jury theorem relies on the restriction to symmetric equilibrium in the above pure common values setting:[3]

**Example.** *There exists a sequence $(\sigma^*)_{n\in\mathbb{N}}$ of asymmetric equilibria for which*

$$\lim_{n\to\infty} \Pr(A \textit{ is elected}|\alpha) = \lim_{n\to\infty} \Pr(B \textit{ is elected } |\beta) = M < 1.$$

The explanation below shows that $M$ as low as $M = [r^3 + 3r^2(1-r)] > 1/2$ is possible in some equilibrium sequence.

Take any voter number $2n+1 > 3$ and split the electorate into 3 experts (the voters $i = 1, 2, 3$) and $2n-2$ non-experts. We show that there is an asymmetric equilibrium in which the experts vote sincerely, but the non-experts behave as follows: Every even-numbered non-expert $i \in \{4, 6, ..., 2n\}$ votes $A$ and every odd-numbered non-expert $i \in \{5, 7, ..., 2n+1\}$ votes for $B$.

[3] The example is from Justus Preusser.

The equilibrium logic is simple. The votes of the non-experts cancel each other. So, an expert is pivotal if one of the two other experts votes $A$ and the other votes $B$, which means that the other experts hold one $a$-signal and one $b$-signal. This makes sincere voting a mutual best response for the experts: Given the symmetric prior and symmetric signal distribution, the posterior probability conditional on the voter being pivotal and conditional on her signal is above the prior 0.5 when her signal is $a$ and below if it is $b$.

For the non-experts, this is different: For example, voter $i = 4$, who is supposed to vote for $A$, is pivotal with an $A$-vote if the experts received 2 $a$-signals and 1 $b$-signal. The pivotal event contains no information about the signals of the remaining voters. The posterior probability conditional on the voter being pivotal and herself having signal $b$ is equal to the prior, 0.5. Thus, conditional on being pivotal with a $b$-signal, the voter is indifferent and willing to vote $A$.

Note that this argument constructs an asymmetric equilibrium for *each* $n$. For each $n$, the full-information outcome is elected in $\alpha$ (in $\beta$) only if at least 2 of the 3 experts receive an $a$-signal (a $b$-signal). This happens with probability $M = [r^3 + 3r^2(1-r)]$. Thus, information does not aggregate as $n \to \infty$. We summarize: The conclusion of the Condorcet jury theorem does not hold for asymmetric equilibria.

**Partisans and Updating Differences.** The example features substantial differences in the non-expert's posteriors conditional on being pivotal; a vote for $A$ is pivotal when there are more $a$-signals among the experts, and a vote for $B$ is pivotal when there are more $b$-signals among them. Thus, a non-expert is more likely to be pivotal with any given voting choice when this choice is in her interest.[4] These posterior differences rationalize some voters voting $A$ deterministically and others with the same preferences simultaneously voting $B$ deterministically.

Our main result will show that equilibria where information aggregation fails no longer exist when there is "noise" in the form of partisan voters. That is, each voter's preferences are such that she prefers $A$ in both states or $B$ in both states, both with strictly positive probability $\varepsilon > 0$. Section 3.3 shows formally that the possibility of being a partisan implies a certain bound on the difference in any two

[4]The opposite phenomenon is called the swing voter's curse; in analogy, one may term our observation "swing-voter's blessing."

voters' posteriors. In the context of the example–modified by assuming that each voter is a partisan for either alternative with a positive probability and otherwise a non-partisan as before—the bound implies that it is not possible that, at the same time, some non-partisans vote $A$ deterministically while others do so for $B$. Thus the equilibrium logic of the example breaks down.

We close this section with some remarks related to the example: First, note that the voter's best equilibrium often has a similar structure in related environments. 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 symmetric equilibrium yields a strictly lower payoff than the voter's best asymmetric equilibrium. The voter's best asymmetric equilibrium features a deterministic set of voters who vote for $B$ independently of their signal, which may be interpreted as an effort to "de-bias" the voting rule; see Ladha, Miller, and Oppenheimer (1996).[5] Second, in the literature, one often considers responsive equilibria where each the voter has a strictly positive probability of voting either way, unlike in our example equilibrium. However, with asymmetric signals and, especially, asymmetric preferences, this requirement is too stringent. One can easily find example settings in which no responsive equilibrium exists. For example, consider an asymmetric environment where signals have bounded precision, and there are some continuously distributed "thresholds of doubts" (as in Section 3 below). Then, in the natural equilibria, voters with thresholds that are known to be sufficiently close to 0 and 1 will vote for their preferred alternative. Third, the example is easily modified so that incentives are strict.[6]

[5]Indeed, these authors report that participants in a voting experiment utilize asymmetric strategies, yielding payoffs above the theoretical maximum across symmetric equilibria; see Ladha, Miller, and Oppenheimer (1996).

[6]Suppose there are 2 voters whose signals have high precision, with

$$\Pr(s_i = a|\alpha) = \Pr(s_i = b|\beta) = r_H > 0.5 \text{ for } i \in \{1, 2, 3\}$$

and the remaining voters $2(n-2)$ voters have low precision

$$\Pr(s_i = a|\alpha) = \Pr(s_i = b|\beta) = r_L > 0.5 \text{ for } i \in \{4, ..., 2n+1\}$$

with

$$0.5 < r_L < r_H < 1.$$

# 2 Large Deviation Theory for a Voting Appplication

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 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 $\gamma n$ 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[7]

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

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 and Touchette (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[8]

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

and that[9]

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

For a sequence of independent Bernoulli variables with distinct success probabili-

[7]Recall that a function $f$ is in $o(n)$ if $\frac{f(n)}{n}$ vanishes to 0 for $n \to \infty$.

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

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

ties $q_i$, 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 $\gamma$,

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

**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.* $F$ *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 $F$ is a step-function, i.e. the corresponding distribution has finite support, $\{p_1, ..., p_D\}$, 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.

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

To evaluate

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

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;

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

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

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

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 $c^{\mathrm{KL-D}}\left(\gamma\right) = \min_{m':\sum_{d=1}^{D} m'_d = \lfloor 2n\gamma \rfloor} \sum_{d=1}^{D} \eta_d \mathrm{KL}_B\left(\frac{m'_d}{n_d}, p_d\right)$. 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.[10]

[10] Note that the minimizing vector $c^{\mathrm{KL-D}}(\gamma)$ 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 in-

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 $F$. ∎

## 2.2 Comparative Statics

**Lemma 1** *Consider two sequences $(X_i)_{i=1}^{\infty}$ and $(X_i')_{i=1}^{\infty}$ of independent Bernoulli random variables with*

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

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

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

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

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."[11] 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,

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

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

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".

fimum $c^{\text{KL}}(\gamma)$ 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 $[0,1]^D$.

[11] 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.

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 $\mathbb{E}_F(q)$ 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.[12] A coupling $v$ is a joint measure that preserves the marginals: For any Lebesgue measurable $U, \tilde{U} \subseteq [0,1]$,[13]

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

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

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 $\tilde{q} \in \mathrm{supp}(F)$.

[12]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

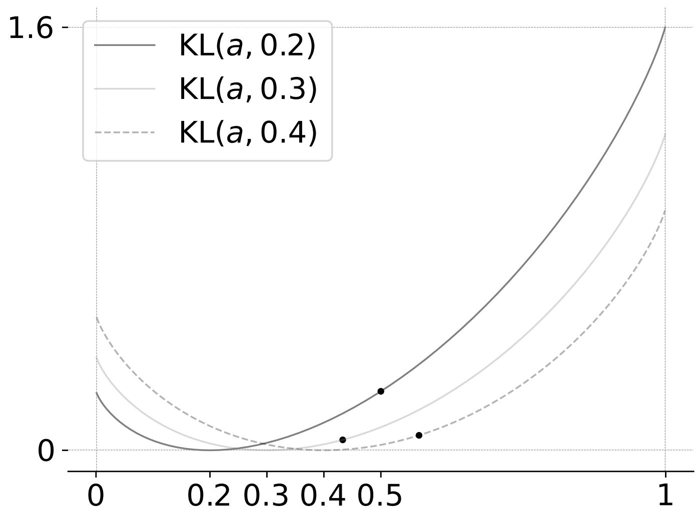


Figure 1: An example where $\gamma = \frac{1}{2}$, $\text{supp}(F) = \{0.2\}$, and $\tilde{F}$ is distributed uniformly on its support $\text{supp}(\tilde{F}) = \{0.3, 0.4\}$. The minimizer of (4) for $F$ is given by $a_F(0.2) = \frac{1}{2}$ and the minimizer for $\tilde{F}$ is given by $a_{\tilde{F}}(0.3) = 1 - a_{\tilde{F}}(0.4) = 0.43303$. We indicate the Kullback-Leibler divergence of the minimizers with the three dots.

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 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 de-

dominance is equivalent to the existence of a monotone coupling defined via (18).

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

terministic 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 (4) 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.

# 3 Condorcet's Jury Theorem without Symmetry

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 $S_i \subset \mathbb{R}$ 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 $\alpha$ is above $y$. Voters with thresholds of doubt $y \in \{0, 1\}$ are "partisans" who prefer $A$ and $B$, respectively, no matter their beliefs.[14] The signal's distribution is given by $\left(\Pr_i(s_i = s|\omega)_{\omega \in \{\alpha,\beta\}}\right)_{s \in S_i}$, has c.d.f $\Psi_{i,\omega}$ in $\omega$, and, conditional on the state, is independent of the preference

[14] 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 $\alpha$ and $-\hat{y}$ in $\beta$ and the payoff from $B$ is 0 in both states. With this specification, a voter prefers $A$ whenever she believes the probability of $\alpha$ 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 $y = \max\{\min\{\hat{y}, 1\}, 0\}$.

type's distribution, which has a continuous and strictly increasing c.d.f. $\Phi_i$. 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 $\frac{1}{2}$,

$$\frac{1}{2n+1}\sum_{i=1}^{2n+1}\Phi_i(0) < \frac{1}{2} - \varepsilon \text{ and } \frac{1}{2} + \varepsilon < \liminf_{n\to\infty} \frac{1}{2n+1}\sum_{i=1}^{2n+1}\Phi_i\left(1^-\right), \tag{23}$$

and from 0 and 1,

$$\varepsilon \leq \Phi_i(0) \text{ and } \Phi_i\left(1^-\right) \leq 1 - \varepsilon \text{ for all } i \in \{1, 2, ...\}. \tag{24}$$

Given (23), under full information about the state, the realized majority preference is $A$ in $\alpha$ and $B$ in $\beta$ 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).

Second, the signals remain boundedly informative, for all $i \in \{1, 2, ...\}$,[15]

$$\varepsilon \leq \Pr_i(s_i = s|\omega) \leq 1 - \varepsilon \quad \text{for all } s \in S_i, \text{ and } \omega \in \{\alpha, \beta\} \tag{25}$$

$$\min_{s\in S_i} \frac{\Pr_i(s_i = s|\alpha)}{\Pr_i(s_i = s|\beta)} < 1 - \varepsilon < 1 + \varepsilon < \max_{s\in S_i} \frac{\Pr_i(s_i = s|\alpha)}{\Pr_i(s_i = s|\beta)} \tag{26}$$

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

$$\Phi_i(x) - \Phi_i(y) \geq L(x - y) \text{ for all } i \in \{1, 2, ...\} \text{ and } x > y. \tag{27}$$

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 $\sigma_n$ 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

[15] The first condition (25) can be replaced by the weaker condition that $\frac{\varepsilon}{1-\varepsilon} \leq \min_{s\in S} \frac{\Pr_i(s_i=s|\alpha)}{\Pr_i(s_i=s|\beta)}$ and $\max_{s\in S} \frac{\Pr_i(s_i=s|\alpha)}{\Pr_i(s_i=s|\beta)} < \frac{1-\epsilon}{\epsilon}$; using this stronger condition simplifies the exposition of our results.

of a tie among any $2n$ voters. Henceforth, we only consider undominated strategies. An " (undominated) equilibrium" is some (undominated) profile $\sigma_n^*$ 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 $\sigma_n^*$ is symmetric if $\sigma_i^* = \sigma_j^*$ for any two voters. Bhattacharya (2013) has shown that, in a symmetric environment, for any sequence of symmetric (undominated) equilibria, the probability that $A$ wins in $\alpha$ and $B$ in $\beta$ converges to 1. Our main result generalizes this.

**Theorem 2** *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 (23) - (27), and any sequence of equilibria $(\sigma_n^*)_{n=1}^\infty$,*

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

## 3.1 Best Response

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

$$q_i(\omega; \sigma_{i,n}) = \int_{s_i, y_i} \sigma(s_i, y_i) d\Phi_i(y_i) d\Psi_{i,\omega}(s_i) \tag{28}$$

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

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

and similarly in state $\beta$,

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

The probability that voter $i$ is pivotal is

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

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

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

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

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

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

## 3.2 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 $p_{i,n} = \Pr\left(\text{piv}_i|\omega;\sigma_{-i,n}\right)$. Then, her posterior conditional on signal $s \in S_i$ is

$$\frac{p_{i,n}\Pr_i(s_i = s|\alpha)}{p_{i,n}\Pr_i(s_i = s|\alpha) + \left(1-p_{i,n}\right)\Pr_i(s_i = s|\beta)}. \tag{29}$$

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

$$\hat{q}_i\left(\omega;p_{i,n}\right) = \sum_{s\in S_i}\Phi_i\left(\frac{p_{i,n}\Pr_i(s_i = s|\alpha)}{p_{i,n}\Pr_i(s_i = s|\alpha) + \left(1-p_{i,n}\right)\Pr_i(s_i = s|\beta)}\right)\Pr_i(s_i = s|\omega) \tag{30}$$

and

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

Given the role of $p_{i,n}$ in (29), Bhattacharya (2013) terms it the induced prior. From

(30), the induced prior is sufficient to determine the best response voting behavior of a voter. 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 (30), we find a new posterior, denoted $\Pr(\alpha|\text{piv}_i; p_n)$. Given this discussion, for any equilibrium $\sigma_n^*$, the induced priors $p_{i,n}^* = \Pr\left(\alpha|\text{piv}_i; \sigma_{-i,n}^*\right)$ must satisfy

$$p_{i,n}^* = \Pr(\alpha|\text{piv}_i; p_n^*) \text{ for all } i \in \{1, 2 \ldots, 2n+1\}. \tag{31}$$

Conversely, any profile of induced priors satisfying (31) induces an equilibrium. We use the fixed-point property (31) to prove existence and, later, the main result.

**Theorem 3** *For every $n$, there exists an equilibrium $\sigma_n^*$.*

**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 (24), 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 (31). Picking, for each voter $i$, the best-response given the induced prior $p_{i,n}^*$ yields an undominated equilibrium profile $\sigma_n^*$. ■

## 3.3 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 (25)

$$\frac{\varepsilon}{1-\varepsilon} \leq \left(\frac{\Pr\left(\text{piv}_i|\alpha;p\right)}{\Pr\left(\text{piv}_i|\beta;p\right)}\right) / \left(\frac{\Pr\left(\text{piv}_j|\alpha;p\right)}{\Pr\left(\text{piv}_j|\beta;p\right)}\right) \leq \frac{1-\varepsilon}{\varepsilon} \text{ for all } i,j \in \{1,2,\ldots,2n+1\}. \tag{32}$$

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,[16]

$$\left(\frac{1-r}{r}\right)^2 < \left(\frac{\Pr\left(\text{piv}_i|\alpha;p_n\right)}{\Pr\left(\text{piv}_i|\beta;p_n\right)}\right) < \left(\frac{r}{1-r}\right)^2 \tag{33}$$

for $r = \Pr\left(s_i = a|\alpha\right) = \Pr\left(s_i = b|\beta\right)$ 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 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

$$p_{4,n} = \frac{r}{1-r} \text{ and } p_{5,n} = \frac{1-r}{r}, \tag{34}$$

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 (34) *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.

In the Appendix, we formally prove that all sequences of asymmetric equilibria aggregate information in the example setting when modified with partisans. In the next Section, we prove this result (Theorem 1) for our model with continuous preference distributions $\Phi_i$. Both proofs build critically on Lemma 3.

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

# 4 Proof: Equilibria Aggregate Information

We now prove Theorem 2. 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 $(p_{i,n})_{n=1,\ldots,\infty}$ 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 1.

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 $i$,

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

Equilibrium induced priors must satisfy the fixed point equation (31),

$$p_{i,n}^* = \Pr\left(\alpha | \text{piv}_i; p_{-i,n}^*, n\right). \tag{35}$$

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

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

Given any $i$ and $p_{i,n} \in (0,1)$, the distribution of the private signals of $i$ induces a distribution of posteriors (29) in states $\alpha$ and $\beta$. Given the bounds on the informativeness of the private signals, (25) and (26), the distribution in $\alpha$ first-order stochastically dominates that in $\beta$. Finally, the claim follows from (30) and since $\Phi_i$ 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, (25) and (26), there is $\delta' > 0$ so that the support of the distributions of posteriors (29) in the states $\alpha$ and $\beta$ is in $[\delta', 1-\delta']$ for all $i$. Finally, given the uniform bounds on the likelihood ratio of the signals, (26), and the uniform Lipschitz bound for $\Phi_i$, (27), we see that the probability to vote $A$ is larger in $\alpha$ than in $\beta$ 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:*

$$0 < \lim_{n\to\infty} \inf p^*_{i,n} \textit{ and } \lim_{n\to\infty} \sup p^*_{i,n} < 1 \textit{ for all } i \in \{1, 2, \ldots, 2n+1\}. \tag{36}$$

We prove the claim by contradiction. Suppose $\lim_{n\to\infty} p^*_{1,n} = 1$.[17] By Lemma 3 (and the implied uniform bound (32)), 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 $\delta$, then $n < (\frac{1}{2} + \varepsilon)2n < \sum_{i=2}^{2n+1} q_i(\beta; p_{i,n})$, given (23). 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 3.1), and $k = n$. This yields that being pivotal is indicative of $\beta$. Consequently, the posterior conditional on being pivotal is below the prior,

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

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) implies that there is a subsequence $(b(n))_{n\in\mathbb{N}}$ for which[18]

$$F^{b(n)}_\omega(q) = \frac{1}{2b(n)+1} | \left\{ i : \hat{q}_i(\omega, p^*_{i,n}) \leq q \text{ and } i \leq 2b(n)+1 \right\} \tag{37}$$

converges pointwise to some c.d.f $F_\omega$ for all states $\omega \in \{\alpha, \beta\}$ and we identify the subsequence with the original sequence to omit the subsequence notation in the following. The expected vote share for $A$ in $\omega$ converges also,

$$\frac{1}{2n+1} \sum_{i=1}^{2n+1} \hat{q}_i(\omega, p^*_i) = \int_q q dF^n_\omega(q) \overset{n\to\infty}{\to} \int_q q dF_\omega(q) = \mathbb{E}_{F_\omega}(q). \tag{38}$$

**Step 4** *The expected vote share of $A$ is strictly larger in $\alpha$ than in $\beta$ for $n$ large enough,*

$$\mathbb{E}_{F_\alpha}(q) > \mathbb{E}_{F_\beta}(q). \tag{39}$$

[17] 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.

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

Suppose $\lim_{n\to\infty} p^*_{1,n} = \bar{p}_1$ (see Footnote 17). From Step 3, $\bar{p}_i \in (0,1)$. Moreover, Lemma 3 implies $\delta > 0$ and $\bar{n} \in \mathbb{N}$ so that $\delta < p^*_{i,n} < 1-\delta$ for all $i$ and $n \geq \bar{n}$. Step 2 then yields some $d > 0$ such that $\hat{q}_{i,n}\left(\alpha; p^*_{i,n}\right) - \hat{q}_{i,n}\left(\beta; p^*_{i,n}\right) \geq d$ for all $i$ and $n \geq \bar{n}$. This implies (39), given (38).

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

$$\mathbb{E}_{F_\alpha}(q) > \frac{1}{2} > \mathbb{E}_{F_\beta}(q). \tag{40}$$

Suppose, for example that $\mathbb{E}_{F_\beta}(q) \geq \frac{1}{2}$. Step 2 - 4 imply that the conditions of Lemma 2 are satisfied for $F = F_\alpha$ and $\tilde{F} = F_\beta$. Lemma 2 and Theorem 1 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 $A$ converges almost surely to the expected share $\mathbb{E}_{F_\omega}(q)$ of votes for $A$ in each state. This together with (40) implies that the full-information outcome is elected,

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

which was the claim of the theorem.

# 5 Conclusion

We revisited the classic Condorcet setting in which agents choose between two alternatives with a simple majority vote. The "modern" Condorcet jury theorem states that the outcome of strategic voting is the same as 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 when considering any sequence of symmetric equilibria.

We lifted these symmetry assumptions, allowing for asymmetric equilibria and voters to draw their types (signals and preferences) from non-identical distributions.

We provided an example of a sequence of non-trivial asymmetric equilibria for which information aggregation fails. That is, the consequence of the theorem does not for asymmetric equilibria, in general. The example equilibrium features a “swing-voter’s blessing’’ where some voters are more likely to be pivotal with any given voting choice when it is in their interest. The logic of the equilibrium rests on this swing voter’s blessing and sufficiently large differences in the voter’s posteriors conditional on being pivotal.

Our main result shows that such failures of information aggregation cannot happen in any non-trivial and possibly asymmetric equilibrium when each voter may be a partisan voter for either alternative, even if just with minimal probability (Condorcet jury theorem without symmetry). A key step is to establish that the partisans imply a tight bound on the swing voter’s blessing (tight in the sense that without partisans it may be violated, as in our example equilibrium.)

On the way, we provide a set of technical results that may be of independent interest to voting theorists.

# 6 Appendix

## 6.1 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 $S_{n\setminus i} = \sum_{j \in \{1,\ldots,2n\} \setminus \{i\}} X_j$,

$$\Pr(S_n = k) = \Pr(X_i = 1) \Pr\left(S_{n\setminus i} = k - 1\right) + \Pr(X_i = 0) \Pr\left(S_{n\setminus i} = k\right). \quad (41)$$

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).[19] 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,

$$\Pr\left[S_{n\setminus i} = k - 1\right] < \Pr\left[S_{n\setminus i} = k\right] \text{ for } k \in \left\{1, ..., \lfloor \mathbb{E}(S_{n\setminus i}) - 1 \rfloor\right\}, \quad (42)$$

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

[19] 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.

Together, (41), (42), and (43) 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.

## 6.2 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 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

$$\Pr(\sum_{i=1}^{2n} X_i^- = \lfloor 2n\gamma \rfloor) \geq \Pr \sum_{i=1}^{2n} X_i = \lfloor 2n\gamma \rfloor) \geq \Pr(\sum_{i=1}^{2n} X_i^+ = \lfloor 2n\gamma \rfloor). \tag{44}$$

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 $F$.[20] Finally, the continuity of the expected Kullback-Leibler divergence

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

$\int_q \mathrm{KL}(a(q), q)dF(q)$ 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 (44), an application of the squeeze lemma yields that the rate function of $(X_i)_{i=1}^{\infty}$ is

$$c^{\mathrm{KL}} = \lim_{D\to\infty} c_D^{\mathrm{KL}+}. \tag{45}$$

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 above by $\frac{1}{2n} = e^{\ln(\frac{1}{2n})} = e^{o(n)}$ for $n$ large.[21]

### 6.3 Proof of Lemma 2

We consider the case when

$$\mathbb{E}_F(q) < \mathbb{E}_{\tilde{F}}(q) \leq \gamma. \tag{46}$$

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

$$m(U \times \tilde{U}) = \lambda([F(q_1^-), F(q_2)] \cap [F(\tilde{q}_1^-), F(\tilde{q}_2)]), \tag{47}$$

where $\lambda$ 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

[21] 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 $\phi$ 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.

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

$$(\tilde{a}, m) \in R(\gamma) \text{ and } \int_{(q,\tilde{q})} \mathrm{KL}(\tilde{a}(q,\tilde{q}), \tilde{q}) dm(q,\tilde{q}) < \inf_{a \in B(\gamma)} \int_q \mathrm{KL}(a(q), q) dF(q). \tag{48}$$

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

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

which satisfies

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

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

$$(\tilde{a}, m) \in R(\gamma) \text{ and } \tilde{q} \leq \tilde{a}(q,\tilde{q}) \leq a_1(q,\tilde{q}) \text{ for all } (q,\tilde{q}). \tag{49}$$

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

$$q < a^*(q). \tag{50}$$

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 (4). 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 (49) and (50) implies that with strictly positive $m$-measure, either

$$q < \tilde{q} \leq a(q,\tilde{q}) \leq a^*(q), \text{ or} \tag{51}$$

$$q < a^*(q) < \tilde{q} = a(q,\tilde{q}). \tag{52}$$

The monotonicity of $m$ further implies that, either (51) holds with the strict inequality $q < \tilde{q}$ replaced by the weak inequality $q \leq \tilde{q}$ $m$-almost everywhere or (52) holds $m$-almost everywhere. Hence,

$$\mathrm{KL}(a^*(q), q) < \mathrm{KL}(\tilde{a}(q, \tilde{q}), \tilde{q}) \tag{53}$$

with strictly positive $m$-measure and a weak inequality $m$-almost everyhwere. Given that the enlarged minimization program has the same solution as (4) (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.

## 6.4 Proof of Lemma 3

Fix $\omega$. 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,

$$\Pr\left(\mathrm{piv}_i|\omega; p_n\right) = P(n-1, n)\, q_j\left(\omega; p_{j,n}\right) + P(n, n-1)\left(1 - q_j\left(\omega; p_{j,n}\right)\right) \tag{54}$$

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 (54) yields

$$\begin{aligned} & \left(\frac{\Pr\left(\mathrm{piv}_i|\alpha; p_n\right)}{\Pr\left(\mathrm{piv}_j|\alpha; p_n\right)}\right) / \left(\frac{\Pr\left(\mathrm{piv}_i|\beta; p_n\right)}{\Pr\left(\mathrm{piv}_j|\beta; p_n\right)}\right) \\ \leq \quad & \max\left\{\frac{1-\hat{q}_i\left(\alpha; p_{i,n}\right)}{1-\hat{q}_j\left(\alpha; p_{j,n}\right)}, \frac{\hat{q}_i\left(\alpha; p_{i,n}\right)}{\hat{q}_j\left(\alpha; p_{j,n}\right)}\right\} / \min\left\{\frac{1-\hat{q}_i\left(\beta; p_{i,n}\right)}{1-\hat{q}_j\left(\beta; p_{j,n}\right)}, \frac{\hat{q}_i\left(\beta; p_{i,n}\right)}{\hat{q}_j\left(\beta; p_{j,n}\right)}\right\}. \end{aligned} \tag{55}$$

Another application of the fact yields

$$\frac{\hat{q}_{i'}\left(\alpha;p_{i',n}\right)}{\hat{q}_{i'}\left(\beta;p_{i',n}\right)}=\frac{\sum_{s\in S_{i'}}\Pr(s_{i'}=s|\alpha)\sigma_i(s)}{\sum_{s\in S_{i'}}\Pr(s_{i'}=s|\beta)\sigma_i(s)}<\max_{s\in S_i}\left(\frac{\Pr(s_{i'}=s|\alpha)}{\Pr(s_{i'}=s|\beta)}\right) \tag{56}$$

and likewise

$$\frac{\hat{q}_{i'}\left(\alpha;p_{i',n}\right)}{1-\hat{q}_{i'}\left(\beta;p_{i',n}\right)}<\max_{s\in S_i}\left(\frac{\Pr(s_{i'}=s|\alpha)}{1-\Pr(s_{i'}=s|\beta)}\right), \tag{57}$$

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

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 (56) here stems from the uniform bound on the likelihood ratio of the signals, (26), and the lower bound on the share of the partisans for each alternative, (24). The upper bounds (55) - (58) jointly prove the upper bound claimed by the lemma. A parallel argument establishes the lower bound.

## 6.5 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.

**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*

$$\Lambda\left(t\right)=\lim_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\left[e^{tnY_n}\right] \tag{59}$$

*exists for all $t\in\mathbb{R}$. If $\Lambda$ 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*

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

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

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 $\Lambda$ 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

$$\Pr(S_n \leq \lfloor \gamma 2n \rfloor) = e^{-2nc^{FL}(\gamma) + o(n)}. \tag{60}$$

In what follows, we connect the point probability $\Pr(S_n = \lfloor \gamma 2n \rfloor)$ to the interval probability $\Pr(S_n \leq \lfloor \gamma 2n \rfloor)$, 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

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

and thus

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

Finally, since $2n = e^{\ln(2n)} = e^{o(n)}$, the interval probability formula (60) implies $\Pr(S_n = \lfloor \gamma 2n \rfloor) \geq e^{-2nc^{FL}(\gamma) + o(n)}$ and therefore also the equality

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

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

$$c^{\mathrm{FL}}(\gamma) = c^{\mathrm{KL}}(\gamma) \text{ for all } \gamma \in (0, 1). \tag{61}$$

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

## 6.6 The Condorcet Jury Theorem in the Example Setting Modified with Partisans

We show that all sequences of asymmetric equilibria aggregate information in the setting from Section 1 when modified with partisans. Formally, the setting is as in Section 1; only the voters do not share a common preference type anymore. We assume that the voters' preference types are drawn independently from an identical distribution where each voter strictly prefers $A$ in both states with probability $\varepsilon > 0$ ($A$-partisan), strictly prefers $B$ in both states with probability $\varepsilon > 0$ ($B$-partisan), and otherwise has a type where she obtains a payoff of 1 when the state matches outcome and 0 otherwise (the previous common preference type). (Clearly, one needs to assume that $\varepsilon < \frac{1}{2}$ for this specification to be well-defined and for all types to have positive probability.)

**Theorem 4** *Consider the setting from Section 1 modified with partisans. For any sequence of (undominated) equilibria* $(\sigma_n^*)_{n=1}^{\infty}$*,*

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

**Proof.** The first part of the proof (Step (1)) parallels Step 4 from the proof of Theorem 1; albeit it requires a different argument.

As in the earlier proof, we use Helly's selection theorem and transition implicitly to a subsequence for which the c.d.f $F_\omega^n(q)$ of the empirical distribution of the first $2n$ voters' probabilities $q_i(\omega; \sigma_{i,n}^*)$ to vote $A$ converges pointwise some c.d.f $F_\omega$ for $\omega \in \{\alpha, \beta\}$. The convergence in distribution implies that the expected vote share for $A$ converges in each state,

$$\frac{1}{2n+1} \sum_{i=1}^{2n+1} q_i(\omega; \sigma_{i,n}^*) = \mathbb{E}_{F_\omega^n}(q) \overset{n\to\infty}{\to} \mathbb{E}_{F_\omega}(q). \tag{62}$$

**Step 1** *The expected vote share of $A$ is strictly larger in $\alpha$ than in $\beta$ for $n$ large enough,*

$$\mathbb{E}_{F_\alpha}(q) > \mathbb{E}_{F_\beta}(q). \tag{63}$$

Suppose not. This implies that the share of sincere voters goes to zero, as $n \to \infty$. Lemma 3 implies that there is at most one type of non-sincere non-partisan voter; either all non-sincere non-partisans are indifferent after a $b$-signal and vote $A$ after an

$a$-signal, or all are indifferent after an $a$-signal and vote $B$ after a $b$-signal (recall the discussion after Lemma 3). Without loss, we consider only the first case here. The proof in the other case is analogous.

Two observations: First, our initial assumption implies that the share of non-sincere non-partisans with strategies converging to always voting $A$ goes to 1, as $n \to \infty$. This means that, the expected vote share of $A$ approximates $\sum_{i=2}^{2n+1} q_i(\omega; \sigma^*_{i,n}) \approx 2n(1-\epsilon)$ in both states. Since the share of $B$-partisans is strictly less than half, $\varepsilon < \frac{1}{2}$, the majority threshold $n$ is below $2n(1-\epsilon)$ and thus below the expected vote share in both states. Second, since all non-sincere non-partisans vote $A$ after $a$ and not always $A$ after $b$, the probability to vote $A$ is weakly larger in $\alpha$ for all voters, $q_i(\alpha; \sigma^*_{i,n}) \geq q_i(\beta; \sigma^*_{i,n})$ (and strictly larger for some $i$ since in any undominated equilibrium not all voters vote deterministically for some alternative).

The two observations allow us to apply Lemma 1, which yields that being pivotal is indicative of $\beta$. Consequently, the posterior conditional on being pivotal is below the prior,

$$\lim_{n\to\infty} \sup \Pr\left(\text{piv}_i | \sigma^*_n, n\right) \leq \Pr\left(\alpha\right) \text{ for all } i.$$

Given the uniform prior $\Pr(\alpha)$, this implies that all non-partisans strictly prefer $B$ after a $b$ signal. We arrive at a contradiction to the earlier implication that almost all non-sincere non-partisans use a strategy that converges to voting $A$ with probability 1.

The remainder of the proof is identical to the final part of the proof of Theorem 1 that follows Step 4 therein. ■

# 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 (2018): "Adverse Selection in Distributive Politics," Working paper.

Alonso, R., and O. Câmara (2016): "Persuading voters," *American Economic Review*, 106(11), 3590–3605.

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

Cramér, H., and H. Touchette (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.

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

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.

Heese, C., and S. Lauermann (2019): "Persuasion and Information Aggregation in Large Elections," CRC TR 224 discussion paper series, University of Bonn and University of Mannheim, Germany, https://ideas.repec.org/p/bon/boncrc/crctr224_2019_128.html.

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

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.

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

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

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

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.

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