To Lead or to be Led: A Generalized Condorcet Jury Theorem under Dependence

Jonas Karge
TU Dresden
Dresden, Germany
jonas.karge@tu-dresden.de

Juliette-Michelle Burkhardt
TU Dresden
Dresden, Germany
juliette-michelle.burkhardt@mailbox.tu-dresden.de

Sebastian Rudolph
TU Dresden
Dresden, Germany
sebastian.rudolph@tu-dresden.de

Dominik Rusovac
TU Dresden
Dresden, Germany
dominik.rusovac@tu-dresden.de

ABSTRACT

Aggregating pieces of information or beliefs held by (abstract) agents is central to a variety of belief merging applications. When the merging process aims at retrieving an underlying *ground truth*, the *Condorcet Jury Theorem* (CJT) allows identifying *voting rules* that almost surely track the true piece of information for large groups of agents, given that specific conditions are met. As essential assumptions, the CJT relies on all agents being equally competent as well as independent from one another. In the search for a generalization of the CJT applicable to real-world scenarios, both aforementioned assumptions were weakened separately. In this work, we provide a generalization of the CJT that allows, at the same time, for heterogeneous competence levels across agents as well as a degree of dependence modeled through an *opinion leader* exerting influence on the electorate. Additionally, we derive a concrete bound on the number of agents necessary to successfully track the underlying ground truth, and examine the bound's tightness by means of statistical simulations.

KEYWORDS

Condorcet Jury Theorem; Opinion Leader; Belief Merging

ACM Reference Format:
Jonas Karge, Juliette-Michelle Burkhardt, Sebastian Rudolph, and Dominik Rusovac. 2024. To Lead or to be Led: A Generalized Condorcet Jury Theorem under Dependence. In *Proc.of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 9 pages.

1 INTRODUCTION

Frequently, it is necessary to gather data from potentially conflicting sources in order to find the underlying *ground truth.* Combining this information by suitable aggregation methods is central to disciplines across computer science such as belief merging, information fusion, as well as multi-agent systems and can be realized through *voting*. As theoretical cornerstone, the *Condorcet Jury Theorem* (CJT) provides probabilistic guarantees for determining the presumed ground truth under specific conditions. Originally, the CJT assumes agents to be equally competent *(homogeneity)*, to be more likely to vote for the correct alternative than for a competitor *(reliability)*, to not influence one another, or to be influenced by an external factor, in the voting process *(independence)*, and to choose exactly one *(completeness)* from two alternatives only *(dichotomy)* under majority voting. With that, the classical CJT [3] states the following.

Theorem 1 (Marquis de Condorcet [3]). *For odd-numbered, homogenous groups of independent and reliable agents in a dichotomic voting setting, the probability that majority voting identifies the correct alternative*

(1) increases monotonically with the number of agents and
(2) converges to 1 *as the number of agents goes to infinity.*

Typically, (1) is referred to as the non-asymptotic, and (2) as the asymptotic part of the CJT. Since most applications to real-world scenarios cannot be guaranteed to adhere to these ideal conditions, it is central to CJT research to find generalizations for the asymptotic part to hold under weakened assumptions. For the non-asymptotic part it was shown, however, that it fails for small numbers of agents when weakening homogeneity [13] and that this failure holds for arbitrarily large numbers of voters [9].

Past Generalizations. Given the failure of the non-asymptotic part, generalizing the CJT typically involves (i) weakening at least one of the imposed assumptions and (ii) proving that the probability to correctly identify the true state converges to 1 as the number of voters approaches infinity under this weakened set of assumptions. This gives rise to various ways of generalizing the CJT.

Weakening Voting Constraints. A first route of generalization is to weaken the dichotomy assumption. In that regard, List and Goodin [12] showed that the asymptotic part continues to hold when dropping dichotomy from the original assumptions, and instead allowing for any finite number of alternatives under *plurality voting*, where the alternative that accumulates the most votes wins. Subsequently, this result was further generalized by giving up completeness: Everaere, Konieczny, and Marquis [4] showed the asymptotic statement to remain valid under *approval voting*, where any agent may simultaneously vote for any number of alternatives.

Weakening Homogeneity. Generalizing towards a different direction, Owen, Grofman, and Feld [13] proved, assuming independence, completeness and dichotomy, the CJT to also hold true when we

Proc.of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

allow for heterogeneous competence levels under majority voting. In this setting, it is only assumed that, when averaging across all agents, they are more likely to vote for the correct alternative than for any other. Finally, Karge and Rudolph [9] showed that the asymptotic statement part of the CJT still holds when allowing for heterogeneous competence levels, and, at the same time, for any finite number of alternatives under approval voting.

Weakening Independence. When allowing for interdependent choices among voters, numerous options to model dependence can be considered. One standard approach introduces the concept of an *opinion leader* (OL) representing an abstract, external influence (such as extreme environmental conditions in sensor fusion scenarios, or actual human agents in political debates like lobbyists or pundits). Typically, the OL herself does not take part in the voting process, but approves or disapproves presented alternatives according to her own competency $\hat{p}$, i.e. the probability to approve the correct alternative. Her choice then influences the agents' vote: each agent votes according to the OL's preference rather than his own "inner voice" with a certain probability π.

In prior work generalizing the CJT under an OL's influence, Boland, Proschan, and Tong [2] assume homogeneous competence levels, p, across agents and an equally competent OL, $\hat{p} = p$, in a dichotomous voting setting under majority voting. Shortly after, Berg [1] extended Boland et al.'s setting to dichotomous weighted voting rules. Finally, the dichotomic voting setting of Goodin and Spiekermann [15] allows for the OL's competency to deviate from the homogeneous competence levels across agents. Additionally, they provide a specific threshold for when the asymptotic part of the CJT breaks depending on π and p.

Other notions of dependence in CJT generalizations include Ladha [11], who allows for pairwise correlations among voters restricting the average correlation coefficients; Kaniovski [7], who characterizes joint distributions that are, depending on the voter's competency and the correlation, beneficial or detrimental to a jury's overall competency; and Pivato [14], who admits dependencies but restricts the average co-variance among voters.

Throughout this work, we model dependence along the OL model as it is developed in Boland, Proschan, and Tong [2] which is sufficiently transparent to be applicable to real-world scenarios, and, at the same time, general enough to capture a variety of applications. We highlight that we consider a single OL instead of multiple external influences as we provide a worst-case analysis of the success probability to identify the correct alternative. In settings with multiple OLs, it is typically assumed that the electorate is partitioned into subgroups each being influenced by exactly one OL. The worst case then occurs when all OLs are perfectly positively correlated in that the votes across all subgroups are swayed in the same direction. However, this case is equivalent to having a single OL influencing the whole group of voters. More generally, it has been argued that having more opinion leaders increases the group competence compared to having a single one [15].

Our Contribution. In this work, we (i) generalize the asymptotic part of the CJT to a setting that simultaneously relaxes all central assumptions underlying the original CJT. This yields, to the best of our knowledge, the most general variant of the CJT thus far. Most prominently, we allow for heterogeneous competence levels across agents as well as interdependence modeled by means of the OL - two assumptions that are typically only weakened individually, but not at the same time. Unlike previous work [14] that does, in fact, weaken both assumptions, we do not require each agent to be reliable, but explicitly allow for unreliable or even malicious agents. Moreover, we (ii) provide a precise threshold for when the asymptotic part breaks. In order to be better applicable to real-world applications, we also provide (iii) a tight, implicit bound on the number of agents necessary to guarantee a prescribed minimal success probability in our setting, and a less tight, explicit bound that requires fewer information on the underlying parameters in an application. With this in mind, our objective is to address a void in the current CJT literature, which predominantly emphasizes the asymptotic behavior in the infinite but lacks specific estimations for real-world scenarios. Finally, we (iv) examine the tightness of both bounds by means of statistical simulations.

2 PRELIMINARIES

Before introducing our formal framework in depth, we informally describe our approval voting setting under OL influence (see Figure 1). In a voting round, the opinion leader and each of the n agents are asked to approve any selection from m given alternatives (also referred to as "worlds"). Then each agent determines his "inner voice", also referred to as "private signal", giving rise to the approval choices he would make if not influenced. The OL reports her approved alternatives as a "public signal", announced to the agents. All agents follow the OL's public signal (rather than their private one) with a certain pre-defined probability π, which characterizes the OL's "influence strength". The final voting result then follows from aggregating all individual votes that either alternative receives, where the world receiving the most approvals wins.

Voting. Let $\mathcal{W} = \{\omega_1, \ldots, \omega_m\}$ denote a finite set of m alternatives, also referred to as *worlds* or *choices*, and, likewise $\mathcal{A} = \{a_1, \ldots, a_n\}$ to be a finite set of n agents. We can then represent a single *approval voting (instance)* by $V \subseteq \mathcal{A} \times \mathcal{W}$ where $(a_i, \omega_j) \in V$ means that agent a_i approves choice ω_j. Subsequently, we define the *score* $\#_V\, \omega$ of some choice $\omega \in \mathcal{W}$ as

$$\#_V\, \omega = |\{a_i \in \mathcal{A}_n \mid (a_i, \omega) \in V\}|.$$

Finally, the winner of V is defined to be the alternative that receives a strictly higher score than any alternative

$$\#_V\, \omega > \max_{\omega' \in \mathcal{W} \setminus \{\omega\}} \#_V\, \omega'.$$

Formal Probabilistic Model. The described voting scenario is modeled by a random process that generates ω_*, the OL's choice as well as V and is govenerd by a joint probability distribution $\mathbb{P}$ over the Bernoulli (i.e., $\{0, 1\}$-valued) random variables $X_*^{\omega_1}, ..., X_*^{\omega_m}$, $X_o^{\omega_1}, ..., X_o^{\omega_m}$ as well as $X_i^{\omega_1}, ..., X_i^{\omega_m}$ for all agents $1, ..., i, ..., n$ and all alternatives $1, ..., j, ..., m$ such that the values taken by these random variables represent the outcome of a voting event as follows:

- $X_*^{\omega_j}$ is 1 if ω_j is the true world state (i.e., $\omega_j = \omega_*$), else 0,
- $X_o^{\omega_j}$ is 1 if the OL approves ω_j, and 0 otherwise,
- $X_i^{\omega_j}$ represents the private signal of the ith agent regarding his approval of the jth world state: it is 1 if a_i privately approves ω_j and otherwise 0.

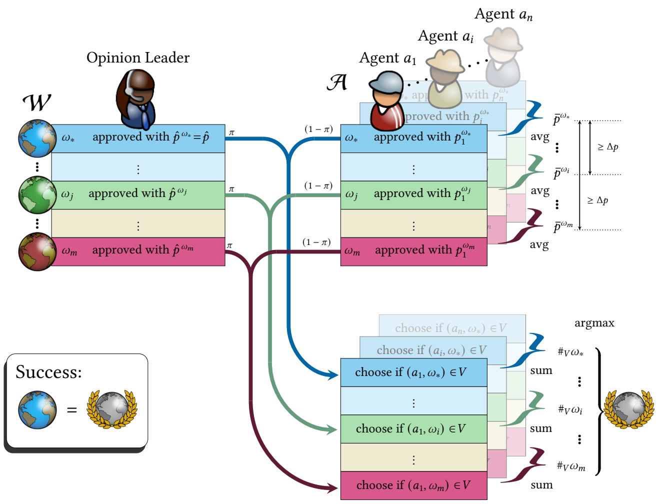

Figure 1: Depiction of the voting process for n agents and m worlds.

Given this joint distribution, we introduce the random variable $V_i^{\omega_j}$ representing the final outcome of an agent's vote, i.e. after the OL potentially exerted influence. According to our assumption, $V_i^{\omega_j}$ is the probabilistic mixture of $X_o^{\omega_j}$ with probability π and of $X_i^{\omega_j}$ with probability $1-\pi$. From this, we obtain for any $x \in \{0,1\}$ that

$$\mathbb{P}(V_i^{\omega_j} = x) = \pi \mathbb{P}(X_o^{\omega_j} = x) + (1-\pi)\mathbb{P}(X_i^{\omega_j} = x).$$

We denote by $p_1^{\omega}, \ldots, p_n^{\omega}$ the Bernoulli parameters of the "inner voice" random variables $X_1^{\omega}, \ldots, X_n^{\omega}$, for all $\omega \in \mathcal{W}$, that is, $p_i^{\omega_j} = \mathbb{P}(X_i^{\omega_j} = 1)$. In a similar vein, for every $\omega \in \mathcal{W}$, we let $\hat{p}^{\omega_1}, \ldots, \hat{p}^{\omega_m}$ denote the Bernoulli parameters of the random variables $X_o^{\omega_1}, \ldots, X_o^{\omega_m}$. Whether the OL approves the correct alternative, i.e. whether $X_o^{\omega_*} = 1$, is governed by the parameter, $\hat{p} = \mathbb{P}(X_o^{\omega_*} = 1)$. For convenience, the expression

$$(X_*^{\omega_j}{=}1) \wedge \bigwedge_{\omega \in \mathcal{W} \setminus \{\omega_j\}} (X_*^{\omega}{=}0)$$

will be abbreviated by $[\omega_* = \omega_j]$.

In the following, we define the two central assumptions regarding the joint distribution that are underlying our result; each of them is also required for all previously presented generalizations of the CJT. Conditioning upon the actual world state, we may then formalize the first central assumption imposed on the joint distribution.

Definition 1. *A joint distribution satisfies* private agent approval independence *if, conditioned on the actual world state, the private decision to approve any given ω_j is made independently across all agents, i.e., for any $\omega, \omega_j \in \mathcal{W}$ and any sequence $v_1, \ldots, v_n$ of values from $\{0,1\}$ the following holds:*

$$\mathbb{P}\Big(\bigwedge_{i=1}^{n} X_i^{\omega_j} = v_i \mid [\omega_*{=}\omega]\Big) = \prod_{i=1}^{n} \mathbb{P}\Big(X_i^{\omega_j} = v_i \mid [\omega_*{=}\omega]\Big).$$

That is, conditioned on the true world state, the joint probability of any given pattern of private approval decisions with respect to a given alternative can be computed by taking the product of the corresponding marginal probabilities.

A further central assumption deals with the "internal competency" p_k^{ω} of the kth agent regarding his capacity to identify the true world state among any number of alternatives *if no influence is exerted.* With

$$\bar{p}^{\omega} = \frac{1}{n}\sum_{k=1}^{n} p_k^{\omega}$$

denoting the average over these "internal competencies", we can formalize this assumption as follows.

Definition 2. *A joint probability distribution satisfies Δp-*group reliability *for some $\Delta p > 0$, if the probability, with respect to the agent's inner voice, to approve the true world state, averaged across all agents, is at least by Δp higher than the averaged probability for approving any other state, i.e., for every n and $\omega_\dagger \in \mathcal{W} \setminus \{\omega_*\}$ holds*

$$\bar{p}^{\omega_*} \geq \Delta p + \bar{p}^{\omega_\dagger}.$$

Now that we have specified the underlying assumptions, we proceed to deriving the asymptotic result.

3 GENERALIZING THE CJT

Fixing Notation. Aiming to represent the proof of our generalization of the CJT more succinctly, we consider ω_* fixed and rename the remaining alternatives accordingly, i.e. $\mathcal{W} = \{\omega_*, \omega_1, \ldots, \omega_{m-1}\}$. As before, we consider $\hat{p}$ to be the opinion leader's competency and π to be her "influence strength", i.e., the (globally fixed) probability for her choice superseding any agent's own private signal.

Conditioning on the OL's Choice. As general strategy to derive our theoretical results, we distinguish the case where the OL chooses the correct alternative from the case where she is wrong. This case distinction allows to view the agents as independent in each subcase.

To prepare the proof of the asymptotic result, we first characterize central properties of the random variables from the aforementioned joint distribution. Recall that, for discrete random variables X and Y, the conditional expectation is given by

$$\mathbb{E}(X \mid Y{=}y) = \sum_x x\,\mathbb{P}(X{=}x \mid Y{=}y)$$

From $\mathbb{P}\big(V_i^{\omega_j} = x\big)$ as well as the conditional expectation for $V_i^{\omega_j}$, for every $\omega_j \in \mathcal{W}$, we obtain the distribution for the score received by ω_j and the corresponding expected values for the two different OL behaviors:

$$V^{\omega_j} = \textstyle\sum_{k=1}^{n} V_k^{\omega_j}$$
$$\mathbb{E}(V^{\omega_j} \mid X_o^{\omega_j}{=}1) = n(\pi + (1{-}\pi)\bar{p}^{\omega_j})$$
$$\mathbb{E}(V^{\omega_j} \mid X_o^{\omega_j}{=}0) = n(1{-}\pi)\bar{p}^{\omega_j}.$$

In the following we will establish lower bounds for the probability that an electorate of agents successfully tracks the correct alternative through approval voting under the assumption of agent approval independence as well as Δp-group reliability. We will first establish these bounds separately for the cases of the OL being right or wrong and combine them afterward, taking into account the OL's competency.

In either case, as general strategy, we consider, agent by agent, the value $V_i^{\omega_*}-V_i^{\omega_\dagger}$. The "composite random variable" $V_i^{\omega_*}-V_i^{\omega_\dagger}$ can have three possible outcomes:

- 1, if agent a_i votes for ω_* but not for $\omega_\dagger$,
- -1, if he votes for $\omega_\dagger$ but not for ω_*, and
- 0 if he votes for both or for none of the two.

We now consider the aggregated random variable $V^{\omega_*-\omega_\dagger}$ defined by

$$V^{\omega_*-\omega_\dagger} = \sum_{k=1}^{n}(V_k^{\omega_*}-V_k^{\omega_\dagger}) = \sum_{k=1}^{n} V_k^{\omega_*} - \sum_{k=1}^{n} V_k^{\omega_\dagger} = V^{\omega_*}-V^{\omega_\dagger}$$

and observe that ω_* wins against $\omega_\dagger$ exactly if $V^{\omega_*-\omega_\dagger} > 0$. In order to find good probability estimates for this, we utilize Hoeffding's inequality [6], which provides a tail estimate for the sum of independent random variables with the property of exhibiting zero probability outside a finite interval.

Lemma 1 (Hoeffding [6]). *Let $X_1, \dots, X_n$ be independent random variables satisfying $\mathbb{P}(l_i \le X_i \le u_i) = 1$ for reals l_i, u_i. Consider the sum of these random variables, $X = \sum_{i=1}^n X_i$. Then for every real number $t > 0$ holds*

$$\mathbb{P}(X - \mathbb{E}(X) \ge t) \le e^{-\frac{2}{\sum_{i=1}^n (u_i - l_i)^2} t^2}.$$

We recall that the agent-wise distributions of $V_i^{\omega_*}-V_i^{\omega_\dagger}$ discussed above are of this type with $l_i = -1$ and $u_i = 1$.

The OL is right. We start by conditioning on the fact that the OL approves the correct alternative. In order to correctly reflect the worst case, we assume that, for any number of alternatives, the OL always approves all competitors of ω_*. In the following, consider an arbitrarily chosen but fixed competing alternative $\omega_\dagger \in \mathcal{W} \setminus \{\omega_*\}$ in the approval vote. In a first step, we derive a lower bound for the probability of ω_* winning against this fixed competitor, $\omega_\dagger$.

Intuitively, the worst-case success probability increases with growing n, since for $X_o^{\omega_*} = 1$, the distributions for $\frac{V^{\omega_*}}{n}$ and $\frac{V^{\omega_\dagger}}{n}$ will get concentrated more and more narrowly around $\pi+(1-\pi)\bar{p}^{\omega_*}$ and $\pi+(1-\pi)\bar{p}^{\omega_\dagger}$, respectively, since the OL approves both alternatives in the worst case. Then, using the strategy described above, we can obtain a lower bound for the probability that in the course of an approval vote, the correct choice ω_* receives more votes than some fixed competitor $\omega_\dagger$. In a next step, we can derive a lower bound for probability for the winning against all competitors (see the top part of Table 1 for the details).

The OL is wrong. In a similar vein, we consider the case of $X_o^{\omega_*} = 0$. Then, we expect the distributions for $\frac{V^{\omega_*}}{n}$ and $\frac{V^{\omega_\dagger}}{n}$ to be concentrated around the values $(1-\pi)\bar{p}^{\omega_*}$ and $\pi+(1-\pi)\bar{p}^{\omega_\dagger}$, respectively.

This also gives rise to a threshold for π for which the asymptotic part holds. In the worst case, we obtain under the group reliability assumption that $\bar{p}^{\omega_*} \ge \bar{p}^{\omega_\dagger} + \Delta p$. Taking into account the expectations of ω_* and $\omega_\dagger$, we have to require that

$$(1-\pi)\bar{p}^{\omega_*} > \pi + (1-\pi)\bar{p}^{\omega_\dagger} \qquad \text{and therefore}$$

$$\pi < \frac{\Delta p}{\Delta p + 1}.$$

That is, the expected values for scores of the correct alternative as well as its strongest competitor are equal or even reversed as soon as the OL's influence level π violates the above condition. Figure 2 illustrates the range of valid π-values for all possible Δp.

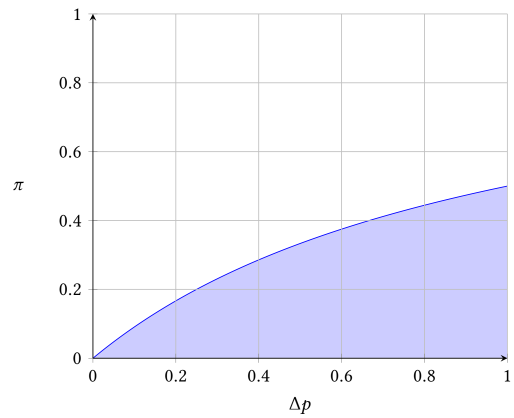

Figure 2: Permissible π-values depending on Δp.

Now given the correct alternative, ω_*, as well as the fixed competitor, ω_*, we again derive a lower bound for the probability that ω_* receives more votes than ω_* in an approval vote as well as for the winning against all competitors (see the middle part of Table 1 for the details).

Aggregating the cases. Now that we have the individual bounds on the success probabilities conditioned on the OL's choice, we can combine both bounds by taking into account the probability for the OL to be either right or wrong, that is, the OL's competency, $\hat{p}$.

As detailed in the bottom part of Table 1, we obtain Equation (5) as bound for the total worst-case success probability for the correct alternative winning the approval vote. Note that this probability approaches 1 with growing n. That is, in the limit, the overall expression converges to 1 for any competence value $\hat{p}$ (under the given restriction for π). Thus, we can state:

Theorem 2. *Consider an approval voting setting with $m > 1$ alternatives, satisfying private agent approval independence (Definition 1) and Δp-group reliability (Definition 2) for some $\Delta p \in (0, 1]$, influenced by an opinion leader with $\pi \in [0, \frac{\Delta p}{\Delta p+1})$ and $\hat{p} \in [0, 1]$. Then the probability that approval voting identifies the correct alternative converges to 1 as the number of agents goes to infinity.*

With this result, we can now also provide upper bounds on the number of agents n needed to guarantee that the success probability exceeds a certain prescribed value $P_{\min}$.

Theorem 3. *Consider an approval voting setting as described in Theorem 2. Then, given a probability $P_{\min} < 1$, it is guaranteed that the success probability of the approval voting process is greater than $P_{\min}$ if the number n of agents obeys any of the below conditions.*

$$\hat{p}e^{-\frac{n}{2}\Delta p^2(1-\pi)^2} + (1-\hat{p})e^{-\frac{n}{2}(\Delta p(1-\pi)-\pi)^2} \le \frac{1-P_{\min}}{m-1} \qquad (7)$$

$$\frac{2}{(\Delta p(1-\pi)-\pi)^2} \ln \frac{m-1}{1-P_{\min}} \le n \qquad (8)$$

Opinion leader is right

$$
\begin{aligned}
&\mathbb{P}(V^{\omega_*} > V^{\omega_\dagger} \mid X_o^{\omega_*} = 1)\\
&= \mathbb{P}(V^{\omega_*} - V^{\omega_\dagger} > 0 \mid X_o^{\omega_*} = 1)\\
&= 1 - \mathbb{P}(V^{\omega_*} - V^{\omega_\dagger} \le 0 \mid X_o^{\omega_*} = 1)\\
&= 1 - \mathbb{P}((V^{\omega_*} - V^{\omega_\dagger}) - \mathbb{E}(V^{\omega_*} - V^{\omega_\dagger} \mid X_o^{\omega_*} = 1) \le -\mathbb{E}(V^{\omega_*} - V^{\omega_\dagger} \mid X_o^{\omega_*} = 1) \mid X_o^{\omega_*} = 1)\\
&= 1 - \mathbb{P}((V^{\omega_\dagger} - V^{\omega_*}) - \mathbb{E}(V^{\omega_\dagger} - V^{\omega_*} \mid X_o^{\omega_*} = 1) \ge \mathbb{E}(V^{\omega_*} \mid X_o^{\omega_*} = 1) - \mathbb{E}(V^{\omega_\dagger} \mid X_o^{\omega_*} = 1) \mid X_o^{\omega_*} = 1)\\
&= 1 - \mathbb{P}((V^{\omega_\dagger} - V^{\omega_*}) - \mathbb{E}(V^{\omega_\dagger} - V^{\omega_*} \mid X_o^{\omega_*} = 1) \ge n(\pi + (1-\pi)\bar{p}^{\omega_*}) - n(\pi + (1-\pi)\bar{p}^{\omega_\dagger}) \mid X_o^{\omega_*} = 1)\\
&\ge 1 - \mathbb{P}((V^{\omega_\dagger} - V^{\omega_*}) - \mathbb{E}(V^{\omega_\dagger} - V^{\omega_*} \mid X_o^{\omega_*} = 1) \ge n(1-\pi)\Delta p \mid X_o^{\omega_*} = 1)\\
&\ge 1 - e^{-\frac{2}{4n}(n(1-\pi)\Delta p)^2} && \text{Hoeffding noting that } u_i - l_i = 2 \text{ for all } i\\
&= 1 - e^{-\frac{1}{2}n\Delta p^2(1-\pi)^2} && (1)
\end{aligned}
$$

Then we obtain for the winning against all competitors:

$$
\begin{aligned}
\mathbb{P}(\bigwedge_{\omega_\dagger \in \mathcal{W}\setminus\{\omega_*\}} V^{\omega_*} > V^{\omega_\dagger} \mid X_o^{\omega_*} = 1) &\ge 1 - \textstyle\sum_{i=1}^{m-1}(1 - \mathbb{P}(V^{\omega_*} > V^{\omega_i} \mid X_o^{\omega_*} = 1))\\
&= 1 - \textstyle\sum_{i=1}^{m-1}(1 - (1 - e^{-\frac{1}{2}n\Delta p^2(1-\pi)^2})) && \text{Equation (1)}\\
&= 1 - (m-1)e^{-\frac{1}{2}n\Delta p^2(1-\pi)^2} && (2)
\end{aligned}
$$

Opinion leader is wrong

$$
\begin{aligned}
&\mathbb{P}(V^{\omega_*} > V^{\omega_\dagger} \mid X_o^{\omega_*} = 0)\\
&= \mathbb{P}(V^{\omega_*} - V^{\omega_\dagger} > 0 \mid X_o^{\omega_*} = 0)\\
&= 1 - \mathbb{P}(V^{\omega_*} - V^{\omega_\dagger} \le 0 \mid X_o^{\omega_*} = 0)\\
&= 1 - \mathbb{P}((V^{\omega_*} - V^{\omega_\dagger}) - \mathbb{E}(V^{\omega_*} - V^{\omega_\dagger} \mid X_o^{\omega_*} = 0) \le -\mathbb{E}(V^{\omega_*} - V^{\omega_\dagger} \mid X_o^{\omega_*} = 0) \mid X_o^{\omega_*} = 0)\\
&= 1 - \mathbb{P}((V^{\omega_\dagger} - V^{\omega_*}) - \mathbb{E}(V^{\omega_\dagger} - V^{\omega_*} \mid X_o^{\omega_*} = 0) \ge \mathbb{E}(V^{\omega_*} \mid X_o^{\omega_*} = 0) - \mathbb{E}(V^{\omega_\dagger} \mid X_o^{\omega_*} = 0) \mid X_o^{\omega_*} = 0)\\
&= 1 - \mathbb{P}((V^{\omega_\dagger} - V^{\omega_*}) - \mathbb{E}(V^{\omega_\dagger} - V^{\omega_*} \mid X_o^{\omega_*} = 0) \ge n(1-\pi)\bar{p}^{\omega_*} - n(\pi + (1-\pi)\bar{p}^{\omega_\dagger}) \mid X_o^{\omega_*} = 0)\\
&\ge 1 - \mathbb{P}((V^{\omega_\dagger} - V^{\omega_*}) - \mathbb{E}(V^{\omega_\dagger} - V^{\omega_*} \mid X_o^{\omega_*} = 0) \ge n\big((1-\pi)\Delta p - \pi\big) \mid X_o^{\omega_*} = 0)\\
&\ge 1 - e^{-\frac{2}{4n}\left(n\left((1-\pi)\Delta p - \pi\right)\right)^2} && \text{Hoeffding with } u_i - l_i = 2 \text{ for all } i\text{, assuming } \pi \le \tfrac{\Delta p}{\Delta p + 1}\\
&\ge 1 - e^{-\frac{1}{2}n\left(\Delta p(1-\pi) - \pi\right)^2} && (3)
\end{aligned}
$$

Then we obtain for the winning against all competitors:

$$
\begin{aligned}
\mathbb{P}(\bigwedge_{\omega_\dagger \in \mathcal{W}\setminus\{\omega_*\}} V^{\omega_*} > V^{\omega_\dagger} \mid X_o^{\omega_*} = 0) &\ge 1 - \textstyle\sum_{i=1}^{m-1}(1 - \mathbb{P}(V^{\omega_*} > V^{\omega_i} \mid X_o^{\omega_*} = 0))\\
&= 1 - \textstyle\sum_{i=1}^{m-1}(1 - (1 - (e^{-\frac{1}{2}n\left(\Delta p(1-\pi) - \pi\right)^2}))) && \text{Equation (3)}\\
&= 1 - (m-1)e^{-\frac{1}{2}n\left(\Delta p(1-\pi) - \pi\right)^2} && (4)
\end{aligned}
$$

Aggregating the cases

$$
\begin{aligned}
\mathbb{P}(\bigwedge_{\omega_\dagger \in \mathcal{W}\setminus\{\omega_*\}} X^{\omega_*} > X^{\omega_\dagger}) &= \mathbb{P}(X_o^{\omega_*} = 1) \cdot \mathbb{P}(\bigwedge_{\omega_\dagger \in \mathcal{W}\setminus\{\omega_*\}} V^{\omega_*} > V^{\omega_\dagger} \mid X_o^{\omega_*} = 1) + \mathbb{P}(X_o^{\omega_*} = 0) \cdot \mathbb{P}(\bigwedge_{\omega_\dagger \in \mathcal{W}\setminus\{\omega_*\}} V^{\omega_*} > V^{\omega_\dagger} \mid X_o^{\omega_*} = 0)\\
&\ge \hat{p}\big(1 - (m-1)e^{-\frac{1}{2}n\Delta p^2(1-\pi)^2}\big) + (1-\hat{p})\big(1 - (m-1)e^{-\frac{1}{2}n(\Delta p(1-\pi)-\pi)^2}\big) && \text{Equation (2) and Equation (4)}\\
&= 1 - (m-1)\Big(\hat{p}e^{-\frac{1}{2}n\Delta p^2(1-\pi)^2} + (1-\hat{p})e^{-\frac{1}{2}n(\Delta p(1-\pi)-\pi)^2}\Big) && (5)\\
&\ge 1 - (m-1)e^{-\frac{1}{2}n(\Delta p(1-\pi)-\pi)^2} && \text{noting that } \Delta p^2(1-\pi)^2 \ge (\Delta p(1-\pi) - \pi)^2 \text{ due to } \pi \le \tfrac{\Delta p}{\Delta p+1} \quad (6)
\end{aligned}
$$

Table 1: Derivations of bounds for success probability, conditioned on the opinion leader's correctness (top and middle) and aggregated (bottom).

The first constraint, following directly from Equation (5), is tighter than the second, which is obtained from Equation (6). The second condition is still useful as it does not require knowledge of the value of $\hat{p}$, and it allows for the direct computation of a lower bound for n.

We note that for $\pi = 0$ (i.e. when the OL exerts no influence), the above bound improves on one that was previously shown for the identical setting without OL [9] for cases with not too big values of Δp, namely

$$\frac{2}{\Delta p^2} \ln \frac{2(m-1)}{1-P_{\min}} \leq n.$$

That is, this framework is indeed a generalization as it dominates this previously shown result.

4 EXPERIMENTS

The theoretical results obtained in the previous section provide upper bounds on the number of agents necessary to successfully track the underlying ground truth with a high enough chance. For further insights into the usefulness of these findings, we conducted experiments [8] to evaluate the tightness of both the implicit and the explicit bound. To this end, we used statistical simulations as described in the following.

Setup. Our experiments were conducted on a high-performance computing system. They were run on four nodes in parallel, each with two 12 core Intel(R) Xeon(R) E5-2680 v3 CPU 2.5GHz with 128GB local memory on SSD.

Experimental Design. Given that our goal was to establish worst-case guarantees, the conditions for our experimental simulations were picked with the goal of reproducing worst-case circumstances. We found this to be the case for homogeneous competence levels among the agents and, toward the highest possible variance, choosing for a given Δp, any agent's probabilities to vote for right/wrong to be centered around $\frac{1}{2}$, that is, we homogeneously (i.e., for all $i \in \{1,\ldots,n\}$) set $p_i^{\omega_*} = \frac{1+\Delta p}{2}$ while we set $p_i^{\omega_\dagger} = \frac{1-\Delta p}{2}$ for all wrong worlds $\omega_\dagger$. Likewise, as in our theoretical considerations, we assume the opinion leader always votes in favor of all false alternatives.

We went on to compare realistic homogeneous experimental settings: we stipulate a high success probability threshold ($P_{min} = 0.9$) and – for the sake of simplicity – consider voting on only $m = 2$ alternatives. We consider different choices for the OL's competency $\hat{p}$ ranging between 0 and 1, where $\hat{p} = 0$ reflects the situation of a "malicious" OL, who has full information of the true world state but always votes exactly the other way, while $\hat{p} = 1$ describes the case where the OL blindly approves all alternatives, right or wrong. For these settings, we iterated through 25 Δp-values across the unit interval, excluding the prohibited case of $\Delta p = 0$. For each Δp-value, we compared our two bounds against the "empirical" value obtained through repeated simulations of the voting process under the given parameters. This series was recorded for four different values of π (the strength of the OL's influence), relative to the threshold of $\frac{\Delta p}{\Delta p+1}$ established in Section 3.

In our simulations, the number n of required agents was empirically determined by consecutive forward-calculations, where supposed values of n were tested by simulating a high number of voting rounds under the given parameter setting. We used binary stochastic search [10] to speed up the identification of the count of agents empirically needed to surpass the given minimal success probability.

Experiments in Pseudocode. The voting simulations were conducted as depicted by Algorithm 1. Identifying the required number of agents is based on a noisy binary search algorithm with backtracking that was developed by Karp and Kleinberg [10]. The search algorithm is adopted to our setting and given by Algorithm 2. The overarching concept involves performing a random walk on an infinite rooted binary tree whose nodes correspond to tuples (a, b) such that $a, b \in \{0, \ldots, \mathsf{ub}\}$ where ub constitutes the upper bound of agents in the tree. The overall objective is to find the exact number of agents within the set $\{0, \ldots, \mathsf{ub}\}$ such that less agents are unlikely and more agents unnessecary to surpass the minimal prescribed success probability, P_{min}. A specific number of agents, referred to as x in line 3 of Algorithm 2, within a node is used to perform voting simulations, until some particular node passes the termination test (lines 5-17), thereby identifying the number of agents required to surpass P_{min}.

The termination test is subject to probabilistic conditions as specified by Karp and Kleinberg [10] that guarantee to identify the exact number of agents efficiently and reliably. Intuitively, the termination test uses Algorithm 1 like a coin toss, calling the procedure $\lceil s \cdot \ln(\mathsf{ub}) \rceil$ times with a probability of $1/\ln(\mathsf{ub})$. Algorithm 1 simulates $k \cdot \ell$ votings with a specified value for each of Δp, $\hat{p}$, $P_{\min}$ and π, and returns a bit, encoding whether to accept the current $n \in \{a, b\}$, viz., 1, if the worst-case probability surpassed $P_{\min}$, and 0 otherwise. More specifically, Algorithm 1 first constructs the agent's private signals (line 5) and subsequently computes each agent's final votes (lines 6-16). This process also assesses whether each agent follows the OL or relies on their private signal (lines 8 -15). Finally, the algorithm evaluates whether the correct alternative won a single voting simulation (lines 17-18), updates wps (line 20) to reflect the proportion of voting rounds won, and checks if this proportion exceeds P_{min} (line 22).

In the event that a node fails the termination test (evaluated in lines 5-17 in Algorithm 2), two possible outcomes are considered.

Firstly, it is assessed whether backtracking has to take place, indicating that the algorithm has gone down the wrong branch as the current node does not contain the required number of agents. Alternatively, two new nodes are constructed. Backtracking occurs (lines 21-22) if, for a given tuple, (a, b), one of two conditions is met: either the lower number of agents, a, successfully surpasses P_{min} twice during simulations, or the greater number of agents, b, fails to exceed P_{min}. If neither of these conditions holds for a given node, the binary tree construction generates two new nodes (lines 24-28). In this process, the tuple (a, b) is divided into two tuples: (a, c) and (c, b). Lines 3-21 and lines 6-16 in Algorithm 1, and the simulations required in lines 7-9 and line 18 in Algorithm 2 ran in parallel, respectively. $n = 0$ returns were mapped to $n = 1$, as we require at least one agent.

Throughout all experiments we set $r = \hat{p}(\frac{2}{\Delta p^2(\pi-1)^2}) \ln(\frac{2(m-1)}{1-P_{\min}}) + (1-\hat{p})(\frac{2}{(\Delta p(\pi-1)+\pi)^2}) \ln(2(\frac{m-1}{1-P_{\min}}))$. This constitutes the upper

Algorithm 1: Simulations with n agents.

Procedure `simulations`$(n, m, k, \ell, \Delta p, \hat{p}, P_{\min}, \pi)$
 `wps` $\leftarrow \emptyset$; $(\bar{p}^{\omega_*}, \bar{p}^{\omega}) \leftarrow (1/2 + \Delta p/2, 1/2 - \Delta p/2)$;
 for $1 \ldots k$ **do**
  u $\leftarrow$ uniform distribution over $[0, 1]$;
  `ps` $\leftarrow \begin{bmatrix} \bar{p}^{\omega_*} & \bar{p}^{\omega} & \ldots & \bar{p}^{\omega} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{p}^{\omega_*} & \bar{p}^{\omega} & \ldots & \bar{p}^{\omega} \end{bmatrix} \in [0, 1]^{n \times m}$;
  for $1 \ldots \ell$ **do**
   `votes` $\leftarrow [v_{ij}] \in \{0, 1\}^{n \times m}$;
   for $p_{ij} \in$ `ps` **do**
    if $X \in \mathsf{u} < \pi$ **then**
     $v_{ij} \leftarrow$ `signal` **if** $j = 1$ **else** 1;
    end
    else
     $v_{ij} \leftarrow 0$ **if** $X \in \mathsf{u} > p_{ij}$ **else** 1;
    end
   end
  end
  if $\sum_{i=1}^{n} v_{i1} > \max_{2 \leq j \leq m} (\sum_{i=1}^{n} v_{ij})$ **then**
   `wins` $\leftarrow$ `wins` $+ 1$;
  end
  `wps` $\leftarrow$ `wps` $\cup \{\mathtt{wins}/\ell\}$;
 end
 return 1 **if** $\sum_{p \in \mathtt{wps}} p/k > P_{\min}$ **otherwise** 0;

bound for the number of agents in Line 1 of Algorithm 2. Parameters k, ℓ and s varied, depending on observed outliers.

Results and Evaluation. Figure 3 plots, side by side for two different OL competency values, the numbers of agents determined via our simulation-based method (solid lines) as well as the corresponding implicit (dashed) and explicit (dotted) bounds according to Theorem 3. As described above, we display curves for four different π-values (0 and 0.7) and have Δp traverse the interval $(0, 1]$. Below these diagrams, we show plots of the factor by which our bounds are higher than the experimentally determined number of agents.

On one hand, we find that throughout all our experiments, the implicit bound stays in the same order of magnitude as the experimental findings, typically between a factor of 4 and 10, very occasionally up to 20. On the other hand, the explicit bound – while acceptable in many settings – can deviate significantly from the empirical value, in particular for large values of $\hat{p}$ and π. This becomes very apparent in Figure 4, where for a high fixed value of π, the agent estimates and the two bounds are displayed for varying $\hat{p}$. Note that, since the explicit bound does not depend on $\hat{p}$ whatsoever, this bound is displayed as a horizontal line.

These findings indicate that the derived implicit bound performs consistently quite well across all considered parameter settings. Still, it is evident that the bound can be improved.

Algorithm 2: Noisy Binary Search with Backtracking.

In : $m, k, \ell, s \in \mathbb{N}$; $\Delta p, \hat{p}, P_{\min}, \pi, r \in \mathbb{R}$
Out : number of agents $n \in \mathbb{N}$
1 `ub` $\leftarrow r + r/10 + 1$; $(a, b) \leftarrow (0, \mathtt{ub})$; array $\mathsf{T} \leftarrow [(a, b)]$;
2 $o \leftarrow \lceil s \cdot \ln(\mathtt{ub}) \rceil$; u $\leftarrow$ uniform distribution over $[0, 1]$;
3 `result`$(x) \leftarrow$ `simulations`$(x, m, k, \ell, \Delta p, \hat{p}, P_{\min})$;
4 **while true do**
5 **if** $1/\ln(\mathtt{ub}) > X \in \mathsf{u}$ **then**
6   $(r_a, r_b) \leftarrow (0, 0)$;
7   **for** $i = 1 \ldots o$ **do**
8    $(r_a, r_b) \leftarrow (r_a + \mathtt{result}(a), r_b + \mathtt{result}(b))$;
9   **end**
10   $(h_a, h_b) \leftarrow (r_a/o, r_b/o)$;
11   **if** $b = a + 1$ **then**
12   **return** a **if** $h_a < 0.5 < h_b$;
13   **end**
14   **else**
15   **return** x **if** $1/4 \leq h_{x \in \{a,b\}} \leq 3/4$;
16   **end**
17 **end**
18 $(r_a, r_b) \leftarrow$
   $(\mathtt{result}(a) + \mathtt{result}(a), \mathtt{result}(b) + \mathtt{result}(b))$;
19 **if** $r_a = 2$ **or** $r_b = 0$ **then**
20   // backtracking
21   remove last element from T;
22   $(a, b) \leftarrow$ last element in T;
23 **end**
24 **else**
25   $c \leftarrow \lfloor a+b/2 \rfloor$;
26   $(a, b) \leftarrow (a, c)$ **if** `result`$(c) = 1$ **otherwise** (c, b);
27   add (a, b) to T **if** (a, b) is not last element in T;
28 **end**
29 **end**

5 CONCLUSION

In this paper, we generalized the asymptotic part of the Condorcet Jury Theorem to a setting that simultaneously weakens all of the central assumptions underlying the CJT in its original form. In more detail, our framework drops *dichotomy* by allowing to choose from any finite number of alternatives. Also, doing away with the requirement of *completeness*, we allow each agent to vote for any subset of those alternatives. Likewise, allowing for different competence levels across agents, we dropped the *homogeneity* assumption, and – in order to better cope with the diversity of competence levels we thereby permit – we weakened *reliability*, in that our setting allows some agents to be unreliable or even malicious as long as the electorate as a whole is reliable on average. Most importantly, we also relaxed the strict *independence* constraint, by showing that a certain level of correlation between agents' decisions can be tolerated, which we demonstrated via the common technique of postulating an opinion leader, exerting influence on all the agents. The opinion leader can be used to model noise systematically interfering with

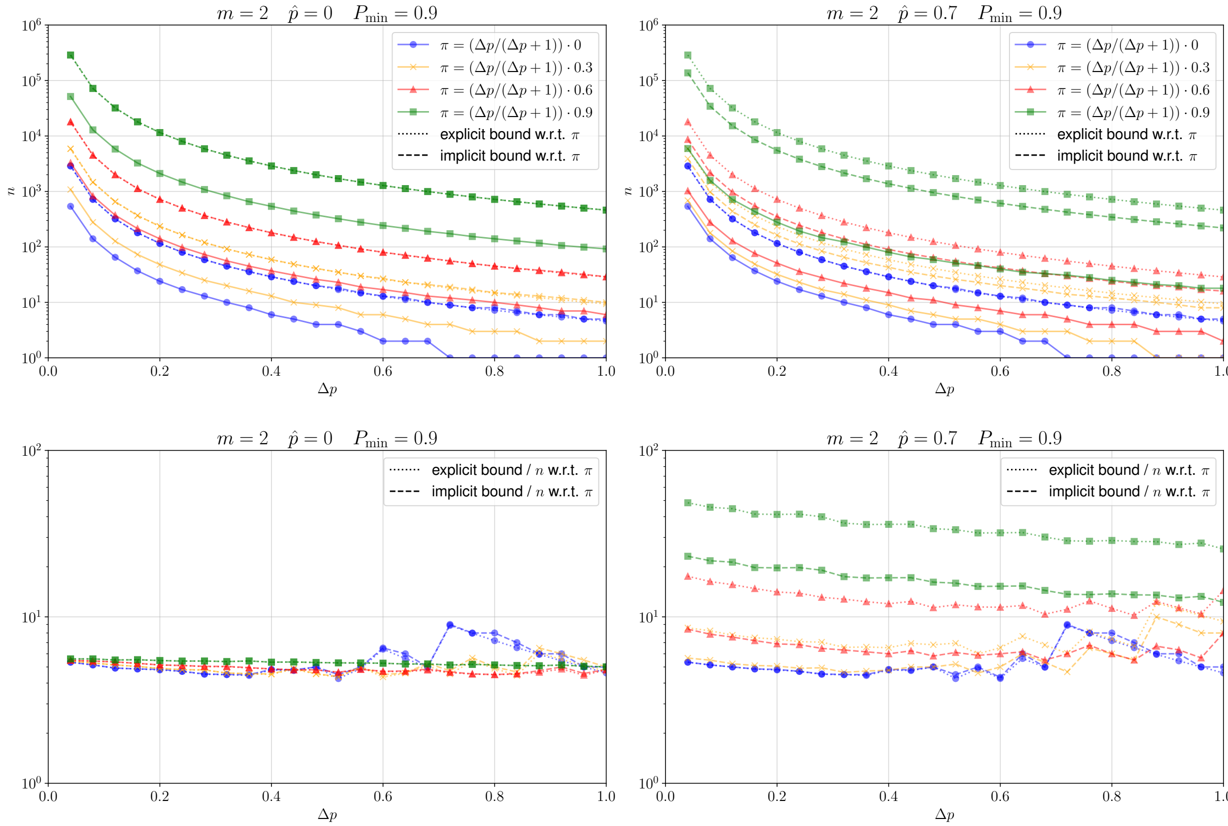

Figure 3: Empirical values and bounds over varying Δp.

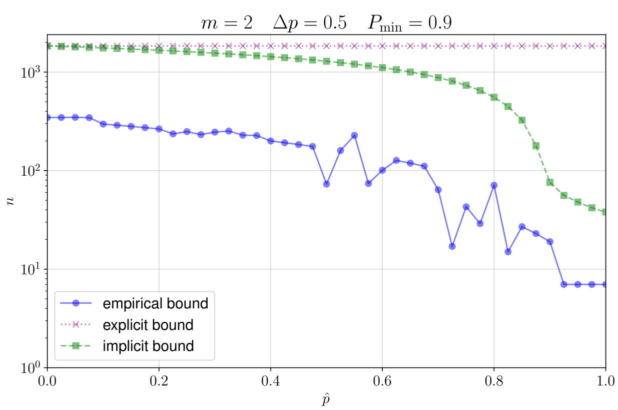

Figure 4: Influence of $\hat{p}$ on the empirical and bound values.

the private signal of the (abstract) agents in the electorate, or even deliberate malicious external influence.

We determined the admissible values for π, that is, the amount of influence the opinion leader may safely exert without breaking the CJT. Finally, we demonstrated by means of statistical simulations that our implicit bound deviates from the experimentally determined values by a factor between 4 and 20, ensuring that throughout all parameter settings, the actual number of agents needed and the conservative worst-case estimate delivered by our findings will be within the same order of magnitude.

Moving forward, we plan to find an improved explicit approximation of the implicit bound given in Theorem 3, better than the explicit bound reported there. To this end, we intend to use approximations by polynomials of an appropriate degree. In a similar vein, we believe that the implicit bound – which, after all, still consistently deviates from the empirical value by a factor of at least 4 – can be further improved by employing better tail probability estimates than the ones currently used, most notably by resorting to a refined variant of Hoeffding's inequality [5].

ACKNOWLEDGMENTS

This work is partly supported by BMBF (Federal Ministry of Education and Research) in DAAD project 57616814 (SECAI, School of Embedded Composite AI, https://secai.org/) as part of the program Konrad Zuse Schools of Excellence in Artificial Intelligence, by the European Research Council through the ERC Consolidator Grant No. 771779 (DeciGUT), and by the BMBF (Federal Ministry of Education and Research), grant 01IS20056_NAVAS.

REFERENCES

[1] Sven Berg. 1994. Evaluation of some weighted majority decision rules under dependent voting. *Mathematical Social Sciences* 28, 2 (1994), 71–83. https://doi.org/10.1016/0165-4896(94)00752-7

[2] Philip J. Boland, Frank Proschan, and Y. L. Tong. 1989. Modelling Dependence in Simple and Indirect Majority Systems. *Journal of Applied Probability* 26, 1 (1989), 81–88. http://www.jstor.org/stable/3214318

[3] Marie Jean Antoine Nicolas Caritat Condorcet. 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie Royale, Paris.

[4] Patricia Everaere, Sébastien Konieczny, and Pierre Marquis. 2010. The Epistemic View of Belief Merging: Can We Track the Truth?. In *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI 2010) (FAIA, Vol. 215)*, Helder Coelho, Rudi Studer, and Michael J. Wooldridge (Eds.). IOS Press, 621–626. https://doi.org/10.3233/978-1-60750-606-5-621

[5] Steven G From. 2022. An improved Hoeffding's inequality of closed form using refinements of the arithmetic mean-geometric mean inequality. *Communications in Statistics-Theory and Methods* (2022), 1–7.

[6] Wassily Hoeffding. 1963. Probability Inequalities for Sums of Bounded Random Variables. *J. Amer. Statist. Assoc.* 58, 301 (1963), 13–30. https://doi.org/10.1080/01621459.1963.10500830

[7] Serguei Kaniovski. 2010. Aggregation of Correlated Votes and Condorcet's Jury Theorem. *Theory and Decision* 69, 3 (2010), 453–468. https://doi.org/10.1007/s11238-008-9120-4

[8] Jonas Karge, Juliette-Michelle Burkhardt, Sebastian Rudolph, and Dominik Rusovac. 2024. *To Lead or to be Led: A Generalized Condorcet Jury Theorem under Dependence (Experiments)*. https://doi.org/10.5281/zenodo.10495388

[9] Jonas Karge and Sebastian Rudolph. 2022. The More the Worst-Case-Merrier: A Generalized Condorcet Jury Theorem for Belief Fusion. In *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning*. 205–214. https://doi.org/10.24963/kr.2022/21

[10] Richard M Karp and Robert Kleinberg. 2007. Noisy binary search and its applications. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Citeseer, 881–890.

[11] Krishna K. Ladha. 1992. The Condorcet Jury Theorem, Free Speech, and Correlated Votes. *American Journal of Political Science* 36, 3 (1992), 617–634. http://www.jstor.org/stable/2111584

[12] Christian List and Robert E. Goodin. 2001. Epistemic Democracy: Generalizing the Condorcet Jury Theorem. *Journal of Political Philosophy* 9, 3 (2001), 277–306. https://doi.org/10.1111/1467-9760.00128 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9760.00128

[13] Guillermo Owen, Bernard Grofman, and Scott L. Feld. 1989. Proving a distribution-free generalization of the Condorcet Jury Theorem. *Mathematical Social Sciences* 17, 1 (1989), 1–16. https://doi.org/10.1016/0165-4896(89)90012-7

[14] Marcus Pivato. 2017. Epistemic democracy with correlated voters. *Journal of Mathematical Economics* 72, C (2017), 51–69. https://doi.org/10.1016/j.jmateco.2017.06

[15] Kai Spiekermann and Robert E. Goodin. 2012. Courts of Many Minds. *British Journal of Political Science* 42, 3 (2012), 555–571. http://www.jstor.org/stable/23274138