Voting for Bins: Integrating Imprecise Probabilistic Beliefs into the Condorcet Jury Theorem

Jonas Karge[1]

[1]*Computational Logic Group, Institute of Artificial Intelligence, Technische Universität Dresden, 01062 Dresden, Germany*

Abstract

In situations where expert committees face significant uncertainty while assessing the likelihood of events, it is crucial to appropriately represent and aggregate their probabilistic opinions. In this work, we adopt the epistemological concept of *imprecise probabilities* to capture an expert's belief and employ the *Condorcet Jury Theorem* (CJT) for aggregating these beliefs through voting. To suit our requirements, we utilize a generalized version of the CJT and integrate imprecise probabilistic beliefs using a framework based on *supervaluationism*, a philosophical theory that addresses vagueness. Drawing inspiration from the field of *expert elicitation* research, we provide an interpretation of the alternatives in the voting process: Experts express their preferences by voting for probability intervals, referred to as *bins*. In this setting, each bin corresponds to a subinterval of the unit interval such that exactly one bin contains the objective probability for the event under evaluation to occur. Finally, we establish a bound on the maximum number of alternatives allowed in an election, which directly corresponds to the achievable precision in the aggregation process.

Keywords

Condorcet Jury Theorem, Imprecise Probabilistic Beliefs, Supervaluationism

1. Introduction

Consider the following four events. There will occur major changes in (i) the Atlantic meridional overturning circulation; (ii) the Greenland ice sheet; (iii) the West Antarctic ice sheet; (iv) the Amazon rainforest and El Nino/Southern Oscillation [1]. What is the probability that at least one of these events will trigger when we face an increase of global mean temperature relative to year 2000 of 2-4 degree Celsius [1]?

Assessing such probabilities is not only crucial for climate policy but also common across various disciplines when determining the likelihood of events under severe uncertainty. Answering questions like the one above often requires aggregating and evaluating the probabilistic beliefs of multiple experts. This latter process is the focal point of a discipline referred to as *expert elicitation*. On a theoretical level, when evaluating events characterized by severe uncertainty, two fundamental questions must be addressed:

(1) How can we appropriately represent the probabilistic beliefs of experts?

(2) What constitutes a reasonable method for aggregation?

KoDis'23: The 2nd International Workshop on Knowledge Diversity, September 3, 2023, Rhodes, Greece
jonas.karge@tu-dresden.de (J. Karge)

In this work, we tackle the first question by modeling experts' beliefs using *imprecise probabilities.* To address the second question, we demonstrate how imprecise probabilistic beliefs can be aggregated through voting, leveraging the *Condorcet Jury Theorem* (CJT). To accomplish this, we initially introduce the original version of the CJT, followed by the generalized variant that is central to our aggregation scenario. Subsequently, we present the imprecise model of belief and illustrate how it can be applied to voting by embedding it into a specific theory of vagueness known as *supervaluationism.* Next, we incorporate imprecise beliefs into the CJT framework, establishing a direct relationship between the number of alternatives experts can vote on and the precision achievable in the aggregation process. Finally, we provide an upper bound on the maximal permissible precision, which depends on the size and competence of the expert committee.

2. The Condorcet Jury Theorem

The CJT is a theoretical cornerstone from voting theory. In a voting setting, a set of agents may vote on a (possibly restricted) set of alternatives. A fundamental assumption of the CJT is that exactly one of those assumptions is correct in that it represents an underlying *ground truth.* The CJT then provides probabilistic guarantees for the capacity of the agents to identify the presumed ground truth through voting. In order to derive these probabilistic guarantees, a few central assumptions are made about the voting setting. In its original variant, it states that for a group of independent, equally competent agents with a better-than-chance probability to identify the true of two alternatives, majority voting best tracks this true state.

Theorem 1 (Marquis de Condorcet 1785)**.** *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 *(non-asymptotic part)*

(2) converges to 1 *as the number of agents goes to infinity* *(asymptotic part).*

Given the inherent limitations and unrealistic assumptions associated with real-world applications of the Condorcet Jury Theorem (CJT), significant research has been devoted to finding generalizations that relax these assumptions while still preserving the core principles. While the asymptotic part of the theorem has been subject to generalizations, it has been demonstrated that the non-asymptotic part fails for small numbers of agents when the assumption of equal voter competence is weakened [3]. Furthermore, this failure persists even for arbitrarily large numbers of voters [4]. The specific generalization of the CJT that we consider for aggregating probabilistic opinions is proposed by Karge and Rudolph 2022. This generalization relaxes many of the aforementioned assumptions. It assumes that the agents in the election are independent but allows for heterogeneous competence levels, including the possibility of totally incompetent voters. Moreover, this generalization permits any finite number of alternatives under approval voting [4].

Our Contribution. In this work, we address a sometimes overlooked, implicit assumption underlying the CJT. Namely, that the agents are assumed to hold binary beliefs where an agent either believes in a proposition and, presumably, votes for it, or does not. However, developing a jury theorem that can aggregate non-binary beliefs, such as probabilistic opinions, remains an open problem [5]. We aim to take a step towards addressing this limitation by incorporating imprecise probabilistic beliefs into the results of Karge and Rudolph 2022.

After providing an intuitive introduction to the CJT and the central generalization in our context, we proceed to present the formal framework of the Karge and Rudolph 2022 generalization in more technical detail. We first elaborate on the approval voting setting, followed by an exposition of the formal probabilistic framework.

Approval Voting. In the approval voting setting, we assume a finite set $\mathcal{W} = \{\omega_1, \ldots, \omega_m\}$ of m items referred to as *alternatives* as well as a finite set $\mathcal{A} = \{a_1, \ldots, a_n\}$ of n *agents*. Note that among the m alternatives there is exactly one alternative that we consider to be *correct*. This correct alternative is denoted by ω_*.

Given $\mathcal{A}$ and $\mathcal{W}$, a single election based on approval voting is given by the relation $V \subseteq \mathcal{A} \times \mathcal{W}$ where $(a_i, \omega_j) \in V$ means that agent a_i approves choice ω_j. Moreover, the *score* $\#_V\, \omega$ of some alternative $\omega \in \mathcal{W}$ is defined as the overall number of votes that ω receives, i.e.,

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

Finally, an alternative $\omega \in \mathcal{W}$ *wins the approval vote* V if it receives strictly higher scores (that is: strictly more votes) than any other alternative, that is,

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

Probabilistic Framework. As usual, the described scenario is modeled as a random process. This process generates ω_* as well as V and is governed by a joint probability distribution $\mathbb{P}$ over the Bernoulli (i.e., $\{0, 1\}$-valued) random variables $X_*^{\omega_1}, \ldots, X_*^{\omega_m}$ as well as $X_i^{\omega_1}, \ldots, X_i^{\omega_m}$ for all agents $1, \ldots, i, \ldots, n$ 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 actual world state (i.e., $\omega_j = \omega_*$), and 0 otherwise, whereas $X_i^{\omega_j}$ is 1 if the ith agent voted for the jth world state (i.e., $(a_i, \omega_j) \in V$) and 0 otherwise.

In a next step, we state the fundamental assumptions necessary for the generalization of the CJT considered here. In the presentation of these assumptions, we let $[\omega_* = \omega_j]$ denote the expression $X_*^{\omega_j}{=}1 \wedge \bigwedge_{\omega \in \mathcal{W} \setminus \{\omega_j\}} X_*^{\omega}{=}0$. The first assumption that needs to be imposed is that the agents do not influence each other in their decision whether or not to approve some choice.

Definition 1. *A joint distribution satisfies* agent approval independence *if, conditioned on the actual world state, the 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:*

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

The second central assumption that we need to impose is that the agents are on average more likely to identify the correct alternative than any competing one.

Definition 2. *A joint probability distribution satisfies Δp-*group reliability *for some $\Delta p > 0$, if the probability 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 $\omega, \omega' \in \mathcal{W}$ with $\omega \neq \omega'$ the following holds:*

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{P}\Big(X_i^{\omega}{=}1 \,|\, [\omega_*{=}\omega]\Big) \geq \Delta p + \frac{1}{n}\sum_{i=1}^{n}\mathbb{P}\Big(X_i^{\omega'}{=}1 \,|\, [\omega_*{=}\omega]\Big).$$

For simplicity, we refer to a voting setting satisfying those two assumptions as *I&R* (for independent and reliable). As final definition for the probabilistic setting, we characterize the chance for the correct alternative to be identified through approval voting.

Definition 3. *Given a family $\mathcal{P}$ of joint probability distributions for n agents and a set $\mathcal{W}$ of m choices, the* approval vote worst-case success probability $P^{\mathrm{wcs}}_{m,n}$ *is defined by*

$$\min_{\substack{\mathbb{P}\in\mathcal{P}\\ \omega\in\mathcal{W}}} \mathbb{P}\Big(\bigwedge_{\omega_\dagger\in\mathcal{W}\setminus\{\omega\}} \textstyle\sum_{k=1}^{n} X_k^{\omega} > \sum_{k=1}^{n} X_k^{\omega_\dagger} \mid [\omega_* = \omega]\Big).$$

With that, we can more formally restate the generalization of the CJT that is central to this work:

Theorem 2 (Generalized CJT)**.** *In any I&R setting with fixed $m \geq 2$ and $\Delta p > 0$ holds $P^{\mathrm{wcs}}_{m,n} \xrightarrow[n\to\infty]{} 1$ [4].*

In words, for any independent and reliable setting with at least two alternatives and a positive Δp-value the worst-case probability for the correct alternative to be identified converges to 1 as the number of agents goes to infinity.

Beyond the convergence behavior in the infinite, it is possible to derive concrete guarantees in the finite. Exploiting Hoeffding's inequality [6] as well as the Chebyshev-Cantelli inequality [7, 8], two ways were established in Karge and Rudolph 2022 to bound the number n of agents needed to guarantee a success probability $P_{\min}$ when choosing among m options. We take the minimum of both.

Theorem 3. *In a Δp-group reliable setting with m choices, the worst case approval vote success probability is at least $P_{\min}$ whenever the number of agents is equal or higher than*

$$\min\Big(\tfrac{2}{\Delta p^2}\ln Q, 1 + \big(\tfrac{1}{\Delta p^2} - 1\big)Q\Big), \tag{1}$$

where $Q = 2\frac{m-1}{1-P_{\min}}$ is the twofold ratio between the number of incorrect alternatives and the admissible error probability.

This bound on the required number of agents will play a crucial role in a later part of this paper, where it will be utilized to estimate the maximal permissible precision when aggregating probabilistic opinions through voting.

In essence, the CJT is a mathematical theorem that provides asymptotic guarantees for effectively identifying an underlying truth through voting. The theorem has been generalized to accommodate more realistic assumptions and enable practical estimates of the number of agents required to track the truth. In the subsequent sections, we establish the groundwork for the generalization that facilitates the aggregation of probabilistic beliefs by introducing a formal model for representing those beliefs.

3. Imprecise Probabilities and Voting

In this section, we first introduce the imprecise probability model of an agent's belief that will be used in our setting. Second, we illustrate how imprecise probabilistic beliefs can be used in decision-making. For this latter objective, we embed the underlying belief model into a philosophical account for vagueness that is referred to as *supervaluationism*. The presentation of both frameworks builds upon and extends the one in Karge 2023.

3.1. Probabilistic Beliefs

When it comes to representing probabilistic beliefs, the classical approach is to use a single probability function to capture an agent's confidence in various propositions. This function assigns a real number between 0 and 1 to each proposition, reflecting the agent's degree of belief in that proposition. This approach can be defined as follows:

Definition 4 (Probability Function)**.** *A probability function $\mathbb{P}$ is a function $\mathbb{P} : 2^{\Omega} \rightarrow \mathbb{R}$, satisfying the probability axioms [10].*

Here, Ω represents a set of possible worlds or states of affairs, and propositions are subsets of Ω. However, when evaluating propositions with severe uncertainty, such as the example of proposition A, which will serve as our running example, it becomes implausible to represent belief states with a single probability function [11].

Example 1 (Proposition A)**.** *Global sea level will rise at least 1,5 meters until the year 2100 above the level of 2000.*

The question arises: What precise probability should the agent assign to proposition A? The traditional approach of representing an agent's belief with a single probability function requires a specific value to be given [12]. However, for propositions characterized by severe uncertainty, using a single probability function seems highly implausible [11]. An alternative approach is to define degrees of belief based on imprecise probabilities:

Definition 5 (Imprecise Probabilities)**.** *Imprecise probabilities are sets of probability functions [13].*

We refer to a specific set of probability functions as the agent's *representor*, denoted by $\mathcal{P}$ [13]. To represent the range of values assigned by the representor to a proposition more compactly, we introduce the concept of an *imprecise degree of belief*:

Definition 6 (Imprecise Degree of Belief)**.** *An agent's imprecise degree of belief in a proposition H is represented by a function, $\mathcal{P}(H)$, with $\mathcal{P}(H) = \{\mathbb{P}(H) : \mathbb{P} \in \mathcal{P}\}$ [14].*

When experts evaluate statements such as proposition A, imprecise probabilities aim at reflecting the unspecific nature of the evidence that experts typically receive. Intuitively, this is achieved by considering all values that are not excluded by the evidence [15]. Following this line of reasoning, an additional assumption is often made when modeling an agent's imprecise degree of belief. More formally, it is assumed that the set of probability functions in the agent's representor is *convex*:

Definition 7. *A set C is convex if, for all $x, y \in C$ and all $t \in [0,1]$ it holds that the point $[(1-t)x + ty] \in C$ [16].*

Convexity implies that for any two points in a set, the line segment connecting them is also within the set. Under this assumption, an agent's imprecise degree of belief can be represented as intervals [16] where the agent's belief is spread over all values not excluded by the evidence she may have. This leads to the following representation of imprecise degrees of belief illustrated at the example of proposition A:

Example 2. *Assume, the agent's representor consists of three probability functions that assign event A values from the set $\{0.4, 0.6, 0.8\}$. By convexity, we may represent the agent's imprecise degree of belief with $\mathcal{P}(A) = [0.6, 0.8]$. Thus, our agent is $60 - 80\%$ confident that event A will occur, i.e., that proposition A is true.*

To summarize, we model an agent's probabilistic beliefs using imprecise probabilities, where the agent assigns an imprecise degree of belief (a sub-interval of the unit interval) to each proposition, reflecting their confidence. Unlike precise probabilistic beliefs, this imprecise model allows for a more natural representation of scenarios involving severe uncertainty. After formalizing the model of an agent's beliefs based on imprecise probabilities, we incorporate imprecise probabilistic beliefs into *supervaluationism*, which provides a convenient framework for handling various decision-making scenarios, including voting, based on this belief model.

3.2. Supervaluationism

Consider a vague predicate such as *tall*. Vague predicates can be made more precise by introducing cutoff points. For example, one person may consider *being tall* to mean a height of at least 180cm, while another person may set the cutoff at 185cm. Each cutoff point represents a precisification of the predicate [17]. When evaluating the truth value of a proposition that contains a vague predicate, supervaluationism requires complete agreement among the precisifications regarding its truth value [18]. In supervaluationism, *complete agreement* refers to a proposition being either *determinately true* or *determinately false*.

In standard supervaluationism, a proposition is deemed *determinately true* if it is true according to all admissible precisifications. Conversely, a proposition is considered *determinately false* if it is false according to all admissible precisifications. Additionally, supervaluationism allows for propositions to have no semantic value, meaning that if a proposition is true according to some, but not all, admissible precisifications, it is considered *indeterminate* whether it is true [12].

In an extension of supervaluationism, referred to as *modified supervaluationism* (MSV), the notion of propositions being *determinately true* is replaced by propositions being *predominantly true*:

Definition 8 (Predominantly True). *A proposition is predominantly true if it is true according to a relative majority of admissible precisifications [19].*

Both in the standard and modified supervaluationistic frameworks, imprecise probabilities can be effectively captured by appropriately defining admissible precisifications [12]. Working on convex sets for the imprecise degree of belief, we define admissible precisifications to be the probability values in the imprecise degree of belief reported by the agent.

To establish a coherent model of belief, the combination of supervaluationism and imprecise probabilities must provide a sound concept for comparing the confidence an agent has in different propositions.

Definition 9 (Predominant Confidence). *Given two propositions, A and B, an agent is considered to be predominantly more confident in proposition A than in proposition B if a greater proportion of elements within the agent's imprecise degree of belief satisfy the condition $Pr(A) > Pr(B)$.*

This definition allows us to compare the relative confidence an agent has in different propositions based on their respective probability values within the agent's imprecise degree of belief.

Observe that determining whether an agent is predominantly more confident in a proposition than in another involves measuring a potentially uncountable number of probability values, which are the precisifications derived from the agent's imprecise degree of belief. In order to achieve this, we note that in our setting, admissible precisifications are confined to the unit interval $[0, 1]$. The standard way to measure the length of an interval is to apply the *Lebesque Measure*. For any closed, $[a, b]$, open, (a, b), or half open, $(a, b]$ or $[a, b)$, interval it holds that its Lebesque measure is of length $l = b - a$. Applying the Lebesque measure to MSV, we can determine the proportion of elements in favor of a proposition by measuring the length of the corresponding interval.

Example 3. *Consider proposition A and its complement B, i.e. global sea level will* ***not*** *rise at least 1,5 meters until the year 2100 above the level of 2000. Suppose we have $\mathcal{P}(H) = [0.4, 1]$ as our agent's imprecise degree of belief. For those elements represented by $(0.5, 1]$ it holds true that $Pr(A) > Pr(B)$. For those represented by $[0.4, 0.5)$ we have $Pr(B) > Pr(A)$. Taking their Lebesque measure, we receive $l(A) = 0.5$ as well as $l(B) = 0.1$. Thus, the agent is predominantly more confident in proposition A.*

With the MSV interpretation of imprecise probabilities, our next objective is to incorporate this model into voting scenarios. Having defined the voting setting and established how imprecise probabilities can be integrated into modified supervaluationism, we now informally describe how agents vote based on their imprecise beliefs.

Definition 10. *Given a set of alternatives $\mathcal{W} = \{\omega_1, \ldots, \omega_m\}$ and set of agents $\mathcal{A} = \{a_1, \ldots, a_n\}$, agent a_i approves alternative ω_j if the agent is predominantly more confident in that alternative than in its competitors.*

However, to provide a more formal definition, we need to specify what constitutes an *alternative* in our setting. This will be addressed in the next section, where we incorporate voting based on imprecise beliefs into the CJT framework.

4. Embedding

Upon establishing an agent's imprecise beliefs and presenting an approach to voting based on those beliefs through modified supervaluationism, as well as delineating the CJT and its underlying voting setting with its probabilistic assumptions, we now integrate these concepts to incorporate imprecise probabilistic beliefs into the CJT voting setting.

4.1. The Alternatives

To recap our main goal, we aim to determine the likelihood of a statement, such as proposition A, in our framework. We have a finite set of agents, each holding an imprecise probabilistic belief in that proposition. These beliefs will be aggregated through voting, where each agent votes for the alternative they are predominantly confident in. In the final step, the alternative that receives the most votes wins the election and represents the aggregated probabilistic assessment of the event, such as proposition A. Following this general aggregation procedure, each alternative represents a probabilistic assessment of the event by itself.

As discussed earlier a central assumption underlying the CJT is that one of the given alternatives is correct. What exactly does that mean in our setting where we aim at aggregating probabilistic opinions? A first idea is: an alternative represents a precise probability for an event to occur assuming an event has an underlying, objective probability. Then, given a set of alternatives where each represents a single probabilistic value, and given the imprecise degrees of beliefs of the agents, these agents vote for one or multiple of the single, precise values based on their beliefs.

However, in real-world applications, agents typically vote for a limited and finite set of alternatives. This introduces a problem with respect to an agent's degree of belief, as probabilities can be arbitrarily precise. If we were to model alternatives as representations of single probabilistic values, the assumption that the correct probability for the event is among the set of alternatives chosen for an election would appear highly unrealistic. To address this issue, we model the alternatives themselves as imprecise probabilities, where each alternative represents an interval of probability values within which the correct probability for an event lies.

The formal modeling of alternatives is detailed in the next subsection and is drawing inspiration from the field of *expert elicitation* research.

Expert Elicitation. In typical expert elicitation frameworks, an expert's belief in a proposition is represented as a precise, subjective probability that characterizes their belief. However, recent advancements have incorporated imprecise probabilities in the form of intervals of subjective probabilities into the analysis of an expert's competence [1]. This approach aims to exclude precise probabilities that are incompatible with the expert's belief, aligning with the concept of distributing an agent's belief over all probabilities not excluded by the evidence, as discussed earlier.

One method for eliciting an agent's imprecise probabilistic beliefs is known as the Linear Programming Imprecise Probabilities Model (LPIPM). Given some unknown real-valued continuous random variable X, applying LPIPM, aims at eliciting an expert's prior distribution of X [20]. For this purpose, the expert is asked to internally evaluate and then announce minimum and maximum values for X that she considers plausible. These values define an interval of the form $[X_{min}, X_{max}]$ within which the expert believes the true value of the variable X lies, with the understanding that the expert assigns a probability of zero to the true value of X lying outside this interval [20].

In the original LPIPM approach, the interval $[X_{\min}, X_{\max}]$ is further divided into subintervals. The expert then evaluates these subintervals through pairwise comparisons, aiming to obtain a more precise estimate of the agent's imprecise degree of belief by eliminating subintervals that the expert deems less likely compared to others. This iterative process allows for a refinement of the expert's belief, resulting in a more nuanced understanding of their imprecise probabilities.

In our setting, we adopt the idea of partitioning an interval into subintervals, but we apply it to the level of alternatives that experts can vote for in an election. More precisely, in an election where experts have to evaluate the likelihood of an event such as proposition A, we regard as simplest set of alternatives the (theoretically excluded, but conceptually noteworthy) case where we have $m = 1$ with $\omega_1 = [0, 1]$. In this case, the only alternative experts can vote for is the entire unit interval. To generate scenarios with more alternatives, we partition the unit interval into $2k$ subintervals of equal Lebesgue measure, denoted as $[X_{j-1}, X_j)$, where $j = 1, 2, ..., 2k$. The value of k depends on the desired precision [20]. For example, if a small group of agents is highly reliable in their estimates, a moderate precision of 10% may be adequate, achieved by setting $k = 5$ (yielding $2k = 10$) and providing 10 subintervals of equal Lebesgue measure.

The division of the unit interval into subintervals is reminiscent of a common practice in the evaluation of probabilistic forecasts known as the binned probability ensemble (BPE) technique. In BPE, a set of forecasts is used to divide the real line into several *bins*, assuming that the true forecast probability falls within one of the bins [21]. In our case, we apply a similar concept to the unit interval, considering each subinterval as an alternative to vote on, referred to as a *bin*.

Voting for Bins. To define the alternatives in the approval vote as bins, representing subintervals of the unit interval as probability values, we proceed to formally outline how agents vote based on their imprecise degrees of belief and MSV.

In the original MSV framework, the notion of *predominant confidence* determined the proposition in which an agent was most confident among *multiple* propositions, considering their imprecise degrees of belief in each proposition.

However, in our voting setting, we focus on an agent's imprecise degree of belief for a *single* proposition. The agent's confidence must then be evaluated with respect to multiple bins by comparing them to her imprecise degree of belief. These bins, in turn, represent probabilistic assessments for the same proposition as the imprecise degree of belief. As the imprecise degree of belief provides an estimate of the likelihood of a proposition being true but not for another probabilistic estimate for the same proposition, we extend the concept of *predominant confidence* to this scenario. Intuitively, we model *predominant confidence* in probabilistic estimates for the same proposition as the imprecise degree of belief as the largest *agreement* between a bin and the imprecise degree of belief. Since both the bin and the belief are probability intervals, we quantify *agreement* as the largest Lebesgue measure between the intersection of the degree of belief and each bin.

Definition 11 (Predominant Confidence - Bins)**.** *Let A be a proposition, $\mathcal{P}(A) = [a, b]$ be an agent's imprecise degree of belief in A, and let $[X_{j-1}, X_j)$, $j = 1, 2, ..., 2k$ be $2k$ bins defined on the unit interval reflecting probability values for A to occur. Given two bins B_1 and B_2, we say that an agent is predominantly more confident in B_1 if the intersection of $\mathcal{P}(A)$ and B_1 is of greater Lebesque measure than the one of $\mathcal{P}(A)$ and B_2. That is, $l(\mathcal{P}(A) \cap B_1) \geq l(\mathcal{P}(A) \cap B_2)$.*

Consider the following example:

Example 4. *Suppose there are only two bins for proposition A with* $B_1 = [0, 0.5)$ *and* $B_2 = [0.5, 1]$ *and let* $\mathcal{P}(A) = [0.3, 0.9]$. *We then have* $\mathcal{P}(A) \cap B_1 = [0.3, 0.5)$ *and* $\mathcal{P}(A) \cap B_2 = [0.5, 0.9]$. *This results in* $l(\mathcal{P}(A) \cap B_1) = 0.2$ *as well as* $l(\mathcal{P}(A) \cap B_2) = 0.4$. *Thus, the agent is predominantly more confident in the second bin.*

Finally, we can define how agents vote based on their imprecise degrees of belief:

Definition 12. *Let* $m = \omega_1, ..., \omega_m$ *be a set of alternatives where each* ω_i *represents a bin of the form* $[X_{j-1}, X_j)$. *Moreover, let* $a_1, ..., a_n$ *be a set of agents and let* V *represent a single election. We say that an agent* a_i *votes for an alternative* ω_j *if she is predominantly confident in that alternative. That is, if* $(l(\mathcal{P}(A) \cap \omega_j) \geq l(\mathcal{P}(A) \cap \omega_k))$ *for all* $j \neq k$ *then* $(a_i, \omega_j) \in V$.

Note that, if the agent is predominantly confident in multiple alternatives (i.e., their intersections are of equal Lebesque measure), she votes for all of them. In particular, if an agent reports the whole unit interval as her imprecise degree of belief, she votes for all alternatives in the approval vote.

4.2. The Probabilistic Assumptions

Having formally defined how agents vote based on their imprecise degree of belief as well as having specified what constitutes an alternative in our setting, we can now combine the imprecise model of belief with the CJT.

To recap, we revisit the previously described joint probability distribution: $X_*^{\omega_1}, ..., X_*^{\omega_m}$ as well as $X_i^{\omega_1}, ..., X_i^{\omega_m}$ for all agents $1, ..., i, ..., n$. While the distribution remains the same, our interpretation of how the random variables take their values changes. Recall that $\omega_1, ..., \omega_m$ represent subintervals of the unit interval and let p^{ω_*} be the true objective probability of an event to occur such that $p^{\omega_*} \in \omega_j$ for exactly one j. 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 bin containing the actual world state (i.e., $p^{\omega_*} \in \omega_j$) and 0 otherwise, whereas $X_i^{\omega_j}$ is 1 if the ith agent voted for the jth bin (i.e., $(l(\mathcal{P}(A) \cap \omega_j) \geq l(\mathcal{P}(A) \cap \omega_k))$ for all $j \neq k$) and 0 otherwise.

With this interpretation of the random variables in the joint probability distribution, as given by Definition 12, we can directly utilize the probabilistic assumptions we mentioned earlier: agent approval independence and Δp-group reliability. Let's provide an intuitive rephrasing of these notions in the language of our setting. First, agent approval independence translates to the idea that for any bin and any agent, the fact that one agent, a_i, is predominantly more confident in a particular bin, B_j, than in all other bins, does not influence whether a different agent, a_k, is also predominantly more confident in B_j. In other words, the confidence of one agent in a bin does not affect the confidence of another agent in the same bin. Second, according to Δp-group reliability, on average, the agents are by the value of Δp more likely to be predominantly more confident in a particular bin than in any other bin. This notion captures the idea that the agents exhibit a certain level of consistency in their confidence assessments, favoring the bin containing the correct probability value over others by a small margin of Δp.

Reasons for Approval Voting. The interpretation of the voting setting and probabilistic framework outlined above highlights why approval voting is a suitable choice.

Let's consider a simpler voting rule like *majority voting*, as in the original CJT. In majority voting, the number of alternatives is limited to two, and agents can vote for only one of them. This voting rule is already unsuitable for our purposes due to its first restriction. Since each alternative corresponds to a bin, which is a subinterval of equal Lebesgue measure on the unit interval, using majority voting would result in a coarse-grained probabilistic aggregation method that only considers two bins, namely $[0, 0.5)$ and $[0.5, 1]$.

Now, let's explore a setting that allows for any finite number of alternatives, but with the restriction that agents can only vote for one alternative, known as *plurality voting*. Recall that an agent votes for an alternative if the Lebesgue measure of the intersection between that bin and the agent's imprecise degree of belief is the largest. Suppose we aim for reasonable high precision, such as 5%, which translates to 20 bins: $Bin_1 = [0, 0.05), Bin_2 = [0.05, 0.1), ..., Bin_{20} = [0.95, 1]$. Let's assume that the correct probability value falls within Bin_2, meaning $p^{\omega_*} \in [0.05, 0.1)$. If we restrict the voting setting to allow agents to vote for only one bin while simultaneously requiring a reasonably high Δp value, we inadvertently impose a limitation on the agents' imprecise degree of belief. In order for an agent to vote for the correct alternative in the plurality setting (i.e., Bin_2 in this case), their imprecise degree of belief must have less overlap with the other bins compared to Bin_2, ensuring that the Lebesgue measure of their intersection is not exceeded by the other bins. This restriction might exclude agents who abstain from making any judgment and report the whole unit interval as their imprecise degree of belief because they would be predominantly confident in all bins. To avoid this artificial restriction, we allow agents to vote for any subset of the finite set of alternatives, which aligns with the concept of approval voting.

Additionally, the embedding presented here offers a small advantage in addressing a conceptual counterargument sometimes raised against the CJT framework. Critics argue that it is unrealistic to assume that the correct alternative is always among the options provided to the agents [22]. However, in our setting, this assumption is naturally satisfied. The correct probability for an event to occur must be an element of the unit interval, which is covered by the partitions into bins. Therefore, the embedding ensures that the correct alternative is always within the set of alternatives considered by the agents.

Having embedded imprecise probabilistic beliefs into the CJT framework, we may moreover invoke the aforementioned estimate on the number of agents required to guarantee a prescribed minimal success probability for the correct alternative to win. Furthermore, given the established correspondence between the number of subintervals of the unit interval and the number of alternatives in the approval vote, which, in turn, reflects the desired precision, we subsequently illustrate how to derive an estimate on the maximal permitted precision.

4.3. Estimating the maximal permitted precision.

In typical scenarios involving the combination of probabilistic expert opinions, the number of voters on the expert board is often known beforehand. Examples of such studies include Kriegler et al. 2009, who investigated the beliefs of 42 climate scientists, and Recchia et al. 2021, who analyzed the forecasting capacities of 140 experts, including epidemiologists and statisticians, regarding COVID-19 outbreaks in the UK. Since the size of expert committees in real-world applications is often fixed, it is not only important to estimate the minimum number of agents necessary to ensure a prescribed minimal success probability, as previously discussed [4], but also to determine the maximal precision that can be allowed. Given a specific Δp-value and a known target minimal success probability, we can derive an upper bound on the maximal permitted precision using Theorem 3. To accomplish this, we consider that (i) each subinterval of equal Lebesgue measure corresponds to a unique alternative denoted by m in the voting process, and (ii) in Theorem 3, we take the minimum of two separate bounds. Therefore, to estimate the maximal allowed precision, we first solve each bound for m individually. As mentioned earlier, we refer to the first bound as the *Hoeffding bound*, and to the second as the *Chebychev-Cantelli* bound, named after the inequalities from which they were derived.

The Hoeffding bound is given by

$$1-(m-1)2e^{-\frac{1}{2}n\Delta p^2} \geq p_{\min}$$

$$(m-1) \leq \frac{(1-p_{\min})}{(2e^{-\frac{1}{2}n\Delta p^2})}$$

Note that $(2e^{-\frac{1}{2}n\Delta p^2})$ is non-zero for all possible values.

$$m \leq \frac{(1-p_{\min})}{(2e^{-\frac{1}{2}n\Delta p^2})} + 1.$$

The Chebychev-Cantelli bound is given by

$$1-\frac{2(m-1)(1-\Delta p^2)}{1+(n-1)\Delta p^2} \geq p_{\min}$$

$$2(m-1)(1-\Delta p^2) \leq (1-p_{\min})(1+(n-1)\Delta p^2)$$

Note that we need to exclude $\Delta p = 1$ for ensuring

that $2(1-\Delta p^2)$ is non-zero.

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

Observe that excluding $\Delta p = 1$ for the Chebychev-Cantelli bound may come as a surprise, especially considering that this bound performed extremely well in the original setting of Karge and Rudolph 2022 where $\Delta p = 1$. Specifically, when estimating the minimal number of agents for any probability of success, it gave the intuitive result that only one agent is needed in the approval vote. However, in our setting, we are providing an upper bound on the number of alternatives. When $\Delta p = 1$, we expect any precision to be allowed since every agent exclusively votes for the correct bin. Thus, in this scenario, there is no finite upper bound as we permit an arbitrary number of bins, which means the number of alternatives can be infinite. Interestingly, when $\Delta p = 1$ and we allow for an unlimited number of bins, we obtain a special case where agents vote not for bins but for precise probability values.

Given this, both bounds provide an estimate for the maximum value of m. In contrast to Theorem 3, we now take the maximum of both bounds to determine this maximum value.

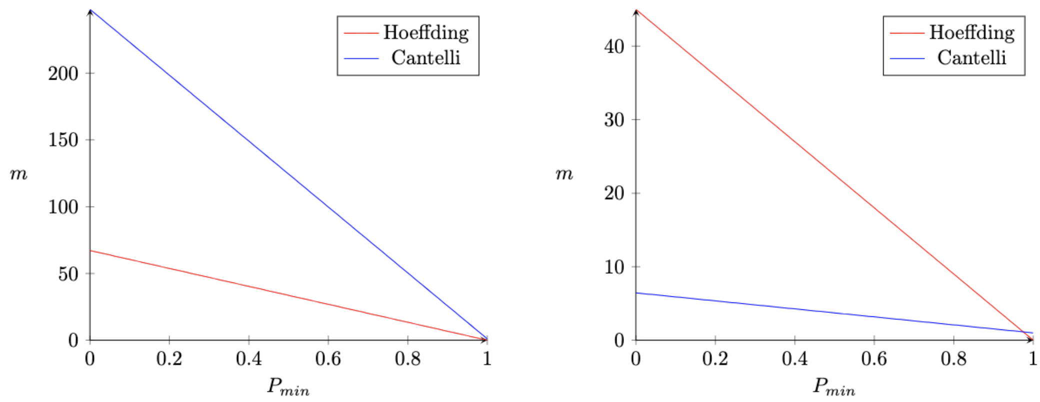


Figure 1: Maximal number of permitted bins for $\Delta p = 0.99, n = 10$ (left) and $\Delta p = 0.3, n = 100$ (right).

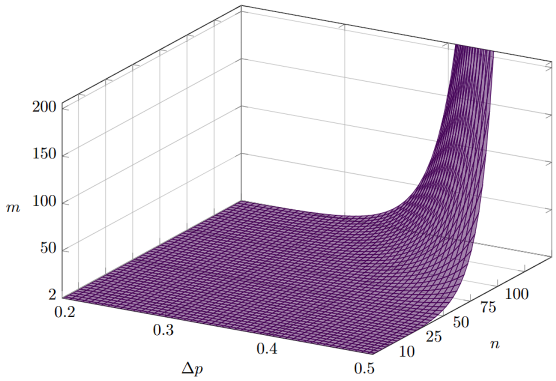
Theorem 4. *In a Δp-group reliable setting where $\Delta p \in (0, 1)$ with n agents, the worst case approval vote success probability is at least $P_{\min}$ whenever the number of alternatives is equal or lower than*

$$max(\frac{(1-p_{\min})}{(2e^{-\frac{1}{2}n\Delta p^2})} + 1, \frac{(1-p_{\min})(1+(n-1)\Delta p^2)}{2(1-\Delta p^2)} + 1). \tag{2}$$

The result obtained from the bounds yields the maximum number of alternatives that guarantee a prescribed minimal success probability $P_{\min}$ for a specific number of agents and a given average reliability assessment represented by Δp (excluding $\Delta p = 1$). This directly translates into the maximal allowed precision in percentage, denoted by C. We define C as the proportion of the unit interval covered by a subinterval, given by $C = \frac{100}{m}$. Therefore, the closer C is to 100, the more imprecise the result of the election will be, while a value closer to 0 indicates higher precision. For example, when $m = 20$, we have a precision of $C = 5\%$, and when $m = 2$, we have $C = 50\%$. It is important to note that when we mention that precision increases, it means that the result of the aggregation procedure will be more precise, even though the precision value in percentage decreases.

Furthermore, it is worth observing that neither of the two bounds dominates the other for all values. This is illustrated in Figure 1, where we consider a fixed average group reliability and number of agents and plot the maximum number of permitted bins against the prescribed minimal success probability. Specifically, for a small number of agents but extremely high Δp-values, the Cantelli bound provides a good estimate, while the Hoeffding bound performs better for small and moderate Δp-values.

To gain a more nuanced understanding of the maximal allowed precision, we refer to Figure 2. This graph illustrates the number of bins as a function of moderate to relatively high Δp-values (0.2-0.5) and small to moderate-sized expert groups (10-100) for a fixed probability of success of $P_{\min} = 0.9$. Several observations can be made from the graph: First, small expert groups require relatively high average competence levels to achieve meaningful precision values for



Selection of data points			
Number of Experts	Δp	Number of Bins	Precision
50	0.3	< 2	N.A.
50	0.4	4	25%
75	0.3	2	50%
75	0.4	21	4.8%
100	0.3	6	16%
100	0.4	150	$0.\bar{6}\%$
150	0.3	44	2.2%
150	0.4	8139	0.01%
200	0.3	406	0.25%
200	0.4	444307	0.0002%

Figure 2: Maximal number of permitted bins for $P_{min} = 0.9$ and varying Δ and $n = 10$ (left) as well as a selection of data points (right).

aggregation. For instance, in order to achieve a precision below 10%, more than 100 agents are needed for a moderate competence margin of $\Delta p = 0.3$. Second, there is a range of values, including $n = 50$ and $\Delta p = 0.3$, where the maximum permitted number of bins is less than two. These values indicate that the average competence margin and number of agents are not sufficient to realize the prescribed success probability. In such cases, we state that the calculated precision is not available since the CJT requires at least two alternatives. Third, the permitted precision grows extremely rapidly once the number of experts and their respective competencies are sufficient to achieve the prescribed success probability. For instance, a group of 150 experts with a competence margin of 0.4 is already sufficient to realize a precision of 0.01%. Increasing the number of experts to 200 for the same Δp value results in a precision of 0.0002%.

5. Summary and Future Work

In this work, we embedded imprecise probabilistic beliefs into a generalization of the Condorcet Jury Theorem that allows for approval voting and heterogeneous competence levels. For this purpose, we combined the epistemological account of imprecise degrees of belief as well as their interpretation in the supervaluationinstic theory of vagueness with a voting setting where each alternative represents an interval of probability assessments (a bin) for the same proposition as the imprecise degree of belief. Furthermore, we established a direct correspondence between the number of bins in the voting process and the maximal permitted precision during aggregation. By solving existing bounds for the number of alternatives, we were able to provide an estimate for the allowed precision in the aggregation procedure. Moving forward, it would be valuable to compare the performance of voting-based aggregation of imprecise probabilistic beliefs with traditional methods of probabilistic opinion pooling. The latter typically involves combining expert opinions through weighted averaging of probabilities to approximate the objective probability of an event. In contrast, our approach assumes the objective probability lies within one of the bins, providing guarantees for hitting that specific bin with a certain probability.

Acknowledgments

Jonas Karge was supported by the Bundesministerium fur Bildung und Forschung (BMBF, Federal Ministry of Education and Research) in the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI).

References

[1] E. Kriegler, J. W. Hall, H. Held, R. Dawson, H. J. Schellnhuber, Imprecise probability assessment of tipping points in the climate system, Proceedings of the National Academy of Sciences 106 (2009) 5041–5046.
[2] M. J. A. N. C. Marquis de Condorcet, Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix, Imprimerie Royale, Paris, 1785.
[3] G. Owen, B. Grofman, S. L. Feld, Proving a distribution-free generalization of the Condorcet Jury Theorem, Mathematical Social Sciences 17 (1989) 1–16.
[4] J. Karge, S. Rudolph, 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, 2022, pp. 205–214.
[5] F. Dietrich, K. Spiekermann, Independent opinions? on the causal foundations of belief formation and jury theorems, Mind 122 (2013) 655–685.
[6] W. Hoeffding, Probability inequalities for sums of bounded random variables, Journal of the American Statistical Association 58 (1963) 13–30.
[7] P. L. Chebyshev, Des valeurs moyennes, Liouville's Journal de Mathématiques Pures et Appliquées 12 (1867) 177–184.
[8] F. P. Cantelli, Sui confini della probabilita, Atti del Congresso Internazional del Matematici 6 (1928) 47 – 59.
[9] J. Karge, Second-order confidence in supervaluationism, Journal for General Philosophy of Science (2023) 1–16.
[10] A. Mahtani, Imprecise probabilities, in: R. Pettigrew, J. Weisberg (Eds.), The Open Handbook of Formal Epistemology, PhilPapers Foundation, 2019, pp. 107–130.
[11] M. Schoenfield, Chilling out on epistemic rationality: A defense of imprecise credences (and other imprecise doxastic attitudes), Philosophical Studies 158 (2012) 197–219.
[12] S. Rinard, A decision theory for imprecise probabilities, Philosopher's Imprint 15 (2015).
[13] S. Bradley, K. Steele, Should subjective probabilities be sharp?, Episteme 11 (2014) 277–289.
[14] S. Bradley, How to choose among choice functions, Episteme 11 (2014) 277–289.
[15] J. M. Joyce, How probabilities reflect evidence, Philosophical Perspectives 19 (2005) 153–178.
[16] K. Steele, Distinguishing indeterminate belief from "risk-averse" preferences, Synthese 158 (2007) 189–205.
[17] R. Keefe, Vagueness: supervaluationism, Philosophy Compass 3 (2008) 315–324.
[18] A. C. Varzi, Supervaluationism and its logics, Mind 116 (2007) 633–676.
[19] J. Karge, A modified supervaluationist framework for decision-making, Logos & Episteme 12 (2021) 175–191.

[20] A. A. Silva, F. C. de Souza, A protocol for the elicitation of imprecise probabilities., in: ISIPTA, volume 1, 2005, pp. 315–321.

[21] J. L. Anderson, A method for producing and evaluating probabilistic forecasts from ensemble model integrations, Journal of climate 9 (1996) 1518–1530.

[22] M. Pivato, Epistemic democracy with correlated voters, Journal of Mathematical Economics 72 (2017) 51–69.

[23] G. Recchia, A. L. Freeman, D. Spiegelhalter, How well did experts and laypeople forecast the size of the covid-19 pandemic?, PloS one 16 (2021).