# The Epistemic View of Belief Merging: Can We Track the Truth?

**Abstract.** Belief merging is often described as the process of defining a belief base which best represents the beliefs of a group of agents (a profile of belief bases). The resulting base can be viewed as a synthesis of the input profile. In this paper another view of what belief merging aims at is considered: the epistemic view. Under this view the purpose of belief merging is to best approximate what is the true state of the world. We point out a generalization of Condorcet's Jury Theorem from the belief merging perspective. Roughly, we show that if the beliefs of sufficiently many reliable agents are merged then in the limit the true state of the world is identified. We introduce a new postulate suited to the truth tracking issue. We identify some merging operators from the literature which satisfy it and other operators which do not.

## 1 Introduction

In many areas of computer science, including distributed databases and multi-agent systems, one needs to synthesize pieces of information issued from several sources. What makes this problem difficult is, among other things, that the information sources typically contradict each other. When the available pieces of information are beliefs represented in propositional logic, this problem is called (propositional) belief merging. Many different belief merging operators have been pointed out so far. In this paper we focus on the purely logical case, i.e., we assume that belief bases are sets of propositional formulae (see e.g. [2, 20, 12, 11, 9]). Other operators have been provided in more general settings, such as weighted logics (possibilistic logic or settings based on ordinal conditional functions) [3, 16, 4], which prove useful when more information are available (especially, when all the pieces of belief are not equally certain). In these more general settings, the merging problem becomes close to preorder (preference) aggregation, as studied in social choice theory [1, 19].

Logical properties of merging operators have been investigated in several works [20, 14, 12]. In [12] a set of logical properties have been put forward to characterize the family of IC (Integrity Constraints) merging operators. IC merging operators have been advocated to be suited to both belief merging and goal merging. Even if it might look strange at a first glance that very different concepts, such as goals and beliefs, can be handled in the same way with respect to aggregation, the adequacy of IC merging operators to propositional merging (whatever goals or beliefs are to be merged) has not been challenged so far. This makes sense since in both cases merging aims at synthesizing the information represented in the given profile of propositional bases.

In this paper we introduce a new point of view about belief merging, that goes beyond the usual synthesis view: the epistemic view.

**Synthesis View:** Under the synthesis view, belief merging aims at characterizing a base which best represents the beliefs of the input profile. This is the view adopted in previous belief merging works.

**Epistemic View:** Under the epistemic view, the purpose of a belief merging process is to best approximate what is the true state of the world.

In the general case, no agent has a perfect view of the real world, her beliefs are pervaded with uncertainty:

- An agent typically does not know which one of the models of her base represents the true state of the world,
- She is not even ensured that the true state of the world is really among the models of her base.[1]

Belief merging under the epistemic view can be considered as a way to circumvent such an uncertainty at the group level, by tracking the true state of the world. Interestingly, the truth tracking issue allows to discriminate belief merging from goal merging. Indeed, the concept of truth tracking is meaningless when goals are considered: there is no notion of "true goal" which would be analogous to the true state of the world in the goal merging setting.

The problem of truth tracking has been studied for centuries in social choice and in political science, in order to justify the foundations of democratic elections or of decisions made by jury trials. The main theoretical result here is Condorcet's Jury Theorem [7]. This theorem states that if a jury is composed of reliable and independent individuals, and if they have to find the right answer to a yes/no question, then the probability that the decision made by the jury is the right one tends to 1 as the size of the jury tends to infinity.

In this paper we formalize the truth tracking issue from a belief merging perspective. We show that some belief merging operators can be used to identify the true state of the world by considering sufficiently many reliable, homogeneous and independent agents. To be more precise, we present a generalization of Condorcet's Jury Theorem under uncertainty (i.e., when each base may have several models). We introduce a Truth Tracking (**TT**) postulate, and we point IC operators satisfying it and other IC operators which do not. This shows that **TT** is independent of the (conjunction of the) IC postulates. We also provide experimental results in order to investigate the convergence speed of truth tracking for the belief merging operator $\Delta^{d_D,\Sigma}$. In most cases the number of agents to be considered for ensuring that the merged base identifies the true state of the world with high probability is not so large.

[1] If one supposes that the agent is ensured that the true state of the world is a model of her belief base, then one talks about "knowledge" — this assumption is the only difference between belief and knowledge — and *knowledge merging* is not so interesting, since the only sensible knowledge merging operator is conjunction.

The rest of the paper is organized as follows. In the next section, we give some formal preliminaries. Then we recall Condorcet's Jury Theorem, and some of its generalizations. In the third section we point out new generalizations of Condorcet's Jury Theorem suited to the belief merging perspective. Then we introduce the **TT** postulate and we show that IC postulates and **TT** are logically independent. Finally we present the experimental results we obtained and we discuss them. For space reasons, we report the proof of the main result (Theorem 3), and omit the other ones.

## 2 Preliminaries

We consider a propositional language $\mathcal{L}$ defined from a finite set of propositional variables $\mathcal{P}$ and the usual connectives.

For any subset $c$ of $\mathcal{P}$, $|c|$ denotes the number of elements of $c$. An interpretation (or state of the world) is a total function from $\mathcal{P}$ to $\{0, 1\}$. The set of all interpretations is noted $\mathcal{W}$. The true state of the world is noted $\omega^\star$. An interpretation $\omega$ is a model of a formula $\phi \in \mathcal{L}$ if and only if it makes it true in the usual truth functional way. $[\phi]$ denotes the set of models of formula $\phi$, i.e., $[\phi] = \{\omega \in \mathcal{W} \mid \omega \models \phi\}$.

A *base* $K$ denotes the set of beliefs of an agent, it is a finite set of propositional formulae, interpreted conjunctively. We identify $K$ with the conjunction of its elements. Basically, a base $K$ represents a set $[K]$ of states of the world.

A *profile* $E$ denotes the beliefs from the group of $n$ agents that are involved in the merging process. In this paper agents express sometimes only a single possible world. In this case a profile is a vector of complete bases. In order to avoid heavy notations, we assimilate each complete base with its model and write such profiles as $E_c = \langle\omega_1, \ldots, \omega_n\rangle$. Elsewhere agents express sets of possible worlds, hence $E$ is a vector of bases $E = \langle K_1, \ldots, K_n\rangle$, as usual in propositional merging. Inclusion of profiles is given by $E = \langle K_1, \ldots, K_n\rangle \sqsubseteq E' = \langle K'_1, \ldots, K'_m\rangle$ iff $n \leq m$ and $\forall i \in 1, \ldots, n$, we have $[K_i] = [K'_i]$.

In the following, agents $1, \ldots, n$ are identified with the corresponding belief bases $K_1, \ldots, K_n$. When unknown, each $K_i$ can also be viewed as a discrete random variable, ranging over $2^{\mathcal{W}}$ (or $\mathcal{W}$ when each agent has to report a complete base). The true state of the world is usually unknown as well so it is also viewed as a random variable $\mathcal{W}^*$, ranging over $\mathcal{W}$.

Two important notions about sets of agents will be considered in the following: *independence* and *homogeneity*.

Agents $1, \ldots, n$ are independent if knowing the true state of the world and a set of states of the world reported by any other agent $j$ does not give any further information on the states of the world given by an agent $i$ (this means that the agents choices are independent conditionally to the true state of the world in a standard Bayesian way [17]). Formally, agents $1, \ldots, n$ are said to be *independent* if $\forall \omega, \omega_1, \ldots, \omega_n \in \mathcal{W}$:

$$P(\bigwedge_{i=1}^{n} \omega_i \models K_i \mid \mathcal{W}^* = \omega) = \prod_{i=1}^{n} P(\omega_i \models K_i \mid \mathcal{W}^* = \omega).$$

Obviously, when agents report complete bases, the formal definition of independence can be stated as follows: agents $1, \ldots, n$ are independent if $\forall \omega, \omega_1, \ldots, \omega_n \in \mathcal{W}$, $P(\bigwedge_{i=1}^{n} [K_i] = \{\omega_i\}) \mid \mathcal{W}^* = \omega) = \prod_{i=1}^{n} P([K_i] = \{\omega_i\} \mid \mathcal{W}^* = \omega)$.

Agents $1, \ldots, n$ are *homogenous* if for every world $\omega_j \in \mathcal{W}$, the probability $P(\omega_j \models K_i)$ that $\omega_j$ is a model of the base $K_i$ of the profile $E$ is the same for all the agents $i \in 1, \ldots, n$ of the set. In particular, the real world $\omega^\star$ has the same probability to appear as a model for each agent.

## 3 Condorcet's Jury Theorem and Extensions

We first consider a profile $E_c$ of $n$ agents where each agent $i$ votes for an alternative, let us say a state of the world $\omega_i \in \mathcal{W}$. Among the possible states of the world is the true one $\omega^\star$.

The hypotheses used in Condorcet's Jury Theorem are that agents are both independent and reliable. Since several notions of reliability will be considered in the following, we call the first one *R1-reliability*:

- The *R1-reliability* $p_i$ of an agent $i$ is the probability that $i$ gives the true state of the world, i.e., $p_i = P([K_i] = \{\omega^\star\})$.
- An agent $i$ is *R1-reliable* if her R1-reliability is strictly greater than $0.5$. **(R1)**

The majority rule simply returns as result the interpretation which receives a strict majority of votes. Formally, let us first define the notion of score of a world with respect to a profile of complete bases: $s(\omega) = |\{\omega_i \in E_c \text{ s.t. } \omega_i = \omega\}|$.

**Definition 1 (Majority)** *Given a profile $E_c$ of $n$ complete bases, the majority rule $m$ is defined as: $m(E_c) = \omega$ if $s(\omega) > n/2$.*

We are now ready to recall Condorcet's Jury Theorem. In this theorem, only two alternatives are considered so that each agent votes for one of them:

**Theorem 1 ([7])** *Consider a set $\mathcal{W} = \{\omega, \omega^\star\}$ consisting of two possible states of the world and a profile $E_c$ of complete bases from a set of $n$ independent and R1-reliable agents sharing the same R1-reliability. The probability that the majority rule on this profile returns the true state of the world $\omega^\star$ tends to $1$ as $n$ tends to infinity, i.e.:*

$$P(m(E_c) = \omega^\star) \xrightarrow[n\to\infty]{} 1.$$

This theorem is a consequence of the (weak) law of large numbers. Roughly, it states that if the individuals in a jury are sufficiently reliable (they perform better than pure randomizers) and independent, then the probability that the jury makes the right decision tends to 1 when the size of the jury tends to infinity. It is interesting to notice that the homogeneity assumption is not used explicitly in Condorcet's Jury Theorem. However, it is implicitly there, just because $R1$-reliability implies homogeneity when only two alternatives are considered (the probability that any agent chooses a world different from $\omega^\star$ is $1 - p$ if $p$ is the agents' $R1$-reliability).

Clearly enough, the assumptions used in Condorcet's Jury Theorem are quite strong. First, usually agents in a jury are not fully independent: they often have a similar background, listen the same opinion leaders, etc. Furthermore, in general, all the agents do not have exactly the same reliability: there are usually agents more competent than others. Interestingly, some extensions of Condorcet's Jury Theorem show that these strong assumptions can be relaxed without questioning the conclusion. Thus, the theorem still holds when the opinions of the individuals are not independent [8]. And as far as reliability is concerned, it is enough to assume that the mean reliability of the individuals is above 0.5 [10].

A further limitation of Condorcet's Jury Theorem is that it considers only two alternatives. A recent result by List and Goodin [15] allows to extend the theorem to any finite number of options. In order to present this result, we first need to recall the definition of the plurality rule:

**Definition 2 (Plurality)** *Given a profile $E_c$ of complete bases, the plurality rule pl is defined as:* $pl(E_c) = \{\omega \text{ s.t. } \forall \omega' \in \mathcal{W} \;\; s(\omega) \geq s(\omega')\}$.

The reliability assumption **(R1)** has to be extended to more than two alternatives. List and Goodin [15] define the following notion of reliability; consider a set $\mathcal{W} = \{\omega_1, \ldots, \omega_{k-1}, \omega^\star\}$ of $k$ possible states of the world:

- An agent is *R2-reliable* if the probability that she votes for $\omega^\star$ is strictly greater than the probability that she votes for another world. **(R2)**

List and Goodin showed that:

**Theorem 2 ([15])** *Consider a set $\mathcal{W} = \{\omega_1, \ldots, \omega_{k-1}, \omega^\star\}$ of $k$ possible states of the world and a profile $E_c$ of complete bases from a set of $n$ independent, homogeneous, and R2-reliable agents.*[2] *The probability that the plurality rule on this profile returns the true state of the world $\omega^\star$ tends to 1 as $n$ tends to infinity, i.e.:*

$$P(pl(E_c) = \{\omega^\star\}) \xrightarrow[n \to \infty]{} 1.$$

This theorem is a generalization of Condorcet's Jury Theorem. When considering only two states of the world, the hypotheses used in List-Goodin's theorem are equivalent to the ones used in Condorcet's Jury Theorem, so that the two theorems are identical in this case, as expected. Observe that the plurality rule is used in List-Goodin's theorem (not the majority rule as in Condorcet's Jury Theorem). and that the reliability assumption only requires that the probability of voting for the true state of the world is strictly greater than the probability of voting for another world, so that the probability of voting for the true state of the world can be less than 0.5. See [15] for more discussions on their theorem and its philosophical consequences, and for a discussion about Condorcet's Jury Theorem.

## 4 A Jury Theorem under Uncertainty

In all these previous works around Condorcet's Jury Theorem, agents are supposed to vote for a unique alternative. This makes them inadequate for our purpose since in belief merging, agents typically give belief bases having several models (and imposing agents to give complete belief bases would be very restrictive since it would deny that the agents beliefs can be uncertain). Thus, from now on, we assume that each agent $i$ gives a belief base $K_i$ which may have several models taken from a finite set $\mathcal{W} = \{\omega_1, \ldots, \omega_{k-1}, \omega^\star\}$.

Let us show how the Jury Theorem can be extended to consider the case when each agent may vote for several alternatives. We first need to define a notion of agent reliability suited to this situation:

- The *R3-reliability* $p_i$ of an agent $i$ is the probability that the true state of the world $\omega^\star$ is among the models of her belief base $K_i$, i.e., $p_i = P(\omega^\star \models K_i)$.
- An agent is *R3-reliable* if $p_i > 0.5$. **(R3)**

Similarly, the notion of score of a world has to be extended : $s_a(\omega) = |\{K_i \in E \text{ s.t. } \omega \models K_i\}|$.

Then it is possible to state the following result:

[2] List and Goodin proposed a notion of reliability which encompasses both our R2-reliability and homogeneity.

**Proposition 1** *Consider a real number $p^\star \in [0, 1[$ and a profile $E$ from a set of $n$ independent agents which have the same R3-reliability $p > p^\star$. The probability that the score of the true state of the world exceeds $np^\star$ tends to 1 when $n$ tends to infinity, i.e.,*

$$P(s_a(\omega^\star) > np^\star) \xrightarrow[n \to \infty]{} 1.$$

This result gives in the limit a lower bound on the score of the true state of the world provided that the agents are equally R3-reliable. It is interesting because it ensures for some voting rules that the true state of the world belongs to the set of states returned by the rule. Consider for instance the following voting rules:

**Definition 3 ($M$ and $Q_p$)** *Let $E$ be a profile from a set of $n$ agents.*

- *The majority rule $M$ is defined as:* $M(E) = \{\omega \text{ s.t. } s_a(\omega) > n/2\}$.
- *More generally, given $k \in ]0, 1[$, the $k$-quota rule $Q_k$ is defined as:* $Q_k(E) = \{\omega \text{ s.t. } s_a(\omega) > kn\}$.

The majority rule $M$ corresponds to the 0.5-quota rule.

As a direct corollary to Proposition 1 we get:

**Proposition 2** *Let $E$ be a profile from a set of $n$ independent agents. If all agents have the same R3-reliability $p > k$, then the true state of the world belongs to the set of states returned by the $k$-quota rule in the limit , i.e.,*

$$P(\omega^\star \in Q_k(E)) \xrightarrow[n \to \infty]{} 1.$$

Let us stress that this proposition only mentions the membership of the true state of the world in the result of the voting process, but it does not exclude that many other states can also appear in this result. Obviously, this is problematic from the truth tracking point of view. In particular, if each agent $i$ gives all the possible worlds ($[K_i] = \mathcal{W}$), then for the corresponding profile $E$ we get all the possible worlds (for instance $Q_k(E) = \mathcal{W}$ whatever $k$), which is not informative at all about the true state of the world.

The problem is due to the notion of R3-reliability that is not strong enough for the truth tracking purpose. Intuitively, asking the agents to give the true state of the world with a high probability is necessary but not sufficient since it does not prevent agents from giving (as models of their bases) too many states. Especially, an agent $i$ whose base is always a tautology ($[K_i] = \mathcal{W}$), obviously carrying no information, is considered fully R3-reliable (i.e., her R3-reliability $p_i$ is equal to 1), which is unexpected. Thus a stronger notion of reliability is necessary. The following notion of R4-reliability is intended to this purpose:

- Let us note $q_{j,i}$ the probability that the world $\omega_j$ belongs to the set of models of the base of an agent $i$, i.e., $q_{j,i} = P(\omega_j \models K_i)$. If there is no ambiguity on the agent then we will simply note $q_j$ instead of $q_{j,i}$.
- The *incompetence* $Q_i$ of an agent $i$ is the maximal probability that a world different from $\omega^\star$ belongs to the set of models of her base, i.e., $Q_i = \max_{\omega_j \in \mathcal{W} \setminus \{\omega^\star\}} q_{j,i}$. The *competence* of an agent is $c_i = 1 - Q_i$.
- An agent is *competent* if $c_i > 0.5$.
- An agent is *R4-reliable* if it is more R3-reliable than incompetent: $p_i = P(\omega^\star \models K_i) > Q_i$. **(R4)**

Intuitively, while R3-reliability expresses the ability of an agent not to miss the true state of the world, the notion of competence deals with the quantity of uncertainty pervading her beliefs. Taken together, R3-reliability and competence are natural and important notions for characterizing the intuitive notion of “reliable agent” in the belief merging setting. While, in the specific case when $\mathcal{W}$ consists only of two alternatives, an agent is competent if and only if she is R3-reliable, in the general case competence and $R3$-reliability are two different notions. Furthermore, it is easy to prove that the notion of R4-reliability extends the previous notions of reliability:

**Proposition 3**

- *When considering only profiles $E_c$ of complete bases, R4-reliability is equivalent to R2-reliability.*
- *When considering only profiles $E_c$ of complete bases and a set $\mathcal{W}$ of interpretations containing only two elements $\{\omega, \omega^\star\}$, R4-reliability, R3-reliability, R2-reliability and R1-reliability are equivalent.*

With the notions of R4-reliability and competence, we can state the following Jury Theorem under Uncertainty:

**Theorem 3** *Let $\mathcal{W} = \{\omega_1, \ldots, \omega_{k-1}, \omega^\star\}$ be a set of possible worlds and let $E$ be a profile from a set of $n$ independent, homogenous and R4-reliable agents. Then $\forall i \in \{1, \ldots, k-1\}$,*

$$P(s_a(\omega^\star) > s_a(\omega_i)) \xrightarrow[n\to\infty]{} 1.$$

**Proof:**

Let $(s_a(\omega_1), \ldots, s_a(\omega_{k-1}), s_a(\omega^\star))$ be a vector of random variables where $s_a(\omega_i) = l$ $(i \in \{1, \ldots, k-1\})$ (resp. $s_a(\omega^\star) = l$) means that the score $s_a(\omega_i)$ (resp. $s_a(\omega^\star)$) is equal to $l$ ($l \in 0 \ldots n$). As the set of agents is homogeneous, we have $q_{j,i} = q_{j,k}$ for every world $\omega_j$ and all agents $i$, $k$. We note $q_j$ this probability, i.e., $q_j = q_{j,i} = P(\omega_j \models K_i)$, for any agent $i$. We note $p$ the probability that an agent gives the true state of the world, i.e., $p = P(\omega^* \models K_i)$, for any agent $i$.

Each of the random variables $s_a(\omega_i)$ $(i \in \{1, \ldots, k-1\})$ (resp. $s_a(\omega^\star)$) follows a binomial distribution with parameters $n$ and $q_i$ (resp. $n$ and $p$). Subsequently, we have that $\forall j \in 0 \ldots n$:

$$P(s_a(\omega_i) = j) = \tbinom{n}{j} q_i^j (1 - q_i)^{n-j}$$

and

$$P(s_a(\omega^\star) = j) = \tbinom{n}{j} p^j (1 - p)^{n-j}.$$

The mean of each $s_a(\omega_i)$ $(i \in \{1, \ldots, k-1\})$ is $nq_i$, its variance is $nq_i(1 - q_i)$, the mean of $s_a(\omega^\star)$ is $np$ and its variance is $np(1 - p)$.

The (weak) law of large numbers applied to $s_a(\omega_i)$ $(i \in \{1, \ldots, k-1\})$ and $s_a(\omega^\star)$ gives that $\forall \epsilon > 0$:

$$P(\mid \frac{s_a(\omega_i)}{n} - q_i \mid \geq \epsilon) \longrightarrow_{n\to\infty} 0, \tag{1}$$

$$P(\mid \frac{s_a(\omega^\star)}{n} - p \mid \geq \epsilon) \longrightarrow_{n\to\infty} 0. \tag{2}$$

Let $q = max_{i \in \{1,\ldots,k-1\}} q_i$ and $\epsilon_1 = \frac{p-q}{2}$. Since each agent is $R4$-reliable, we have that $q_i < p$ for each $i \in \{1, \ldots, k-1\}$, so $q < p$. As a consequence, we get that $\epsilon_1 > 0$. Using inequations (1) and (2), one concludes that for each $i \in \{1, \ldots, k-1\}$:

$$P(\mid \frac{s_a(\omega_i)}{n} - q_i \mid \geq \epsilon_1) \longrightarrow_{n\to\infty} 0,$$

and

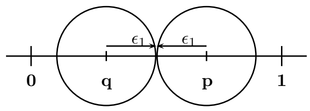
$$P(\mid \frac{s_a(\omega^\star)}{n} - p \mid \geq \epsilon_1) \longrightarrow_{n\to\infty} 0.$$

It easily gives that:

$$P(\frac{s_a(\omega_i)}{n} > q_i + \epsilon_1) \longrightarrow_{n\to\infty} 0, \tag{3}$$

and

$$P(\frac{s_a(\omega^\star)}{n} < p - \epsilon_1) \longrightarrow_{n\to\infty} 0. \tag{4}$$




The picture above explains the idea of the proof: when the weak law of large numbers can be used for a random variable, the values of this variable are close to its mean with a high probability. Schematically, the probability that all the values of the variable are in a sphere with the mean as center and $\epsilon_1$ as radius tends to 1 in the limit. As a consequence, as $p > q$, the probability that the two spheres intersect tends to 0 in the limit.

The problematic case for the proof is when $\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^\star)}{n}$. Suppose that $\frac{s_a(\omega_i)}{n} \leq q_i + \epsilon_1$ and that $\frac{s_a(\omega^\star)}{n} \geq p - \epsilon_1$. Then, as $\forall i \in 1 \ldots k-1$, by definition of $\epsilon_1$, $q_i + \epsilon_1 \leq p - \epsilon_1$, we get $\frac{s_a(\omega_i)}{n} \leq q_i + \epsilon_1 \leq p - \epsilon_1 \leq \frac{s_a(\omega^\star)}{n}$. As a consequence, $\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^\star)}{n}$ may happen only if $\frac{s_a(\omega_i)}{n} > q_i + \epsilon_1$, or $\frac{s_a(\omega^\star)}{n} < p - \epsilon_1$. In this case, we have:

$P(\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^\star)}{n}) = P(\frac{s_a(\omega_i)}{n} > q_i + \epsilon_1$ and $\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^\star)}{n}) + P(\frac{s_a(\omega^\star)}{n} < p - \epsilon_1$ and $\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^\star)}{n}) - P(\frac{s_a(\omega_i)}{n} > q_i + \epsilon_1$ and $\frac{s_a(\omega^\star)}{n} < p - \epsilon_1$ and $\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^\star)}{n})$.

Since $P(\frac{s_a(\omega^\star)}{n} < p - \epsilon_1$ and $\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^\star)}{n}) \leq P(\frac{s_a(\omega^\star)}{n} < p - \epsilon_1)$, and $P(\frac{s_a(\omega_i)}{n} > q_i + \epsilon_1$ and $\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^\star)}{n}) \leq P(\frac{s_a(omega_i)}{n} > q_i + \epsilon_1)$, we get:

$$P(\frac{s_a(\omega_i)}{n} > \frac{s_a(\omega^\star)}{n}) \leq P(\frac{s_a(\omega^\star)}{n} < p - \epsilon_1) + P(\frac{s_a(\omega_i)}{n} > q_i + \epsilon_1).$$

Finally, with assertions (3) and (4), we obtain:

$$P(\frac{s_a(\omega_i)}{n} \geq \frac{s_a(\omega^\star)}{n}) \xrightarrow[n\to\infty]{} 0,$$

or, equivalently

$$P(s_a(\omega^\star) > s_a(\omega_i)) \xrightarrow[n\to\infty]{} 1.$$

□

Theorem 3 is a generalization of Condorcet's Jury Theorem to an uncertainty framework, where the agents can give a set of worlds instead of a single one. Indeed, Condorcet's Jury Theorem is recovered when each agent reports a complete base and only two states of the world are possible. Notice that, in Theorem 3, agents are not required to be R3-reliable or competent. Indeed, the conclusion holds as soon as the R3-reliability of each agent is greater than her incompetence.

Interestingly, allowing the agents to vote for any number of worlds, and to choose as result the worlds with the greatest score is just approval voting [6]:

**Definition 4 (Approval)** *Given a profile of bases $E$, the approval rule $av$ is defined as: $av(E) = \{\omega$ s.t. $\forall \omega' \in \mathcal{W}\ s_a(\omega) \geq s_a(\omega')\}$.*

Thus, Theorem 3 shows that approval voting ensures to track the true world in the uncertain framework, just as plurality voting does in the “standard” framework.

## 5 Truth Tracking for Belief Merging

The ability of a merging operator $\Delta$ to achieve the truth tracking issue can be modeled as a new postulate, called **T**ruth **T**racking postulate:

**TT** Let $\omega^\star$ be the true state of the world. Let $(E_n)_{n\in\mathbb{N}}$ be any sequence of widening[3] profiles from a set of $n$ independent, homogenous and R4-reliable agents. Then $P([\Delta(E_n)] = \{\omega^\star\}) \xrightarrow[n\to\infty]{} 1$.

This postulate is satisfied by a merging operator when it allows to identify the true state of the world by listening sufficiently many homogeneous independent agents who are more R3-reliable than incompetent.

Let us now investigate the behaviour of some well-known belief merging operators with respect to this postulate. We first recall the definition of distance-based merging operators (see [12] for details).

**Definition 5 (distance-based merging operators)** *Let $d$ be a pseudo-distance between worlds and $f$ be an aggregation function. The merging operator $\Delta^{d,f}(E)$ is defined by: $[\Delta^{d,f}_{\mu}(E)] = \min([\mu], \leq_E)$*
*where the pre-order $\leq_E$ on $\mathcal{W}$ induced by $E$ is defined by:*

- *$\omega \leq_E \omega'$ if and only if $d(\omega, E) \leq d(\omega', E)$, where*
- *$d(\omega, E) = f_{K\in E}(d(\omega, K))$, where*
- *$d(\omega, K) = min_{\omega'\models K} d(\omega, \omega')$.*

Usual (pseudo-)distances are $d_D$ the drastic distance ($d_D(\omega, \omega') = 0$ if $\omega = \omega'$ and 1 otherwise), and $d_H$ the Hamming distance ($d_D(\omega, \omega') = n$ if $\omega$ and $\omega'$ differ on $n$ variables). Usual aggregation functions are $\Sigma$, $Gmax$ (see [13]) and $Gmin$ (see [9]).

We obtained the following results:

**Proposition 4**

- *$\Delta^{d_D,\Sigma}$ ($= \Delta^{d_D,Gmax}$) satisfies* **TT**.
- *For each pseudo-distance $d$, $\Delta^{d,Gmin}$ satisfies* **TT**.
- *$\Delta^{d_H,\Sigma}$ does not satisfy* **TT**.
- *$\Delta^{d_H,Gmax}$ does not satisfy* **TT**.

This proposition shows that **TT** and (the conjunction of) the IC postulates are independent in the sense that there exist IC merging operators satisfying **TT** (e.g. $\Delta^{d_D,\Sigma}$) but it is not the case that each IC merging operator satisfies it (e.g. $\Delta^{d_H,\Sigma}$).

## 6 Some Experimental Results

The results about truth tracking we pointed out in the previous sections all concern the identification of the true state of the world $\omega^\star$ *in the limit*. None of them gives any information about truth tracking from the pratical side, in the sense of a bound on the number of bases from which the identification is achieved with high probability.

In order to investigate this issue, we performed a number of experiments using $\Delta^{d_D,\Sigma}$, that is an IC merging operator which satisfies **TT** and that is easy to implement. We investigated the convergence speed of truth tracking using $\Delta^{d_D,\Sigma}$, depending on the agents R3-reliability $p$ and incompetence $Q = q$; for simplicity reasons, we made the assumption that all worlds $\omega_i$ (different from the true world) have the same probability (i.e., $\forall\omega_i \in \mathcal{W} \setminus \{\omega^\star\}$, $q_i = q$). We considered sets of interpretations of various sizes (up to $2^{15}$), we fixed the true state $\omega^\star$ as the world mapping each propositional variable to 0, and we generated profiles $E$ from $n$ homogeneous and independent agents, with R3-reliability $p$ and incompetence $q$, for different values of $n$. For each value of $n$, we computed 1000 profiles $E$. For each $E$, we computed $\Delta^{d_D,\Sigma}(E)$ and check whether $[\Delta^{d_D,\Sigma}(E]) = \{\omega^\star\}$ holds. The proportion of the 1000 profiles for which $[\Delta^{d_D,\Sigma}(E]) = \{\omega^\star\}$ holds gives an estimate of the probability of success of truth tracking.

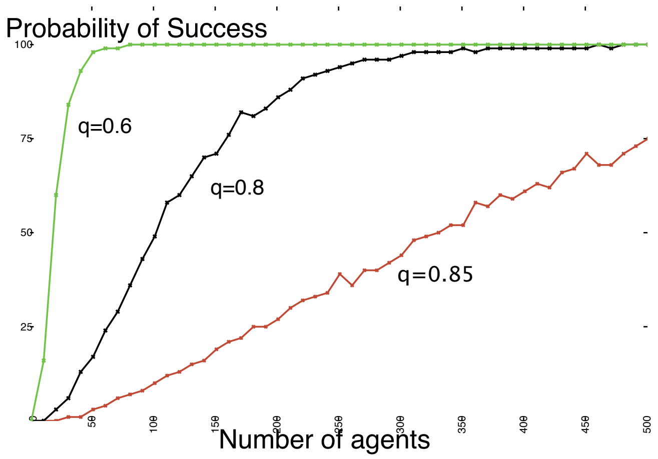



**Figure 1.** Convergence speed (7 variables, p=0.9)

Figure 1 gives the probability that $[\Delta^{d_D,\Sigma}(E]) = \{\omega^\star\}$ given the number $n$ of agents, when $p = 0.9$ and $|\mathcal{W}| = 2^7$ worlds, for several values of $q$. Interestingly, we can observe on Figure 1 that the convergence speed is high: to get $[\Delta^{d_D,\Sigma}(E)] = \{\omega^\star\}$ with a probability greater than 90%, 800 agents are necessary for $q = 0.85$, 230 agents are necessary for $q = 0.8$ and only 40 agents are necessary for $q = 0.6$.

We report the curves only for one value of $p$, but the general shape of the curves for other values of $p$ is very closed to the one reported. More precisely, the figure obtained for other values of $p$, even for $p$ very low, is quite the same that in figure 1. Empirically, it turns out that the "level of R4-reliability" of agents, i.e., the value $p - q$ seems to have more impact on the convergence speed of truth tracking using $\Delta^{d_D,\Sigma}$ than the fact that these values of $p$ and $q$ are rather high or rather low.

Figure 2 gives the probability of success of truth tracking given the number of propositional variables (hence the number of worlds) with $p = 0.7$ and $q = 0.4$.

As expected the complexity of discriminating the true state of the world increases with the number of possible states of the world. But, interestingly, the number of agents to be considered in order to achieve the truth tracking issue with high probability is not that huge compared to the number of interpretations. For instance, one can observe on Figure 2 that, when 10 variables are considered, less than 50 agents are enough to ensure that $[\Delta^{d_D,\Sigma}(E)] = \{\omega^\star\}$ with probability greater than 90%, despite the fact that a single state has to be discriminated among 1024 ones and that the agents R3-reliability

---

[3] $(E_n)_{n\in\mathbb{N}}$ satisfies $\forall i \in \mathbb{N}$, $E_i \sqsubseteq E_{i+1}$.

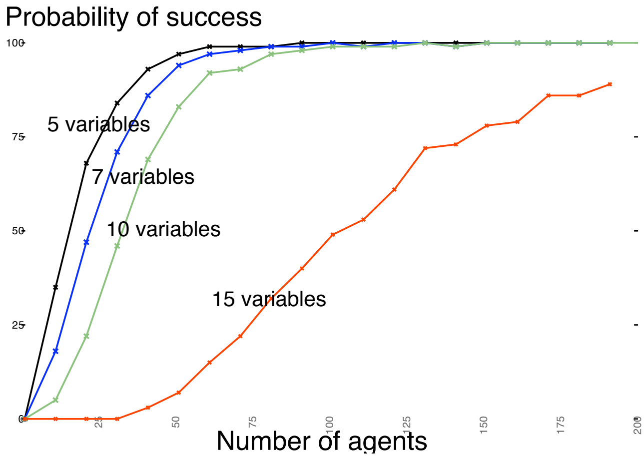



**Figure 2.** Convergence speed (p=0.7, q =0.4)

and competence are not so high.

## 7 Conclusion

In this work we have discriminated two possible interpretations of what merging aims at: the synthesis view and the epistemic view. The synthesis view is the usual view of belief merging; it aims at finding out the base that best ("most faithfully") reflects the given profile. The epistemic view that we have introduced amounts to tracking the true state of the world.

The contribution of the paper is manyfold. First, the epistemic view allows to draw a clear distinction between goal merging and belief merging. Indeed, there cannot be an epistemic view for goal merging, because there is no notion of true ("objectively correct") goal, echoing the notion of true state of the world. As far as we know, this is the first time that belief merging and goal merging are separated on formal grounds, namely, by a logical property (**TT**) advocated to hold for only one of these two cases (belief merging).

We have also provided a generalization of Condorcet's Jury Theorem under uncertainty, and thanks to it, pointed out some IC merging operators satisfying **TT**; in addition, we have proved that other IC merging operators do not satisfy **TT**, showing that this new postulate and the conjunction of the IC postulates are independent. Finally we studied the convergence speed of a specific operator ($\Delta^{d_D,\Sigma}$).

The problem of truth tracking has been investigated in the related framework of judgment aggregation [5]. In [18] the authors take advantage of the operator $\Delta^{d_H,\Sigma}$ for this purpose, and give empirical evidence that it performs quite well (typically better than other judgment aggregation procedures) for mildly reliable agents. This work departs from our own one by many aspects; especially, it focuses on specific profiles (representing the doctrinal paradox) leading to specific results, while in our work we did not put any constraints on the admissible profiles.

The hypotheses considered in the Jury Theorem under Uncertainty may seem quite demanding. In most cases it is hard to ensure that agents are both reliable, independent and homogeneous. Nonetheless these assumptions are exactly the same ones as those used in previous works (Condorcet and List-Goodin theorems). Furthermore, they are not "too strong" since they allow to discard some IC merging operators (e.g. $\Delta^{d_H,Gmax}$) from those which could be suited to the truth tracking issue. Finally, like Condorcet's Jury Theorem for which generalizations have been obtained by relaxing for instance the independence assumption [8] or the reliability one [10], one can expect similar generalizations to hold for the Jury Theorem under Uncertainty. Searching for such generalizations is an issue for further research.

## REFERENCES

[1] K. J. Arrow, *Social Choice and Individual Values*, Wiley, New York, second edn., 1963.
[2] C. Baral, S. Kraus, and J. Minker, 'Combining multiple knowledge bases', *IEEE Transactions on Knowledge and Data Engineering*, **3**(2), 208–220, (1991).
[3] S. Benferhat, D. Dubois, S. Kaci, and H. Prade, 'Possibilistic merging and distance-based fusion of propositional information', *Annals of Mathematics and Artificial Intelligence*, **34**(1–3), 217–252, (2002).
[4] S. Benferhat, S. Lagrue, and J. Rossit, 'An egalitarist fusion of incommensurable ranked belief bases under constraints', in *Proceedings of AAAI'07*, pp. 367–372, (2007).
[5] L. Bovens and W. Rabinowicz, 'Democratic answers to complex questions', *Synthese*, **150**, 131–153, (2006).
[6] S. J. Brams and P. C. Fishburn, *Approval voting*, Springer Verlag, 1983.
[7] 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.
[8] D. Estlund, 'Opinion leaders, independence and condorcet's jury theorem', *Theory and Decision*, **36**(2), 131–162, (January 1994).
[9] P. Everaere, S. Konieczny, and P. Marquis, 'Quota and Gmin merging operators', in *Proceedings of IJCAI'05*, pp. 424–429, (2005).
[10] O. Guillermo, B. Grofman, and S. Feld, 'Proving a distribution-free generalization of the Condorcet jury theorem', *Mathematical Social Sciences*, **17**(1), 1–16, (February 1989).
[11] S. Konieczny, J. Lang, and P. Marquis, 'DA$^2$ merging operators', *Artificial Intelligence*, **157**, 49–79, (2004).
[12] S. Konieczny and R. Pino Pérez, 'Merging information under constraints: a logical framework', *Journal of Logic and Computation*, **12**(5), 773–808, (2002).
[13] S. Konieczny and R. Pino Pérez, 'On the frontier between arbitration and majority', in *Proceedings of KR'02*, pp. 109–118, (2002).
[14] P. Liberatore and M. Schaerf, 'Arbitration (or how to merge knowledge bases)', *IEEE Transactions on Knowledge and Data Engineering*, **10**(1), 76–90, (1998).
[15] C. List and R. Goodin, 'Epistemic democracy: Generalizing the Condorcet jury theorem', *Journal of Political Philosophy*, **9**(3), 277–306, (2001).
[16] T. Meyer, 'On the semantics of combination operations', *Journal of Applied Non-Classical Logics*, **11**(1-2), 59–84, (2001).
[17] Judea Pearl, *Causality: models, reasoning, and inference*, Cambridge University Press, 2000.
[18] G. Pigozzi and S. Hartmann, 'Judgment aggregation and the problem of truth-tracking', in *Proceedings of TARK'07*, pp. 248–252, (2007).
[19] M. S. Pini, F. Rossi, K. B. Venable, and T. Walsh, 'Aggregating partially ordered preferences: impossibility and possibility results', in *Proceedings of TARK'05*, pp. 193–206, (2005).
[20] P. Z. Revesz, 'On the semantics of arbitration', *International Journal of Algebra and Computation*, **7**(2), 133–160, (1997).