Representation of belief in relation to randomness

Thierry Viéville, Chloé Mercier

RESEARCH REPORT
N° 9493
December 2022
Project-Team Mnemosyne

ISSN 0249-6399 ISRN INRIA/RR--9493--FR+ENG

Representation of belief in relation to randomness*

Thierry Viéville[†‡], Chloé Mercier[†]

Project-Team Mnemosyne

Research Report n° 9493 — December 2022 — 20 pages

Abstract: In this technical note, we develop a belief representation that generalizes the standard possibility theory, extending it to a bi-dimensional representation where randomness is taken into account. This way, we revisit and extend the usual quantitative implementation of belief level.
We state the basic requirements and discuss the difference with the standard possibility theory before investigating the compatibility with Boolean and Kleene's three-valued logic calculus. We then relate this representation to the probability theory, based on the assumption that necessity and possibility provide bounds for probability values, and finally discuss how to define a relevant projection of real values onto an admissible representation.
This provides the basic ingredients to implement such extension of belief representation in existing mechanisms using the basic representation. The operational part of this development is available as open-source code.

Key-words: Possibility Theory, Modal Logic, Representation of Randomness

* Supported by Inria, AEx AIDE `https://team.inria.fr/mnemosyne/en/aide`.
† Mnemosyne Inria Research Team
‡ LINE laboratory, Université Côte d'Azur

RESEARCH CENTRE BORDEAUX – SUD-OUEST

200 avenue de la Vieille Tour
33405 Talence Cedex

Representation du niveau de croyance en lien avec la notion d'aléa

Résumé : Dans cette note technique, nous développons une représentation de la notion de niveau de croyance qui généralise la théorie des possibilités standard, en l'étendant à une représentation bidimensionnelle où l'aléatoire est pris en compte. De cette façon, nous revisitons et étendons l'implémentation numérique habituelle du niveau de croyance.

Nous énonçons les contraintes de base et discutons de la différence avec la théorie des possibilités usuelle avant d'étudier la compatibilité avec le calcul logique booléen à trois valeurs de Kleene. Nous relions ensuite cette représentation à la théorie des probabilités, basée sur l'hypothèse que la nécessité et la possibilité fournissent des bornes pour les valeurs de probabilité, et discutons enfin de la manière de définir une projection pertinente des valeurs réelles sur une représentation admissible.

Cela fournit les ingrédients de base pour mettre en œuvre une telle extension de la représentation des croyances dans les mécanismes existants utilisant la représentation de base. La partie opérationnelle de ce développement est disponible en code open-source.

Mots-clés : Théorie des possibilités, Logique modale, Représentation du hasard

Contents

1 Introduction **3**

2 Basic requirements **4**

3 Relation with vanilla possibility theory **5**

4 Compatibility with Boolean and three-valued logic calculus **7**
4.1 Compatibility requirements 7
4.2 Compatible interpolation 9

5 Probability requirements and compatibility **10**
5.1 Belief deriving from a probability distribution 11
5.2 Probability deriving from belief distributions 12

6 Mapping the belief in the 2D plane **15**
6.1 About complex representation 15
6.2 Polar parameterization 15
6.3 Projection on admissible representation 16

1 Introduction

We often need to parameterize the notion of partial knowledge and its related degree of belief, i.e., the estimation of the available knowledge by a subject. As formalized since the beginning of the 20th century (see e.g., [Smithson, 1970] for a presentation), human evaluation of knowledge is deeply driven by a weak notion of probability, with two key points: The idea that the degree of belief can be expressed by a number between 0 and 1 (often given as a percentage between 0% and 100% of "plausibility", "chance", or any other qualification of a belief evaluation), and the idea that for a given event, the count of its observed occurrences allows acquiring some knowledge about the possibility and necessity of this event occurrence. In [Raufaste et al., 2003], experimental results in cognitive psychology suggest that there are situations where people reason about uncertainty using the rules or possibility theory, rather than with those of probability theory, but still in connection to it. While almost all partially known information is related to probability, the human "level of truth" is more subtle and related to possibility and necessity, as formalized in the possibility theory, as discussed in [Denœux et al., 2020a] and [Denœux et al., 2020b]. Possibility theory is related to modal logic, i.e., something true in a "given context" [Fischer, 2018], which is also considered as representative of what is modeled in educational science and philosophy [Rusawuk, 2018], because it corresponds to commonsense reasoning in Piaget's sense [Smith, 1994], taking exceptions into account, i.e., considering non-monotonic reasoning.

Possibility theory is devoted to the modeling of incomplete information, for example in the context of an observer's belief regarding a potential event and surprise after the event occurrence (please refer to [Denœux et al., 2020a] for a general introduction). We would like to not only consider something that is partially true or false but also partially known. As proposed in [Mercier et al., 2021], we consider a value between -1 if false, 1 if true, and, in between if neither true nor false, we consider it is unknown, with a 0 value if totally unknown. This means that we consider being in an open world, where all that is not true, is not necessarily false, but unknown, so that the classical logic principle of excluded middle can not be used, yielding to a weaker logic. As a consequence, anything might be true unless it can be proven false, prompting a need to characterize unknown statements.

Furthermore, in symbolic artificial intelligence, i.e., knowledge representation and reasoning, a link has been drawn between this necessity/possibility dual representation and ontology formalisms [Tettamanzi et al., 2017]. The key point is that we are in an open world, as developed in this document, and the notion of negation is either not defined (as in the RDFS model yielding monotonic reasoning only), or defined at a higher level (as in OWL and more generally in description logics) that we also consider as a perspective of this work. At a more operational level, this leads to a multi-valued logic.

An important aspect of modal logic is its semantic interpretation as frame semantics[1] (also called relational semantics or Kriple semantics), i.e., the fact that a given set of assertions is satisfied in given contexts (say, given worlds), connected by accessibility relations. Accessible worlds are those compatible with, e.g. the knowledge of the agent (epistemic modality), and/or its moral norms (deontic modality), and/or its belief (doxastic modality), but also physical constraints (alethic modality) and/or time constraints. This covers many issues of knowledge representation in cognition. Given a context, an assertion is *necessary true if it is true in all accessible contexts and* possibly true if true in some of them. On the one hand, this allows considering what might be true, what should be true, what one believes to be true, and so forth. On the other hand, this formalizes the knowledge structures that capture the typical features of a situation (see e.g., [Fikes and Kehler, 1985] for a review of these notions). However, as pointed out in [Johnson-Laird and Ragni, 2019], human reasoning is not compatible with formal modal logic but is built on an internal model enumerating possible alternatives.

We are going to take these elements into account in this study, introducing a numerical quantification of the degree of necessity or possibility, allowing us to specify a class of operators that can perform deduction from a given set of facts, with the ambition to represent both rather formal deductions and more approximate analogy reasoning.

Let us develop the technical aspects of the belief representation proposed here, and extend it to a bi-dimensional case where randomness is to be taken into account. The operational part of these developments is available[2] as open-source code.

2 Basic requirements

We consider a "universe of discourse" Ω and events E which are subsets[3] of Ω. Given an event E occurrence, revisiting the possibility theory[4], related to modal logic[5], we consider the three

[1]The terminology regarding "frame" semantics/logic is a bit puzzling. Frame semantics is a formal semantics for non-classical logic systems, i.e., a construction that allows to "make sense" beyond an abstract syntactic presentation. Linguistic frame semantics relates linguistic semantics to encyclopedic knowledge, with the idea the meaning of a single word does not exist unless accessing all the essential knowledge that pertains to that word. This is related to Minsky's notion of "frame´´ to introduce the notion of context in reasoning, as discussed in [Fikes and Kehler, 1985]. A step further, frame logic is a language for knowledge representation, combining conceptual modeling with object-oriented, frame-based languages. It offers a syntax and well-defined semantics of a logic-based language.

[2] The implementation is available here:
`https://line.gitlabpages.inria.fr/aide-group/symboling/ModalType.html`,
as a component of the `symboling` package of the `aide-lib` middleware, still in development, but already available for sharing and co-development, and available at `softwareheritage.org` as `https://archive.softwareheritage.org/browse/origin/directory/?origin_url=https://gitlab.inria.fr/line/aide-group/symboling`.

[3]Here for the sake of simplicity we consider as σ-algebra $\mathcal{P}(\Omega)$ the power set, i.e. set of all subsets, of Ω. All our developments seem generalizable to σ-algebra, i.e., a subset of the power set closed by union and complementation. We also implicitly consider being in a universe that is finite but not enumerable in a tractable way.

[4]https://en.wikipedia.org/wiki/Possibility_theory

[5]https://en.wikipedia.org/wiki/Modal_logic

distributions of possibility π, probability p and necessity ν, such that:

$$0 \leq \nu(E) \leq p(E) \leq \pi(E) \leq 1$$

with the minimal requirement for both ν and π to be monotonic in the sense that:

$$E \subseteq E' \Rightarrow \nu(E) \leq \nu(E') \text{ and } \pi(E) \leq \pi(E')$$

known as "belief functions", yielding:

$$\begin{array}{rcccccc} \pi(E \cap E') & \leq & \min(\pi(E), \pi(E')) & \leq & \max(\pi(E), \pi(E')) & \leq & \pi(E \cup E') \\ \nu(E \cap E') & \leq & \min(\nu(E), \nu(E')) & \leq & \max(\nu(E), \nu(E')) & \leq & \nu(E \cup E') \end{array}$$

which is straightforward to verify, since $E \cap E' \subset E \subset E \cup E'$.

We also can state, without loss of generality[6], that:

$$\nu(\emptyset) = \pi(\emptyset) = 0 \text{ and } \nu(\Omega) = \pi(\Omega) = 1,$$

allowing to be coherent with the probability distribution. The probability distribution $p(E)$ follows the usual axiomatic definition and is thus also monotonic.

3 Relation with vanilla possibility theory

What differs from the original possibility theory is twofold.

On the one hand, we consider being in an open world, i.e., the fact that we do not explicitly know Ω. In other words, the state space is not exhaustive, i.e., the Ω we consider and with respect to which we have belief may vary with our own knowledge.

On the other hand, we consider another notion of necessity, i.e., defined as a lower bound of probability, in addition to the fact that the possibility distribution can be considered as an upper-bound of the probability distribution [Denœux et al., 2020a], which thus slightly differs from the vanilla theory semantics.

These two points are related: indeed since we are in an open world, we are not able to consider the complement $\Omega - E$ of an event, so we can not define the necessity as $\nu(E) = 1 - \pi(\Omega - E)$ as in the vanilla theory. Furthermore, in the vanilla theory, we consider that there exists a state value $\{e_\bullet\}$ with $\pi(\{e_\bullet\}) = 1$ corresponding to the actual world, which is a very reasonable assumption if Ω is a closed world, thus must include the actual world, but not if it is an open world, so that the actual world may be "outside" of our actual universe of knowledge. The consequence of considering this $\{e_\bullet\}$ implies[7] either $\nu(E) = 0$ or $\pi(E) = 1$ which means that we restrain the representation to a unique value, writing:

$$\tau(E) \stackrel{\text{def}}{=} \nu(E) + \pi(E) - 1 \text{ with } \nu(E)\,(1 - \pi(E)) = 0$$

in one to one correspondence with π and ν since[8]:

---

[6] Given any monotonic distributions ν and π the transformation:

$$\nu(E) \leftarrow \frac{\nu(E) - \nu(\emptyset)}{\nu(\Omega) - \nu(\emptyset)}, \pi(E) \leftarrow \frac{\pi(E) - \pi(\emptyset)}{\pi(\Omega) - \pi(\emptyset)}$$

allows to obtain $\nu(\emptyset) = \pi(\emptyset) = 0$ and $\nu(\Omega) = \pi(\Omega) = 1$ as soon as $\nu(\Omega) > \nu(\emptyset)$ and $\pi(\Omega) > \pi(\emptyset)$, while if $\nu(\Omega) = \nu(\emptyset)$ or $\pi(\Omega) = \pi(\emptyset)$ we are in the uninteresting trivial case where all values are the same.

[7] To put it simply, we must have either $e_\bullet \in E$ or $e_\bullet \in \Omega - E$, but because the distribution is monotonic and $\pi(\{e_\bullet\}) = 1$:

$$e_\bullet \in E \Rightarrow \{e_\bullet\} \subseteq E \Rightarrow \pi(\{e_\bullet\}) \leq \pi(E) \leq 1 \Rightarrow \pi(E) = 1,$$

we thus must have either $\pi(E) = 1$ or $\pi(\Omega - E) = 1$, i.e., $\nu(E) = 0$.

[8] Using the Heaviside function:

$$H(\tau) \stackrel{\text{def}}{=} \left\{ \begin{array}{rl} 1 & \text{if } \tau > 0 \\ 1/2 & \text{if } \tau = 0 \\ 0 & \text{if } \tau < 0 \end{array} \right.$$

$$\begin{cases} \pi(E) &= 1 + H(-\tau(E))\,\tau(E) \\ \nu(E) &= H(\tau(E))\,\tau(E) \end{cases}$$

and we easily verify what is proposed in Fig. 1, introducing partial belief between the discrete notion of binary true or false values, from either side of the unknown value. We also verify that τ is monotonic in the sense that:

$$E \subseteq E' \Rightarrow \tau(E) \leq \tau(E'), \tau(\emptyset) = -1, \tau(\Omega) = 1.$$

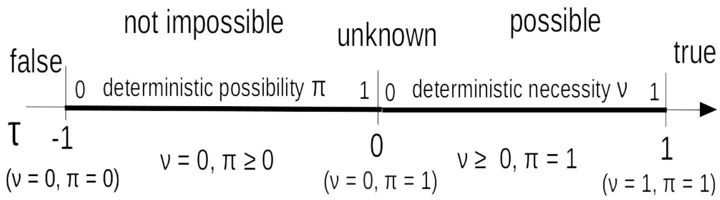

Figure 1: The 1D representation of necessity and possibility in the deterministic case. The representation using τ makes explicit the fact we implicitly consider "to what extent" a value is either true or false with respect to being unknown. The true value corresponds to 1 (fully possible and necessary), the false value to -1 (neither possible nor necessary) and the unknown value to 0, which corresponds to a fully possible but absolutely not necessary value. In the $]-1,0[$ interval the value is more or less possible but not necessary, thus partially false, and in the $]0,1[$ interval the value is fully possible and more or less necessary, thus partially true.

The restriction of such 1D representation is discussed in [Dubois and Prade, 2015]. For instance, they explain that "a subjective distribution would be uniform in both cases where the agent is fully ignorant and when he perfectly knows that the stochastic process generating the events is pure randomness" and there is no clue about the precise underlying epistemic state. Here we would like to make the distinction between randomness and (full or partial) ignorance.

To this end, in our case, the necessity is another "independent" distribution. We can thus make the distinction between the *deterministic case where the relation to probability is not considered and the* stochastic case, where the probability is estimated through an interval, as illustrated in Fig. 2. The latter includes the former, which corresponds to stating belief "up to the best current knowledge of the universe".

In both cases, we are driven by the principle of minimal specificity, stating that any hypothesis not known to be impossible cannot be ruled out, while given some knowledge we must consider the smallest possibility and largest necessity given this knowledge. This idea is directly related to the entropy principle, i.e., to consider the distribution with maximum randomness compatible with the given knowledge or assumption.

A step further, in the scope of the vanilla theory, it seems natural to assume, for exclusive events $E \cap E' = \emptyset$, the possibility $\pi(E \cup E')$ must only be a function of $\pi(E)$ and $\pi(E')$, and they further propose to use the max operator, obtaining an explicit formula $\pi(E) = \max_{e \in \Omega} \pi(\{e\})$. This is perfectly reasonable in a closed world, i.e., given that Ω is fixed. In our case, we have no such explicit formula, only constraints of monotonicity, and additional constraints to guarantee compatibility with Boolean and probability calculus, now developed.

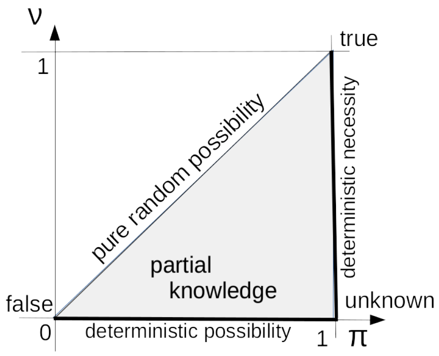

Figure 2: The 2D representation of necessity and possibility in the (π, ν) space. The deterministic case corresponds to values on the thick black segments. The diagonal with $\pi = \nu$ corresponds to a known probability. Admissible values are inside the drawn triangle.

4 Compatibility with Boolean and three-valued logic calculus

4.1 Compatibility requirements

Back to our extension of the vanilla possibility theory, we are going to analyze how this could be compatible with usual Boolean calculus, and its extension to "unknown" values, which can be done in several ways, here we adopt the restriction, to an open world, of the Kleene logic[9], where the unknown state can be thought of as neither true nor false, as made explicit in Table. 1.

The link between a probabilistic estimation as a measure of the occurrence of an event, and another interpretation as a degree of confidence in a proposition within the scope of three-valued logic has been already studied in, e.g., [Negri, 2013]. In the deterministic case, three-valued representations of imperfect information with respect to belief function representations have been discussed in detail in [Ciucci et al., 2014]. Here we simply analyze at the implementation level, to what extent we can consider three-valued logic interpretation as a limit case of the proposed representation.

Regarding implication, for a given property $\mathcal{P}(E)$ on an event E, the usual definition of $\mathcal{P}(E_1) \Rightarrow \mathcal{P}(E_2)$ corresponds to the occurrence of the event $E_1 \cup (\Omega - E_2)$. As in modal logic,

[9]https://en.wikipedia.org/wiki/Three-valued_logic

[10]Thanks to the fact that the distributions are monotonic, since:

$$\tau(E_1 \cap E_2) \leq \min(\tau(E_1), \tau(E_2)) \leq \max(\tau(E_1), \tau(E_2)) \leq \tau(E_1 \cup E_2)$$

we easily verify, than, if one value $\tau(E_i) = -1$ then $\tau(E_1 \cap E_2) = -1$, if one value $\tau(E_i) = 1$ then $\tau(E_1 \cup E_2) = 1$, while if both values $\tau(E_1) = \tau(E_2) = -1$ then $\tau(E_1 \cup E_2)$ is unconstrained. Furthermore, in the unknown cases with one value $\tau(E_i) = 0$ we observe that the value is only partially constrained as stated in the table.
Regarding the symmetric difference $E_1 \ominus E_2 = (E \cup E') - (E \cap E')$, if one value is true and one value is false we obtain $\tau(E \cap E') = -1$ and $\tau(E \cup E') = 1$ which means that E_1 and E_2 do not intersect while their union occurred so that the symmetric difference is true. If both are true, then $\tau(E \cup E') = 1$ and $\tau(E \cap E')$ is unconstrained, so that $E_1 \ominus E_2$ can be either true or false, thus is unconstrained. There is a similar situation if both are false. If one value is unknown then $\tau(E \cap E') \leq 0$ and $\tau(E \cup E') \geq 0$ and we observe with a similar reasoning yields to the fact that $\tau(E_1 \ominus E_2) \leq 0$ in this case.
Regarding the set difference, since $(E_1 - E_2) \subset E_1$ thus $\tau(E_1 - E_2) \leq \tau(E_1)$, thus if $\tau(E_1) = -1$ then $\tau(E_2 - E_1) = -1$, while if $\tau(E_1) = 0$ then $\tau(E_2 - E_1) \leq 0$. Up to our best understanding, there is no further constraint in this case.

		conjonction	disjunction	exclusive disjuction	set difference
$\tau(E_1)$	$\tau(E_2) =$	$\tau(E_1 \cap E_2) =$ $\min(\tau(E_1), \tau(E_2))$	$\tau(E_1 \cup E_2) =$ $\max(\tau(E_1), \tau(E_2))$	$\tau(E_1 \ominus E_2) =$ $-\tau(E_1)\,\tau(E_2)$	$\tau(E_1 - E_2)$ (see text)
1	1	1	**1**	-1	-1
1	0	0 ($\leq$ **0**)	**1**	0 ($\geq$ **0**)	0
1	-1	**-1**	**1**	**1**	1
0	1	0 ($\leq$ **0**)	**1**	0 ($\geq$ **0**)	0 ($\leq$ **0**)
0	0	0 ($\leq$ **0**)	0 ($\geq$ **0**)	0 ($\geq$ **0**)	0 ($\leq$ **0**)
0	-1	**-1**	0 ($\geq$ **0**)	0 ($\geq$ **0**)	0 ($\leq$ **0**)
-1	1	**-1**	**1**	**1**	**-1**
-1	0	**-1**	0 ($\geq$ **0**)	0 ($\geq$ **0**)	**-1**
-1	-1	**-1**	-1	-1	**-1**

Table 1: Three-valued logic computation rules for conjunction, disjunction, symmetric difference, and difference of two events. Both Boolean ($\mathcal{P}(E)$ stating that the occurrence of the event E is either true, false, or unknown) and ensemble notations are given, with an explicit formula valid for the $\{-1, 0, 1\}$ values. Bold values correspond to the fact that the value is implied by the distribution monotonicity[10].

since we are in a situation where we can not consider Ω, we only can define a restrained implication operator, i.e. in relation to inclusion: If $E_1 \supseteq E_2$ is given a priory, then $\tau(E_1) \geq \tau(E_2)$ and:
- Given values of $\tau(E_1)$ and $\tau(E_2)$ we may evaluate if the assertion $E_1 \supseteq E_2$ (or $E_1 \supset E_2$) is contradicted, thus false (i.e., if $\tau(E_1) \not\geq \tau(E_2)$), otherwise it is plausible, but we can not decide, given τ values, if it is true.
- Given the fact that $E_1 \supseteq E_2$ (or $E_1 \supset E_2$), if $\tau(E_1)$ is given, we may deduce some constraint on $\tau(E_2)$ (e.g., if $\tau(E_1) = -1$ then $\tau(E_2) = -1$), with dual deductions if the $\tau(E_2)$ is given (e.g., if $\tau(E_2) = 0$, then $\tau(E_1) \neq -1$).
- We can also define the set-difference $E_1 - E_2$ and obtain[11]:

$$\tau(E_1 - E_2) = \text{ if } \tau(E_1) = 1 \text{ then } -\tau(E_2) \text{ else } \tau(E_1)$$

ans some additional values (e.g., $E_2 - E_1 = \emptyset$ and $\tau(E_2 - E_1) = -1$, with similar results for $\tau((E_2 - E_1) \cap E_2) = \tau(\emptyset) = -1$, $\tau((E_2 - E_1) \cup E_2) = \tau(E_1)$, etc).

Conversely, if we consider a closed world and the vanilla theory, allowing to state $\pi(E1 \cup E2) = \max(\pi(E1), \pi(E2))$ and define $\pi(\Omega - E) = 1 - \nu(E)$ we easily obtain (see, e.g., [Denœux et al., 2020a]) the compatibility with the Boolean calculus, and partially regarding the three value logic (i.e., if $\mathcal{P}(E)$ is unknown so is $\mathcal{P}(\Omega - E)$, while it is not determinate[12] for $\mathcal{P}(E1 \cup E2)$ if $\mathcal{P}(E_1)$ or $\mathcal{P}(E_2)$ are unknown).

To summarize, while the vanilla theory is fully compatible with Boolean calculus, our extended theory or the vanilla theory with respect to three-valued logic, requires additional deduction rules to be stated, in order to fit the required conditions.

[11] We also can look for the simpler semi-algebraic expression (see [Bochnak et al., 2013] for details) equivalent to this conditional expression on $\{-1, 0, 1\}$ and obtain:

$$\tau(E_1 - E_2) = \frac{\tau(E_1)}{2}\left(1 - \tau(E_1)\,\tau(E_2) - \tau(E_1) - \tau(E_2)\right).$$

[12] If, for instance, both $\mathcal{P}(E_1)$ or $\mathcal{P}(E_2)$ are unknown, we obtain:

$$\pi(E_1) = \pi(E_2) = 1, \nu(E_1) = \nu(E_2) = 0 \Rightarrow \pi(E1 \cup E2) = 1, \nu(E1 \cup E2) \geq 0$$

so that the result can technically be either true or unknown.

1					
$\tau_2 = \nu_2$		$\tau_{and}(\tau_1, \tau_2) \leq \pi_1 - 1 \leq 0$ $0 \leq \nu_2 \leq \tau_{or}(\tau_1, \tau_2)$		$0 \leq \tau_{and}(\tau_1, \tau_2) = \min(\nu_1, \nu_2)$ $0 \leq \max(\nu_1, \nu_2) \leq \tau_{or}(\tau_1, \tau_2)$	
0					
$\tau_2 = \pi_2 - 1$		$\tau_{and}(\tau_1, \tau_2) \leq \min(\pi_1, \pi_2) - 1 \leq 0$ $\tau_{or}(\tau_1, \tau_2) = \max(\pi_1, \pi_2) - 1 \leq 0$		$\tau_{and}(\tau_1, \tau_2) \leq \pi_2 - 1 \leq 0$ $0 \leq \nu_1 \leq \tau_{or}(\tau_1, \tau_2)$	
-1					
τ_2 τ_1	-1	$\tau_1 = \pi_1 - 1$	0	$\tau_1 = \nu_1$	1

Table 2: Bounds obtained with the vanilla assumptions for conjunction and disjunction.

Moreover, under the standard possibility theory [Denœux et al., 2020a], we obtain[13] interesting bounds reported in Table 2, that will design the interpolation, as proposed now.

4.2 Compatible interpolation

The next point is to interpolate values in the $[-1, 1]$ interval so that it is compatible with the previous three-valued logic table. We can easily see that the min and max operators interpolate conjunction and disjunction, while exclusive disjunction corresponds to a product and set difference to a polynomial expression, as shown in Fig.3.

The situation is however not so simple for conjunction and disjunction. Let us consider any regular function to estimate an N-ary operator, constrained by three-valued logic values on $\{-1, 0, 1\}$. Such a function has a series expansion of the form:

$$\tau_{op}(\tau_1, \cdots \tau_N) = \sum_{d_1 \geq 0, \cdots d_N \geq 0} k_{d_1, \cdots d_N} \, \tau_1^{k_1} \, \cdots \, \tau_2^{k_N}$$

In fact, since:

$$\tau_i \in \{-1, 0, 1\} \Rightarrow \forall d > 2, (\tau_i)^d = \left\{ \begin{array}{ll} (\tau_i)^2 & \text{if } d \text{ is even} \\ (\tau_i)^1 & \text{if } d \text{ is odd} \end{array} \right. ,$$

we are left with a polynomial of degree 2 on the constrained values. Considering higher degree constrained by the three-valued logic values $\{-1, 0, 1\}$ will thus only generate redundant unknowns.

We thus may consider interpolating between the $\{-1, 0, 1\}$ values in the $[-1, 1]$ interval using a polynomial of degree two of the form:

$$\tau_{op}(\tau_1, \cdots \tau_N) \stackrel{\text{def}}{=} \sum_{(d_1, \cdots d_N) \in \{0,1,2\}^N} a_{d_1, \cdots d_N} \, \prod_{i=1}^N (\tau_i)^{d_i}$$

thus with 3^N unknown coefficients $a_{d_1, \cdots d_N}$, we obtain a linear system of 3^N equations, the matrix of which writes $M_{(\tau_1, \cdots \tau_N)}^{(d_1, \cdots d_N)} = \prod_{i=1}^N (\tau_i)^{d_i}$ and it is straightforward to verify that all matrix rows are independent, yielding a unique solution.

Thanks to this design choice, in our case, we obtain[14]:

[13] In standard possibility theory:

$$\begin{array}{ccccccc} \pi(E \cap E') & \leq & \min(\pi(E), \pi(E')) & \leq & \max(\pi(E), \pi(E')) & = & \pi(E \cup E') \\ \nu(E \cap E') & = & \min(\nu(E), \nu(E')) & \leq & \max(\nu(E), \nu(E')) & \leq & \nu(E \cup E') \end{array}$$

while $\tau = \pi - 1$ if $\tau \leq 0$ and $\tau = \nu$ if $\tau \geq 0$, yielding the following results.
Furthermore:

$$\left. \begin{array}{lll} \tau_1 \leq 0 & \Rightarrow & \tau_1 = \pi_1 - 1, \nu_1 = 0 \\ \tau_2 \geq 0 & \Rightarrow & \tau_2 = \nu_2, \pi_2 = 1 \end{array} \right\} \Rightarrow \left\{ \begin{array}{lll} \tau_{and} \leq 0 & \Rightarrow & \tau_{and} = \pi_{and} - 1 \leq \min(\pi_1, \pi_2) - 1 = \pi_1 - 1 \\ \tau_{and} \geq 0 & \Rightarrow & \tau_{and} = \nu_{and} = \min(\nu_1, \nu_2) = 0 \\ \tau_{or} \leq 0 & \Rightarrow & \tau_{or} = \pi_{or} - 1 = \max(\pi_1, \pi_2) - 1 = 1 - 1 = 0 \\ \tau_{or} \geq 0 & \Rightarrow & \tau_{or} = \nu_{or} \geq \max(\nu_1, \nu_2) = \nu_2 \end{array} \right.$$

allowing to derive the proposed bounds, with similar derivations in other cases.

[14] We also obtain, in coherence with the previous developments:

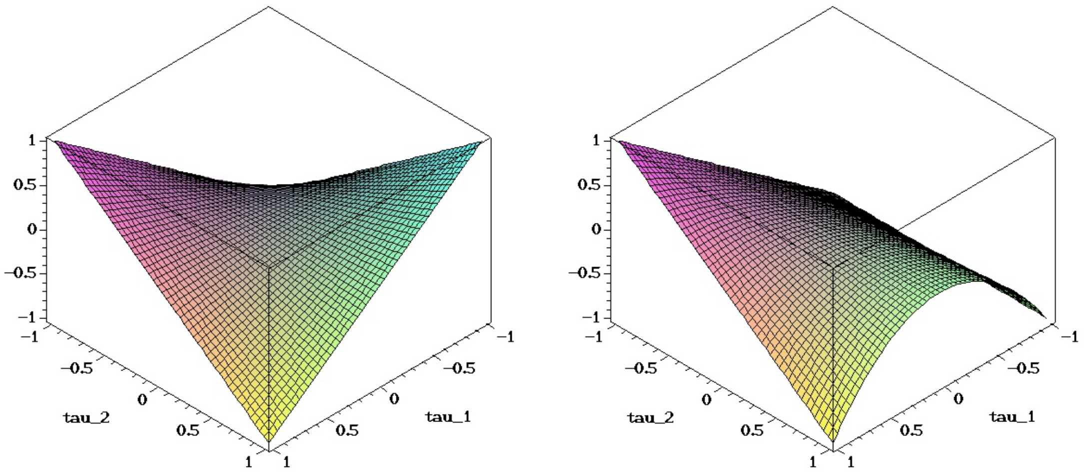

Figure 3: The τ_{xor} polynomial interpolation on the left and the τ_{minus} polynomial interpolation on the right, showing the regularity of both interpolations.

$$\begin{array}{rcl} \tau_{and}(\tau_1, \tau_2) & = & \frac{1}{2}\left(\tau_1^2\,\tau_2^2 - \tau_1^2 + \tau_1\,\tau_2 - \tau_2^2 + \tau_1 + \tau_2\right) \\ \tau_{or}(\tau_1, \tau_2) & = & \frac{1}{2}\left(-\tau_1^2\,\tau_2^2 + \tau_1^2 - \tau_1\,\tau_2 + \tau_2^2 + \tau_1 + \tau_2\right) \end{array}$$

This is directly derived from algebraic considerations[15], and Fig. 4 shows that this solution is coherent with the vanilla min operator.

However, there is a caveat because the obtained operators are not monotonic, but have unexpected behaviors with respect to the min and max bounds, as reported in Table 3, using symbolic algebra derivations[15]. In our case conjunction and disjunction is always bounded by the min and max value contrary to what is expected for a monotonic operator. We thus propose simply using the min and max operators. A step further, the present development may be used in future work in order to design better operators, for instance considering expansion of degree higher than 2, if additional optimization requirements are introduced.

This interpolation implementation is available[2] as open-source code.

5 Probability requirements and compatibility

The relation between vanilla possibility and probability has been extensively studied, including the relation and distinction with respect to the Dempster–Shafer theory[16], where probability is bounded by "belief" and "plausibility", for which the semantic interpretation differs, while the formal definition is well defined, from a notion of "mass" on the event power set. Further study of relations between possibility as supremum preserving normalized measure and probability distributions is available in [Troffaes et al., 2011], while a discussion on the semantic implications

$$\begin{array}{rcl} \tau_{xor}(\tau_1, \tau_2) & = & -\tau_1\,\tau_2 \\ \tau_{minus}(\tau_1, \tau_2) & = & \frac{\tau_1}{2}\,(1 - \tau_1\,\tau_2 - \tau_1 - \tau_2) \end{array}$$

[15] Derivations are available here: `https://gitlab.inria.fr/line/aide-group/symboling/-/raw/master/src/RR-modal/modal-combination.mpl`

[16] https://en.wikipedia.org/wiki/Dempster-Shafer_theory

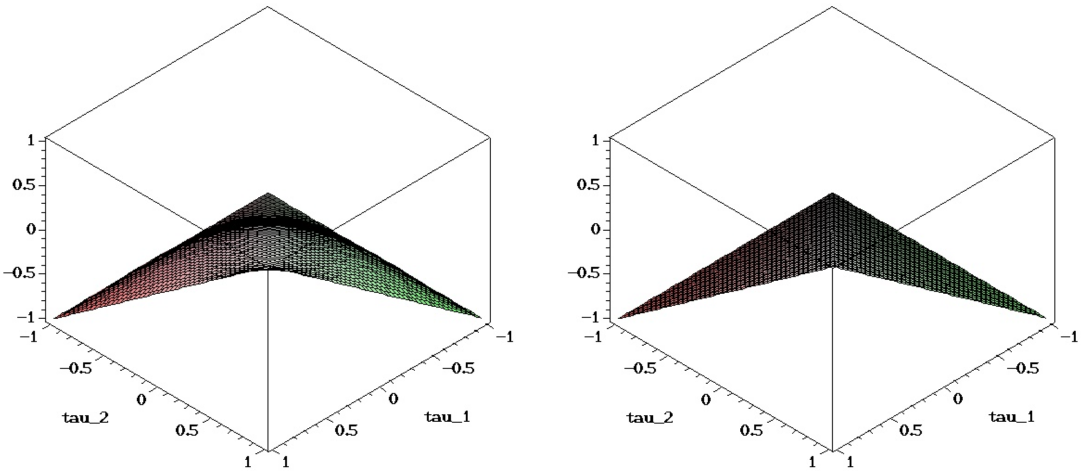

Figure 4: The τ_{and} polynomial interpolation on the left compared to the min interpolation on the right. The former is a simple regularized version of the latter. Numerically the mean-square error between both function is lower than 6% on $[0,1] \times [-1,0]$ and $[-1,0] \times [0,1]$ and lower than 11% on $[-1,0] \times [-1,0]$ and $[0,1] \times [0,1]$. A dual result is obtained regarding τ_{or} polynomial interpolation and the max interpolation.

is given here [Denœux et al., 2020a]. The relation between probability interval via the notion of probability boxes and possibility measures is also well-studied [Troffaes et al., 2011] and we will not redevelop these two aspects, only consider two precise issues: (i) how to define our stochastic necessity and possibility given a probabilistic framework, and (ii) given a couple of necessity and probability distributions, how to calculate the closest distribution couple compatible with a probability distribution.

To the first end, inspired by a [Vallaeys, 2021] formalism proposal, we consider that an agent wants to get some information about its knowledge of an event E, given a set of parameterized a-priory knowledge $\mathcal{K}$ about the universe and some observations $O \subseteq \mathcal{O}$. We consider that its partial knowledge may be due to either to *incomplete information*, i.e., imprecision or *stochastic information*, i.e., uncertainty due to randomness, as developed in [Denœux et al., 2020a]. Its goal is to explore all $\kappa \in \mathcal{K}$ given the observations to estimate its belief regarding an event E, with this key idea to separate the agent knowledge from the randomness of the observable.

5.1 Belief deriving from a probability distribution

In order to relate the belief function to a probability distribution, following [Vallaeys, 2021], on track is to consider that there exists an underlying implicit agent conditional probability underneath the user belief so that:

$$\begin{array}{rcl} \nu(E) & = & \min_{\kappa \in \mathcal{K}} P_\kappa(E|O) \\ \pi(E) & = & \max_{\kappa \in \mathcal{K}} P_\kappa(E|O) \end{array}$$

We refer to this as the H_{cond} hypothesis, in the sequel.

This differs from [Denœux et al., 2020a] who proposed to consider likelihood functions, i.e., $\nu(E) = \min_{o \in O} \tilde{P}(E|o)$, for a non-parametric underlying implicit conditional probability, with

τ_2					
1					
		$\min(\tau_1,\tau_2) \leq \tau_{and}(\tau_1,\tau_2) \leq \max(\tau_1,\tau_2)$ $\min(\tau_1,\tau_2) \leq \tau_{or}(\tau_1,\tau_2) \leq \max(\tau_1,\tau_2)$ $\min(\tau_1,\tau_2) \leq \tau_{xor}(\tau_1,\tau_2) \leq \max(\tau_1,\tau_2)$ $\min(\tau_1,\tau_2) \leq \tau_{minus}(\tau_1,\tau_2) \leq \max(\tau_1,\tau_2)$		$\min(\tau_1,\tau_2) \simeq \tau_{and}(\tau_1,\tau_2) \leq \max(\tau_1,\tau_2)$ $\min(\tau_1,\tau_2) \leq \tau_{or}(\tau_1,\tau_2) \simeq \max(\tau_1,\tau_2)$ $\tau_{xor}(\tau_1,\tau_2) \leq \min(\tau_1,\tau_2)$ $\tau_{minus}(\tau_1,\tau_2) \leq \max(\tau_1,\tau_2)$	
0					
		$\min(\tau_1,\tau_2) \simeq \tau_{and}(\tau_1,\tau_2) \leq \max(\tau_1,\tau_2)$ $\min(\tau_1,\tau_2) \leq \tau_{or}(\tau_1,\tau_2) \simeq \max(\tau_1,\tau_2)$ $\max(\tau_1,\tau_2) \leq \tau_{xor}(\tau_1,\tau_2)$ $\min(\tau_1,\tau_2) \leq \tau_{minus}(\tau_1,\tau_2)$		$\min(\tau_1,\tau_2) \leq \tau_{and}(\tau_1,\tau_2) \leq \max(\tau_1,\tau_2)$ $\min(\tau_1,\tau_2) \leq \tau_{or}(\tau_1,\tau_2) \leq \max(\tau_1,\tau_2)$ $\min(\tau_1,\tau_2) \leq \tau_{xor}(\tau_1,\tau_2) \leq \max(\tau_1,\tau_2)$ $\min(\tau_1,\tau_2) \leq \tau_{minus}(\tau_1,\tau_2) \leq \max(\tau_1,\tau_2)$	
-1					
τ_1	-1		0		1

Table 3: Bounds obtained with the polynomial interpolation of three-valued logic operators, see text for details.

the corresponding definition for $\pi(E)$. This, on one hand, does not explicitize the fact that belief relies on previous knowledge, and, on the other hand, highly depends on the granularity of the observable, i.e. the way we decide to segment O in elements o.

This being stated after [Vallaeys, 2021], we observe that the algebra derivations are the same as in [Denœux et al., 2020a], i.e., by construction ν and π are monotonic in the previous sense. Furthermore, they are dual[17] with respect to the complementation:

$$\pi(E) + \nu(\Omega - E) = \pi(\Omega - E) + \nu(E) = 1,$$

but we do not either $\nu(E) = 0$ or $\pi(E) = 1$, being beyond the vanilla theory. We also obtain using the fact that $P_\kappa(E \cup E'|O) = P_\kappa(E|O) + P_\kappa(E'|O) - P_\kappa(E \cap E'|O)$:

$$\begin{array}{l} \nu(E) + \nu(E') \leq \nu(E \cap E') + \pi(E \cup E') \leq \pi(E) + \pi(E') \\ \max(0, \nu(E) + \nu(E') - 1) \leq \nu(E \cap E') \leq \pi(E \cup E') \leq \min(1, \pi(E) + \pi(E') - 1) \end{array}$$

while if E' and E'' are independent, from $P_\kappa(E' \cap E''|O) = P_\kappa(E'|O)\, P_\kappa(E''|O)$ we obtain:

$$\begin{array}{rcccccl} \nu(E')\,\nu(E'') & \leq & \nu(E' \cap E'') & \leq & \pi(E' \cap E'') & \leq & \pi(E')\,\pi(E'') \\ \nu(E') + \nu(E'') - \nu(E')\,\nu(E'') & \leq & \nu(E' \cup E'') & \leq & \pi(E' \cup E'') & \leq & \pi(E') + \pi(E'') - \pi(E')\,\pi(E''). \end{array}$$

5.2 Probability deriving from belief distributions

On the reverse, without the H_{cond} hypothesis, considering Ω as a summable set of singletons $\{e\}$ and given two distributions ν and π, a probability distribution is defined on all events and is compatible with the necessity and possibility distribution, if and only if:

$$\nu(E) \leq \textstyle\sum_{e \in E} p(\{e\}) \leq \pi(E),$$

which is the main condition of *admissibility* for the ν and π.

A double necessary condition[18] for $p(\Omega) = 1$ is that:

$$\textstyle\sum_{e \in \Omega} \nu(\{e\}) \leq 1 \leq \sum_{e \in \Omega} \pi(\{e\}).$$

- The counter-fact that $\sum_{e \in \Omega} \pi(\{e\}) < 1$ can be interpreted as the fact that some possible events are missing and it is always possible to add a fully unknown $\nu(U) = 0, \pi(U) = 1$ unknown

[17] We, e.g., derive:

$$\nu(\Omega - E) = \min_{\kappa \in \mathcal{K}} (1 - P_\kappa(E|O)) = 1 - \max_{\kappa \in \mathcal{K}} P_\kappa(E|O) = 1 - \Pi(E).$$

[18] Otherwise, if $\sum_{e \in \Omega} \nu(\{e\}) > 1$, this implies $P(\Omega) = \sum_{e \in \Omega} p(\{e\}) > 1$ which is not compatible with a probability distribution. There is a symmetric observation for $\sum_{e \in \Omega} \pi(\{e\}) < 1$.

"otherwise" event in order to fulfill the condition. This trick explains why, as pointed out in [Denœux et al., 2020a] that assumes that $\pi(\{e\}_\bullet) = 1$ for some punctual event $\{e\}_\bullet$, we always can consider that a possibility distribution can be considered as the upper-bound of a probability.
- The counter-fact that $\sum_{e\in\Omega} \nu(\{e\}) > 1$ must be interpreted differently, as the fact that necessity is over-evaluated because impossible events have not to be taken into account; this means that the distribution has to be adjusted as developed in this section. Given this double necessary condition, all probability distributions, parameterized by $\alpha = [\cdots \alpha_e \cdots]$, such that:

$$p_\alpha(\{e\}) \stackrel{\text{def}}{=} \alpha_e\, \nu(\{e\}) + (1-\alpha_e)\, \pi(\{e\}), 0 \leq \alpha_e \leq 1, \sum_{e\in\Omega} p(\{e\}) = 1$$

are well defined singletons on $\{e\}$. Such a α exists[19]. However, in order to have a distribution defined on all events E, we need a stronger condition[20] and, if not, we must adjust[21] the distribution ν and π in order the condition to be verified.

Among all distributions, we may consider the maximal entropy[22] which is the one that makes the fewest assumptions about the data distribution, given the present constraints, thus best represents the current state of knowledge.

This can be written as an iterative linear programming problem [23]:

[19] It is easy to verify that a solution writes:

$$\alpha_e = \alpha_\bullet \in [0,1], \quad \alpha_\bullet \stackrel{\text{def}}{=} \frac{\sum_{e\in\Omega} \pi(\{e\})-1}{\sum_{e\in\Omega} \pi(\{e\})-\nu(\{e\})}, \sum_{e\in\Omega} p_{\alpha_\bullet}(\{e\}) = 1, \nu(\{e\}) \leq p_{\alpha_\bullet}(\{e\}) \leq \pi(\{e\}).$$

[20] A sufficient but not necessary condition is the fact that, on a given event E, the possibility is subadditive and the necessity is superadditive, in the following sense:

$$\underbrace{\nu(E) \leq \sum_{e\in E} \nu(\{e\})}_{superadditive} \leq p(E) \leq \underbrace{\sum_{e\in E} \pi(\{e\}) \leq \pi(E)}_{subadditive},$$

which is a rather strong assumption.

[21] Another alternative could have been, given a necessity and possibility distribution couple, to exclude events with incompatible probabilities from the probability distribution, in order to obtain a partially defined probability, but this seems to be intractable, because the probability must be defined on a σ-algebra Σ, so that, as soon as an event $E \notin \Sigma$ is excluded, we must exclude $\Omega - E$ and recursively exclude at least one event E' for which there exists a $E'' \in \Sigma$ with either $E = E' \cup E''$ or $E = E' \cap E''$ which seems obviously to be a combinatory explosive intractable ill-defined problem, and may result to only consider the trivial σ-algebra.

[22] Considering the discrete probability distribution it writes:

$$\max_{\{\cdots p(\{e\})\cdots\}} H \stackrel{\text{def}}{=} -\sum_{e\in\Omega} p(\{e\}) \log_2 (p(\{e\}))\,.$$

The distribution with the maximum entropy corresponds to the uniform distribution u in the absence of constraint. Maximizing this entropy corresponds also to minimizing the KL-divergence between the p distribution and the uniform distribution, as easily verified.

[23] Let us consider:

$$\begin{array}{rcl} \mathcal{L} & = & \min_{p,\tilde\nu,\tilde\pi} \max_{\lambda_\bullet,\lambda_e,\mu_e} \sum_{e\in\Omega} p(\{e\}) \log_2 (p(\{e\})) + \frac{1}{\upsilon} \sum_E (\nu(E) - \tilde\nu(E)) + (\tilde\pi(E) - \pi(E)) \\ & + & \lambda_\bullet(\sum_{e\in E} p(\{e\}) - 1) + \lambda_E(\sum_{e\in E} p(\{e\}) - \nu(E)) + \mu_E(\pi(E) - \sum_{e\in E} p(\{e\})) \end{array}$$

where $\lambda_\bullet$ is the Lagrangian multiplier allowing to verify the normalization equation and λ_E and μ_E the Karush–Kuhn–Tucker multipliers corresponding to the inequality constraints.

In order to turn this into a linear programming problem, we consider an initial value $\bar p(\{e\}) \stackrel{\text{def}}{=} p_{\alpha_\bullet}(\{e\})$ of the probability distribution, as defined in footnote[19], and initial admissible values :

$$\bar\nu(E) \stackrel{\text{def}}{=} \min(\nu(E), \sum_{e\in E} \bar p(\{e\})), \bar\pi(E) \stackrel{\text{def}}{=} \max(\pi(E), \sum_{e\in E} \bar p(\{e\})),$$

thus either equal to the initial distribution or adjusted in order for all admissible constraints to be verified. We then optimize the derogate criterion:

$$\sum_{e\in E} p(\{e\})\, log_2\, (\bar p(\{e\})) + \text{constraints} \cdots$$

with respect to $p(\{e\}), \tilde\nu(E), \tilde\pi(E)$ which defines a linear programming problem with an admissible initial solution, and the initial solution can be refined using, e.g., a simplex algorithm. Then we can iterate this linear programming problem reintroducing the obtained probability value, and readjusting the distributions accordingly. During the iteration, we adjust the hyper-parameter $\upsilon \to 0$ in order to obtain a minimal adjustment of the ν and π

$$\min_{p(\{e\}),\tilde{\nu}(E),\tilde{\pi}(E)} \sum_{e\in E} p(\{e\})\, log_2\left(p(\{e\})\right) + \frac{1}{\upsilon} \sum_E (\nu(E) - \tilde{\nu}(E)) + (\tilde{\pi}(E) - \pi(E))$$

with:

$$\sum_{e\in\Omega} p(\{e\}) = 1 \text{ and } \forall E,\ 0 \leq \tilde{\nu}(E) \leq \nu(E) \leq \pi(E) \leq \tilde{\pi}(E) \leq 1 \text{ and } \tilde{\nu}(E) \leq \sum_{e\in E} p(\{e\}) \leq \tilde{\pi}(E)$$

where we lower ν to $\tilde{\nu}$ and increase π to $\tilde{\pi}$ until the inequality is verified, while we minimize the KL-divergence of the p distribution with respect to the uniform distribution. The hyper-parameter $\upsilon \rightarrow 0$ allows finding the admissible distribution, as close as possible to the original one. The main aspects of this mechanism are illustrated in Fig. 5 for the 2D case, while the 1D case is trivial[24].

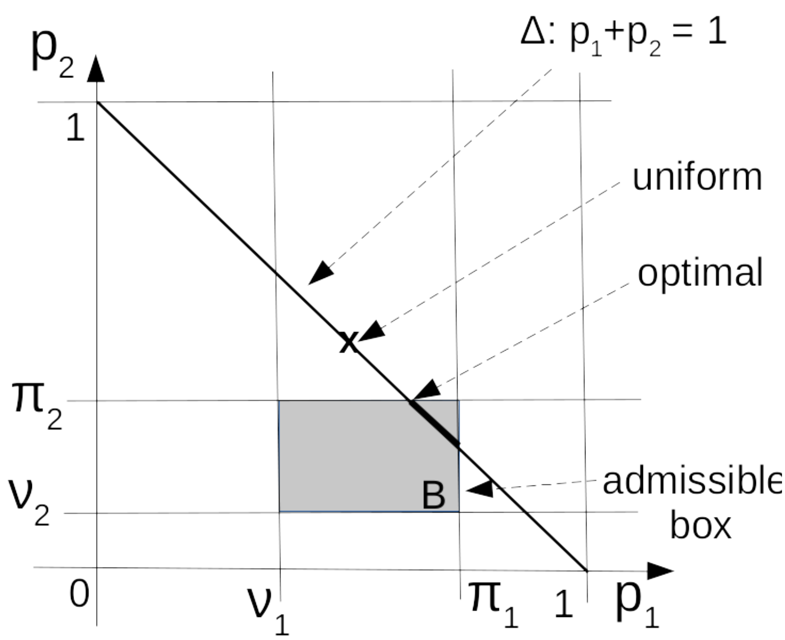

Figure 5: Illustrating the conjoint estimation of an optimal probability distribution and the necessity and possibility adjustment for the problem to be well-defined, in the 2D case. In order some solution to exist, the admissible rectangle $B = [\nu_1, \pi_1] \times [\nu_2, \pi_2]$ must intersect the Δ line of equation $p_1 + p_2 = 1$. When this condition is fulfilled, among all solutions, the probability distribution with the highest entropy is the closest to the uniform distribution.

---

distribution.

To interpret this setup let us consider a sub-optimal algorithm, where we only consider $\bar{\nu}$ and $\bar{\pi}$ so that we obtain admissible distributions, only adjusting the probability distribution on the singleton, so that $\lambda_E = \nu_E = 0$ for the event that is not a singleton. In this restrained case the normal equations write:

$$0 = \partial_{p(\{e\})}\mathcal{L} = \log_2\left(p(\{e\})\right) + 1 + \lambda_\bullet + \lambda_{\{e\}} - \mu_{\{e\}}.$$

and if $p(\{e\})$ is either constrained to be equal to $\nu(\{e\})$ or $\pi(\{e\})$ or unconstrained, so that $\lambda_{\{e\}} = \mu_{\{e\}} = 0$ and we are left with $\log_2\left(p(\{e\})\right) + 1 + \lambda_\bullet = 0$ for the unconstrained value of $p(\{e\})$, which means that for these values, the optimal value is the same, i.e., the optimal value is a uniform value on unconstrained dimensions. In other words, we look for the probability distribution as close as possible to the uniform distribution, given the constraints.

[24] In the 1D case, the maximum entropy is obtained for $p = 1/2$ so that we obtain $p = \max(\nu, \min(\pi, 1/2))$.

6 Mapping the belief in the 2D plane

6.1 About complex representation

Since $\zeta \stackrel{\text{def}}{=} (\pi, \nu)$ is a 2D quantity, we may wonder if using complex numbers could bring some benefit to manipulate the related information, including at the numerical implementation level, using for instance complex-valued neural networks (see, e.g., [Guberman, 2016] for a general presentation).

Obviously, the complex plane can be considered as a 2D real plane, enriched by the complex multiplication as an additional algebraic operation. Up to our best analysis, multiplication has no use in our development, while holomorphicity would be an obstacle to defining the following mapping.

We thus considered being "inspired" by complex representation, in particular with the idea to introduce a polar representation, as follows, but no more.

6.2 Polar parameterization

In order to estimate numerically $\nu \leq p \leq \pi$ in 2D we are going to map (π, ν) in such a way that the deterministic value $\tau = \nu + \pi - 1 \in [-1, 1]$, while in the second dimension we would like to represent the precision on the randomness which is related to $\pi - \nu$ which is the probability interval length. We propose:

$$\left\{ \begin{array}{rcccl} \tau & \stackrel{\text{def}}{=} & \nu + \pi - 1 & \in & [-1, 1] \\ \rho & \stackrel{\text{def}}{=} & \nu - \pi + 1 & \in & [0, 1] \end{array} \right. \Leftrightarrow \text{ with } \left\{ \begin{array}{rcl} \pi & = & \frac{\tau - \rho}{2} + 1 \\ \nu & = & \frac{\tau + \rho}{2}, \end{array} \right.$$

introducing ρ, with $\rho = 1$ if $\pi = \nu$ thus corresponding to a precise probability estimation and $\rho = 0$ if $\pi = 1$ and $\nu = 0$ thus the probability is unbounded.

A step further this is in direct relation with a polar parameterization, where τ is the abscissa, while the orientation θ corresponds to the probability: 0 for a true value, i.e., a probability of 1 and Π for a false value, i.e., a probability of 0, we can propose that the 2D vector orientation could correspond to the probability when defined, and its magnitude to the amplitude of randomness, thus 1 if completely random with $\nu = \pi$ and 0 if undefined, i.e., when $\pi - \nu = 1$. Considering these requirement, as illustrated in Fig. 6, we propose to define:

$$\left\{ \begin{array}{rcl} \tau & \stackrel{\text{def}}{=} & \rho \cos(\theta) = \nu + \pi - 1 \\ \xi & \stackrel{\text{def}}{=} & \rho \sin(\theta) = 2\sqrt{\nu\,(1 - \pi)} \\ \rho & = & \nu - \pi + 1 \\ \theta & = & \arccos \frac{\nu + \pi - 1}{\nu - \pi + 1} \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rcl} \pi & = & 1 + \frac{1}{2}(\tau - \rho) \\ \nu & = & \frac{1}{2}(\tau + \rho) \\ \rho^2 & \stackrel{\text{def}}{=} & \tau^2 + \xi^2 \end{array} \right.$$

and we easily verify using computer algebra symbolic derivation[25] that the mapping is one-to-one and regular[26] from the triangle $\mathcal{T}$ to the unary positive half disk $\mathcal{D}$:

$$\mathcal{T} \stackrel{\text{def}}{=} \{(\pi, \nu), 0 \leq \nu \leq \pi \leq 1\} \leftrightarrow \mathcal{D} \stackrel{\text{def}}{=} \{(\tau, \xi), 0 \leq \xi, -1 \leq \tau \leq 1, \tau^2 + \xi^2 \leq 1\}.$$

When in a pure deterministic case, i.e., when $\xi = 0$, we obtain:

$$\tau = \nu + \pi - 1 \text{ with } \left\{ \begin{array}{rcl} \pi & = & 1 + H(-\tau)\,\tau \\ \nu & = & H(\tau)\,\tau \end{array} \right.$$

and in a pure random case, with $\nu = p = \pi$, thus $\tau^2 + \xi^2 = 1$, writing $\theta \stackrel{\text{def}}{=} \arctan(\xi, \tau)$

[25] See: `https://gitlab.inria.fr/line/aide-group/symboling/-/raw/master/src/RR-modal/possibility-polar-representation.mpl`.

[26] More precisely the function $(\pi, \nu) \to (\tau, \xi)$ is one-to-one and its Jacobian matrix is definite on all points including at $(0, 0)$ with a positive Jacobian determinant, as stated by symbolic calculus.

$$p = \frac{\cos(\theta)+1}{2} = \frac{\tau+1}{2} = \left.\frac{\nu+\pi}{2}\right|_{\pi=\nu}$$

while if $\nu < \pi$ the former formula is the closest probability estimation of p in the (ρ, ξ) space, while the latter is the closest probability estimation of p in the (π, ν) space, this latter estimation being called the *credibility* when considering the vanilla theory [Peykani et al., 2018].

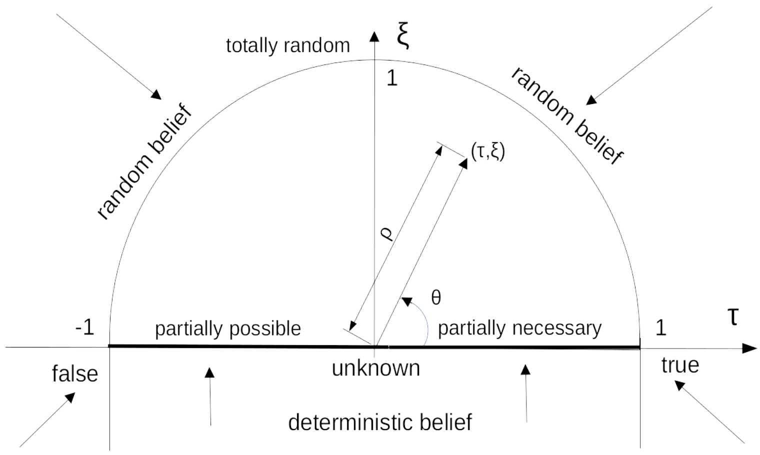

Figure 6: Representation of partial truth $(\tau, \xi) \in \mathcal{R}^2$, with $\tau \in [-1, 1]$ and $\xi \in [0, 1]$, in relation with necessity and possibility. If $\xi = 0$ (deterministic belief) the belief is entirely deterministic, which corresponds to a real grounding. If $\tau^2 + \xi^2 = 1$ (random belief) the belief is purely random and corresponds to a known probability. In between, partially known probability is considered, in the upper unitary semi-disk, with $\nu \leq p \leq \pi$.

This polar representation is interesting because it allows a homogeneous representation of the randomness as compared to (π, ν) or (τ, ρ) representation shown in Fig.7, where the surface of the admissible region is reduced in the neighborhood the false and true value, with respect to the unknown central area. This may be useful for future use of this representation.

From this definition, we obtain coherent values as explicitized in Table 4.

6.3 Projection on admissible representation

We also can define the projection of any value (π, ν), using:

$$\tau \leftarrow \max(-1, \min(1, \nu + \pi - 1)), \rho \leftarrow \max(|\tau|, \min(1, \nu - \pi + 1))$$

which partition the (π, ν) space in four oblique regions of false, partially possible, partially necessary, and true regions, as illustrated in Fig. 7. Here the design choice on the τ axis is to re-project any value beyond the bounds on the corresponding false or true value, and to re-project any ρ value on the closest admissible value.

This projection implementation is available[2] as open-source code.

This allows finding for any approximate value (π, τ) the closest admissible value, yielding an interpretation of each portion of the space outside the semi-disk of admissible values. This may for instance be used when estimating a value through a neural network, as experimented in [Vallaeys, 2021] for an alternative representation. The design choice here is not to consider a sigmoid-like profile, but more something related to a "bounded ReLu function" (i.e., $u \to \max(0, u)$). This is to be used when implementing the non-linearity of the neural network

Event	(π, ν)	(p, ρ)	(τ, ξ)	θ	
$\emptyset$	$(0,0)$	$(0,0)$	$(1,0)$	0	Corresponds to a certainly *false* event, that will necessarily not occur.
Ω	$(1,1)$	$(1,0)$	$(1,0)$	Π	Corresponds to a certainly *true* event, that will necessarily occur.
	$(1,0)$	$(?,1)$	$(0,0)$	$?$	Corresponds to a totally *unknown* event, i.e., fully possible (i.e., not surprising to occur), but totally not necessary (i.e., not surprising not to occur), which means that we have no belief about it, and its probability is entirely unknown.
	$(\pi,0)$	$p \leq \pi$ $\rho = 1 - \pi$	$(\tau,0), \tau < 0$	0	Corresponds to partially known *deterministic* event, which is totally not necessary ($\nu = 0$).
	$(1,\nu)$	$\nu \leq p$ $\rho = \nu$	$(\tau,0), \tau > 0$	Π	Corresponds to partially known *deterministic* event, which is fully possible ($\pi = 1$).
	(p,p)	$(p,0)$	$(\tau,\xi), \tau^2 + \xi^2 = 1$	θ	Corresponds to an event known as *random* for which we precisely know its probability p, otherwise the probability is assumed to be only approximately known.
	$(1/2,1/2)$	$(\frac{1}{2},0)$	$(0,1)$	$\Pi/2$	Corresponds to a totally *random* binary event, with a totally known probability.

Table 4: Noticeable values of possibility, necessity, and probability for this representation.

that will not only evaluate the stochastic precision of the measure but also the presence or lack of information (i.e. evaluate the presence or absence of data, as illustrated in Fig. 8 using a different non-linearity, from [Vallaeys, 2021], who also successfully experiment also on the MINST data set. Other experimental results on neural networks estimating necessity and possibility are available, e.g., pattern classification using an interval arithmetic perceptron [Drago and Ridella, 1999] where $\nu \leq \pi$ is fulfill considering min and max operators on the network output, which is qualitatively the same idea as proposed in [Vallaeys, 2021], and even earlier, e.g., [Ishibuchi et al., 1992].

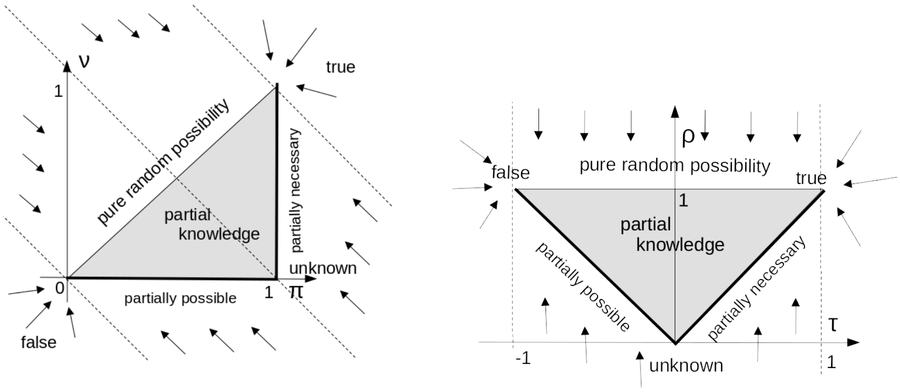

Figure 7: Projection of any value in the (π, ν) space on the left and in the (τ, ρ) on the right. Interestingly enough the two representations are related by a rotation and some re-scaling.

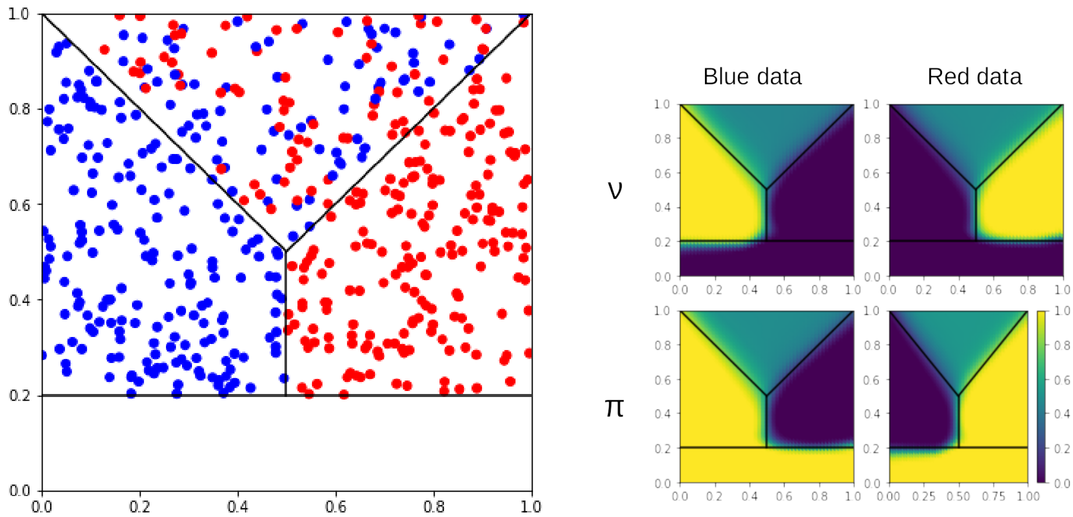

Figure 8: A binary synthetic data experiment from [Vallaeys, 2021]: The input data on the left draws four zones, one with blue points, one with red points, one with mixed points, and one without data. The network result is on the right where in this simple case, we obtain true, false, unknown, or random values as expected. A three-layer small-size standard perception-like network with soft-max output has been used, adjusted using a cross-entropy criterion with constraint quadratic on the output to guarantee a correct representation with $\nu \leq \pi$. This numerical experiment should be considered a simple illustration.

References

[Bochnak et al., 2013] Bochnak, J., Coste, M., and Roy, M.-F. (2013). *Real Algebraic Geometry*. Springer Science & Business Media. Google-Books-ID: GJv6CAAAQBAJ.

[Ciucci et al., 2014] Ciucci, D., Dubois, D., and Lawry, J. (2014). Borderline vs unknown: comparing three-valued representations of imperfect information. *International Journal of Approximate Reasoning*, 55(9):1866–1889.

[Denœux et al., 2020a] Denœux, T., Dubois, D., and Prade, H. (2020a). Representations of Uncertainty in AI: Beyond Probability and Possibility. In Marquis, P., Papini, O., and Prade, H., editors, *A Guided Tour of Artificial Intelligence Research: Volume I: Knowledge Representation, Reasoning and Learning*, pages 119–150. Springer International Publishing, Cham.

[Denœux et al., 2020b] Denœux, T., Dubois, D., and Prade, H. (2020b). Representations of Uncertainty in AI: Probability and Possibility. In Marquis, P., Papini, O., and Prade, H., editors, *A Guided Tour of Artificial Intelligence Research: Volume I: Knowledge Representation, Reasoning and Learning*, pages 69–117. Springer International Publishing, Cham.

[Drago and Ridella, 1999] Drago, G. and Ridella, S. (1999). Possibility and Necessity Pattern Classification using an Interval Arithmetic Perceptron. *Neural Computing & Applications*, 8(1):40–52.

[Dubois and Prade, 2015] Dubois, D. and Prade, H. (2015). Practical Methods for Constructing Possibility Distributions. *International Journal of Intelligent Systems*, vol. 31(n° 3):pp. 215–239. Number: n° 3 Publisher: Wiley.

[Fikes and Kehler, 1985] Fikes, R. and Kehler, T. (1985). The role of frame-based representation in reasoning. *Communications of the ACM*, 28(9):904–920.

[Fischer, 2018] Fischer, B. (2018). Modal Epistemology: Knowledge of Possibility & Necessity.

[Guberman, 2016] Guberman, N. (2016). On Complex Valued Convolutional Neural Networks. *arXiv:1602.09046 [cs]*. arXiv: 1602.09046.

[Ishibuchi et al., 1992] Ishibuchi, H., Fujioka, R., and Tanaka, H. (1992). Possibility and necessity pattern classification using neural networks. *Fuzzy Sets and Systems*, 48(3):331–340.

[Johnson-Laird and Ragni, 2019] Johnson-Laird, P. and Ragni, M. (2019). Possibilities as the foundation of reasoning. *Cognition*, 193:103950.

[Mercier et al., 2021] Mercier, C., Chateau-Laurent, H., Alexandre, F., and Viéville, T. (2021). Ontology as neuronal-space manifold: towards symbolic and numerical artificial embedding. In *KRHCAI-21@KR2021*.

[Negri, 2013] Negri, M. (2013). Partial Probability and Kleene Logic. *arXiv:1310.6172 [math]*. arXiv: 1310.6172.

[Peykani et al., 2018] Peykani, P., Mohammadi, E., Pishvaee, M., Rostamy-Malkhalifeh, M., and Jabbarzadeh, A. (2018). A Novel Fuzzy Data Envelopment Analysis Based on Robust Possibilistic Programming: Possibility, Necessity and Credibility-Based Approaches. *RAIRO - Operations Research*, 52:1445–1463.

[Raufaste et al., 2003] Raufaste, E., da Silva Neves, R., and Mariné, C. (2003). Testing the descriptive validity of possibility theory in human judgments of uncertainty. *Artificial Intelligence*, 148(1):197–218.

[Rusawuk, 2018] Rusawuk, A. L. (2018). Possibility and Necessity: An Introduction to Modality.

[Smith, 1994] Smith, L. (1994). The development of modal understanding: Piaget's possibility and necessity. *New Ideas in Psychology*, 12(1):73–87.

[Smithson, 1970] Smithson, M. (1970). Human Judgment And Imprecise Probabilities.

[Tettamanzi et al., 2017] Tettamanzi, A., Zucker, C. F., and Gandon, F. (2017). Possibilistic testing of OWL axioms against RDF data. *International Journal of Approximate Reasoning*.

[Troffaes et al., 2011] Troffaes, M. C. M., Miranda, E., and Destercke, S. (2011). On the connection between probability boxes and possibility measures. In *Proceedings of the 7th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-2011)*, France. Atlantis Press.

[Vallaeys, 2021] Vallaeys, T. (2021). *Généraliser les possibilités-nécessités pour l'apprentissage profond*. report, Inria. Pages: 1.

RESEARCH CENTRE
BORDEAUX – SUD-OUEST

200 avenue de la Vieille Tour
33405 Talence Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399