Paraconsistent probabilistic reasoning: applied to scenario recognition and voting theory

Lionel Daniel

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

MINES ParisTech

École doctorale n°84 :
Sciences et technologies de l'information et de la communication

Doctorat européen ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

l'École nationale supérieure des mines de Paris

Spécialité « Informatique temps-réel, robotique et automatique »

présentée et soutenue publiquement par

Lionel DANIEL

le 05 février 2010

Définition d'une logique probabiliste tolérante à l'inconsistance

– appliquée à la reconnaissance de scénarios & à la théorie du vote –

~ ~ ~

Paraconsistent probabilistic reasoning

– applied to scenario recognition & voting theory –

Directrice de thèse : **Valérie ROY**
Co-encadrement de la thèse : **Annie RESSOUCHE**

Jury

Jonathan LAWRY, Professeur, Dept. of Engineering Mathematics, Univ. de Bristol		Président & Rapporteur
Gabriele KERN-ISBERNER, Professeur, Dept. of Computer Science, Univ. de Dortmund		Rapporteur
Pierre MARQUIS, Professeur, Centre de Recherche en Informatique de Lens, Univ. d'Artois		Examinateur
Valérie ROY, Maitre de recherche, Centre de Mathématiques Appliquées, MINES ParisTech		Examinateur
Annie RESSOUCHE, Directeur de recherche, PULSAR, INRIA Sophia-Antipolis		Examinateur

T H È S E

MINES ParisTech
Centre de Mathématiques Appliquées
Rue Claude Daunesse B.P. 207, 06904 Sophia Antipolis Cedex, France

Contents

Preface

Remerciements

Je destine mes premiers remerciements à mes directrices de thèse, Valérie Roy et Annie Ressouche, pour m'avoir initié au monde de la recherche dès mon stage de Maîtrise en 2005 et m'avoir encadré ces cinq dernières années passées au Centre de Mathématiques Appliquées (CMA) sur le site MINES-ParisTech de Sophia-Antipolis. Valérie sut tisser un cocon, propice à la maturation de ma thèse, me protégeant des pressions industrielles, des charges administratives et d'enseignement, tout en me supportant financièrement et scientifiquement. Bien sûr, le tissage n'eût été réussi sans l'approbation de notre directrice de centre Nadia Maïzi, l'assistance bienveillante de notre secrétaire Dominique Micollier, et l'aide logistique de Gilles Guerassimoff et Jean-Charles Bonin.

Les dix premiers mois passés dans ce cocon me permirent d'abstraire mon sujet de thèse initial sur la "reconnaissance de comportements menaçant le port de Marseille", puis d'atteindre les travaux de Jeff Paris sur le raisonnement incertain. Je tiens à lui exprimer ma gratitude pour m'avoir accueilli et supervisé cinq mois à l'université de Manchester dans le cadre du programme Mathlogaps ; je remercie aussi Yves Bertot et Michel Auguin pour avoir appuyé ma candidature à ce programme, et Dugald Macpherson pour l'avoir acceptée. Ainsi, Jeff me révéla la voie féconde du *raisonnement probabiliste paraconsistant*.

De retour dans mon cocon sophipolitain, j'ai continué le développement de cette théorie grâce au soutien essentiel de Jean-Paul Marmorat et Yves Rouchaleau ; leurs compétences scientifiques et pédagogiques m'inspirent une infinie estime. Aussi, les dernières corrections de ce manuscrit de thèse résultent des critiques instructives de Gabriele Kern-Isberner et Jonathan Lawry, mes rapporteurs, mais aussi des questions et remarques de Pierre Marquis, membre du jury ; je leur exprime donc une grande reconnaissance.

Enfin, je remercie très fort tous mes collègues du CMA pour tous ces bons moments passés ensemble, et j'adresse ma plus profonde gratitude à ma famille pour leur irremplaçable et permanent soutien.

Résumé étendu

Cette thèse introduit un cadre de travail théorique pour raisonner à partir de bases de connaissances propositionnelles probabilistes éventuellement inconsistantes. Par raisonner, il faut comprendre non seulement inférer (c'est-à-dire déduire) des informations à partir d'une base de connaissances donnée, mais aussi mesurer la qualité de cette base. Inférer et mesurer sont deux activités s'inscrivant dans un processus de décision global, comme présenté dans l'exemple suivant.

Supposons que Tom arrête sa voiture à une intersection pour décider s'il va tourner à Droite ou à Gauche. Ignorant la bonne direction à suivre, il demande conseil à ses amies: Sophia lui réponds qu'elle penche pour aller à Droite, tandis que Mia lui assure que la bonne direction est Gauche. Tom aurait peut-être ainsi tourné à Gauche s'il ne savait pas que Mia est une farceuse, contrairement à Sophia qui fait souvent preuve de sagesse; mais Tom est bien conscient de la fiabilité de ses amies. La connaissance de Tom est alors la suivante:

Source	*Fiabilité*	*Proposition*	*Probabilité*
Sophia	90% (sage)	Droite	[60%:100%]
Mia	20% (farceuse)	Gauche	100%

Formellement, soit v la variable propositionnelle signifiant que "Droite est la bonne direction"; la valeur de cette variable est soit **v**rai, soit **f**aux. Je note $\omega(v)$ la probabilité que v vaille **v**rai, et $\omega(\neg v)$ la probabilité que v vaille **f**aux. Le mot probabilité fait ici référence à une fonction ω satisfaisant les deux axiomes de Kolmogorov:
(P1) si $\models \theta$ alors $\omega(\theta) = 1$;
(P2) si $\models \neg(\theta \wedge \phi)$ alors $\omega(\theta \vee \phi) = \omega(\theta) + \omega(\phi)$,
où θ et ϕ sont des phrases d'un langage propositionnel, comme $\Theta ::= (\Theta \wedge \Theta) \mid (\Theta \vee \Theta) \mid (\neg\Theta) \mid v$ par exemple, et où $\models \theta$ représente une tautologie. Tom reçoit ainsi de Sophia l'information $K_{\text{Sophia}} \stackrel{\text{def}}{=} \{\text{“}60\% \leq \omega(v) \leq 100\%\text{”}\}$, et de Mia l'information $K_{\text{Mia}} \stackrel{\text{def}}{=} \{\text{“}\omega(\neg v) = 100\%\text{”}\}$. La connaissance de Tom est donc la fusion de ces deux informations pondérées par leur niveau confiance respectif: $K_{\text{Tom}} \stackrel{\text{def}}{=} K^{90\%}_{\text{Sophia}} \Cup K^{20\%}_{\text{Mia}}$. Supposons[1] que

[1] Tom adhère ici à la seconde maxime de R. Descartes [8]: "lorsqu'il n'est pas en notre pouvoir de discerner les plus vraies opinions, nous devons suivre les plus probables, et en-

Tom tourne à Droite ssi $\omega(v) \geq 50\%$ et tourne à Gauche ssi $\omega(v) < 50\%$. Je présente maintenant les principales questions abordées dans ma thèse:

1. Quelle direction Tom devrait-il prendre, c'est-à-dire, quelle unique valeur de $\omega(v)$ peut-il *inférer* à partir de sa connaissance K_{Tom}?

   Réponse: Tom discrédite Mia car il pense qu'elle est bien moins fiable que Sophia. Tom devrait donc tourner à Droite car la probabilité que Droite soit la bonne direction est supérieure à 50%, comme indiquée par le processus d'inférence $\mathcal{I}^E_{\text{ME}}$ défini au chapitre 4: $\omega(v) = (\mathcal{I}^E_{\text{ME}}(K_{\text{Tom}}))(v) \approx 60\%$. Ce processus d'inférence élit la distribution de probabilité qui représente au mieux le monde réel d'après une connaissance donnée (ex: K_{Tom}) et un ensemble de principes. Ces principes sont une tentative de définition du sens commun sous-jacent au raisonnement paraconsistant; un exemple de principe est que "la valeur de $\omega(v)$ ne devrait pas dépendre de la musique que Tom écoute", car écouter de la musique ne fournit aucune information concernant la bonne direction à prendre.

2. À quel point Tom est-il *confiant* que tourner à Droite soit la bonne décision?

   Réponse: La confiance qu'a Tom en la décision de tourner à Droite est $\mu^{\text{conf}}(K_{\text{Tom}}, \text{Droite}) \approx 16\%$. La raison est que chaque distribution de probabilité ω est vue comme un argument soutenant soit la décision de prendre à Droite (si $\omega(v) \geq 50\%$) soit la décision de prendre à Gauche (si $\omega(v) < 50\%$) avec une force dépendante de K_{Tom}; je calcule que la distribution de probabilité soutenant au plus la décision de prendre à Droite (Gauche) a une force d'environ 87% (71%), d'où je conclue que la confiance de Tom en la décision de tourner à Droite est d'environ $16\% = 87\% - 71\%$, et de tourner à Gauche est d'environ $-16\% = 71\% - 87\%$. Ainsi, Tom devrait plutôt tourner à Droite qu'à Gauche, mais comme sa confiance n'est que de 16%, il devrait retarder sa prise de décision de tourner afin d'obtenir d'avantage d'informations sur la direction à prendre.

3. À quel point ce que dit Sophia est *incohérent* par rapport à ce que dit Mia, c'est-à-dire à quelle distance sont-elles d'atteindre un consensus sur la valeur de $\omega(v)$?

   Réponse: L'incohérence (en tant que gap) entre les dires de Sophia "$60\% \leq \omega(v)$" et ceux de Mia "$\omega(v) = 0\%$" est $\mu^{\text{icoh}}_{\mathcal{G}}(K_{\text{Sophia}}, K_{\text{Mia}}) = 60\% - 0\% = 60\%$. Si Mia était moins certaine qu'il faille tourner à Gauche, "$\omega(v) \in$ [0%:40%]" par exemple, alors l'incohérence aurait été moindre, $60\% - 40\% = 20\%$ par exemple, et si Sophia avait été encore moins certaine qu'il faille tourner à Droite, "$\omega(v) \in$ [30%:100%]" par exemple, alors l'incohérence aurait été minimale, c'est-à-dire égale à 0%, car [0%:40%] intersecte [30%:100%] ce qui signifie qu'un consensus est atteint.

4. À quel point la connaissance de Tom est-elle *inconsistante*, et qui de Mia ou Sophia en est le plus *coupable*?

   Réponse: Puisque Tom fait bien plus confiance à Sophia qu'à Mia, sa connaissance est faiblement inconsistante, c'est-à-dire $\mu^{\text{icst}}(K_{\text{Tom}}) \approx 13\%$, et cette inconsistance est principalement imputable à Mia: sa part de responsabilité est de $\mu^{\text{culp}}_{K_{\text{Tom}}}(K_{\text{Mia}}) \approx 13\%$ alors que celle de Sophia est de $\mu^{\text{culp}}_{K_{\text{Tom}}}(K_{\text{Sophia}}) \approx 0\%$.

5. Qui de Mia ou Sophia procure à Tom l'information la plus *précise*?

   Réponse: Mia fournit à Tom une valeur précise, "$\omega(v) = 0\%$", alors que Sophia lui fournit seulement un intervalle de valeurs, "$\omega(v) \in$ [60%:100%]". La mesure de précision de l'information procurée par Mia est alors plus élevée que celle de l'information procurée par Sophia: $\mu^{\text{pre}}(K_{\text{Mia}}) > \mu^{\text{pre}}(K_{\text{Sophia}})$.

Le chapitre 2 décrit une nouvelle représentation de la connaissance appelée *candidacy fonction* permettant de relâcher les contraintes exprimées sur les distributions de probabilité; une notion de *fiabilité* y est introduite. Le chapitre 3 approfondit les précédentes questions 2 à 5, puis le chapitre 4 traite la question 1. Le chapitre 5 propose une potentielle application à la théorie du vote.

General notations

Let $\mathbb{R}$ be the set of real numbers. Let $\mathbb{N}$ be the set of natural numbers, 0 included. Let $\{0, 1, 2\}$ be a set (or a multiset if explicitly stated). Let $[0, 1, 2]$ be a list, or a horizontal vector. Let $[0; 1; 2]$ be a vertical vector. Let $[A, B]$ and $[A; B]$ respectively be the horizontal and vertical concatenation of A and B, where A and B are matrices, vectors, or scalars with coherent dimensions. Let [0:1] and]0:1[respectively be a closed and open interval of real numbers. Let $\prod v$ and $\sum v$ respectively be $\prod_{i=1}^{|v|} v_i$ and $\sum_{i=1}^{|v|} v_i$, where v_i is the i^{th} element of a vector v having $|v|$ elements. Let $\mathcal{L}_2(x, y) \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{|x|}(x_i - y_j)^2}$ be the Euclidean distance between two points x and y. Let ln and exp be respectively the logarithm and the exponential to the natural base.

tre plusieurs opinions également reçues, [nous devons suivre] les plus modérées."

Chapter 1

Introduction

> Reasoning is nothing but reckoning
>
> Thomas Hobbes, in [15, chapter V]

Humans commonly reason from uncertain even inconsistent knowledge. *Common sense* underlying paraconsistent reasoning should thus exist; its axiomatisation is then prospected.

This thesis introduces a theoretical framework to *reason* from inconsistent probabilistic propositional knowledge bases; by reasoning, we[1] mean inferring (ie deducing) information from a given (possibly inconsistent) uncertain knowledge base, where uncertainty is identified[2] with imprecise probability over propositional sentences. By reasoning, we also mean appraising knowledge, like quantifying the inconsistency of a knowledge base, or measuring the dissimilarity (ie the distance) between two knowledge bases, even when such bases are inconsistent.

Introductory example

Suppose Tom stops his car at an intersection to decide if he will turn Left or Right. Unaware of the correct direction, Tom asks his friends: Sophia leans towards turning Right, while Mia claims certainty in Left. Tom would thus turn Left if he ignored Sophia's wisdom and Mia's inveterate jokiness, but he is conscious of his friends' reliability. Tom's knowledge is then depicted as follows:

Source	*Reliability*	*Proposition*	*Probability*
Sophia	90% (sage)	turn Right	[60%:100%]
Mia	20% (joker)	turn Left	100%

More formally, let v be the propositional variable that means "Right is the correct direction", of which the value is either **t**rue or **f**alse. We denote by $\omega(v)$ the probability that v is **t**rue, and by $\omega(\neg v)$ the probability that v is **f**alse. Hence, $0\% \leq \omega(v) = 1 - \omega(\neg v) \leq 100\%$. Sophia provides Tom with the knowledge base $K_{\text{Sophia}} \stackrel{\text{def}}{=} \{\text{“}60\% \leq \omega(v) \leq 100\%\text{”}\}$. Mia provides Tom with the knowledge base $K_{\text{Mia}} \stackrel{\text{def}}{=} \{\text{“}\omega(\neg v) = 100\%\text{”}\}$, which means that v is **f**alse in the real world. Tom's knowledge is thus represented as the mergence of Sophia's and Mia's knowledge bases, weighted by reliabilities: $K_{\text{Tom}} \stackrel{\text{def}}{=} K^{90\%}_{\text{Sophia}} \Cup K^{20\%}_{\text{Mia}}$. Suppose[3] Tom turns Right iff $\omega(v) \geq 50\%$, and turns Left iff $\omega(v) < 50\%$. The main questions addressed in this thesis are:

1. Which direction Tom should take, ie which unique value of $\omega(v)$ can be *inferred* from knowledge base K_{Tom}?

   Answer: Tom almost disbelieves Mia because she is much less reliable than Sophia. Tom should thus turn Right; the probability that Right is the correct direction is greater than 50%, as indicated by our inference process $\mathcal{I}^E_{\text{ME}}$ defined in chapter 4: $\omega(v) = (\mathcal{I}^E_{\text{ME}}(K_{\text{Tom}}))(v) \approx 60\%$.

2. How *confident* is Tom in turning Right?

   Answer: Tom's confidence in turning Right is $\mu^{\text{conf}}(K_{\text{Tom}}, \text{Right}) \approx 16\%$. The reason is that each probability distribution ω is seen as an argument for supporting either Right (if $\omega(v) \geq 50\%$) or Left (if $\omega(v) < 50\%$) with a strength depending on K_{Tom}; we compute that the probability distribution most supporting Right (Left) has strength about 87% (71%), whence Tom's confidence in turning Right is about $16\% = 87\% - 71\%$, and in turning Left is about $-16\% = 71\% - 87\%$. Thus, Tom should rather turn Right than Left, but since his confidence is only 16%, he should also consider postponing his decision in turning Right to acquire more knowledge about the correct direction.

[1] In this thesis, pronoun *we* indifferently refers to, sometimes the author with a thought to his surrounding helpful people, sometimes to both the reader and the author.

[2] The reader interested in the justification for identifying uncertainty with probability is invited to read [33, chapter 2]

[3] Tom adheres to R. Descartes's second maxim: "when it is not in our power to determine what is true, we ought to act according to what is most probable; and even though we do not remark a greater probability in one opinion than in another, we ought notwithstanding to choose one or the other" [8].

3. How *incoherent* is Sophia's statement with respect to Mia's statement, ie how far are they from reaching a consensus about the value of $\omega(v)$?

   Answer: The incoherence (as the gap) between Sophia's statement "$60\% \leq \omega(v)$" and Mia's statement "$\omega(v) = 0\%$" is $\mu^{\text{icoh}}_{\mathcal{G}}(K_{\text{Sophia}}, K_{\text{Mia}}) = 60\% - 0\% = 60\%$. If Mia was less certain about turning Left, eg "$\omega(v) \in [0\%{:}40\%]$", then the incoherence would have been lower, eg $60\% - 40\% = 20\%$, and if Sophia was even less certain about turning Right, eg "$\omega(v) \in [30\%{:}100\%]$", then the incoherence would have been minimal, ie 0%, since [0%:40%] overlaps [30%:100%].

4. How *inconsistent* is Tom's knowledge, and who of his friends is the most *culpable* for making his knowledge inconsistent?

   Answer: Since Tom almost disbelieves Mia but strongly trusts Sophia, his knowledge is slightly inconsistent, ie $\mu^{\text{icst}}(K_{\text{Tom}}) \approx 13\%$, and he almost entirely imputes the inconsistency of his knowledge to Mia: $\mu^{\text{culp}}_{K_{\text{Tom}}}(K_{\text{Mia}}) \approx 13\%$ whereas $\mu^{\text{culp}}_{K_{\text{Tom}}}(K_{\text{Sophia}}) \approx 0\%$.

5. Is Mia's statement more *precise* than Sophia's one?

   Answer: Yes, because Mia provides Tom with a precise probability measure of v, ie "$\omega(v) = 0\%$", whereas Sophia only provides an interval, ie "$\omega(v) \in [60\%{:}100\%]$". The precision measure of the knowledge provided by Mia is thus higher than that of Sophia: $\mu^{\text{pre}}(K_{\text{Mia}}) > \mu^{\text{pre}}(K_{\text{Sophia}})$.

Question 1 is answered in chapter 4 by applying to K_{Tom} a commonsensical inference process, ie by deducing from K_{Tom} the value of $\omega(v)$ while adhering to some *principles* intended to define common sense. Such a principle could be "the value of $\omega(v)$ should not depend on whether Tom is listening music", because listening music is *irrelevant* to determine the correct direction.

Related work

Paraconsistent reasoning not only tolerates inconsistency, but also considers it as informative. Therefore, when reasoning from a possibly inconsistent knowledge base, every knowledge item deserves consideration. In this thesis, we do not restore[4] consistency: we rather *live* with inconsistency.

For example, if we follow a paraconsistent logic, we should infer from the propositional knowledge base $\text{kb} \stackrel{\text{def}}{=} \{\neg v_1, v_1, v_1, v_2\}$ that the value of the propositional variable v_2 is **t**rue (and we should not infer $\neg v_2$); we may furthermore infer $\neg v_1$ since $\neg v_1 \in \text{kb}$, but in which case, we should also infer v_1 since $v_1 \in \text{kb}$: this small *explosion*, ie inferring a fact and its contrary, is not desirable. In this thesis, we consider a knowledge base as a multiset. We thus infer v_1 rather than $\neg v_1$ because v_1 appears twice in kb, where $\neg v_1$ appears only once. Each knowledge item is considered as a vote (here, there are two votes for v_1 and one vote for $\neg v_1$). We may also infer from kb a third value for v_1, which could mean both **t**rue and **f**alse, or could mean unknown value; such a logic with several values is called a many-valued logic.

In this thesis, the many-valued logic we employ is the probabilistic logic. This logic gives to each propositional sentence, eg $\neg v_1$, a probability value in [0:1], where 0 means **f**alse and 1 means **t**rue. There exist several probabilistic techniques to reason from kb. For example, [39] provides an entailment relation ${}^{\eta}\!\rhd_{\varsigma}$ that considers as consequence of a propositional knowledge base (where the subjective probability of each item is known to be greater than η) any propositional sentence having, by applying the probabilistic logic, a probability greater than ς. In this thesis, we instead focus on inferring one probability distribution *best* satisfying a (possibly inconsistent) probabilistic propositional knowledge base, like $K \stackrel{\text{def}}{=} \{\omega(\neg v_1) \geq \eta, \omega(v_1) \geq \eta, \omega(v_1) \geq \eta, \omega(v_2) \geq \eta\}$ with $\eta > \frac{1}{2}$; in chapter 4, we extend to inconsistent knowledge bases the inference process axiomatised by J.B. Paris and A. Vencovská in [38]. Our *knowledge bases*, defined in chapter 2, are multisets of algebraic constraints over a probability distribution. Such a knowledge base is inconsistent iff no probability distribution satisfies the multiset of constraints. Inconsistency might thus be conceived as another notion of uncertainty beyonds the uncertainty represented by probability (see [4, chapter 3] for a rich survey about uncertainty representations, and [31, 43] for an introduction then a survey about upper and lower previsions). In order to elect the probability distributions that *best* satisfy an inconsistent knowledge base, we blur its constraints. For example, the blur version of the constraint "$\omega(v_1) \geq 1$" means that $\omega(v_1)$ should be near 1, but may equal 0.92, or might equal 0.74. We propose in §2.3.5 a principled blur function that blurs every constraints in a knowledge base; such a blur function resembles a membership function from the standpoint of fuzzy set theory.

This thesis also addresses a problem originating from voting theory, which belongs to the field of social choice theory. We show that paraconsistent probabilistic reasoning is the solution for reaching a consensus among conflicting agents' opinions about a distribution (of a financial investment, see section 2.1, or of a limited resource, see §5.1.2).

[4] In §2.3.5 page 11, we define the consistent version of an inconsistent knowledge base

Chapter 2

A new knowledge representation to resolve inconsistency

> Essentially, all models are wrong,
> but some are useful.
>
> George E. P. Box, 1987

According to George E. P. Box, the models of knowledge discussed in this chapter are all *wrong*. The first section will hopefully convince the reader that probabilistic models are *useful*, as they capture problems of consensus decision making, which originate from voting theory. A convinced reader may then appreciate, in the second section, a general probabilistic knowledge representation: candidacy functions. As complex numbers were a new representation of numbers to deal with square roots of negative numbers, candidacy functions are a new representation of knowledge to deal with *paraconsistent* reasoning. Finally, the last section suggests a stream of assumptions and principles to construct the candidacy function corresponding to a given knowledge base; analogously, we construct the complex number corresponding to a given real number.

2.1 Motivation: how to reach a rational consensus in financial investments?

Voting theory is a theory of electing a societal preference from individual preferences. In the following motivating example, which is excerpted from § 5.1.2 on page 42, we aim to reduce voting theory to paraconsistent probabilistic reasoning.

A society composed of $I \in \mathbb{N}$ individuals has to invest 1 dollar in $J \in \mathbb{N}$ companies $\{ \alpha_1, \alpha_2, \ldots, \alpha_J \}$. An investment distribution ω is a function that maps each company to the amount of money invested in this company by the society; furthermore, an investment distribution satisfies these two assumptions: (A1) the society invests 1 dollar in J companies, ie $\$1 = \sum_{j=1}^{J} \omega(\alpha_j)$, and (A2) the society does not borrow money from these companies, ie $\forall j \in \{ 1, 2, \ldots, J \}, \omega(\alpha_j) \geq \$0$.

Each individual i independently expresses a multiset K_i of wishes for the investment distribution ω. For example, i may wish to invest twice more in company α_1 than in company α_2, ie $\omega(\alpha_1) = 2 * \omega(\alpha_2)$, and may wish that the total amount invested in α_1 and α_2 be within 0.2 and 0.3 dollar, ie $\$0.2 \leq \omega(\alpha_1) + \omega(\alpha_2) \leq \$0.3$. Besides, each individual i is given a reliability level $\sigma_i \in]0{:}1[$, which tends towards 1 as the society deems i more reliable.

Thus, the society seeks a voting system $\mathcal{I}$ yielding the investment distribution $\hat{\omega}$ that *best* conciliates the wishes K_i of each individual i, according to their reliability level σ_i and some *common sense*; formally, $\hat{\omega} \stackrel{\text{def}}{=} \mathcal{I}(\cup_{i=1}^{I} K_i^{\sigma_i})$, where $\mathcal{I}$ must satisfy several principles intended to define common sense. By interpreting assumptions (A1) and (A2) as Kolmogorov's axioms for probability, investment distributions can be identified with *probability* distributions. We therefore take the *probabilistic* standpoint to define $\mathcal{I}$ as an inference process, of which a definition will be given in chapter 4; this motivates us to study paraconsistent probabilistic reasoning.

2.2 Two knowledge representations to juggle inconsistent and imprecise probabilities

In this section, we define several probabilistic languages allowing us to express constraints on a probability distribution on sentences of a propositional language. We then define a knowledge base as a multiset of such constraints; a knowledge base is consistent iff the multiset is satisfiable. After, we show that conditional probabilities are expressible in $\mathbb{K}^L$, which is the set of knowledge bases containing only linear constraints. Finally, we introduce candidacy functions: the probability distributions maximising a candidacy function are nominated to be the best candidates for representing the real world. Candidacy functions are a general probabilistic knowledge representation; we therefore explain

in section 2.3 the construction of the candidacy function corresponding to a given knowledge base.

2.2.1 Probability distribution $\omega \in \Omega$ underlain by a propositional language Θ

Let Θ be a propositional language generated by $\Theta ::= (\Theta \wedge \Theta) \mid (\Theta \vee \Theta) \mid (\neg\Theta) \mid \mathit{vars}$, where *vars* is a finite set of propositional variables being either **t**rue or **f**alse, and where logical connectives $\wedge$, $\vee$, and $\neg$ have their respective classical semantics[1] *and*, *or*, and *not*. In the sequel, we suppose Θ fixed and propositions in Θ, unless otherwise stated. Propositions are usually denoted by θ, ϕ, or ψ. We denote a tautology θ by $\models \theta$. Furthermore, let $\alpha_\Theta \stackrel{\text{def}}{=} \{ \alpha_j \mid j = 1, 2, \ldots, J \}$ be the set of minterms[2] of Θ, where $J \stackrel{\text{def}}{=} 2^{|vars|}$ with $|vars|$ being the number of propositional variables. Also, let $\alpha_\theta \stackrel{\text{def}}{=} \{ \alpha_j \mid \models (\neg\alpha_j \vee \theta) \}$ be the minterms of a proposition θ. Finally, each proposition θ is supposed to be in the canonical disjunctive normal form, ie $\theta = \bigvee_{\alpha\in\alpha_\theta} \alpha$.

Definition 1. *Kolmogorov's axioms for probability are:*
(P1) if $\models \theta$ then $\omega(\theta) = 1$;
(P2) if $\models \neg(\theta \wedge \phi)$ then $\omega(\theta \vee \phi) = \omega(\theta) + \omega(\phi)$,
where ω is a function from Θ to [0:1]*, $\theta, \phi \in \Theta$.*

Definition 2. *A probability distribution ω is a function that satisfies Kolmogorov's axioms for probability. We denote by Ω the set of probability distributions.*

Notice that the minterms of Θ are mutually exclusive, ie $\models \neg(\alpha_i \wedge \alpha_j)$ for any two distinct minterms α_i and α_j. Since θ is a disjunction of minterms, $\omega(\theta)$ equals $\omega(\bigvee_{\alpha\in\alpha_\theta} \alpha)$ by definition, and equals $\sum_{\alpha\in\alpha_\theta} \omega(\alpha)$ by axiom (P2). Thus, a probability distribution ω can be seen as a function from α_Θ to [0:1], hence as a point $[\omega(\alpha_1); \omega(\alpha_2); \ldots; \omega(\alpha_J)]$ in a Euclidean space of dimension J such that its j^{th} coordinate $\omega_j \in$ [0:1] equals $\omega(\alpha_j)$. Furthermore, [33, pages 13–15] shows that a point $\omega \in \mathbb{R}^J$ in a Euclidean space of dimension J denotes a probability distribution iff $\omega \geq \vec{0}$ and[3] $1 = \sum_{j=1}^{J} \omega_j$. Thus, writing $\Omega \subset \mathbb{R}^J$ makes sense.

θ	ϕ	$\theta \wedge \phi$	$\theta \vee \phi$	$\neg\theta$
false	**f**alse	**f**alse	**f**alse	**t**rue
false	**t**rue	**f**alse	**t**rue	**t**rue
true	**f**alse	**f**alse	**t**rue	**f**alse
true	**t**rue	**t**rue	**t**rue	**f**alse

[1] Classical semantics

[2] A *minterm* is a sentence of a propositional language. A minterm has the form $\bigwedge_{v\in vars} \pm v$, where *vars* is the set of propositional variables and where $\pm v$ means either $\neg v$ or v. A minterm is an *atom* in J.B. Paris's terminology (see [33]).

[3] Remember the assumption (A1) $1 = \sum_{j=1}^{J} \omega(\alpha_j)$, and (A2) $\forall j \in \{ 1, 2, \ldots, J \}, \omega(\alpha_j) \geq 0$ in the motivating example about voting theory at section 2.1 on the previous page.

2.2.2 The probable worlds: the models Ω_K of a probabilistic knowledge base $K \in \mathbb{K}$

In this thesis, we identify knowledge with a possibly unsatisfiable multiset of constraints on a probability distribution ω.

Definition 3 (Constraint)**.** *A constraint c is an inequality of the following general form: "$b \geq f(\omega)$", where $b \in \mathbb{R}$, $f : \mathbb{D} \mapsto \mathbb{R}$ is a function such that $\Omega \subseteq \mathbb{D} \subseteq \mathbb{R}^J$, and $\exists x, x' \in \mathbb{D}, f(x) > b \geq f(x')$.*

Definition 4 (Knowledge base)**.** *A knowledge base K is a finite multiset of constraints. We denote by $\mathbb{K}$ the set of knowledge bases.*

If f is a linear function, ie if a constraint c has the form "$b \geq [a_1, a_2, \ldots, a_J] * \omega$", where a_j are real numbers such that $1 = \sqrt{\sum_{j=1}^{J} {a_j}^2}$, then c is said to be a *linear* constraint. If f is a polynomial, ie if c has the form "$b \geq \sum_{i=1}^{I\in\mathbb{N}} a_i * \prod_{j=1}^{J} \omega_j^{d_{ij}}$", where a_i and d_{ij} are real numbers, then c is said to be a *polynomial* constraint. Let $\mathbb{K}^P \subset \mathbb{K}$ be the set of polynomial knowledge bases, which are multisets of polynomial constraints, and let $\mathbb{K}^L \subset \mathbb{K}^P$ be the set of linear knowledge bases, which are multisets of linear constraints. Also, let $\mathbb{K}^= \subset \mathbb{K}^L$ be the set of linear knowledge bases containing only equality constraints, which are pairs of linear constraints of the form $\{ b \geq f(\omega), -b \geq -f(\omega) \}$, or equivalently $b = f(\omega)$.

Let Sol_c be the set of solutions of a constraint c; notice that $\emptyset \neq Sol_c \subset \mathbb{D}$ since $\exists x, x' \in \mathbb{D}, f(x) > b \geq f(x')$ (see Def. 3). Let $Sol_K \stackrel{\text{def}}{=} \bigcap_{c\in K} Sol_c$ be the set of solutions of a knowledge base K. Let $\Omega_c \stackrel{\text{def}}{=} \Omega \cap Sol_c$ be the set of probability distributions satisfying c.

In addition, we denote by $\mathbb{K}^*$ the set of knowledge bases such that, for each of their constraints c, Sol_c is a union of pairwise-disjoint convex sets $\{S_1, S_2, \ldots, S_m\}$, and the probability distributions have a common nearest set S_i, ie $\exists i, \forall j \in \{1, 2, \ldots, m\}, \forall \omega \in \Omega, \mathcal{G}(\omega, S_i) < \mathcal{G}(\omega, S_j)$, where we denote by $\mathcal{G}(\omega, S)$ the smallest Euclidean distance between a point ω and any point in a set S. Notice that $\mathbb{K}^L \subset \mathbb{K}^*$ since Sol_c, which is a halfspace, is convex for any linear constraint c.

Definition 5 (Models of K)**.** *A model of a knowledge base K is a probability distribution satisfying all the constraints in K. We denote by $\Omega_K \stackrel{\text{def}}{=} \Omega \cap Sol_K$ the set of models of K.*

Definition 6 (Consistency)**.** *A knowledge base K is consistent iff $\Omega_K \neq \emptyset$, ie, iff there exists a probability distribution satisfying all the constraints in K; otherwise, K is said to be* inconsistent.

Definition 7. *A maximal consistent subset Q of a knowledge base K is a consistent subset of K such that no constraint $c \in K \setminus Q$ can be added to Q without yielding $Q \cup \{c\}$ inconsistent. The set of maximal consistent subsets of K is defined as follows:*

$$\mathrm{MCS}_K \stackrel{\text{def}}{=} \left\{ Q \;\middle|\; \begin{array}{l} \forall c \in K \setminus Q, \Omega_{Q\cup\{c\}} = \emptyset \\ \text{and } \Omega_Q \neq \emptyset \text{ and } Q \subseteq K \end{array} \right\}$$

Definition 8. *A kernel Q of a consistent knowledge base K is a smallest subset of K such that $\Omega_Q = \Omega_K \neq \emptyset$ and Q does not contain tautologies, ie $\forall c \in Q, \Omega_c \subset \Omega$. We denote by $\heartsuit K$ the set of kernels of K.*

2.2.3 Expressiveness of $\mathbb{K}$: conditional probability and stochastic independence

In this thesis, we focus on linear knowledge bases $\mathbb{K}^L$. Our motivation is that a knowledge base $K \in \mathbb{K}^L$ is simply a matrix inequality $B \geq A * \omega$, where $\omega \in \mathbb{R}^J$, where B and A are defined as follows, with $I \in \mathbb{N}$ being the number of constraints of K:

$$B \stackrel{\text{def}}{=} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_I \end{bmatrix} \quad A \stackrel{\text{def}}{=} \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,J} \\ a_{2,1} & a_{2,2} & \dots & a_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I,1} & a_{I,2} & \dots & a_{I,J} \end{bmatrix}$$

Despite their simplicity, linear knowledge bases generalise widely used bases such as sets of propositions or sets of conditional probabilities.

Conditional probability: expressible in $\mathbb{K}^L$

We now show that inequalities with conditional probabilities are expressible in $\mathbb{K}^L$. Let $\omega(\psi \mid \phi) \stackrel{\text{def}}{=} \frac{\omega(\psi \wedge \phi)}{\omega(\phi)}$ denote the probability of $\psi \in \Theta$ being true assuming $\phi \in \Theta$ is true, where ω is a probability distribution on Θ. Let $a, b, d \in \mathbb{R}$, and suppose $\omega(\phi) > 0$. We thus have $d \geq \sum_{k=1}^{m\in\mathbb{N}} a_k * \omega(\psi_k \mid \phi)$ iff $d \geq \sum_{k=1}^{m} a_k * \frac{\omega(\psi\wedge\phi)}{\omega(\phi)}$ iff $d * \omega(\phi) \geq \sum_{k=1}^{m} a_k * \omega(\psi_k \wedge \phi)$ iff $0 \geq -d * \omega(\phi) + \sum_{k=1}^{m} a_k * \omega(\psi_k \wedge \phi)$ iff $0 \geq \sum_{k=1}^{m+1} a_k * \omega(\theta_k)$, where $a_{m+1} * \omega(\theta_{m+1}) \stackrel{\text{def}}{=} -d * \omega(\phi)$. Notice that this last inequality is a linear combination of marginal probabilities expressible in $b' \geq \sum_{i=1}^{n\in\mathbb{N}} a'_i * \omega(\theta_i)$. We showed in §2.2.1 that $\omega(\theta) = \sum_{\alpha\in\alpha_\theta} \omega(\alpha)$ for any proposition $\theta \in \Theta$. Hence, $b' \geq \sum_{i=1}^{n} a'_i * \omega(\theta_i)$ iff $b' \geq \sum_{i=1}^{n} a'_i * \sum_{\alpha\in\alpha_{\theta_i}} \omega(\alpha)$ iff $b' \geq \sum_{j=1}^{J} a''_j * \omega(\alpha_j)$ after expanding and collecting probabilities over the minterms. Finally, in order to ease some forthcoming computations, the coefficients are normalised: $b \stackrel{\text{def}}{=} \frac{b'}{\text{norm}}$, $a_j \stackrel{\text{def}}{=} \frac{a''_j}{\text{norm}}$, $\text{norm} \stackrel{\text{def}}{=} \sqrt{\sum_{j=1}^{J} a''^2_j}$. Hence, any linear combination of probabilities can be rewritten in this normalised form $b \geq \sum_{j=1}^{J} a_j * \omega(\alpha_j)$, which is a linear constraint.

Stochastic independence: expressible in $\mathbb{K}^P$

As explained in [33, chapter 9], stochastic independence is not expressible in $\mathbb{K}^L$. For example, the following polynomial constraint is not linear: $\omega(\theta \wedge \phi) = \omega(\theta) * \omega(\phi)$, where $\theta, \phi \in \Theta$ and $\omega \in \Omega$. Stochastic independence can nevertheless be expressed in polynomial knowledge bases $\mathbb{K}^P$. The issue we must face when reasoning with such bases is the apparent impossibility to define a commonsensical inference process returning a unique probability distribution, as required by *uniqueness* (see principle $\mathbf{P}^{\mathcal{I}}_{\alpha}$); for example, how could we infer a unique probability of rain if we only know that such a probability equals either 0.4 or 0.6 (this is a polynomial knowledge base)? Notice that [38] provides a principled inference process for consistent polynomial knowledge bases which suffers from this issue. In §2.2.2, we attempted to define $\mathbb{K}^*$ as the set of knowledge bases on which applying a commonsensical inference process always returns a unique probability distribution. Thus, $\mathbb{K}^*$ not only contains the linear knowledge bases, but also bases made of polynomial constraints like $0.2 \leq x * y$, where $x, y \in \mathbb{R}^J$: the reason is that all the probability distributions are nearer to the same parabolic set, namely $\{[x; y] \in \mathbb{R}^{+J} \times \mathbb{R}^{+J} \mid 0.2 \leq x * y\}$, which is a convex set.

2.2.4 The most probable worlds: the best candidates $\hat{\Omega}_C$ of a candidacy function $C \in \mathbb{C}$

Knowledge is intended to represent the real world. As George E. P. Box wrote, *all models are wrong, but some are useful*, where we now interpret *useful models* as *good candidates for representing the real world*. Among all the probability distributions, those in Ω_K are *useful*, wrt a knowledge base K. But when K is inconsistent, Ω_K is empty; does it mean that no probability distribution is *useful*? Probably not. In section 2.3, we thus propose to convert K into a candidacy function C_K, which gives to each probability distribution ω a degree in [0:1] of candidacy for representing the real world wrt K; in other words, $C_K(\omega)$ returns 1 iff ω satisfies K, otherwise it returns the degree to which ω satisfies K.

Definition 9. *A candidacy function C is a function from Ω to [0:1] such that $C(\omega) > 0$ means ω is a candidate for representing the real world.*

Definition 10 (Best candidates wrt C)**.** *The non-empty set of probability distributions that are the best candidates (or the nominees) for representing the real world, wrt a candidacy function C, is defined as follows:*

$$\hat{\Omega}_C \stackrel{\text{def}}{=} \arg\max_{\omega\in\Omega} C(\omega)$$

2.3 From knowledge bases to candidacy functions: a bridge to paraconsistency

In the previous section, we defined a knowledge base in $\mathbb{K}$ as a multiset of constraints because it is natural to express knowledge in terms of constraints. In order to benefit from both the convenience of $\mathbb{K}$ and the expressiveness of candidacy functions $\mathbb{C}$, we propose several principles to construct the candidacy function corresponding to a given knowledge base. Stating these principles requires some notations and assumptions. We thus study an operator enriching the propositional language underlying a knowledge base. We then propose a definition for the intended meaning conveyed by a knowledge base. A natural definition of equivalence between knowledge bases follows. After explaining the merging of knowledge bases and candidacy functions, we define in §2.3.5 a function that blurs any knowledge base $K \in \mathbb{K}$ to return its corresponding candidacy function C_K. Such a blur satisfies principles like *language invariance* (see $\mathbf{P}_9^{\mathbb{C}}$), which states, roughly, that knowledge written in a certain language should not change when this language becomes more expressive. We furthermore define another equivalence relation between knowledge base that extends the relation used by J.B. Paris in [33, pages 89–91] to inconsistent knowledge bases.

Notations

We consider a probability distribution $\omega : \Theta \mapsto [0{:}1]$ underlain by a propositional language Θ having n variables to be a point p in a Euclidean space $\mathbb{R}^{2^n}$, where each axes of $\mathbb{R}^{2^n}$ is uniquely labelled with one minterm of Θ. A point p in a such labelled space is thus said to be underlain by language Θ.

Let $\Theta(\diamond)$ return the underlying propositional language of $\diamond$, where $\diamond$ can be a knowledge base K, a point $p \in \mathbb{R}^{2^n}$, a set of points $\mathfrak{k} \subset \mathbb{R}^{2^n}$, a constraint c, the set of probability distributions Ω, a probability distribution ω, or a candidacy function C; if $\diamond$ is a language, then $\Theta(\diamond)$ returns $\diamond$.

Let $vars(\diamond)$ return the finite set of propositional variables of $\Theta(\diamond)$.

We define the gap between two sets of points $\mathfrak{k}_1$ and $\mathfrak{k}_2$ as $\mathcal{G}(\mathfrak{k}_1, \mathfrak{k}_2) \stackrel{\text{def}}{=} \inf \{ \mathcal{L}_2(x, y) \mid x \in \mathfrak{k}_1, y \in \mathfrak{k}_2 \}$, ie the Euclidean distance between the two nearest points in $\mathfrak{k}_1$ and $\mathfrak{k}_2$; we furthermore denote by $\mathcal{G}(x, \mathfrak{k}) \stackrel{\text{def}}{=} \mathcal{G}(\{x\}, \mathfrak{k})$ the gap between a point x and a set of points $\mathfrak{k}$, and by $\mathcal{G}(x, y) \stackrel{\text{def}}{=} \mathcal{G}(\{x\}, \{y\})$ the gap between two points x and y, which is also the distance $\mathcal{L}_2(x, y)$.

2.3.1 Extending the set of propositional variables with the language enrichment operator $\diamond \oplus v$

We now define a polymorphic operator $\diamond \oplus v$ (read it as "$\diamond$ enriched with v") that adds a *new* propositional variable v in the underlying propositional language of an object $\diamond$ having n variables, where object $\diamond$ can be:

- a propositional language Θ, defined as a set of minterms α_Θ:

$$\alpha_{\Theta \oplus v} \stackrel{\text{def}}{=} \{ \alpha \wedge \neg v, \alpha \wedge v \mid \alpha \in \alpha_\Theta \}$$

  A propositional language Θ enriched with a new variable v, denoted by $\Theta \oplus v$, is strictly more expressive than Θ: any sentence in Θ is expressible in $\Theta \oplus v$ because each sentence can be written as a disjunction of minterms, and each minterms α of Θ can be expressed as a conjunction of two minterms of $\Theta \oplus v$, ie $\alpha = (\alpha \wedge v) \vee (\alpha \wedge \neg v)$.

- a linear constraint c of the form "$b \geq A * x$", where $x \in \mathbb{R}^{2^n}$ and $x' \in \mathbb{R}^{2^{n+1}}$ (notice that since c is normalised, the norm of vector A is 1, hence the norm of vector $[A, A]$ is $\sqrt{2}$):

$$c \oplus v \stackrel{\text{def}}{=} \text{“}\frac{b}{\sqrt{2}} \geq \frac{[A, A]}{\sqrt{2}} * x'\text{”}$$

  For example, the enrichment with v_2 of the linear constraint "$20\% \geq \omega(v_1)$" underlain by a propositional language with one variable v_1 is performed as follows:

$$\text{“}0.2 \geq [0, 1] * x\text{”} \oplus v_2 = \text{“}\frac{0.2}{\sqrt{2}} \geq \frac{[0, 1, 0, 1]}{\sqrt{2}} * x'\text{”}$$

  where $\omega : \Theta \mapsto [0{:}1]$ and $\omega : \Theta \oplus v_2 \mapsto [0{:}1]$ are two probability distributions, $vars(\Theta) = \{v_1\}$, and

$$x \stackrel{\text{def}}{=} \begin{bmatrix} \omega(\neg v_1) \\ \omega(v_1) \end{bmatrix} \text{ and } x' \stackrel{\text{def}}{=} \begin{bmatrix} \omega'(\neg v_1 \wedge \neg v_2) \\ \omega'(v_1 \wedge \neg v_2) \\ \omega'(\neg v_1 \wedge v_2) \\ \omega'(v_1 \wedge v_2) \end{bmatrix}$$

- a knowledge base K:

$$K \oplus v \stackrel{\text{def}}{=} \{ c \oplus v \mid c \in K \}$$

  For example, the enrichment with v_2 of the linear knowledge base containing the two constraints "$20\% \geq \omega(v_1)$" and "$40\% \geq \omega(\neg v_1)$" and underlain by a propositional language with one variable v_1 is performed as follows (where x and x' are defined as above):

$$\text{“}\begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix} \geq \begin{bmatrix} 0, 1 \\ 1, 0 \end{bmatrix} * x\text{”} \oplus v_2 = \text{“}\frac{\begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix}}{\sqrt{2}} \geq \frac{\begin{bmatrix} 0, 1, 0, 1 \\ 1, 0, 1, 0 \end{bmatrix}}{\sqrt{2}} * x'\text{”}$$

- a point $p \in \mathbb{R}^{2^n}$, defined as an intersection of 2^n hyperplanes expressed as a linear knowledge base K such that $Sol_K = \{p\}$. Informally, K is simply the set of constraints "$p = x$" where $x \in \mathbb{R}^{2^n}$. Formally, let A be the identity matrix of dimension 2^n, and let $B \stackrel{\text{def}}{=} p$ be the coordinates of p. Enriching a point p is thus enriching the knowledge base $K \stackrel{\text{def}}{=}$ "$[B; -B] \geq [A; -A] * x$":

$$p \oplus v \stackrel{\text{def}}{=} Sol_{K \oplus v}$$

For example, figure 2.1 illustrates that the enrichment of a point $x \in \mathbb{R}^{2^n}$ (the yellow dot) with a new variable v yields an infinite set of points $x \oplus v$ (the yellow box).

- a set of points $\mathfrak{k} \subset \mathbb{R}^{2^n}$:

$$\mathfrak{k} \oplus v \stackrel{\text{def}}{=} \bigcup_{p \in \mathfrak{k}} p \oplus v$$

- a (non-linear) constraint c, which is identified with its set of solutions (notice that the following definition is implicit, and so is the definition of $K \oplus v$ when K is not a linear knowledge base):

$$Sol_{c \oplus v} \stackrel{\text{def}}{=} \bigcup_{p \in Sol_c} p \oplus v$$

- the set of probability distributions Ω, which is seen as the linear knowledge base constraining the probability value of each minterm to sum up to 1 while being positive real number. Let x' be a point in $\mathbb{R}^{2^{n+1}}$, $One \stackrel{\text{def}}{=} [1, 1, \ldots, 1]$ be the one-row matrix made of 2^{n+1} ones, and A be set to the identity matrix of dimension 2^{n+1}. The enrichment of the set of probability distributions Ω is thus performed as follows:

$$\Omega \oplus v \stackrel{\text{def}}{=} Sol_{\text{“}\left[\frac{[1;-1]}{\sqrt{2^{n+1}}};[0;0;\ldots;0]\right] \geq \left[\frac{[One;-One]}{\sqrt{2^{n+1}}};-A\right] * x'\text{”}}$$

- a probability distribution ω, where p_ω is the point corresponding to ω, which is thus seen as an intersection of hyperplanes:

$$\omega \oplus v \stackrel{\text{def}}{=} (p_\omega \oplus v) \cap (\Omega \oplus v)$$

and such that, for each probability distribution $\omega' \in \omega \oplus v$, $\Theta(\omega') = \Theta(\omega) \oplus v$, and for each $j \in \{1, 2, \ldots, 2^n\}$, if $\omega_j = \omega(\alpha_j)$ then $\omega'_j \stackrel{\text{def}}{=} \omega'(\alpha_j \wedge \neg v)$ and $\omega'_{j+2^n} \stackrel{\text{def}}{=} \omega'(\alpha_j \wedge v)$. Thus, $\omega_j = \omega'_j + \omega'_{j+2^n}$;

- a candidacy function C:

$$\forall \omega \in \Omega, \forall \omega' \in \omega \oplus v, (C \oplus v)(\omega') \stackrel{\text{def}}{=} C(\omega)$$

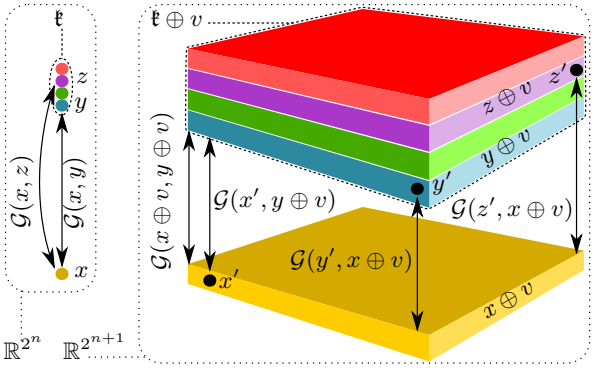


Figure 2.1: The gaps between two points, eg $\mathcal{G}(x, y)$, and between a point and a set, eg $\mathcal{G}(x, \mathfrak{k})$, are invariant by language enrichment, up to a factor of $\sqrt{2}$ (see propositions 2 and 3).

Finally, we recursively define the addition of a set of variables *vars* to an object $\diamond$ as follows:

$$\diamond \oplus vars \stackrel{\text{def}}{=} \begin{cases} (\diamond \oplus v) \oplus (vars \setminus \{v\}) & \text{if } v \notin \Theta(\diamond), \\ \diamond \oplus (vars \setminus \{v\}) & \text{if } v \in \Theta(\diamond), \\ \diamond & \text{if } vars = \emptyset. \end{cases}$$

Proposition 1. *A probability distribution in $\Omega \oplus v$ induces a unique probability distribution in Ω.*

Proof. $\omega' \in \Omega \oplus v$ induces a probability distribution in Ω iff, firstly, for each minterm α of $\Theta(\Omega)$, $0 \leq \omega'(\alpha)$, which is true since $\omega' : \Theta(\Omega \oplus v) \mapsto [0{:}1]$ and $\alpha \in \Omega \oplus v$, and secondly, the probabilities given by ω' to the minterms of $\Theta(\Omega)$ sum up to 1, ie $\sum_{\alpha \in \alpha_{\Theta(\Omega)}} \omega'(\alpha) = \sum_{\alpha \in \alpha_{\Theta(\Omega)}} \omega'(\alpha \wedge v) + \omega'(\alpha \wedge \neg v) = \sum_{\alpha' \in \alpha_{\Theta(\Omega \oplus v)}} \omega'(\alpha') = 1$, which is true by marginalisation, ie $\forall \alpha \in \alpha_{\Theta(\Omega)}, \omega'(\alpha) = \omega'(\alpha \wedge (v \vee \neg v)) = \omega'(\alpha \wedge v \vee \alpha \wedge \neg v) = \omega'(\alpha \wedge v) + \omega'(\alpha \wedge \neg v)$. □

Proposition 2. *The gap $\mathcal{G}$ between two points x and y is proportional to the gap between any point $x' \in x \oplus v$ and the set of points $y \oplus v$:*

$$\mathcal{G}(x, y) = \sqrt{2} * \mathcal{G}(x', y \oplus v)$$

Proof. Let $x, y, s, t \in \mathbb{R}^J$ be any four points. Let $x' \stackrel{\text{def}}{=} [\frac{x}{2} + s; \frac{x}{2} - s]$ and $y' \stackrel{\text{def}}{=} [\frac{y}{2} + t; \frac{y}{2} - t]$. Notice that x' and y' are any two points in $x \oplus v$ and $y \oplus v$, respectively, since $x' \in x \oplus v$ iff $x_j = x'_j + x'_{j+J}$ and $y' \in y \oplus v$ iff $y_j = y'_j + y'_{j+J}$, for any $j \in \{1, 2, \ldots, J\}$. Notice also that $\mathcal{G}(x', y')^2 = \sum_{j=1}^J |(\frac{x_j}{2} + s_j) - (\frac{y_j}{2} + t_j)|^2 + |(\frac{x_j}{2} - s_j) - (\frac{y_j}{2} - t_j)|^2 = \sum_{j=1}^J |\frac{x_j - y_j}{2} + (s_j - t_j)|^2 + |\frac{x_j - y_j}{2} - (s_j - t_j)|^2 = \sum_{j=1}^J 2 * (\frac{x_j - y_j}{2})^2 + 2 * (s_j - t_j)^2 = \frac{1}{2} * \sum_{j=1}^J (x_j - y_j)^2 + 2 * \sum_{j=1}^J (s_j - t_j)^2 = \frac{1}{2} * \mathcal{G}(x, y)^2 + 2 *$

$\mathcal{G}(s,t)$[2]. Thus, for two given points x and y, $\mathcal{G}(s,t) = 0$ iff $\mathcal{G}(x',y')$ is minimal, ie $\forall a' \in x \oplus v, \forall b' \in y \oplus v, \mathcal{G}(x',y') \leq \mathcal{G}(a',b')$. Furthermore, for a given point x', $\mathcal{G}(x', y \oplus v) = \mathcal{G}(x',y') = \frac{1}{\sqrt{2}} * \mathcal{G}(x,y)$, where y' is such that $\mathcal{G}(x',y')$ is minimal, ie $\forall b' \in y \oplus v, \mathcal{G}(x',y') \leq \mathcal{G}(x',z')$. Therefore, $\sqrt{2} * \mathcal{G}(x', y \oplus v) = \mathcal{G}(x,y)$. □

Proposition 3. *The gap $\mathcal{G}$ between a point x and a set of points $\mathfrak{k}$ is proportional to the gap between any point $x' \in x \oplus v$ and the set $\mathfrak{k} \oplus v$:*

$$\mathcal{G}(x,\mathfrak{k}) = \sqrt{2} * \mathcal{G}(x', \mathfrak{k} \oplus v) = \sqrt{2} * \mathcal{G}(x \oplus v, \mathfrak{k} \oplus v)$$

Proof. Let $y \in \mathfrak{k}$ be a point such that $\mathcal{G}(x,y) = \mathcal{G}(x,\mathfrak{k})$, ie $\forall z \in \mathfrak{k}, \mathcal{G}(x,y) \leq \mathcal{G}(x,z)$. According to Prop. 2, $\forall x' \in x \oplus v, \mathcal{G}(x,y) = \sqrt{2} * \mathcal{G}(x', y \oplus v)$ hence we have $\forall z \in \mathfrak{k}, \forall x', x'' \in x \oplus v, \mathcal{G}(x', y \oplus v) \leq \mathcal{G}(x'', z \oplus v)$. Notice that the gap between a point x'' and a set of points $z \oplus v$ is lower than the gap between x'' and a point z' in $z \oplus v$, ie $\forall z' \in z \oplus v, \mathcal{G}(x'', z \oplus v) \leq \mathcal{G}(x'', z')$. We therefore obtain $\forall z \in \mathfrak{k}, \forall z' \in z \oplus v, \forall x', x'' \in x \oplus v, \mathcal{G}(x', y \oplus v) \leq \mathcal{G}(x'', z')$. By definition of $\mathfrak{k} \oplus v$, it follows that $\forall z' \in \mathfrak{k} \oplus v, \forall x', x'' \in x \oplus v, \mathcal{G}(x', y \oplus v) \leq \mathcal{G}(x'', z')$. We then deduce $\forall x' \in x \oplus v, \mathcal{G}(x', y \oplus v) = \mathcal{G}(x', \mathfrak{k} \oplus v) = \mathcal{G}(x \oplus v, \mathfrak{k} \oplus v)$ and conclude that $\forall x' \in x \oplus v, \mathcal{G}(x,\mathfrak{k}) = \mathcal{G}(x,y) = \sqrt{2} * \mathcal{G}(x', y \oplus v) = \sqrt{2} * \mathcal{G}(x', \mathfrak{k} \oplus v) = \sqrt{2} * \mathcal{G}(x \oplus v, \mathfrak{k} \oplus v)$. □

2.3.2 Assumptions about the knowledge content $\mathfrak{K}_K$ to decipher a knowledge base K

A knowledge base, written in a certain language like $\mathbb{K}^L$, is a vehicle for knowledge rather than the knowledge itself; what we call *knowledge content* is the intended meaning conveyed by a knowledge base, ie its very essence, freed from any syntactic or linguistic consideration. We furthermore distinguish two levels of knowledge content: the *internal* level, at which knowledge management occurs (eg: knowledge merging), and the *external* level, at which the inferences (hence the decisions) are performed (see chapter 4 about inference processes). The internal (external) level is related to the notion of *credal* (*pignistic*) level introduced in [41, §3.2]. We now propose different assumptions about the knowledge content $\mathfrak{K}_K$ of a knowledge base K. The next five assumptions focus on the *internal* level, while the additional assumption stated in § 2.3.6 on page 15 focuses on the *external* level. Let $\mathfrak{K}_\emptyset \stackrel{\text{def}}{=} \{\Omega\}$.

Each element $\mathfrak{k}$ of multiset $\mathfrak{K}_K$ is called a knowledge item; the sources of knowledge providing these items are supposed to be mutually independent. Each assumption defines what K is a description of.

① $\mathfrak{K}_K \stackrel{\text{def}}{=} \Omega_K$. This assumption is made by J.B. Paris in his book (see [33, pages 89–91]) when he employs the Hausdorff distance (see Def 23 on page 19) between Ω_{K_1} and Ω_{K_2} as the distance between knowledge contents, where K_1 and K_2 are consistent knowledge bases. However, this assumption is undefined for inconsistent knowledge bases since the Hausdorff distance from or to an empty set is undefined. Therefore, assumption ① is not acceptable.

② $\mathfrak{K}_K \stackrel{\text{def}}{=} \{\, \Omega_Q \mid Q \in \text{MCS}_K \,\}$. This assumption is not only equivalent to ① when K is consistent because $\text{MCS}_K = \{K\}$, but also defined when K is inconsistent because $\Omega_Q \neq \emptyset$. Notice that ② identifies contradictions, eg "$-1 \geq \omega(\alpha_1)$", with tautologies, eg "$1 \geq \omega(\alpha_1)$", because tautologies appear in every $Q \in \text{MCS}_K$ without impacting Ω_Q, while contradictions never appear in $Q \in \text{MCS}_K$ hence never impact Ω_Q. More generally, accepting ② is ignoring all the constraints in K that do not belong to a kernel (see Def. 8 on page 5) of a maximal consistent subset, ie the knowledge of K is the same as that of $\bigcup_{Q \in \text{MCS}_K} \bigcup_{P \in \heartsuit_Q} P$.

③ $\mathfrak{K}_K \stackrel{\text{def}}{=} \{\, \Omega_c \mid c \in K \,\}$. To accept this assumption is to consider as equivalent all the contradictions (if c is a contradiction, then $\Omega_c = \emptyset$ by definition), and is to consider as equivalent all the tautologies (if c is a tautology, then $\Omega_c = \Omega$ by definition). Also, a knowledge base K is identified here with a set of constraints on a probability distribution, whereas in assumptions ① or ②, K is rather considered as a whole.

④ $\mathfrak{K}_K \stackrel{\text{def}}{=} \{\, Sol_c \mid c \in K \,\}$. Accepting ④ is deeming $\mathfrak{K}_K$ to be the extensional multiset of an intensional multiset K (each constraint c in K is here the intensional version of an extensional set Sol_c). Under assumption ④, we denote by $\mathfrak{k}_c \stackrel{\text{def}}{=} Sol_c$ the knowledge item corresponding to a constraint c. If c is a linear constraint, then $\mathfrak{k}_c$ is a halfspace of $\mathbb{R}^J$, where J is the number of minterms of the underlying propositional language of c. Notice that a solution $x \in Sol_c$ of a constraint c may not be a probability distribution, ie x may be in $\mathbb{R}^J \setminus \Omega$. Then, accepting assumption ④ is considering a knowledge base as a multiset of constraints on a point in $\mathbb{R}^J$ rather than on a probability distribution in Ω, like in assumption ③. Thus, ④ is more paraconsistent than ③ because ④ deals with constraints that are inconsistent with the axioms for probabilities.

⑤ $\mathfrak{K}_K \stackrel{\text{def}}{=} \{\, Sol_c \mid c \in K \text{ and } \Omega \not\subseteq Sol_c \,\}$. Accepting ⑤ is assuming ④ while deeming tautologies void of knowledge content.

In order to deal with paraconsistent probabilistic reasoning, we require $\mathfrak{K}_K$ to be defined when K is inconsistent (not like ①), and to distinguish not only between contradictions and tautologies (not like ②), but also between two contradictions (not like ③). Moreover, we consider a tautology as redundant with the axioms for probability, hence being void of knowledge content (not like ④): in this thesis, we thus accept ⑤ as a definition for the *internal* level of knowledge content (assumption ⑥ on page 15 will be accepted as a definition for the *external* level). We furthermore adhere to the following principle, which is named as *Watts assumption* in [33, pages 67 and 134].

$\mathbf{P}_1^{\mathbb{C}}$ *Watts assumption.* Under assumption ⑤, The knowledge content $\mathfrak{K}_K$ of a knowledge base K is essentially all the relevant knowledge that we have.

2.3.3 Comparing knowledge at *credal* level: the internal equivalence $\overset{\text{i}}{\equiv}$

Definition 11 (Internal equivalence)**.** *Two knowledge bases K_1 and K_2 are internally equivalent iff they have the same knowledge content wrt assumption ⑤.*

$$K_1 \overset{\text{i}}{\equiv} K_2 \overset{\text{def}}{=} \begin{cases} \left(\mathfrak{K}_{K_1}^{⑤} = \mathfrak{K}_{K_2}^{⑤}\right) & \text{if } \Theta(K_1) = \Theta(K_2), \\ \mathbf{t}\text{rue} & \text{if } \exists v, K_1 = K_2 \oplus v, \\ \mathbf{f}\text{alse} & \text{otherwise.} \end{cases}$$

Two candidacy functions C_1 and C_2 are internally equivalent iff they return the same value for each probability distribution.

$$C_1 \overset{\text{i}}{\equiv} C_2 \overset{\text{def}}{=} \begin{cases} (C_1 = C_2) & \text{if } \Theta(C_1) = \Theta(C_2), \\ \mathbf{t}\text{rue} & \text{if } \exists v, C_1 = C_2 \oplus v, \\ \mathbf{f}\text{alse} & \text{otherwise.} \end{cases}$$

We denote by $\emptyset_{\mathbb{K}}$ any tautological knowledge base in $\{ K \in \mathbb{K} \mid \Omega_K = \Omega \}$. We furthermore denote by $1_{\mathbb{C}}$ any tautological candidacy function in $\{ C \in \mathbb{C} \mid \forall \omega \in \Omega, C(\omega) = 1 \}$. Notice that the two sets are the equivalent classes of tautologies wrt $\overset{\text{i}}{\equiv}$.

2.3.4 Merging knowledge from independent sources: the merging operator $\Cup$

In order to lighten the notation, we assume knowledge bases and candidacy functions to be underlain by the same propositional language. A merging operator $\Cup$ is a binary operator that should satisfies the following principles.

$\mathbf{P}_2^{\mathbb{C}}$ *Closure.* Merging two knowledge bases, or two candidacy functions, should yield a knowledge base, or a candidacy function, respectively:

$$K_1 \Cup K_2 \in \mathbb{K} \qquad C_1 \Cup C_2 \in \mathbb{C}$$

$\mathbf{P}_3^{\mathbb{C}}$ *Associativity & symmetry.* A merging operator should be indifferent to the order in which knowledge bases, or candidacy functions, are merged:

$$(K_1 \Cup K_2) \Cup K_3 = K_1 \Cup (K_2 \Cup K_3)$$
$$(C_1 \Cup C_2) \Cup C_3 = C_1 \Cup (C_2 \Cup C_3)$$

$$K_1 \Cup K_2 = K_2 \Cup K_1$$
$$C_1 \Cup C_2 = C_2 \Cup C_1$$

$\mathbf{P}_4^{\mathbb{C}}$ *Identity element.* A merging operator should be indifferent to tautological knowledge bases, or tautological candidacy functions, because they are redundant with the axioms for probability:

$$K \Cup \emptyset_{\mathbb{K}} = K \qquad C \Cup 1_{\mathbb{C}} = C$$

Dependent versus independent sources. Suppose one person provides us with a knowledge base K_1 contradicting a knowledge base K_2 that we obtained from a second person; our knowledge is thus $K_1 \Cup K_2$. Further suppose that a third person provides us with K_2; has our knowledge changed? Formally, does $K_1 \Cup K_2$ differ from $K_1 \Cup K_2 \Cup K_2$? If yes, then we are considering people as independent sources of knowledge; *in this thesis, we assume each knowledge item to be independent on the others.* Otherwise, $K_1 \Cup K_2 = K_1 \Cup K_2 \Cup K_2$ means we are considering people as dependent sources, which happens if we know that the second and the third person have acquired K_2 by watching the same television program, for example. In such case, we may consider these two people as one single source of knowledge; K_2 should thus equal $K_2 \Cup K_2$, which means that $\Cup$ should be idempotent. Since in this thesis we assume the sources of knowledge to be independent, *idempotence* is undesirable.

$\mathbf{P}_5^{\mathbb{C}}$ *Non-idempotence.* The more numerous the independent sources supporting a probability distribution ω are, the higher the candidacy level of ω should be for representing the real world. Such a statement implies the rejection of *idempotence*, where $K \neq \emptyset_{\mathbb{K}}$ and $C \neq 1_{\mathbb{C}}$:

$$K \Cup K \neq K \qquad C \Cup C \neq C$$

Definition 12 (Merging operator)**.** *The merging operators for knowledge bases $\Cup : \mathbb{K} \times \mathbb{K} \mapsto \mathbb{K}$ and for candidacy functions $\Cup : \mathbb{C} \times \mathbb{C} \mapsto \mathbb{C}$ are respectively defined as follows, where $\cup$ is the union of two multisets $K_1, K_2 \in \mathbb{K}$, and $*$ is the pointwise product of two real functions $C_1, C_2 \in \mathbb{C}$:*

$$K_1 \Cup K_2 \overset{\text{def}}{=} K_1 \cup K_2 \qquad C_1 \Cup C_2 \overset{\text{def}}{=} C_1 * C_2$$

Proposition 4. *$\Cup$ is non-idempotent (see $\mathbf{P}_5^{\mathbb{C}}$). $\langle \mathbb{K}, \Cup \rangle$ and $\langle \mathbb{C}, \Cup \rangle$ are commutative monoids, ie semigroups with identity (see $\mathbf{P}_2^{\mathbb{C}}$, $\mathbf{P}_3^{\mathbb{C}}$, and $\mathbf{P}_4^{\mathbb{C}}$).*

Proof. Firstly, the defined merging operators are non-idempotent: $K \uplus K \neq K$ because K is a multiset, and $C \uplus C \neq C$ because $\forall x \in]0{:}1[, x * x \neq x$. Secondly, $\langle \mathbb{K}, \uplus \rangle$ and $\langle \mathbb{C}, \uplus \rangle$ are commutative monoids: $\mathbb{K}$ and $\mathbb{C}$ are closed under mergence because the union $\cup$ of two multisets is a multiset and the product $*$ of two functions is a function, $\uplus$ is associative and symmetric because $\cup$ and $*$ are thus, and $\emptyset_{\mathbb{K}}$ and $1_{\mathbb{C}}$ are identity elements because the empty multiset $\emptyset$ and $1 \in \mathbb{R}$ respectively serve as identity element for $\cup$ and $*$. □

Thus, we merge *dependent* candidacy functions by using the only idempotent t-norm[4] called *minimum* t-norm (see [21, page 103]), ie $(C_1 \uplus_{\text{idempotent}} C_2)(\omega) \stackrel{\text{def}}{=} \min(C_1(\omega), C_2(\omega))$, whereas we merge *independent* candidacy functions by using *algebraic product* t-norm, ie $(C_1 \uplus_{\text{product}} C_2)(\omega) \stackrel{\text{def}}{=} C_1(\omega) * C_2(\omega)$, for any probability distribution ω. We could define a continuum of merging operators $\uplus^\lambda$ where λ runs through [0:1] such that $\uplus^0 = \uplus_{\text{idempotent}}$ and $\uplus^1 = \uplus_{\text{product}}$; for example, $\uplus^\lambda$ may be based on the Dubois-Prade t-norm (see [22, page 74]), ie $(C_1 \uplus^\lambda_{\text{DB}} C_2)(\omega) \stackrel{\text{def}}{=} \frac{C_1(\omega) * C_2(\omega)}{\max(C_1(\omega), C_2(\omega), \lambda)}$, or may be based on the Frank t-norm (see [21, page 108]). As a perspective, we should further study t-norms in order to characterise the merging operator for independent candidacy functions, and to characterise the continuum between dependence and independence of sources of knowledge.

Notice that a candidacy function $0_{\mathbb{C}}$ returning 0 for any probability distribution is an absorbing element for the monoid $\langle \mathbb{C}, \uplus \rangle$ because $\forall C \in \mathbb{C}, 0_{\mathbb{C}} \uplus C = 0_{\mathbb{C}}$. Merging such an absorbing candidacy function with another one is nominating all the probability distributions to be the best candidates, ie $\Omega = \hat{\Omega}_{C \uplus 0_{\mathbb{C}}}$. A paraconsistent probabilistic logic should avoid this effect, which is similar to the *explosion* in classical logic, ie "anything follows from a contradiction". We shall thus establish a principle (see $\mathbf{P}_7^{\mathbb{C}}$) to eschew such absorbing candidacy functions.

2.3.5 The paraconsistent representation of a knowledge base K: the candidacy function C_K

We now propose several principles for a blur function $\mathcal{B} : \mathbb{K} \mapsto \mathbb{C}$, which returns the unique candidacy function C_K intended to represent the knowledge content $\mathfrak{K}_K$ of a given knowledge base K: $C_K \stackrel{\text{def}}{=} \mathcal{B}(K)$. We also denote by $C_{\mathfrak{k}}$ the candidacy function representing a given knowledge item $\mathfrak{k} \in \mathfrak{K}_K$.

[4] A triangular norm, t-norm for short, is a function $T : [0{:}1] \times [0{:}1] \mapsto [0{:}1]$ satisfying four axioms: *commutativity* $T(x, y) = T(y, x)$, *associativity* $T(x, T(y, z)) = T(T(x, y), z)$, *monotonicity* $T(x, y) \leq T(x, z)$ whenever $y \leq z$, and *boundary condition* $T(x, 1) = 1$.

$\mathbf{P}_6^{\mathbb{C}}$ *Homomorphism.* Merging the candidacy functions representing two knowledge bases should yield the candidacy function representing the mergence of the two knowledge bases. Furthermore, the candidacy function corresponding to a tautological knowledge base is tautological. Formally, $\mathcal{B} : \mathbb{K} \mapsto \mathbb{C}$ is a homomorphism between monoids $\langle \mathbb{K}, \uplus \rangle$ and $\langle \mathbb{C}, \uplus \rangle$:

$$C_{K_1} \uplus C_{K_2} = C_{K_1 \uplus K_2} \text{ and } C_{\emptyset_{\mathbb{K}}} \stackrel{\text{i}}{\equiv} 1_{\mathbb{C}}$$

$\mathbf{P}_7^{\mathbb{C}}$ *Proximity.* From two probability distributions $\omega_1, \omega_2 \in \Omega$, the best candidate, wrt a given knowledge item $\mathfrak{k} \in \mathfrak{K}_K$, is the one nearer to $\mathfrak{k}$, wrt the gap $\mathcal{G}$ between a point and a set of points[5]:

$$\text{if } \mathcal{G}(\omega_1, \mathfrak{k}) < \mathcal{G}(\omega_2, \mathfrak{k}) \text{ then } C_{\mathfrak{k}}(\omega_1) > C_{\mathfrak{k}}(\omega_2) > 0$$

$\mathbf{P}_8^{\mathbb{C}}$ *Unanimity.* The models of a knowledge base K are the sole probability distributions to be unanimously designated (by the knowledge items in $\mathfrak{K}_K$) as candidates for representing the real world, ie $\omega \in \Omega_K$ iff $C_K(\omega) = 1$. Thus, if K is inconsistent, no probability distribution ω is unanimously designated as candidate, ie $C_K(\omega) < 1$:

$$\Omega_K = \{ \omega \in \Omega \mid C_K(\omega) = 1 \}$$

We furthermore extend this requirement to each knowledge item $\mathfrak{k} \in \mathfrak{K}_K$:

$$\Omega \cap \mathfrak{k} = \{ \omega \in \Omega \mid C_{\mathfrak{k}}(\omega) = 1 \}$$

$\mathbf{P}_9^{\mathbb{C}}$ *Language invariance.* The candidacy function representing a knowledge base K should be invariant by additions of new variables in the underlying propositional language of K:

$$C_K \stackrel{\text{i}}{\equiv} C_{K \oplus v}$$

We furthermore extend this requirement to each knowledge item $\mathfrak{k} \in \mathfrak{K}_K$:

$$C_{\mathfrak{k}} \stackrel{\text{i}}{\equiv} C_{\mathfrak{k} \oplus v}$$

[5] We recall that the gap between two points ω_1 and ω_2 in a Euclidean space is defined as the Euclidean distance $\mathcal{L}_2(\omega_1, \omega_2)$. In case ω_1 and ω_2 represent probability distributions, ie when they are positive vectors such that their elements sum up to 1, J. Lawry suggests to use an information-based distance such as the Kullback-Leibler divergence (or relative entropy) instead of $\mathcal{L}_2(\omega_1, \omega_2)$; this means that, under assumption ⑤, each constraint c in a knowledge base $K \in \mathbb{K}$ must satisfy $Sol_c \cap \Omega \neq \emptyset$, ie we would not be able to deal with probabilistic contradictions like "$\omega(\theta) \geq 101\%$", where θ is a proposition. We would thus obtain a weaker paraconsistent knowledge representation. An information-based gap between a probability distribution $\omega : \Theta \mapsto [0{:}1]$ and a knowledge item $\mathfrak{k}$, which must contain at least one probability distribution, may be defined as follows:

$$\mathcal{G}_{\text{KL}}(\omega, \mathfrak{k}) \stackrel{\text{def}}{=} \min_{\omega' \in \Omega \cap \mathfrak{k}} \sum_{\alpha \in \alpha_\Theta} \omega(\alpha) * \ln \left(\frac{\omega(\alpha)}{\omega'(\alpha)} \right), \text{ if } \Omega \cap \mathfrak{k} \neq \emptyset$$

These principles guide us towards the definition of $\mathcal{B}$. Firstly, *homomorphism* suggests that $C_K \stackrel{\text{def}}{=} \mathcal{B}(K) \stackrel{\text{def}}{=} \mathbb{U}_{\mathfrak{k}\in\mathfrak{K}_K} C_{\mathfrak{k}}$ under assumptions ④ or ⑤. Secondly, *proximity* indicates that $C_{\mathfrak{k}}(\omega) \stackrel{\text{def}}{=} g_{\mathfrak{k}}(n, \mathcal{G}(\omega, \mathfrak{k}))$ for any probability distribution ω underlain by a propositional language with $n \stackrel{\text{def}}{=} |vars(\omega)|$ variables, where $g_{\mathfrak{k}} : \mathbb{N} \times \mathbb{R}^+ \mapsto]0{:}1]$ is strictly decreasing and continuous wrt its second argument. Suppose that $g_{\mathfrak{k}}$ is independent on $\mathfrak{k}$: we denote this common function by g ; we drop this hypothesis in §2.3.5 when we introduce the *reliability level* of a knowledge item. By requiring g to be strictly positive, *proximity* prevents $\mathcal{B}$ from yielding a (partially) absorbing candidacy function, ie a candidacy function returning 0 for all (or some) probability distributions. Thirdly, *unanimity* states that $C_{\mathfrak{k}}(\omega) = 1$ iff $\omega \in \mathfrak{k}$ iff $\mathcal{G}(\omega, \mathfrak{k}) = 0$, hence $g(n, 0) \stackrel{\text{def}}{=} 1$. Fourthly, *language invariance* states that $C_{\mathfrak{k}} \stackrel{\text{i}}{\equiv} C_{\mathfrak{k}\oplus v}$, which means $C_{\mathfrak{k}} \oplus v = C_{\mathfrak{k}\oplus v}$, hence requires g to satisfy $\forall \omega' \in \omega \oplus v, g(n, \mathcal{G}(\omega, \mathfrak{k})) = g(n+1, \mathcal{G}(\omega', \mathfrak{k} \oplus v))$. Because we assume ⑤, and since proposition 3 states that $\forall \omega' \in \omega \oplus v, \mathcal{G}(\omega, \mathfrak{k}) = \sqrt{2} * \mathcal{G}(\omega', \mathfrak{k} \oplus v)$, we define g as $g(n, \mathcal{G}(\omega, \mathfrak{k})) \stackrel{\text{def}}{=} h(\sqrt{2^n} * \mathcal{G}(\omega, \mathfrak{k}))$, where $h : \mathbb{R}^+ \mapsto]0{:}1]$ is strictly decreasing and continuous, and such that $h(0) \stackrel{\text{def}}{=} 1$, which ensures the continuity of $g(n, x)$ when $x = 0$. Finally, we constrain the possible definitions for h by stating a principle concerning linear knowledge bases.

$\mathbf{P}^{\mathbb{C}}_{10}$ *Convexity.* The set of best candidates wrt the candidacy function representing a linear knowledge base K should be convex:

$$\forall \lambda \in [0{:}1], \forall \omega_1, \omega_2 \in \hat{\Omega}_{C_K}, (1-\lambda)\omega_1 + \lambda\omega_2 \in \hat{\Omega}_{C_K}$$

We see the following two motivations for requiring $\hat{\Omega}_{C_K}$ to be convex when $K \in \mathbb{K}^L$.

- *Restoration of consistency.* The first motivation is that we may want to convert any inconsistent linear knowledge base K_1 into a consistent knowledge base K_2 such that $\Omega_{K_2} = \hat{\Omega}_{C_{K_1}}$. In which case, $\mathbb{K}^L$ is closed by the conversion operation iff K_2 is a linear knowledge base iff $\hat{\Omega}_{C_{K_1}}$ is a convex polyhedron (notice that assuming ④ or ⑤ causes the manifold $\hat{\Omega}_{C_{K_1}}$ to have flat faces).

- *Entropy-based inference.* The second motivation appears when we adhere to the principles stated in chapter 4, where an inference process called $\mathcal{I}^E_{\text{ME}}$ is defined as the arguments of the maximisation of a *strictly concave* function E over $\hat{\Omega}_{C_K}$. One of these principles, namely $\mathbf{P}^{\mathcal{I}}_{\alpha}$, requires an inference process to return a unique probability distribution: principle $\mathbf{P}^{\mathcal{I}}_{\alpha}$ is thus satisfied by $\mathcal{I}^E_{\text{ME}}$ if $\hat{\Omega}_{C_K}$ is convex. For example, applying $\mathcal{I}^E_{\text{ME}}$ to the linear knowledge base K described in Fig 2.2 would return one probability distribution (namely $[0.4; 0.6]$ for $C^{h_1}_K$ or $C^{h_2}_K$, and $[0.5; 0.5]$ for $C^{h_5}_K$), but two distinct probability distributions for $C^{h_3}_K$ and $C^{h_4}_K$. The reason is that only $C^{h_1}_K$, $C^{h_2}_K$, and $C^{h_5}_K$ are based on log-concave function, which implies their unimodality. We recall that a function h is log-concave on a domain $\mathbb{D}$ iff $\forall x, y \in \mathbb{D}, h(\lambda x + (1-\lambda)y) \geq h(x)^\lambda * h(y)^{1-\lambda}$, where λ runs through $[0{:}1]$.

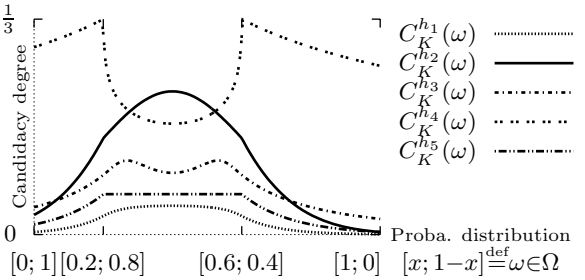


Figure 2.2: Five candidacy functions $C^{h_i}_K$ corresponding to the inconsistent knowledge base K made of the following two linear constraints: "$x \leq 0.2$" and "$0.6 \geq x$", where $\alpha_{\Theta(K)} \stackrel{\text{def}}{=} \{v, \neg v\}$ and $x \stackrel{\text{def}}{=} \omega(v)$. Although the underlying function h_i of each $C^{h_i}_K$ satisfies all the principles stated before *convexity* (see $\mathbf{P}^{\mathbb{C}}_{10}$ and Fig. 2.3), only $C^{h_1}_K$, $C^{h_2}_K$, and $C^{h_5}_K$ are unimodal. This unimodality comes from the log-concavity of h_i, which implies the convexity of $\hat{\Omega}_{C^{h_i}_K}$ (see Prop 5). Functions $C^{h_i}_K$ are such that $h_1(x) \stackrel{\text{def}}{=} \sin(\frac{\pi}{2}e^{-x})$, $h_2(x) \stackrel{\text{def}}{=} h^{0.5}_{\text{HG}}(x)$, which is the half-Gaussian cumulative distribution function (see Def. 16 on page 14), $h_3(x) \stackrel{\text{def}}{=} \sin(\frac{\pi}{2(1+x^2)})$, $h_4(x) \stackrel{\text{def}}{=} 1 - \frac{\sqrt{x}}{1+\sqrt{x}}$, and $h_5(x) \stackrel{\text{def}}{=} h^{0.5}_{\text{exp}}(x)$, which is the exponential blur (see Def. 15 on page 14). For graphical reasons, functions $C^{h_i}_K$ are plotted with this knowledge base: $\{$"$x \leq 2$", "$6 \geq x$"$\}$.

Proposition 5. *Principle $\mathbf{P}^{\mathbb{C}}_{10}$ is satisfied iff h is log-concave, ie $\hat{\Omega}_{C_K}$ is convex iff h is log-concave, for any linear knowledge base K.*

Proof. Before proving that $\hat{\Omega}_{C_K}$ is convex iff C_K is log-concave, we prove that C_K is log-concave iff h is log-concave. A constant function, a multiplication of two log-concave functions, and the cumulative distribution function of a Gaussian density function are log-concave (see [6, section 3.5] and [2]). Since K is linear, each knowledge item $\mathfrak{k}$ in $\mathfrak{K}_K$ is a halfspace. Function $\mathcal{G}(\omega, \mathfrak{k})$ is thus convex when ω varies and $\mathfrak{k}$ is set. Hence, $h(\sqrt{2^n} * \mathcal{G}(\omega, \mathfrak{k}))$ is log-concave iff h is log-concave and decreasing. By defining $C_{\mathfrak{k}}$ accordingly to principles stated in §2.3.5, $C_{\mathfrak{k}}(\omega) \stackrel{\text{def}}{=} h(\sqrt{2^n} * \mathcal{G}(\omega, \mathfrak{k}))$ is log-concave

iff h is log-concave. Since $C_K \stackrel{\text{def}}{=} \Cup_{\mathfrak{k}\in\mathfrak{K}_K} C_{\mathfrak{k}} = \prod_{\mathfrak{k}\in\mathfrak{K}_K} C_{\mathfrak{k}}$, the candidacy function C_K corresponding to a linear knowledge base K is log-concave iff h is log-concave.

We now prove that $\hat{\Omega}_{C_K}$ is convex iff C_K is log-concave. Let λ run through [0:1]. Let x and y be two probability distributions belonging to $\hat{\Omega}_{C_K}$, and let $z \stackrel{\text{def}}{=} \lambda x + (1-\lambda)y$ be a probability distribution. Thus, $\hat{\Omega}_{C_K}$ is convex iff $z \in \hat{\Omega}_{C_K}$ iff $C_K(z) = C_K(x)$ iff $C_K(z) \leq C_K(x)$ and $C_K(z) \geq C_K(x)$. We have $C_K(z) \leq C_K(x)$ since x maximises C_K by definition. Also, notice that $C_K(x) = C_K(y)$ since both x and y maximises C_K by definition; hence $C_K(x) = C_K(x)^\lambda * C_K(y)^{1-\lambda}$. Therefore, $C_K(z) \geq C_K(x)$ iff $C_K(\lambda x + (1-\lambda)y) \geq C_K(x)^\lambda * C_K(y)^{1-\lambda}$ by definition of z, iff C_K is log-concave. □

Definition 13 (Log-concave blur)**.** *The candidacy function C_K corresponding to a knowledge base K is defined as follows:*

$$C_K \stackrel{\text{def}}{=} \mathcal{B}(K) \stackrel{\text{def}}{=} C_{\mathfrak{K}_K}$$

$$C_{\mathfrak{K}} \stackrel{\text{def}}{=} \Cup_{\mathfrak{k}\in\mathfrak{K}} C_{\mathfrak{k}} \qquad C_{\mathfrak{k}}(\omega) \stackrel{\text{def}}{=} h(\sqrt{2^n} * \mathcal{G}(\omega, \mathfrak{k}))$$

where $h : \mathbb{R}^+ \mapsto]0{:}1]$ is a strictly decreasing, positive, and continuous log-concave function such that $h(0) \stackrel{\text{def}}{=} 1$.

Proposition 6. *The log-concave blur satisfies principles $\mathbf{P}_6^{\mathbb{C}}$ to $\mathbf{P}_{10}^{\mathbb{C}}$.*

Proof. See the previous discussion in §2.3.5 where we constructively defined the log-concave blur in order to satisfy principles $\mathbf{P}_6^{\mathbb{C}}$ to $\mathbf{P}_{10}^{\mathbb{C}}$.

□

On considering reliability levels

All the results of this thesis hold if h is log-concave (eg, see $h(x) \stackrel{\text{def}}{=} \sin(\frac{\pi}{2}e^x)$ in Fig. 2.3). Suppose each knowledge item $\mathfrak{k} \in \mathfrak{K}_K$ of a knowledge base K is given a reliability level $\sigma_{\mathfrak{k}} \in [0{:}1]$, which tends towards 1 as $\mathfrak{k}$ deems reliable; $\{\, \sigma_{\mathfrak{k}} \mid \mathfrak{k} \in \mathfrak{K}_K \,\}$ may represent the reliability of sensory data K, or the credence given to an agent's knowledge base K. Let K^σ be the knowledge base K such that $\forall \mathfrak{k} \in \mathfrak{K}_K, \sigma = \sigma_{\mathfrak{k}}$; we denote its corresponding candidacy function by both C_{K^σ} and C_K^σ.

Definition 14 (Blur wrt reliability levels)**.** *This definition extends Def. 13 by considering reliability levels. The candidacy function $C_{\mathfrak{k}}^\sigma$ corresponding to a knowledge item $\mathfrak{k} \in \mathfrak{K}$ with reliability level σ is defined as follows (where $h^\sigma : \mathbb{R}^+ \mapsto]0{:}1]$ is a strictly decreasing, positive, and continuous log-concave function such that $h^\sigma(0) \stackrel{\text{def}}{=} 1$):*

$$C_{\mathfrak{k}}^\sigma(\omega) \stackrel{\text{def}}{=} \begin{cases} h^\sigma(\sqrt{2^n} * \mathcal{G}(\omega, \mathfrak{k})) & \text{if } \sigma \in]0{:}1[, \\ 1 & \text{if } \sigma = 1 \text{ and } \mathcal{G}(\omega, \mathfrak{k}) = 0, \\ 0 & \text{if } \sigma = 1 \text{ and } \mathcal{G}(\omega, \mathfrak{k}) > 0, \\ 1 & \text{if } \sigma = 0. \end{cases}$$

The candidacy function C_K^σ corresponding to a knowledge base K with reliability level $\sigma \in [0{:}1]$ is defined as follows:

$$C_K^\sigma \stackrel{\text{def}}{=} C_{\mathfrak{K}_K}^\sigma \qquad C_{\mathfrak{K}}^\sigma \stackrel{\text{def}}{=} \Cup_{\mathfrak{k}\in\mathfrak{K}} C_{\mathfrak{k}}^\sigma$$

We defined $C_{\mathfrak{k}}^\sigma$ as 0 when $\sigma = 1$ in order to allow the specification of hard constraints in knowledge bases. For example, if the knowledge item $\mathfrak{k}_{\{c\}}$ corresponding to a linear constraint c is considered as reliable, then $C_{\mathfrak{k}_c}^1(\omega)$ equals 1 iff ω satisfies c, 0 otherwise. $C_{\mathfrak{k}_c}^1$ is then a step function: in Fig. 2.3, h_{exp}^σ tends towards a step function when the reliability level σ tends towards 1. Thus, except for defining *consistent* hard constraints, the reasonable values for σ are [0:1[.

In order to complete Def. 14, we shall suggest two definitions for h^σ: the exponential blur (see Def. 15) and the half-Gaussian blur (Def. 16). Before, we state a principle concerning flat knowledge contents, ie knowledge contents of which all the knowledge items are equally reliable. We then construct a blur function (see Def. 15) satisfying it, but behaving in an extreme manner: if an item of knowledge $\mathfrak{k}_1$ contradicts another one $\mathfrak{k}_2$, and if $\mathfrak{k}_1$ is strictly more reliable that $\mathfrak{k}_2$, then the candidacy function $C_{\{\mathfrak{k}_1,\mathfrak{k}_2\}}$ is maximal for a probability distribution satisfying $\mathfrak{k}_1$ (hence not satisfying $\mathfrak{k}_2$). We then propose a more conciliatory blur function (see Def. 16 on page 14): in which case, $C_{\{\mathfrak{k}_1,\mathfrak{k}_2\}}$ is maximal for a probability distribution satisfying neither $\mathfrak{k}_1$ nor $\mathfrak{k}_2$, but being nearer to satisfy $\mathfrak{k}_1$ than to satisfy $\mathfrak{k}_2$. However, this more conciliatory blur function fails to satisfy the following principle.

$\mathbf{P}_{11}^{\mathbb{C}}$ *Reliability invariance.* The set of best candidates $\hat{\Omega}_{C_{K^\sigma}}$ should be σ-invariant, if the common reliability level σ of all the knowledge items of $\mathfrak{K}_K$ varies in]0:1[.

$$\hat{\Omega}_{C_{K^{\sigma_1}}} = \hat{\Omega}_{C_{K^{\sigma_2}}}, \text{ for all } \sigma_1, \sigma_2 \in]0{:}1[$$

Adhering to *reliability invariance* constrains the possible definitions of h^σ as follows. Let $\sigma, \sigma_1, \sigma_2 \in]0{:}1[$ be three reliability levels, and let $a_1, a_2 \in]0{:}{+}\infty[$ respectively represent a non-normalised version of σ_1 and σ_2. Let $h^\sigma : \mathbb{R}^+ \mapsto [0{:}1]$ be a strictly decreasing and smooth function; hence h^σ is invertible. Let $\text{inv}(f)$ be the inverse function of an invertible function f. We denote by $C_K^{h^\sigma}$ the candidacy function of K when K is blurred with h^σ.

Proposition 7. *Principle* $\mathbf{P}_{11}^{\mathbb{C}}$ *holds if*

$$\left(h^{\sigma_1} \circ \mathrm{inv}(h^{\sigma_2})\right)(x) = x^a$$

where a *is a constant, which only depends on the given reliability levels* σ_1 *and* σ_2*.*

Proof. $\mathbf{P}_{11}^{\mathbb{C}}$ is satisfied iff

$$\hat{\Omega}_{C_K^{h^{\sigma_1}}} = \hat{\Omega}_{C_K^{h^{\sigma_2}}}$$

iff

$$\arg\max_{\omega\in\Omega} C_K^{h^{\sigma_1}}(\omega) = \arg\max_{\omega\in\Omega} C_K^{h^{\sigma_2}}(\omega)$$

if

$$C_K^{h^{\sigma_1}} = f \circ C_K^{h^{\sigma_2}}$$

where $f : [0{:}1] \mapsto \mathbb{R}$ is a strictly increasing function, iff

$$\prod_{\mathfrak{k}\in\mathfrak{K}_K} C_{\mathfrak{k}}^{h^{\sigma_1}}(\omega) = f\Big(\prod_{\mathfrak{k}\in\mathfrak{K}_K} C_{\mathfrak{k}}^{h^{\sigma_2}}(\omega)\Big)$$

for each probability function $\omega \in \Omega$, iff

$$\prod_{\mathfrak{k}\in\mathfrak{K}_K} h^{\sigma_1}(\sqrt{2^n} * \mathcal{G}(\omega,\mathfrak{k})) = f\Big(\prod_{\mathfrak{k}\in\mathfrak{K}_K} h^{\sigma_2}(\sqrt{2^n} * \mathcal{G}(\omega,\mathfrak{k}))\Big)$$

iff

$$\prod_{\mathfrak{k}\in\mathfrak{K}_K} g(z_{\mathfrak{k}}) = f\Big(\prod_{\mathfrak{k}\in\mathfrak{K}_K} z_{\mathfrak{k}}\Big) \tag{2.1}$$

where $z_{\mathfrak{k}} \stackrel{\text{def}}{=} h^{\sigma_2}(\sqrt{2^n} * \mathcal{G}(\omega,\mathfrak{k}))$ is in $[0{:}1]$, and $g \stackrel{\text{def}}{=} h^{\sigma_1} \circ \mathrm{inv}(h^{\sigma_2})$; hence, $g : [0{:}1] \mapsto [0{:}1]$ is a strictly increasing function such that $g \circ h^{\sigma_2} = h^{\sigma_1}$. In case $\mathfrak{K}_K$ contains one knowledge item $\mathfrak{k}$, (2.1) requires f and g to satisfy $g(z_{\mathfrak{k}}) = f(z_{\mathfrak{k}})$, or equivalently if $x \in [0{:}1]$

$$f(x) = g(x) \tag{2.2}$$

In case $\mathfrak{K}_K$ contains two knowledge items $\mathfrak{k}_1$ and $\mathfrak{k}_2$, (2.1) together with (2.2) requires f (hence g) to satisfy the following equality, where $x \stackrel{\text{def}}{=} z_{\mathfrak{k}_1}$ and $y \stackrel{\text{def}}{=} z_{\mathfrak{k}_2}$ are in $[0{:}1]$.

$$f(x) * f(y) = f(x * y) \tag{2.3}$$

Suppose that f has a derivative denoted by f' and defined on $[0{:}1]$. By differentiating both sides of equality (2.3) wrt x then y, we obtain that f must satisfy

$$\begin{cases} f'(x) * f(y) = f'(x * y) * y \\ f(x) * f'(y) = f'(x * y) * x \end{cases}$$

By multiplying the first line by x and the second line by y, we have

$$x * \frac{f'(x)}{f(x)} = y * \frac{f'(y)}{f(y)} \tag{2.4}$$

if $\forall x \in [0{:}1], f(x) \neq 0$. Since equality (2.4) must hold for all $x, y \in [0{:}1]$, $a \stackrel{\text{def}}{=} x * \frac{f'(x)}{f(x)}$ is a constant real number. After integrating both sides of $\frac{a}{x} = \frac{f'(x)}{f(x)}$ wrt x with $x \neq 0$, we obtain $a * \ln(x) + \lambda_1 = \ln(f(x)) + \lambda_2$, where λ_1 and λ_2 are two constant real numbers. Hence, f must satisfy $f(x) = \exp(a * \ln(x) + \lambda_1 - \lambda_2)$, which is equivalent to $f(x) = b * x^a$ with $b \stackrel{\text{def}}{=} \exp(\lambda_1 - \lambda_2)$. Furthermore, equality (2.3) requires $bx^a * by^a = b(xy)^a$, which implies $b = 1$. Therefore, f must satisfy

$$f(x) = x^a$$

which implies, by definition of g and (2.2), that h^{σ_1} and h^{σ_2} must satisfy

$$\left(h^{\sigma_1} \circ \mathrm{inv}(h^{\sigma_2})\right)(x) = x^a \tag{2.5}$$

□

We can rewrite equation (2.5) as $\exp(\mathbf{a}(\sigma_1) * \frac{\ln(x)}{\mathbf{a}(\sigma_2)}) \stackrel{\text{def}}{=} x^a = \left(h^{\sigma_1} \circ \mathrm{inv}(h^{\sigma_2})\right)(x)$ where $a = \frac{\mathbf{a}(\sigma_1)}{\mathbf{a}(\sigma_2)}$ is the ratio of $\mathbf{a}(\sigma_1)$ to $\mathbf{a}(\sigma_1)$, and where $\mathbf{a} :]0{:}1[\mapsto]-\infty{:}0[$ is a decreasing bijection that scales a given reliability level. Notice that the inverse function of $\exp(d * x)$ is $\frac{\ln(x)}{d}$, where $d \in \mathbb{R} \setminus \{0\}$. We thus suggest to define h^{σ} as $\exp(\mathbf{a}(\sigma)) * x)$, where $\mathbf{a}(\sigma)$ could be defined as $\frac{\sigma}{\sigma-1}$ or $\ln(1-\sigma)$ for examples. To constrain the possible definitions for $\mathbf{a}(\sigma)$, we adhere to the following principle.

$\mathbf{P}_{12}^{\mathbb{C}}$ *Reliability reinforcement.* Increasing the redundancy reinforces the reliability. More specifically, merging several candidacy functions of the same knowledge content yields one candidacy function corresponding to that knowledge content with higher reliability. Formally, it must exist a strictly increasing and symmetric aggregation function $F :]0{:}1[^m \mapsto]0{:}1[$ such that the following equality holds:

$$\Cup_{i=1}^{m\in\mathbb{N}} C_{\mathfrak{K}}^{\sigma_i} = C_{\mathfrak{K}}^{F(\sigma_1,\sigma_2,\ldots,\sigma_m)}$$

Proposition 8. *If* $h^{\sigma}(x)$ *is defined as* $\exp(\mathbf{a}(\sigma) * x)$*, then principle* $\mathbf{P}_{12}^{\mathbb{C}}$ *implies*

$$\mathbf{a}(F(\sigma_1,\sigma_2,\ldots,\sigma_m)) = \sum_{i=1}^{m} \mathbf{a}(\sigma_i)$$

where $F :]0{:}1[^m \mapsto]0{:}1[$ *and* $\mathbf{a} :]0{:}1[\mapsto]-\infty{:}0[$ *respectively aggregate and scale a set of reliability levels* σ_i*.*

Proof. Principle $\mathbf{P}_{12}^{\mathbb{C}}$ states that $\Cup_{i=1}^{m} C_{\mathfrak{K}}^{\sigma_i} = C_{\mathfrak{K}}^{F(\sigma_1,\sigma_2,\ldots,\sigma_m)}$, which is equivalent to $\prod_{i=1}^{m}\prod_{\mathfrak{k}\in\mathfrak{K}} C_{\mathfrak{k}}^{\sigma_i}(\omega) = \prod_{\mathfrak{k}\in\mathfrak{K}} C_{\mathfrak{k}}^{F(\sigma_1,\sigma_2,\ldots,\sigma_m)}(\omega)$ for any probability distribution $\omega \in \Omega$, and to $\prod_{\mathfrak{k}\in\mathfrak{K}}\prod_{i=1}^{m} h^{\sigma_i}(x_{\mathfrak{k}}) = \prod_{\mathfrak{k}\in\mathfrak{K}} h^{F(\sigma_1,\sigma_2,\ldots,\sigma_m)}(x_{\mathfrak{k}})$ for any $x_{\mathfrak{k}} \stackrel{\text{def}}{=} \sqrt{2^n} * \mathcal{G}(\omega,\mathfrak{k})$ in $]0{:}+\infty[$, which implies $\prod_{i=1}^{m} h^{\sigma_i}(x) = h^{F(\sigma_1,\sigma_2,\ldots,\sigma_m)}(x)$ for any x in $]0{:}+\infty[$. Suppose $h^{\sigma}(x)$ defined as $\exp(\mathbf{a}(\sigma) * x)$.

Principle $\mathbf{P}_{12}^{\mathbb{C}}$ thus implies $\prod_{i=1}^{m} \exp(\mathbf{a}(\sigma_i) * x) = \exp(\mathbf{a}(F(\sigma_1, \sigma_2, \ldots, \sigma_m)) * x)$, hence $\exp(x * \sum_{i=1}^{m} \mathbf{a}(\sigma_i)) = \exp(x * \mathbf{a}(F(\sigma_1, \sigma_2, \ldots, \sigma_m)))$, from where we conclude that $\sum_{i=1}^{m} \mathbf{a}(\sigma_i) = \mathbf{a}(F(\sigma_1, \sigma_2, \ldots, \sigma_m))$. □

After having assumed $h^{\sigma}(x)$ defined as $\exp(\mathbf{a}(\sigma) * x)$, we now suggest to define $\mathbf{a}(\sigma)$ as $\ln(1-\sigma)$; we thus suggest to define h^{σ} as $\exp(\ln(1-\sigma) * x)$, or equivalently, as $(1-\sigma)^x$. With such a definition for h^{σ}, the function that aggregates reliability levels in principle $\mathbf{P}_{12}^{\mathbb{C}}$ equals $F(\sigma_1, \sigma_2, \ldots, \sigma_m) = 1 - \exp(\sum_{i=1}^{m} \ln(1-\sigma_i)) = 1 - \prod_{i=1}^{m} 1 - \sigma_i$. Thus, the candidacy function corresponding to a knowledge content $\mathfrak{K}$ reliable up to a non-normalised level $m * \ln(1-\sigma)$ is equal to the mergence of m candidacy function of $\mathfrak{K}$ reliable up to $\ln(1-\sigma)$; roughly, the non-normalised reliability level of m knowledge sources that are reliable up to $\ln(1-\sigma)$ equals $m * \ln(1-\sigma)$.

Definition 15 (Exponential blur)**.** *This definition completes Def. 14.*

$$h_{\exp}^{\sigma}(x) \stackrel{\text{def}}{=} (1-\sigma)^x$$

Proposition 9. *The exponential blur satisfies principles* $\mathbf{P}_6^{\mathbb{C}}$ *to* $\mathbf{P}_{12}^{\mathbb{C}}$

Proof. Since exp is a log-concave function, the exponential blur is a log-concave blur, which satisfies principles $\mathbf{P}_6^{\mathbb{C}}$ to $\mathbf{P}_{10}^{\mathbb{C}}$ by Prop. 6. Furthermore, principle $\mathbf{P}_{11}^{\mathbb{C}}$ is satisfied if $\left(h^{\sigma_1} \circ \text{inv}(h^{\sigma_2})\right)(x) = x^a$ (see Prop. 7), which holds when $h^{\sigma_1}(x) \stackrel{\text{def}}{=} \exp(\mathbf{a}(\sigma_1) * x)$ and $\text{inv}(h^{\sigma_2})(x) = \frac{\ln(x)}{\mathbf{a}(\sigma_2)}$ and $a = \frac{\mathbf{a}(\sigma_1)}{\mathbf{a}(\sigma_2)}$. Besides, principle $\mathbf{P}_{12}^{\mathbb{C}}$ is satisfied since $\mathbb{U}_{i=1}^{m} C_{\mathfrak{K}}^{\sigma_i}$ equals $\prod_{i=1}^{m} \prod_{\mathfrak{k} \in \mathfrak{K}} h^{\sigma_i}(x_{\mathfrak{k}})$ where $x_{\mathfrak{k}} \stackrel{\text{def}}{=} \sqrt{2^n} * \mathcal{G}(\omega, \mathfrak{k})$, equals $\prod_{\mathfrak{k} \in \mathfrak{K}}^{x_{\mathfrak{k}}} \left(\prod_{i=1}^{m} (1-\sigma_i)\right)^{x_{\mathfrak{k}}}$ equals $\prod_{\mathfrak{k} \in \mathfrak{K}} (1 - F(\sigma_1, \sigma_2, \ldots, \sigma_m))^{x_{\mathfrak{k}}}$ equals $\prod_{\mathfrak{k} \in \mathfrak{K}} C_{\mathfrak{k}}^{F(\sigma_1, \sigma_2, \ldots, \sigma_m)}$ equals $C_{\mathfrak{K}}^{F(\sigma_1, \sigma_2, \ldots, \sigma_m)}$. □

Throughout this thesis, we employ the exponential blur. Due to the *reliability invariance* (also called σ-invariance) of the best candidates, we simply denote by C_K the candidacy function C_K^{σ} corresponding to a knowledge base K reliable up to a level $\sigma \in]0{:}1[$.

On assuming h is half-Gaussian

Definition 16 (Half-Gaussian blur)**.** *This definition completes Def. 14. We here suppose that a reliability level* $\sigma \in]0{:}1[$ *can be encoded as the standard deviation* $-4\ln(\sigma_{\mathfrak{k}})$ *of a half-Gaussian cumulative distribution function* h_{HG}^{σ} *defined as*

$$h_{\text{HG}}^{\sigma}(x) \stackrel{\text{def}}{=} 1 + \text{erf}\left(\frac{-x}{-4\ln(\sigma_{\mathfrak{k}}) * \sqrt{2}}\right)$$

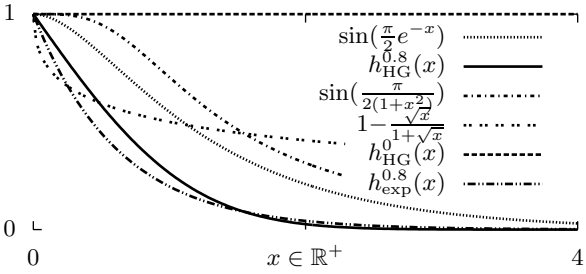


Figure 2.3: Potential definitions for h (the first two are log-concave contrary to the next two). $h_{\text{HG}}^{\sigma_{\mathfrak{k}} \in]0:1]}$ is the half-Gaussian cumulative distribution function with $-4\ln(\sigma_{\mathfrak{k}})$ as standard deviation. If $x \stackrel{\text{def}}{=} \sqrt{2} * \mathcal{G}(\omega, \mathfrak{k})$ is the gap between a probability distribution ω and a knowledge item $\mathfrak{k}$, then these functions are intended to return the candidacy degree of ω to represent the real world, wrt $\mathfrak{k}$ and a reliability level $\sigma_{\mathfrak{k}}$.

where the error function is defined as

$$\text{erf}(x) \stackrel{\text{def}}{=} \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

Graphs of function $h_{\text{HG}}^{\sigma_{\mathfrak{k}}}$ are drawn in Fig. 2.3 for several values of reliability level $\sigma_{\mathfrak{k}}$, ie for several values of standard deviation $-4\ln(\sigma_{\mathfrak{k}})$. Notice that a knowledge item $\mathfrak{k}$ being unreliable, ie having a reliability level $\sigma_{\mathfrak{k}}$ equal to 0, is considered as void of knowledge content, because $h_{\text{HG}}^{0} = 1$ hence $C_{\mathfrak{k}}$ is a tautological candidacy function $1_{\mathbb{C}}$. If $\mathfrak{k}$ is reliable, ie $\sigma_{\mathfrak{k}} = 1$, then $C_{\mathfrak{k}}$ could be an absorbing candidacy function $0_{\mathbb{C}}$, which violates *proximity* (see $\mathbf{P}_7^{\mathbb{C}}$).

We now sketch a tentative argument for defining h_{HG} as a half-Gaussian cumulative distribution. Let c be a linear constraint of the form "$b \geq [a_1, a_2, \ldots, a_J] * x$" where x runs through $\mathbb{R}^J$. Let us consider c as a random polynomial inequality $\mathfrak{P}_c \stackrel{\text{def}}{=}$ "$b + \epsilon_b \geq ([a_1, a_2, \ldots, a_J] + \epsilon_A) * x$", where ϵ_b is a real random variable (with expected value 0) having probability density functions PDF_b, and where $\epsilon_A \stackrel{\text{def}}{=} [\epsilon_{a_1}, \epsilon_{a_2}, \ldots, \epsilon_{a_J}]$ is a vector of real random variables (with expected value 0) having probability density functions PDF_{a_j}. Thus, $\mathfrak{P}_c(x)$ is intended to return the probability that x satisfies c, wrt PDF_b and PDF_{a_j}. Before defining $C_{\mathfrak{k}_c}(x)$ according to $\mathfrak{P}_c(x)$, we now explain why only ϵ_b should be considered as random contrarily to ϵ_{a_j}, which should be set to 0.

$\mathbf{P}_{13}^{\mathbb{C}}$ *Characterisation.* A candidacy function $C_{\mathfrak{k}_c}$ should characterise its corresponding linear constraint c, which is identified with the solutions Sol_c of c.

$$Sol_c = \mathfrak{k}_c = \left\{ x \in \mathbb{R}^J \;\middle|\; C_{\mathfrak{k}_c}(x) = 1 \right\}$$

Such a principle is a strengthened version of $\mathbf{P}_8^{\mathbb{C}}$ and assumes ⑤. Under assumption ⑤, for each constraint c in a knowledge base K, $\mathfrak{K}_K$ contains a knowledge item $\mathfrak{k}_c$ that is the non-empty set of points $Sol_c \subset \mathbb{R}^J$ satisfying c. Since c is a linear constraint, Sol_c characterises c, hence $\mathfrak{k}_c$ characterises c. $\mathbf{P}_{13}^{\mathbb{C}}$ means that $x \in Sol_c$ iff $C_{\mathfrak{k}_c}(x) = 1$, ie point x satisfies linear constraint c iff $\mathfrak{P}_c(x) = 1$. It therefore states that $x \in Sol_c$ should satisfy "$b+\epsilon_b \geq ([a_1, a_2, \ldots, a_J]+\epsilon_A)*x$". However, if $\epsilon_A \neq \vec{0}$ or if $\epsilon_b < 0$, then there exists $x \in Sol_c$ that does not satisfy the latter inequality. Roughly, the hyperplane separating Sol_c from $\mathbb{R} \setminus Sol_c$ can be randomly translated in one direction (ie ϵ_b is a random variable with a *half* distribution) but not tilted (ie ϵ_{a_j} must be set to 0). If we further require PDF_b to have a Gaussian-like shape, then PDF_b must be the half-Gaussian probability density function. Thus, the (language invariant) probability that $x \in \mathbb{R}^J$ satisfies c is given by $h^{\sigma}_{\mathrm{HG}}(\sqrt{2^n} * \mathcal{G}(\omega, Sol_c))$, wrt standard deviation $-4\ln(\sigma_{\mathfrak{k}})$. Criticisms against the previous argumentation (for assuming h is half-Gaussian) may be raised. For example, assuming PDF_b to have a Gaussian-like shape might be irrelevant since the random value of ϵ_b might not range over the whole set $\mathbb{R}$; rather, the random value x of ϵ_b might range over a subset $S \subset \mathbb{R}$ such that the hyperplane separating Sol_c from $\mathbb{R} \setminus Sol_c$ is randomly translated by $x \in S$ and intersects the set of probability distributions Ω.

2.3.6 Comparing knowledge at *pignistic* level: the external equivalence $\overset{\mathrm{e}}{\equiv}$

C_K being defined, we are now ready to propose a new assumption about the knowledge content of a knowledge base K; other assumptions are stated in § 2.3.2 on page 8. The following assumption is intended to define the *external* level of knowledge content.

⑥ $\mathfrak{K}_K \overset{\mathrm{def}}{=} \{\hat{\Omega}_{C_K}\}$. To accept this assumption, which extends ① to inconsistent knowledge bases, is to consider the set of best candidates for representing the real world as the only relevant knowledge.

Definition 17 (External equivalence)**.** *Two knowledge bases K_1 and K_2 are externally equivalent iff they have the same knowledge content wrt assumption ⑥.*

$$K_1 \overset{\mathrm{e}}{\equiv} K_2 \overset{\mathrm{def}}{=} \begin{cases} \left(\mathfrak{K}^{⑥}_{K_1} = \mathfrak{K}^{⑥}_{K_2}\right) & \text{if } \Theta(K_1) = \Theta(K_2), \\ \mathbf{t}\text{rue} & \text{if } \exists v, K_1 = K_2 \oplus v, \\ \mathbf{f}\text{alse} & \text{otherwise.} \end{cases}$$

Two candidacy functions C_1 and C_2 are externally equivalent iff they have the same best candidates.

$$C_1 \overset{\mathrm{e}}{\equiv} C_2 \overset{\mathrm{def}}{=} \begin{cases} \left(\hat{\Omega}_{C_1} = \hat{\Omega}_{C_2}\right) & \text{if } \Theta(C_1) = \Theta(C_2), \\ \mathbf{t}\text{rue} & \text{if } \exists v, C_1 = C_2 \oplus v, \\ \mathbf{f}\text{alse} & \text{otherwise.} \end{cases}$$

Remark that two knowledge bases K_1 and K_2 are externally equivalent iff their corresponding candidacy functions are externally equivalent, ie $K_1 \overset{\mathrm{e}}{\equiv} K_2$ iff $C_{K_1} \overset{\mathrm{e}}{\equiv} C_{K_2}$.

2.4 Conclusions and perspectives

In this chapter, we extend several notions defined in [33, 38] in order to deal with inconsistent knowledge bases. According to definitions 3 and 4, our knowledge bases are multisets (instead of sets) of inequalities (instead of equalities) having a general form (instead of a linear form or a polynomial form, although results in [38] seem not limited to polynomial equalities). We then introduce our new probabilistic knowledge representation as a function returning for each probability distribution its candidacy degree for representing the real world. However, knowledge is more naturally expressible through a knowledge base than through a candidacy function. In section 2.3, we thus exhibit the construction of the candidacy function C_K corresponding to a given knowledge base K; if we dare draw an analogy with fuzzy set theory, we would say that the knowledge base is to the crisp set what the candidacy function is to the fuzzy set.

Candidacy functions are fundamental objects for paraconsistent probabilistic reasoning: inconsistency does not exist, though the candidacy function $0_{\mathbb{C}}$ constantly equal to zero might express that *There is no real world*[6]. We eschew such absorbing candidacy functions by requiring the construction of C_K to follow several principles. The three key principles are *proximity*, *language invariance*, and *reliability invariance*. *Proximity* (see $\mathbf{P}_7^{\mathbb{C}}$) states that a probability distribution close to satisfy a knowledge base should be close to representing the real world; by adhering to this principle, we avoid C_K being absorbing, even partially absorbing (see §2.3.5). *Language invariance* (see $\mathbf{P}_9^{\mathbb{C}}$) requires a candidacy function to be invariant by language enrichment: roughly, this means that when a new word appears in a language, then the knowledge previously expressed in terms of this language remains unchanged.

[6] $0_{\mathbb{C}}$ resembles a contradiction: if $0_{\mathbb{C}}$ means that *There is no real world*, then no probability distribution should be nominated for representing the real world. However, the best candidates are all the probability distributions: $\Omega = \hat{\Omega}_{0_{\mathbb{C}}}$ (this resembles *explosion* in classical logic). Hence, an inference process based on $\hat{\Omega}_C$ interprets $0_{\mathbb{C}}$ as a tautology.

Reliability invariance (see $\mathbf{P}_{11}^{\mathbb{C}}$) ensures that the nomination of the best candidates is independent on the global reliability level $\sigma \in]0{:}1[$ given to a knowledge base K, ie, if each knowledge item of $\mathfrak{K}_K$ are reliable to the same level σ, then the best candidates $\hat{\Omega}_{C_K^\sigma}$ are invariant when only σ varies in $]0{:}1[$; we thus denote C_K^σ by C_K. In the motivating example about voting theory (see section 2.1), such reliability levels may capture the society's confidence in its individuals, who have to elect an investment distribution; in which case, the reliability levels could depend on the individuals' investment skills. If the society refuses the use of reliability levels to elect the investment distribution, it thus suffices to set all these levels to a common value in $]0{:}1[$.

In this chapter, we show that the candidacy function C_K corresponding to a knowledge base K can be construct while satisfying all our principles (see Prop. 9). However, further investigations into the set of principles are needed to characterise the construction of C_K.

Chapter 3

Four measures to appraise knowledge

> You can't control what you can't measure.
>
> Tom DeMarco, in [7, page 6]

Having defined our general probabilistic knowledge representation, ie the candidacy functions $\mathbb{C}$ (see Def. 9 on page 5), we now desire to exploit its expressiveness in order to draw paraconsistent inferences (see chapter 4). But before, in the light of Tom DeMarco's statement, this chapter introduces several measures enabling us to discuss about candidacy functions. These principled measures quantify important notions such as the dissimilarity, the inconsistency, the incoherence, the surprise, the precision, and the confidence.

3.1 Common definitions

The reader may skip this section, which presents three entailment relations used to express several principles, then introduces a geometric notion employed in the definitions of the culpability measure (see Def. 28) and the precision measure (see §3.5.3).

3.1.1 From tautological deductions to the explosion: a continuum of entailment relations $\models$

In this section, we define two entailment relations between a knowledge content $\mathfrak{K}$ and a knowledge item $\mathfrak{k}$, and an entailment relation between two candidacy functions C_1 and C_2. Informally, $\mathfrak{K} \models \mathfrak{k}$ means $\mathfrak{K}$ entails (or implies) $\mathfrak{k}$, or $\mathfrak{k}$ is a consequence of (or deducible from) $\mathfrak{K}$. We suppose $\Theta(\mathfrak{K}) = \Theta(\mathfrak{k}) = \Theta(C_1) = \Theta(C_2)$. There is a continuum (and a complete partial ordering) of entailments for a given knowledge content: the strongest (or more precautionary) is $\mathfrak{K} \models_{\text{s}} \mathfrak{k}$ iff $\mathfrak{k} \supseteq \Omega$ (ie, $\mathfrak{K}$ entails only tautologies), and the weakest is $\mathfrak{K} \models_{\text{w}} \mathfrak{k}$ iff $\mathfrak{k} \supseteq \emptyset$ (ie, $\mathfrak{K}$ entails every knowledge items: the explosion). We now suggest several entailment relations weaker than $\models_{\text{s}}$ but stronger than $\models_{\text{w}}$.

Definition 18 (Inevitable consequences)**.** *A knowledge item $\mathfrak{k}$ is an inevitable consequence of a knowledge content $\mathfrak{K}$, denoted by $\mathfrak{K} \models_{\text{ic}} \mathfrak{k}$, iff $\mathfrak{k}$ contains not only the best candidates of $C_{\mathfrak{K}}$, but also all the points belonging to the knowledge items of $\mathfrak{K}$: $\mathfrak{K} \models_{\text{ic}} \mathfrak{k}$ iff $\mathfrak{k} \supseteq \hat{\Omega}_{C_{\mathfrak{K}}} \cup \bigcup_{\mathfrak{t} \in \mathfrak{K}} \mathfrak{t}$.*

Definition 19 (Free formulae)**.** *A knowledge item $\mathfrak{k}$ is a free formula of $\mathfrak{K}$, denoted by $\mathfrak{K} \models_{\text{ff}} \mathfrak{k}$, iff $\mathfrak{k}$ contains not only the probability distributions belonging to every maximal consistent subset of $\mathfrak{K}$, but also the best candidates of $C_{\mathfrak{K}}$: $\mathfrak{K} \models_{\text{ff}} \mathfrak{k}$ iff $\mathfrak{k} \supseteq \hat{\Omega}_{C_{\mathfrak{K}}} \cup \bigcup_{Q \in \text{MCS}_{\mathfrak{K}}} \bigcap_{\mathfrak{t} \in Q} \mathfrak{t}$.*

The set $\{ \mathfrak{k} \in \mathfrak{K} \mid \mathfrak{K} \models_{\text{ff}} \mathfrak{k} \}$ resembles the *free formulae* defined the literature, which are the propositions belonging to all the maximal consistent subsets of a set of propositions (see [18, page 2] for a definition in terms of minimal inconsistent subsets). The free formulae are not culpable for making the knowledge content inconsistent. Furthermore, we denote by $\mathfrak{k}_1 \models \mathfrak{k}_2$ the entailment of a knowledge item $\mathfrak{k}_2$ by the knowledge content $\{\mathfrak{k}_1\}$: $\mathfrak{k}_1 \models \mathfrak{k}_2$ iff $\{\mathfrak{k}_1\} \models_{\text{ic}} \mathfrak{k}_2$ iff $\{\mathfrak{k}_1\} \models_{\text{ff}} \mathfrak{k}_2$ iff $\mathfrak{k}_2 \supseteq \hat{\Omega}_{C_{\mathfrak{k}_1}} \cup \mathfrak{k}_1$.

Definition 20. *A candidacy function C_1 entails another one C_2, denoted by $C_1 \models C_2$, iff $\forall \hat{\omega} \in \hat{\Omega}_{C_1}, C_2(\hat{\omega}) = 1$ and $\forall \omega \in \Omega, C_1(\omega) \leq C_2(\omega)$.*

Thus, ($C_1 \models C_2$ and $C_2 \models C_1$) iff ($C_1 = C_2$ and $\exists \hat{\omega} \in \Omega, C_1(\hat{\omega}) = 1$). Also, if $C_1 \models C_2$ then $C_1 \Cup C_2 \models C_2$ and $C_1 \overset{\text{e}}{\equiv} C_1 \Cup C_2$ and $\exists C_3 \in \mathbb{C}, C_1 = C_2 \Cup C_3$. Besides, if $\mathfrak{K} \models_{\text{ic}} \mathfrak{k}$ then $\mathfrak{K} \models_{\text{ff}} \mathfrak{k}$ and $C_{\mathfrak{K}} \models C_{\mathfrak{k}}$ (if the blur function used to obtain $C_{\mathfrak{K}}$ is the same as the one used to obtain $C_{\mathfrak{k}}$). When we write an expression containing an entailment relation which is independent on the choice between $\models_{\text{ic}}$ and $\models_{\text{ff}}$, we then simply denote this relation by $\models$.

3.1.2 A geometrical property for the best candidates: the solo-dimensionality of a manifold

In the section 3.5, a precision measure "counts" the best candidates of a given candidacy function. The set

of such best candidates forms a manifold in a Euclidean space of which we compute the Lebesgue measure.

Intuitively, a manifold M is solo-dimensional iff its Lebesgue measure takes into account each point of M (ie each best candidates). For example, a convex manifold is solo-dimensional. However, if a manifold $M \subset \mathbb{R}^3$ is the disjoint union of a cylinder with a square, then the Lebesgue measure in dimension 3 of the square is null. Therefore, the Lebesgue measure of M, which is the Lebesgue measure of the cylinder, does not take into account the points of the square. Thus, M is said to be *not* solo-dimensional.

Definition 21 (Solo-dimensional manifold)**.** *A manifold $M \subset \mathbb{R}^d$ in a Euclidean space of dimension $d \in \mathbb{N}$ is* solo-dimensional *iff M is Lebesgue measurable and, for each ball B centred on any point of M with any strictly positive radius, the Lebesgue measure of $M \cap B$ in dimension d is strictly positive; d is thus the "sole" dimension of M.*

3.2 Dissimilarity measure μ^{dis}: a metric on candidacy functions $\mathbb{C}$

Essentially similar problems should have essentially similar solutions. This unifying principle[1], called the *Symmetry principle* in [34], motivates us to formalise the notion of similarity. In this section, we thus define two measures that quantify the dissimilarity between two candidacy functions. These principled measures will serve as metrics for defining two notions of convergences underlying *continuity* principles for the forthcoming measures and inferences processes.

3.2.1 Introduction

A real-life multisensor system continuously updates a knowledge base with possibly contradictory and uncertain information. We can improve either the fault tolerance or the sensor coverage of such a system by reconfiguring its sensors in order to either increase or decrease the sensor redundancy. The more redundant two sensor groups are, the less dissimilar their information is, where their information is represented as candidacy functions.

Our problem is thus to measure how dissimilar two candidacy functions are. In this section, we define two principled dissimilarity measures founded upon two different assumptions about knowledge contents. The first one, the internal dissimilarity measure, considers that merging a candidacy function with separately two candidacy functions decreases their dissimilarity measures; metaphorically, pouring colour paint into two paint-pots makes the paint in these pots look less dissimilar. However, this behaviour might be undesirable: if a message is hidden using steganography into an image that looks similar to the original image, then knowing the stegokey makes these two images look more dissimilar. Therefore, we define the external dissimilarity measure, which generalises the Hausdorff metric used in [33] to compare consistent linear knowledge bases.

Before defining these two dissimilarity measures, we present seven principles to be satisfied by such measures.

3.2.2 Principles

In this section, we state several principles to be satisfied by a dissimilarity measure $\mu^{\text{dis}}(C_1, C_2)$ returning a real number when applied to two candidacy functions $C_1, C_2 \in \mathbb{C}$. By satisfying these principles, μ^{dis} non-trivially (principle $\mathbf{P}^{\text{dis}}_{\text{vi}}$) measures the distance (principles $\mathbf{P}^{\text{dis}}_{\text{ii}}$, $\mathbf{P}^{\text{dis}}_{\text{iv}}$, and $\mathbf{P}^{\text{dis}}_{\text{iii}}$) between two candidacy functions.

$\mathbf{P}^{\text{dis}}_{\text{i}}$ *Language invariance.* A dissimilarity measure is invariant by language enrichment.

$$\mu^{\text{dis}}(C_1, C_2) = \mu^{\text{dis}}(C_1 \oplus v, C_2 \oplus v)$$

$\mathbf{P}^{\text{dis}}_{\text{ii}}$ *Separation.* Two candidacy functions are not dissimilar iff they are equivalent, wrt a certain equivalence relation (see $\overset{\text{i}}{\equiv}$ at Def. 11 on page 9 and $\overset{\text{e}}{\equiv}$ at Def. 17 on page 15).

$$\mu^{\text{dis}}(C_1, C_2) = 0 \text{ iff } C_1 \equiv C_2$$

$\mathbf{P}^{\text{dis}}_{\text{iii}}$ *Triangle inequality.* If knowledge C is very similar to C_1 and C_2, then knowledge C_1 should not be too dissimilar from C_2.

$$\mu^{\text{dis}}(C_1, C) + \mu^{\text{dis}}(C, C_2) \geq \mu^{\text{dis}}(C_1, C_2)$$

$\mathbf{P}^{\text{dis}}_{\text{iv}}$ *Symmetry.* A dissimilarity measure is commutative.

$$\mu^{\text{dis}}(C_1, C_2) = \mu^{\text{dis}}(C_2, C_1)$$

$\mathbf{P}^{\text{dis}}_{\text{v}}$ *Paint-pot.* Two candidacy functions are less dissimilar after being separately merged with two equivalent candidacy functions. Metaphorically, pouring colour paint into two paint-pots makes the paint in these pots look less dissimilar. Notice that *paint-pot*, which is a kind of translational symmetry, implies *symmetry* by taking $C = C' = 1_{\mathbb{C}}$ then swapping C_1 and C_2.

$$\begin{array}{l}\text{if } C \equiv C' \\ \text{then } \mu^{\text{dis}}(C_1, C_2) \geq \mu^{\text{dis}}(C_2 \Cup C, C_1 \Cup C')\end{array}$$

[1] The commonsensical principles for inference processes (see chapter 4) are presented in [34] as special cases of the *Symmetry principle*.

$\mathbf{P}^{\text{dis}}_{\text{vi}}$ *Continuum.* There always exists a candidacy function C less dissimilar than another one C_1, wrt a given candidacy function C_2. This principle avoids trivial dissimilarity measure that constantly returns the same value for any two non-equivalent candidacy functions, wrt a certain equivalence relation (see $\overset{\text{i}}{\equiv}$ or $\overset{\text{e}}{\equiv}$).

$$\begin{array}{l}\text{if } C_1 \not\equiv C_2 \text{ then } \exists C, C_1 \not\equiv C \not\equiv C_2\\ \text{and } \mu^{\text{dis}}(C, C_2) < \mu^{\text{dis}}(C_1, C_2)\end{array}$$

$\mathbf{P}^{\text{dis}}_{\text{vii}}$ *Consequence invariance.* The dissimilarity measure between C_1 and C_2 equals the one between C_2 and the mergence of C_1 with one of its consequences C.

$$\text{if } C_1 \models C \text{ then } \mu^{\text{dis}}(C_1, C_2) = \mu^{\text{dis}}(C_1 \Cup C, C_2)$$

3.2.3 Internal dissimilarity measure $\mu^{\text{dis}}_{\mathcal{L}_\infty}$: the uniform norm of two candidacy functions

The internal dissimilarity measure $\mu^{\text{dis}}_{\mathcal{L}_\infty}$ is founded upon the uniform norm $\mathcal{L}_\infty$ of two candidacy functions:

$$\mathcal{L}_\infty(C_1, C_2) \overset{\text{def}}{=} \max_{\omega\in\Omega} |C_1(\omega) - C_2(\omega)|, \text{ if } \Theta(C_1) = \Theta(C_2)$$

We qualify $\mu^{\text{dis}}_{\mathcal{L}_\infty}$ as *internal* because it separates candidacy functions wrt the internal equivalence (see $\mathbf{P}^{\text{dis}}_{\text{ii}}$).

Definition 22 (Internal dissimilarity measure)**.**

$$\mu^{\text{dis}}_{\mathcal{L}_\infty}(C_1, C_2) \overset{\text{def}}{=} \mathcal{L}_\infty(C_1, C_2)$$

Proposition 10. *$\mu^{\text{dis}}_{\mathcal{L}_\infty}$ satisfies principles $\mathbf{P}^{\text{dis}}_{\text{i}}$, $\mathbf{P}^{\text{dis}}_{\text{ii}}$ wrt $\overset{\text{i}}{\equiv}$, $\mathbf{P}^{\text{dis}}_{\text{iii}}$, $\mathbf{P}^{\text{dis}}_{\text{iv}}$, $\mathbf{P}^{\text{dis}}_{\text{v}}$ wrt $\overset{\text{i}}{\equiv}$, and $\mathbf{P}^{\text{dis}}_{\text{vi}}$, but not $\mathbf{P}^{\text{dis}}_{\text{vii}}$.*

Proof. $\mathbf{P}^{\text{dis}}_{\text{i}}$ *Language invariance.* $\mu^{\text{dis}}_{\mathcal{L}_\infty}(C_1 \oplus v, C_2 \oplus v) = \mathcal{L}_\infty(C_1 \oplus v, C_2 \oplus v) = \max_{\omega'\in\Omega\oplus v} |(C_1 \oplus v)(\omega') - (C_2 \oplus v)(\omega')|$. By definition of $\diamond \oplus v$ (see §2.3.1), $\forall \omega \in \Omega, \forall \omega' \in \omega \oplus v, (C \oplus v)(\omega') \overset{\text{def}}{=} C(\omega)$. Hence, $\mu^{\text{dis}}_{\mathcal{L}_\infty}(C_1 \oplus v, C_2 \oplus v) = \max_{\omega\in\Omega} |C_1(\omega) - C_2(\omega)| = \mu^{\text{dis}}_{\mathcal{L}_\infty}(C_1, C_2)$.

$\mathbf{P}^{\text{dis}}_{\text{ii}}$ *Separation.* $\mu^{\text{dis}}_{\mathcal{L}_\infty}(C_1, C_2) = 0$ iff $\forall \omega \in \Omega, C_1(\omega) - C_2(\omega) = 0$ iff $C_1 = C_2$, which is equivalent to $C_1 \overset{\text{i}}{\equiv} C_2$ in case $\Theta(C_1) = \Theta(C_2)$.

$\mathbf{P}^{\text{dis}}_{\text{iii}}$ *Triangle inequality.* Since $\mathcal{L}_\infty$ is a metric, $\mu^{\text{dis}}_{\mathcal{L}_\infty}$ also satisfies $\mathbf{P}^{\text{dis}}_{\text{iii}}$.

$\mathbf{P}^{\text{dis}}_{\text{iv}}$, $\mathbf{P}^{\text{dis}}_{\text{v}}$ *Symmetry, Paint-pot.* Since $\mathcal{L}_\infty$ is symmetric, μ^{dis} is symmetric. Also, $\mu^{\text{dis}}(C_1, C_2) \geq \mu^{\text{dis}}(C_2 \Cup C, C_1 \Cup C')$ iff $\mathcal{L}_\infty(C_1, C_2) \geq \mathcal{L}_\infty(C_2 \Cup C, C_1 \Cup C')$ iff $\mathcal{L}_\infty(C_1, C_2) \geq \max_{\omega\in\Omega} |C_2(\omega) * C(\omega) - C_1(\omega) * C'(\omega)|$ by definition of $\Cup$ and by replacing C' by C since $C \equiv C'$, iff $\mathcal{L}_\infty(C_1, C_2) \geq \max_{\omega\in\Omega} |C_2(\omega) - C_1(\omega)| * C(\omega)$ since $C(\omega) \geq 0$. Furthermore, for any probability distribution $\omega \in \Omega$, we have $|C_2(\omega) - C_1(\omega)| \geq |C_2(\omega) - C_1(\omega)| * C(\omega)$ since $1 \geq C(\omega) \geq 0$. Let ω be a probability distribution maximising $|C_2(\omega) - C_1(\omega)| * C(\omega)$. Thus, $\mathcal{L}_\infty(C_1, C_2) \geq |C_2(\omega) - C_1(\omega)| \geq |C_2(\omega) - C_1(\omega)| * C(\omega) = \mathcal{L}_\infty(C_2 \Cup C, C_1 \Cup C')$.

$\mathbf{P}^{\text{dis}}_{\text{vi}}$ *Continuum.* We construct C as the pointwise convex combination of C_1 with C_2, where $\lambda \in]0{:}1[$:

$$C(\omega) \overset{\text{def}}{=} \lambda C_1(\omega) + (1 - \lambda) C_2(\omega)$$

$\mathbf{P}^{\text{dis}}_{\text{vii}}$ *Consequence invariance.* The following counterexample shows that $\mu^{\text{dis}}_{\mathcal{L}_\infty}$ does not satisfy this principle. Let $C_1 = C_2$ and $C_1 \models C$ such that there exists a probability distribution $\omega \notin \hat{\Omega}_{C_1}$ where $C(\omega) < 1$. Thus, $(C_1 \Cup C)(\omega) = C_1(\omega) * C(\omega) < C_1(\omega)$. Therefore, $\mu^{\text{dis}}_{\mathcal{L}_\infty}(C_1, C_2) = 0 < C_1(\omega) - C_1(\omega) * C(\omega) = \mu^{\text{dis}}_{\mathcal{L}_\infty}(C_1 \Cup C, C_2)$.

□

Notice that the following dissimilarity measure satisfies neither *language invariance* nor *separation*:

$$\mu^{\text{dis}}_{\int}(C_1, C_2) \overset{\text{def}}{=} \frac{\int_\Omega |C_1(\omega) - C_2(\omega)| \mathrm{d}\omega}{\int_\Omega 1 \mathrm{d}\omega}$$

3.2.4 External dissimilarity measure $\mu^{\text{dis}}_{\mathcal{H}}$: the Hausdorff distance of the best candidates

The external dissimilarity measure $\mu^{\text{dis}}_{\mathcal{H}}$ is a metric on the candidacy functions having equal underlying language. It is founded upon the Hausdorff distance $\mathcal{H}$.

Definition 23. *The Hausdorff distance $\mathcal{H}$ of two nonempty compact (bounded and closed) sets X and Y of points*[2] *in a Euclidean space is defined as follows:*

$$\mathcal{H}(X, Y) \overset{\text{def}}{=} \inf \left\{ \delta \;\middle|\; \begin{array}{l}\forall x \in X, \exists y \in Y, \delta \geq \mathcal{L}_2(x, y)\\ \text{and } \forall y \in Y, \exists x \in X, \delta \geq \mathcal{L}_2(x, y)\end{array} \right\}$$

We qualify the following dissimilarity measure $\mu^{\text{dis}}_{\mathcal{H}}$ as *external* because it separates candidacy functions wrt the external equivalence (see $\mathbf{P}^{\text{dis}}_{\text{ii}}$).

Definition 24 (External dissimilarity measure)**.**

$$\mu^{\text{dis}}_{\mathcal{H}}(C_1, C_2) \overset{\text{def}}{=} \mathcal{H}(\hat{\Omega}_{C_1}, \hat{\Omega}_{C_2})$$

Proposition 11. *$\mu^{\text{dis}}_{\mathcal{L}_\infty}$ satisfies principles $\mathbf{P}^{\text{dis}}_{\text{ii}}$ wrt $\overset{\text{e}}{\equiv}$, $\mathbf{P}^{\text{dis}}_{\text{iii}}$, $\mathbf{P}^{\text{dis}}_{\text{iv}}$, $\mathbf{P}^{\text{dis}}_{\text{vi}}$, and $\mathbf{P}^{\text{dis}}_{\text{vii}}$, but neither $\mathbf{P}^{\text{dis}}_{\text{i}}$, nor $\mathbf{P}^{\text{dis}}_{\text{v}}$.*

[2] After considering footnote 5 on page 10, we could replace $\mathcal{L}_2$ by $\mathcal{G}_{\text{KL}}$ in Def. 23 when X and Y are sets of probability distributions.

Proof. $\mathbf{P}_{\text{i}}^{\text{dis}}$ *Language invariance.* The following counter-example shows that $\mu_{\mathcal{H}}^{\text{dis}}$ does not satisfy this principle. Let C_1 and C_2 be two candidacy functions underlain by a language with one propositional variable v. Let $\omega_1 \stackrel{\text{def}}{=} [0;1]$ and $\omega_2 \stackrel{\text{def}}{=} [1;0]$ be two probability distributions such that $\omega_1(v) = 0$ and $\omega_2(v) = 1$. Suppose that $C_1(\omega_1) = 1$ and $\hat{\Omega}_{C_1} = \{\omega_1\}$, and that $C_2(\omega_2) = 1$ and $\hat{\Omega}_{C_2} = \{\omega_2\}$. Let $C_1' \stackrel{\text{def}}{=} C_1 \oplus v'$ and $C_2' \stackrel{\text{def}}{=} C_2 \oplus v'$, where $v' \neq v$. In figure 3.1 on page 31, $\omega_1 \oplus v'$ and $\omega_2 \oplus v'$ are graphically represented by the two colour-filled rectangles with $a = 0$ and $a = 1$. Thus, $\mu_{\mathcal{H}}^{\text{dis}}(C_1, C_2) = \mathcal{H}(\{\omega_1\}, \{\omega_2\}) = \mathcal{L}_2(\omega_1, \omega_2) = \sqrt{2}$ whereas $\mu_{\mathcal{H}}^{\text{dis}}(C_1', C_2') = \mathcal{H}(\{\omega_1'\}, \{\omega_2'\}) = \mathcal{L}_2(\omega_1', \omega_2') = \sqrt{\frac{3}{2}}$, where $\omega_1' \in \omega_1 \oplus v'$ and $\omega_2' \in \omega_2 \oplus v'$ such that $\mathcal{L}_2(\omega_1', \omega_2') = \mathcal{H}(\omega_1 \oplus v', \omega_2 \oplus v')$. For example, take $\omega_1' \stackrel{\text{def}}{=} [0;0;0;0]$ and $\omega_2' \stackrel{\text{def}}{=} [\frac{1}{2};\frac{1}{2};0;0]$. In figure 3.1 on page 31, ω_1' and ω_2' respectively correspond to the points $\frac{1}{2} * H * \omega_1' = [\frac{1}{2};\frac{1}{2};\frac{1}{2};\frac{1}{2}]$ and $\frac{1}{2} * H * \omega_2' = [0;-\frac{1}{2};0;\frac{1}{2}]$, where H is a Hadamard matrix. Thus $\mu_{\mathcal{H}}^{\text{dis}}(C_1, C_2) \neq \mu_{\mathcal{H}}^{\text{dis}}(C_1', C_2')$ in this counter-example.

$\mathbf{P}_{\text{ii}}^{\text{dis}}$ *Separation.* $\mu_{\mathcal{L}_\infty}^{\text{dis}}(C_1, C_2) = 0$ iff $\hat{\Omega}_{C_1} = \hat{\Omega}_{C_2}$ iff $C_1 \stackrel{\text{e}}{\equiv} C_2$ in case $\Theta(C_1) = \Theta(C_2)$.

$\mathbf{P}_{\text{iii}}^{\text{dis}}$ *Triangle inequality.* Since $\mathcal{H}$ is a metric on Ω, $\mu_{\mathcal{L}_\infty}^{\text{dis}}$ satisfies $\mathbf{P}_{\text{iii}}^{\text{dis}}$.

$\mathbf{P}_{\text{iv}}^{\text{dis}}$ *Symmetry.* $\mu_{\mathcal{H}}^{\text{dis}}$ is symmetric since $\mathcal{H}$ is symmetric.

$\mathbf{P}_{\text{v}}^{\text{dis}}$ *Paint-pot.* The following counter-example shows that $\mu_{\mathcal{H}}^{\text{dis}}$ does not satisfy this principle. Let Θ be a language with two propositional variables. Let $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \alpha_\Theta$ be the four minterms of Θ. Let $\epsilon \in]0{:}1[$. Let a, b, c, and d be four linear constraints underlain by Θ and defined as follows: $a \stackrel{\text{def}}{=}$ "$1 \geq \frac{1}{1-\epsilon} * \omega(\alpha_1) + *\omega(\alpha_2)$", $b \stackrel{\text{def}}{=}$ "$1 - \epsilon \geq (1-\epsilon) * \omega(\alpha_1) - \omega(\alpha_2)$", $c \stackrel{\text{def}}{=}$ "$0 \geq \omega(\alpha_3)$", and $d \stackrel{\text{def}}{=}$ "$1 \leq \omega(\alpha_1) + \omega(\alpha_2)$". We denote by $C_{\{a,c\}}$ the candidacy function corresponding to the knowledge base $\{a, c\}$. Thus, $\mathcal{H}(C_{\{a,c\}}, C_{\{b,c\}}) = \epsilon$ whereas $\mathcal{H}(C_{\{b,c,d\}}, C_{\{a,c,d\}}) = \sqrt{2}$. Therefore, $\mu_{\mathcal{H}}^{\text{dis}}(C_{\{a,c\}}, C_{\{b,c\}}) = \epsilon < \sqrt{2} = \mu^{\text{dis}}(C_{\{b,c\}} \Cup C_{\{d\}}, C_{\{a,c\}} \Cup C_{\{d\}})$.

$\mathbf{P}_{\text{vi}}^{\text{dis}}$ *Continuum.* Notice that $C_1 \not\equiv C_2$ iff $\exists \delta \in \mathbb{R}, 0 < \delta < \mu^{\text{dis}}(C_1, C_2)$. It suffices to construct C such that $0 < \mathcal{H}(\hat{\Omega}_C, \hat{\Omega}_{C_2}) \leq \delta$. We thus propose to construct $C \stackrel{\text{def}}{=} (C_2 \setminus \mathcal{B}_\delta^\omega) \cup \omega_\delta$ such that C is a copy of C_2, from which we remove a ball $\mathcal{B}_\delta^\omega$ of probability distributions centred in $\omega \in C_2$ with a radius of δ, then to which we add a probability distribution ω_δ such that $\mathcal{L}_2(\omega, \omega_\delta) = \delta$. Instead of adding ω_δ to $(C_2 \setminus \mathcal{B}_\delta^\omega)$, we could have added the probability distributions belonging to the frontier[3] $\mathcal{F}$ of $\mathcal{B}_\delta^\omega$: $C \stackrel{\text{def}}{=} (C_2 \setminus \mathcal{B}_\delta^\omega) \cup \mathcal{F}(\mathcal{B}_\delta^\omega)$.

$\mathbf{P}_{\text{vii}}^{\text{dis}}$ *Consequence invariance.* By definition of $C_1 \models C$, $\forall \hat{\omega} \in \hat{\Omega}_{C_1}, C(\hat{\omega}) = 1$, hence $\forall \hat{\omega} \in \hat{\Omega}_{C_1}, C_1(\hat{\omega}) = C_1(\hat{\omega}) * C(\hat{\omega}) = (C_1 \Cup C)(\hat{\omega})$ then $\hat{\Omega}_{C_1} = \hat{\Omega}_{C_1 \Cup C}$. Therefore, $\mu_{\mathcal{H}}^{\text{dis}}(C_1, C_2) = \mu_{\mathcal{H}}^{\text{dis}}(C_1 \Cup C, C_2)$

□

Furthermore, $\mu_{\mathcal{H}}^{\text{dis}}$ is σ-invariant, ie $\mu_{\mathcal{H}}^{\text{dis}}(C_{K_1^\sigma}, C_{K_2^\sigma})$ equals $\mu_{\mathcal{H}}^{\text{dis}}(C_{K_1^{\sigma'}}, C_{K_2^{\sigma'}})$ for any two reliability levels σ and σ' in $]0{:}1[$; th reason is that the best candidates are σ-invariant (see Prop. 9 on page 14).

On considering candidacy degrees We also propose an intermediate metric $\mu_{\bar{\mathcal{H}}}^{\text{dis}}$ between $\mu_{\mathcal{L}_\infty}^{\text{dis}}$ and $\mu_{\mathcal{H}}^{\text{dis}}$, ie $\mu_{\bar{\mathcal{H}}}^{\text{dis}}$ is such that $\mu_{\mathcal{L}_\infty}^{\text{dis}}(C_1, C_2) = 0$ implies $\mu_{\bar{\mathcal{H}}}^{\text{dis}}(C_1, C_2) = 0$ implies $\mu_{\mathcal{H}}^{\text{dis}}(C_1, C_2) = 0$:

$$\mu_{\bar{\mathcal{H}}}^{\text{dis}}(C_1, C_2) \stackrel{\text{def}}{=} \bar{\mathcal{H}}(C_1, C_2)$$

where

$$\bar{\mathcal{H}}(C_1, C_2) \stackrel{\text{def}}{=} \mathcal{H}\left(\left\{ [\hat{\omega}; C_1(\hat{\omega})] \;\middle|\; \hat{\omega} \in \hat{\Omega}_{C_1} \right\}, \left\{ [\hat{\omega}; C_2(\hat{\omega})] \;\middle|\; \hat{\omega} \in \hat{\Omega}_{C_2} \right\}\right)$$

$\mu_{\bar{\mathcal{H}}}^{\text{dis}}$ satisfies the same principles as $\mu_{\mathcal{H}}^{\text{dis}}$ (wrt a stronger equivalence relation), but takes into account the candidacy degree of the best candidates, which is not σ-invariant, hence $\mu_{\bar{\mathcal{H}}}^{\text{dis}}$ is not σ-invariant. When the best candidates $\hat{\Omega}_{C_1}$ and $\hat{\Omega}_{C_2}$ have the same candidacy degree, then $\mu_{\bar{\mathcal{H}}}^{\text{dis}}(C_1, C_2)$ equals $\mu_{\mathcal{H}}^{\text{dis}}(C_1, C_2)$.

3.2.5 Notions of convergence for a sequence of knowledge

Each previously defined dissimilarity measure, which is a metric, induces a notion of convergence of a sequence of candidacy functions. Such a notion of convergence is necessary to establish the concept of *continuity* for the forthcoming tools performing on $\mathbb{C}$, ie the next measures and inference processes.

Definition 25. *A sequence of candidacy functions $S\colon \mathbb{N} \mapsto \mathbb{D}$, where $\mathbb{D} \subseteq \mathbb{C}$, converges to a candidacy function $C \in \mathbb{D}$ wrt a metric μ^{dis} on $\mathbb{D}$, noted $\lim_{i\to\infty} \mu^{\text{dis}}(S(i), C) = 0$, iff $\forall \varepsilon \in \mathbb{R}^+, \exists N_\varepsilon \in \mathbb{N}, \forall i \geq N_\varepsilon, \mu^{\text{dis}}(S(i), C) < \varepsilon$.*

Notice that if S converges to C wrt $\mu_{\bar{\mathcal{H}}}^{\text{dis}}$, then S converges to C wrt $\mu_{\mathcal{H}}^{\text{dis}}$. Moreover, notice that the following proposition shows that $\hat{\Omega}_C$ may change discontinuously wrt $\mathcal{H}$ when C changes continuously wrt $\mu_{\mathcal{L}_\infty}^{\text{dis}}$.

[3] The *frontier* $\mathcal{F}(S)$, or boundary, of a set $S \subseteq \mathbb{D}$ is a subset of $\mathbb{D}$ containing every frontier point of S. A frontier point s of S is such that any open set containing s intersects both S and its complement $\mathbb{D} \setminus S$.

Proposition 12. *A sequence of candidacy functions S does* not necessary *converge to $C \in \mathbb{C}$ wrt $\mu^{\text{dis}}_{\mathcal{H}}$ when S converges to C wrt $\mu^{\text{dis}}_{\mathcal{L}_\infty}$.*

Proof. Suppose that a propositional language with two variables underlies the following three (non-normalised) linear constraints: $c_1 \stackrel{\text{def}}{=}$ "$\omega(\alpha_1) \leq 0.1$", $c_2^\epsilon \stackrel{\text{def}}{=}$ "$0.2 \leq \omega(\alpha_1) - \epsilon * \omega(\alpha_2)$", where $\epsilon \in \mathbb{R}$, and $c_3 \stackrel{\text{def}}{=}$ "$\omega(\alpha_3) \leq 0$". Thus, $\hat{\Omega}_{C_{\{c_1,c_2^0,c_3\}}}$ equals $\left\{ \left[\frac{0.1+0.2}{2}; y; 0; 1 - \left(\frac{0.1+0.2}{2} + y\right)\right] \mid y \in [0{:}1 - \frac{0.1+0.2}{2}] \right\}$ whereas $\lim_{\epsilon\to 0,\epsilon>0} \hat{\Omega}_{C_{\{c_1,c_2^\epsilon,c_3\}}}$ equals $\left\{\left[\frac{0.1+0.2}{2}; 0; 0; 1 - \frac{0.1+0.2}{2}\right]\right\}$. Therefore, if $S(i) \stackrel{\text{def}}{=} C_{\{c_1,c_2^{1/i},c_3\}}$ and $C \stackrel{\text{def}}{=} C_{\{c_1,c_2^0,c_3\}}$, then $\lim_{i\to\infty} \mu^{\text{dis}}_{\mathcal{L}_\infty}(S(i), C) = 0$ whereas $\lim_{i\to\infty} \mu^{\text{dis}}_{\mathcal{H}}(S(i), C) = \sqrt{1.7}$, which is $\mathcal{L}_2(\left[\frac{0.1+0.2}{2}; 0; 0; 1 - \frac{0.1+0.2}{2}\right], \left[\frac{0.1+0.2}{2}; 1 - \frac{0.1+0.2}{2}; 0; 0\right])$. □

3.2.6 Conclusions

Our main contribution is twofold: firstly, we establish seven principles to be satisfied by a dissimilarity measure when applied to candidacy functions. Secondly, we define two principled dissimilarity measures founded upon two distinct notions of equivalence between knowledge contents (see Def. 11 and Def. 17). To our knowledge, no measure for inconsistent linear knowledge bases, hence for $\mathbb{C}$, exists in the literature.

Proofs of propositions 10 and 11 not only show that our measures are metrics for $\mathbb{C}$ (principles $\mathbf{P}^{\text{dis}}_{\text{ii}}$, $\mathbf{P}^{\text{dis}}_{\text{iv}}$, and $\mathbf{P}^{\text{dis}}_{\text{iii}}$), but also that our internal measure can be utilised to compare candidacy functions about different topics, since it satisfies *language invariance* (see principle $\mathbf{P}^{\text{dis}}_{\text{i}}$).

In future research, our dissimilarity measure should be compared to the ones defined for fuzzy sets (see [11]), and those for propositional knowledge bases (ie knowledge bases involving only categorical probabilities), which are a special case of our knowledge bases. This may conduce to a coherence measure (see [12], although proposition 8 page 232 states the opposite of *paint-pot*).

3.3 Inconsistency measure μ^{icst}: to spot the defects in a knowledge content

Inconsistencies may arise when a knowledge base is a collection of items coming from different sources. In this section, we propose both an inconsistency measure that quantifies the global inconsistency of the knowledge base, and a culpability measure that evaluates the blame on each item for making the whole knowledge base inconsistent.

3.3.1 Introduction

A real-life multisensor system continuously updates a knowledge base with possibly contradictory and uncertain information. We can improve such a system by reconfiguring its sensors in order to minimise the amount of inconsistency brought by each sensor or sensor group. Our problem is thus to measure how inconsistent the knowledge content of a knowledge base is.

In this section, we define an inconsistency measure together with its culpability measure, which evaluates the responsibility of each knowledge item for making the whole knowledge inconsistent. Our measures satisfy principles ensuring not only their robustness against knowledge fluctuations (see principle $\mathbf{P}^{\text{icst}}_d$), but also their tolerance to the enrichment of their knowledge domain with new topics, which happens when new sensors are dynamically plugged into the system (see principle $\mathbf{P}^{\text{icst}}_a$).

Recently, several inconsistency measures appeared in the literature (see [18] for set of propositions, and [42] who extends it to conditional probabilities). However, to our knowledge, none of them are defined for the whole set of linear knowledge bases $\mathbb{K}^L$. Thus, our principled measures seem to be the first to quantify the inconsistency of knowledge bases in $\mathbb{K} \supset \mathbb{K}^L$.

The remainder of this section starts by a presentation of several principles to be satisfied by an inconsistency measure. Then, we propose an inconsistency measure, together with its culpability measure, that satisfies these principles.

3.3.2 Principles

In this section, we state several principles to be satisfied by an inconsistency measure μ^{icst} returning a positive real number when applied to a knowledge content. We then defined μ^{icst} over the knowledge content of a knowledge base in $\mathbb{K}$ instead of $\mathbb{C}$ because the notion of (in)consistency is meaningless for candidacy functions. Let $\Omega_{\mathfrak{K}} \stackrel{\text{def}}{=} \Omega \cap \bigcap_{\mathfrak{k}\in\mathfrak{C}} \mathfrak{k}$ be the probability functions belonging to every knowledge item of $\mathfrak{K}$. Hence, $\Omega_{\mathfrak{K}_K} = \Omega_K$ under assumption ⑤ or ⑥. A knowledge content $\mathfrak{K}$ is thus said to be consistent iff $\Omega_{\mathfrak{K}} \neq \emptyset$, otherwise $\mathfrak{K}$ is said to be inconsistent. Being principled, μ^{icst} non-trivially (principle $\mathbf{P}^{\text{icst}}_f$) separates (principles $\mathbf{P}^{\text{icst}}_b$) between consistent or inconsistent knowledge contents (hence knowledge bases).

While $\mu^{\text{icst}}(\mathfrak{K})$ is the inconsistency measure of a knowledge content $\mathfrak{K}$, $\mu^{\text{culp}}_{\mathfrak{K}}(\mathfrak{k})$ is the culpability measure of a knowledge item $\mathfrak{k}$ belonging to $\mathfrak{K}$: this last measure quantifies the degree to which $\mathfrak{k}$ can be held responsible for making $\mathfrak{K}$ inconsistent. Notice that the concept of inconsistency is dual to the holistic conception of coherence (see [5, page 81]), which measures the degree

to which the knowledge items of $\mathfrak{K}$ fit together when $\mathfrak{K}$ is consistent. Also, these concepts share desiderata, like *consequence invariance* (see principle $\mathbf{P}_h^{\text{icst}}$ and [9, page 418]).

$\mathbf{P}_a^{\text{icst}}$ *Language invariance.* An inconsistency measure is invariant by language enrichment.

$$\mu^{\text{icst}}(\mathfrak{K}) = \mu^{\text{icst}}(\mathfrak{K} \oplus v)$$

$\mathbf{P}_b^{\text{icst}}$ *Separation* (extends *consistency* of [18, Def. 8]). A knowledge content $\mathfrak{K}$ is consistent iff its inconsistency measure is null.

$$\Omega_{\mathfrak{K}} \neq \emptyset \text{ iff } \mu^{\text{icst}}(\mathfrak{K}) = 0$$

$\mathbf{P}_c^{\text{icst}}$ *Equivalence.* Equivalent knowledge contents have equal inconsistency measure, wrt a certain equivalence relation (see $\overset{\text{i}}{\equiv}$ at Def. 11 on page 9 and $\overset{\text{e}}{\equiv}$ at Def. 17 on page 15).

$$\text{if } C_{\mathfrak{K}_1} \equiv C_{\mathfrak{K}_2} \text{ then } \mu^{\text{icst}}(\mathfrak{K}_1) = \mu^{\text{icst}}(\mathfrak{K}_2)$$

$\mathbf{P}_d^{\text{icst}}$ *Continuity.* When a knowledge content changes continuously, so its inconsistency measure does, wrt a certain notion of convergence (see Def. 25 on page 20). This principle ensures a certain robustness for μ^{icst} in the face of minor fluctuations in the knowledge content.

$$\begin{array}{l}\text{if } \lim_{i\to\infty} \mu^{\text{dis}}(C_{\mathfrak{K}_i}, C_{\mathfrak{K}}) = 0 \\ \text{then } \lim_{i\to\infty} \mu^{\text{icst}}(\mathfrak{K}_i) = \mu^{\text{icst}}(\mathfrak{K})\end{array}$$

$\mathbf{P}_e^{\text{icst}}$ *Monotonicity* (extends *monotonicity* of [18, Def. 8]). Merging two knowledge contents increases the degree of inconsistency. We recall that a knowledge content is a multiset.

$$\mu^{\text{icst}}(\mathfrak{K}_1) \leq \mu^{\text{icst}}(\mathfrak{K}_1 \cup \mathfrak{K}_2)$$

$\mathbf{P}_f^{\text{icst}}$ *Strict monotonicity.* The inconsistency measure of $\mathfrak{K}_1$ is strictly lower than that of its mergence with an inconsistent knowledge content $\mathfrak{K}_2$.

$$\text{if } 0 < \mu^{\text{icst}}(\mathfrak{K}_2) \text{ then } \mu^{\text{icst}}(\mathfrak{K}_1) < \mu^{\text{icst}}(\mathfrak{K}_1 \cup \mathfrak{K}_2)$$

$\mathbf{P}_g^{\text{icst}}$ *Minimality* (extends *minimality* of [18, Def. 12]). A consequence $\mathfrak{k}$ of a knowledge content $\mathfrak{K}$ does not bring any contradiction to $\mathfrak{K}$.

$$\text{if } \mathfrak{K} \models \mathfrak{k} \text{ then } \mu_{\mathfrak{K}}^{\text{culp}}(\mathfrak{k}) = 0$$

$\mathbf{P}_h^{\text{icst}}$ *Consequence invariance* (adapts *free formula independence* of [18, Def. 8]). Merging a knowledge content with its consequences leaves invariant the inconsistency measure.

$$\text{if } \mathfrak{K} \models \mathfrak{k} \text{ then } \mu^{\text{icst}}(\mathfrak{K}) = \mu^{\text{icst}}(\mathfrak{K} \cup \{\mathfrak{k}\})$$

$\mathbf{P}_i^{\text{icst}}$ *Dominance* (extends *dominance* of [18, Def. 8]). Stronger knowledge items potentially bring more contradictions, where a knowledge item $\mathfrak{k}_1$ is said stronger than $\mathfrak{k}_2$ iff $\mathfrak{k}_1$ entails $\mathfrak{k}_2$.

$$\text{if } \mathfrak{k}_1 \models \mathfrak{k}_2 \text{ then } \mu^{\text{icst}}(\mathfrak{K} \cup \{\mathfrak{k}_1\}) \geq \mu^{\text{icst}}(\mathfrak{K} \cup \{\mathfrak{k}_2\})$$

$\mathbf{P}_j^{\text{icst}}$ *Equitable distribution* (extends *distribution* of [18, Def. 12]). An inconsistency measure of a knowledge content $\mathfrak{K}$ only depends on the culpability measures of each of its $m \in \mathbb{N}$ knowledge items, without preference for some of them. Thus, the inconsistency measure is equitably distributed among the culpability measures. Let $f\colon \mathbb{R}^{+m} \mapsto \mathbb{R}^+$ be a symmetric function for all its arguments that aggregates the culpability measure of each merged knowledge item. Notice that the symmetry of f formalises the equity of the distribution.

$$\mu^{\text{icst}}(\mathfrak{K}) = f(\mu_{\mathfrak{K}}^{\text{culp}}(\mathfrak{k}_1), \ldots, \mu_{\mathfrak{K}}^{\text{culp}}(\mathfrak{k}_m))$$

Notice that [17, Def. 8] and [18, Def. 8 and Def. 12] also defines other properties, like *decomposability* and *MinInc*. Thus, further investigations on inconsistency measures should consider these properties.

3.3.3 Inconsistency measure μ^{icst}: the candidacy degree of the best candidates

In this section, we define our inconsistency measure $\mu^{\text{icst}}(\mathfrak{K})$ as the amount of contradictions inside a knowledge content $\mathfrak{K}$. The associated culpability measure $\mu_{\mathfrak{K}}^{\text{culp}}(\mathfrak{k})$ quantifies the amount of contradictions brought to a knowledge content $\mathfrak{K}$ by one of its knowledge items $\mathfrak{k}$. After, we demonstrate that our measure is principled.

Definition 26 (Inconsistency measure)**.**

$$\mu^{\text{icst}}(\mathfrak{K}) \overset{\text{def}}{=} 1 - \max_{\omega\in\Omega} C_{\mathfrak{K}}(\omega)$$

Notice that $\mu^{\text{icst}}(\mathfrak{K})$ equals $1 - \mu^{\text{dis}}_{\mathcal{L}_\infty}(C_{\mathfrak{K}}, 0_{\mathbb{C}})$, which is the complementary distance between $C_{\mathfrak{K}}$ and the absorbing candidacy function $0_{\mathbb{C}}$, where $0_{\mathbb{C}}$ is seen as absolutely inconsistent. Thus, $\mu^{\text{icst}}(\mathfrak{K})$ measures how far $\mathfrak{K}$ is to be consistent.

We also want to define a measure $\mu_{\mathfrak{K}}^{\text{culp}}(\mathfrak{k})$ that evaluates the culpability of each knowledge item $\mathfrak{k}$ for making $\mathfrak{K}$ inconsistent. According to *equitable distribution* (see principle $\mathbf{P}_j^{\text{icst}}$), the inconsistency measure of $\mathfrak{K}$ should only depend on the culpability measures of each knowledge item $\mathfrak{k} \in \mathfrak{K}$; we will thus define the culpability measure such that $\mu^{\text{icst}}(\mathfrak{K})$ is distributed among each $\mu_{\mathfrak{K}}^{\text{culp}}(\mathfrak{k})$. Furthermore, for any best candidate $\hat{\omega}$ of $C_{\mathfrak{K}}$ such that $C_{\mathfrak{K}}(\hat{\omega}) \in]0{:}1]$, we have $C_{\mathfrak{K}}(\hat{\omega}) = \prod_{\mathfrak{k}\in\mathfrak{K}} C_{\mathfrak{k}}(\hat{\omega})$, hence $\ln(C_{\mathfrak{K}}(\hat{\omega})) = \ln(\prod_{\mathfrak{k}\in\mathfrak{K}} C_{\mathfrak{k}}(\hat{\omega}))$, hence $1 = \sum_{\mathfrak{k}\in\mathfrak{K}} \frac{\ln(C_{\mathfrak{k}}(\hat{\omega}))}{\ln(C_{\mathfrak{K}}(\hat{\omega}))}$.

Definition 27 (Culpability distribution)**.** *The culpability distribution $\varkappa_{\mathfrak{K}}(\hat{\omega})$ of a knowledge content $\mathfrak{K}$ among its $m \in \mathbb{N}$ knowledge items $\{\mathfrak{k}_1, \mathfrak{k}_2, \ldots, \mathfrak{k}_m\}$ wrt a given best candidate $\hat{\omega}$ of $C_{\mathfrak{K}}$ is defined as the following vector:*

$$\varkappa_{\mathfrak{K}}(\hat{\omega}) \stackrel{\text{def}}{=} \left[\varkappa^{\mathfrak{k}_1}_{\mathfrak{K}}(\hat{\omega}); \varkappa^{\mathfrak{k}_2}_{\mathfrak{K}}(\hat{\omega}); \ldots; \varkappa^{\mathfrak{k}_m}_{\mathfrak{K}}(\hat{\omega})\right]$$

where $\varkappa^{\mathfrak{k}}_{\mathfrak{K}} : \Omega \mapsto [0{:}1]$ is the normalised culpability of $\mathfrak{k}$ in making $\mathfrak{K}$ inconsistent wrt a given best candidate $\hat{\omega}$:

$$\varkappa^{\mathfrak{k}}_{\mathfrak{K}}(\hat{\omega}) \stackrel{\text{def}}{=} \begin{cases} \frac{\ln(C_{\mathfrak{k}}(\hat{\omega}))}{\ln(C_{\mathfrak{K}}(\hat{\omega}))} & \text{if } 0 < \mu^{\text{icst}}(\mathfrak{K}) < 1, \\ \frac{1}{|\mathfrak{K}|} & \text{if } \mu^{\text{icst}}(\mathfrak{K}) = 0, \\ 0 & \text{if } \mu^{\text{icst}}(\mathfrak{K}) = 1 \text{ and } C_{\mathfrak{k}}(\hat{\omega}) \neq 0, \\ \frac{1}{|\{\ \mathfrak{k} \in \mathfrak{K} \mid C_{\mathfrak{k}}(\hat{\omega}) = 0\ \}|} & \text{if } C_{\mathfrak{k}}(\hat{\omega}) = 0. \end{cases}$$

Remark that Def. 27 simply states that the knowledge items of $\mathfrak{K}$ are equally culpable in making $\mathfrak{K}$ inconsistent in case $\mathfrak{K}$ is consistent, ie $\mu^{\text{icst}}(\mathfrak{K}) = 0$. In case $\mu^{\text{icst}}(\mathfrak{K}) = 1$, ie when $C_{\mathfrak{K}}$ is absorbing, then the culpability of making $\mathfrak{K}$ inconsistent is only distributed among the knowledge items $\{\ \mathfrak{k} \in \mathfrak{K} \mid C_{\mathfrak{k}}(\hat{\omega}) = 0\ \}$ that make $C_{\mathfrak{K}}$ absorbing; the other knowledge items, which are such that $C_{\mathfrak{k}}(\hat{\omega}) > 0$, are considered non-culpable, ie $\varkappa^{\mathfrak{k}}_{\mathfrak{K}}(\hat{\omega}) = \lim_{C_{\mathfrak{K}}(\hat{\omega}) \to 0} \frac{\ln(C_{\mathfrak{k}}(\hat{\omega}))}{\ln(C_{\mathfrak{K}}(\hat{\omega}))} = 0$.

Remark also that $\varkappa_{\mathfrak{K}}(\hat{\omega})$ might be considered as a *probability distribution* since it is a positive vector of which its elements sum up to one: $\forall \mathfrak{k} \in \mathfrak{K}, \varkappa^{\mathfrak{k}}_{\mathfrak{K}}(\hat{\omega}) \geq 0$ and $1 = \sum_{\mathfrak{k} \in \mathfrak{K}} \varkappa^{\mathfrak{k}}_{\mathfrak{K}}(\hat{\omega})$. The *least biased* probability distributions are known to maximise the entropy $E(\vec{x})$ when adhering to several commonsensical principles defined in [38] and restated in chapter 4. We may thus defined the *least biased culpability distributions* as $\{\varkappa_{\mathfrak{K}}(\hat{\omega}) \mid \hat{\omega} \in \arg\max_{\hat{\omega} \in \hat{\Omega}_{C_{\mathfrak{K}}}} E(\varkappa_{\mathfrak{K}}(\hat{\omega}))\}$, where $E(\varkappa_{\mathfrak{K}}(\hat{\omega}))$ is the entropy of a culpability distribution $\varkappa_{\mathfrak{K}}(\hat{\omega})$.

Definition 28 (Culpability measure)**.** *Let $\widetilde{\Omega}_{\mathfrak{K}}$ be the set of best candidates of $C_{\mathfrak{K}}$ such that $\{\varkappa_{\mathfrak{K}}(\widetilde{\omega}) \mid \widetilde{\omega} \in \widetilde{\Omega}_{\mathfrak{K}}\}$ are the least biased culpability distributions:*

$$\widetilde{\Omega}_{\mathfrak{K}} \stackrel{\text{def}}{=} \arg\max_{\hat{\omega} \in \hat{\Omega}_{C_{\mathfrak{K}}}} E(\varkappa_{\mathfrak{K}}(\hat{\omega})) \qquad E(\vec{x}) \stackrel{\text{def}}{=} -\sum_{j=1}^{|\vec{x}|} \vec{x}_j * \ln(\vec{x}_j)$$

Suppose that $\widetilde{\Omega}_{\mathfrak{K}}$ can be partitioned as $\{\widetilde{\Omega}_0, \widetilde{\Omega}_1, \ldots, \widetilde{\Omega}_p\}$ with $p \in \mathbb{N}$ minimal, where $\widetilde{\Omega}_0$ is a (possibly empty) finite set of points, and for each part $\widetilde{\Omega}_{i \geq 1}$, $\widetilde{\Omega}_i$ is a solo-dimensional manifold (see Def. 21 on page 18), and $C_{\mathfrak{k}}$ is Lebesgue integrable on $\widetilde{\Omega}_i$, and $\forall \omega \in \widetilde{\Omega}_i, C_{\mathfrak{k}}(\omega) > 0$. Then, we define the culpability measure of a knowledge item $\mathfrak{k}$ in a knowledge content $\mathfrak{K}$ as the following continuous geometric mean:

$$\mu^{\text{culp}}_{\mathfrak{K}}(\mathfrak{k}) \stackrel{\text{def}}{=} 1 - \sqrt[|\widetilde{\Omega}_0|+p]{\left(\prod_{\widetilde{\omega} \in \widetilde{\Omega}_0} C_{\mathfrak{k}}(\widetilde{\omega})\right) * \left(\prod_{i=1}^{p} \vartheta(\widetilde{\Omega}_i, \mathfrak{k})\right)}$$

where $\vartheta(\widetilde{\Omega}_i, \mathfrak{k}) \stackrel{\text{def}}{=} \exp\left(\frac{\int_{\widetilde{\Omega}_i} \ln(C_{\mathfrak{k}}(\widetilde{\omega}))\mathrm{d}\widetilde{\omega}}{\int_{\widetilde{\Omega}_i} 1\mathrm{d}\widetilde{\omega}}\right)$.

Remember that $\mu^{\text{culp}}_{\mathfrak{K}}$ is not defined when $\widetilde{\Omega}_{\mathfrak{K}}$ cannot be partitioned into solo-dimensional manifolds. Also, we should further study the dependence of the culpability measure on the chosen partition for $\widetilde{\Omega}_{C_{\mathfrak{K}}}$. However, if K is a linear knowledge base and if the blur function is h_{HG} (see Def. 16), we then conjecture that $\forall \widetilde{\omega}_1, \widetilde{\omega}_2 \in \widetilde{\Omega}_{\mathfrak{K}_K} \forall \mathfrak{k} \in \mathfrak{K}_K, C_{\mathfrak{k}}(\widetilde{\omega}_1) = C_{\mathfrak{k}}(\widetilde{\omega}_2)$. Therefore, if this conjecture holds, then $\mu^{\text{culp}}_{\mathfrak{K}_K}(\mathfrak{k}) = 1 - C_{\mathfrak{k}}(\widetilde{\omega})$, where $\widetilde{\omega} \in \widetilde{\Omega}_{\mathfrak{K}}$.

Notice that assuming ⑥, which states that $\mathfrak{K}_K \stackrel{\text{def}}{=} \{\hat{\Omega}_{C_K}\}$, is ignoring the candidacy degree of the best candidates, ie is ignoring the inconsistency measure. Thus, employing μ^{icst} while assuming ⑥ seems to us somewhat irrational. Therefore, $\equiv$ and μ^{dis} in principles $\mathbf{P}^{\text{icst}}_c$ and $\mathbf{P}^{\text{icst}}_d$ should *not* be instantiated by $\stackrel{\text{e}}{\equiv}$ and $\mu^{\text{dis}}_{\mathcal{H}}$, hence should be instantiated by $\stackrel{\text{i}}{\equiv}$ and $\mu^{\text{dis}}_{\mathcal{L}_\infty}$ (or even by $\mu^{\text{dis}}_{\mathcal{H}}$); μ^{icst} is thus an internal measure rather than an external one.

Proposition 13. *μ^{icst} satisfies principles $\mathbf{P}^{\text{icst}}_a$ to $\mathbf{P}^{\text{icst}}_j$, unless $\stackrel{\text{e}}{\equiv}$ instantiates $\equiv$ in $\mathbf{P}^{\text{icst}}_c$ and $\mu^{\text{dis}}_{\mathcal{H}}$ instantiates μ^{dis} in $\mathbf{P}^{\text{icst}}_d$; in which case, μ^{icst} satisfies all the principles beside $\mathbf{P}^{\text{icst}}_c$ and $\mathbf{P}^{\text{icst}}_d$.*

Proof. $\mathbf{P}^{\text{icst}}_a$ *Language invariance.* Since the blur function is language invariant (see principle $\mathbf{P}^{\mathbb{C}}_9$ on page 10), ie $C_{\mathfrak{K}} \stackrel{\text{i}}{\equiv} C_{\mathfrak{K} \oplus v}$, we have $\max_{\omega \in \Omega} C_{\mathfrak{K}}(\omega) = \max_{\omega' \in \Omega \oplus v} C_{\mathfrak{K}}(\omega')$, hence $\mu^{\text{icst}}(\mathfrak{K}) = \mu^{\text{icst}}(\mathfrak{K} \oplus v)$.

$\mathbf{P}^{\text{icst}}_b$ *Separation.* Since the blur function satisfies *unanimity* (see principle $\mathbf{P}^{\mathbb{C}}_8$ on page 10), ie $\forall \mathfrak{k} \in \mathfrak{K}_K, \Omega \cap \mathfrak{k} = \{\ \omega \in \Omega \mid C_{\mathfrak{k}}(\omega) = 1\ \}$, we have $\Omega \cap \bigcap_{\mathfrak{k} \in \mathfrak{K}_K} \mathfrak{k} = \{\ \omega \in \Omega \mid \Cup_{\mathfrak{k} \in \mathfrak{K}} C_{\mathfrak{k}}(\omega) = 1\ \} = \{\ \omega \in \Omega \mid C_{\mathfrak{K}}(\omega) = 1\ \}$. Thus, $\{\ \omega \in \Omega \mid C_{\mathfrak{K}}(\omega) = 1\ \} \neq \emptyset$ iff $\max_{\omega \in \Omega} C_{\mathfrak{K}}(\omega) = 1$ iff $\mu^{\text{icst}}(\mathfrak{K}_K) = 0$.

$\mathbf{P}^{\text{icst}}_c$ *Equivalence.* If $C_{\mathfrak{K}_1} \stackrel{\text{i}}{\equiv} C_{\mathfrak{K}_2}$, then $C_{\mathfrak{K}_1} = C_{\mathfrak{K}_2}$, hence $\max_{\omega \in \Omega} C_{\mathfrak{K}_1}(\omega)$ equals $\max_{\omega \in \Omega} C_{\mathfrak{K}_2}(\omega)$; this is not necessary the case when $C_{\mathfrak{K}_1} \stackrel{\text{e}}{\equiv} C_{\mathfrak{K}_2}$, ie $\hat{\Omega}_{C_{\mathfrak{K}_1}} = \hat{\Omega}_{C_{\mathfrak{K}_2}}$. Hence, μ^{icst} satisfies *equivalence* wrt $\stackrel{\text{i}}{\equiv}$, but not wrt $\stackrel{\text{e}}{\equiv}$. Notice that $\mu^{\text{icst}}(\mathfrak{K}_1) = \mu^{\text{icst}}(\mathfrak{K}_2)$ if $\mu^{\text{dis}}_{\mathcal{H}}(C_{\mathfrak{K}_1}, C_{\mathfrak{K}_2}) = 0$.

$\mathbf{P}^{\text{icst}}_d$ *Continuity.* We have $\lim_{i \to \infty} \mu^{\text{icst}}(\mathfrak{K}_i) = \mu^{\text{icst}}(\mathfrak{K})$ iff $\lim_{i \to \infty} \max_{\omega \in \Omega} C_{\mathfrak{K}_i}(\omega) = \max_{\omega \in \Omega} C_{\mathfrak{K}}(\omega)$ if $\lim_{i \to \infty} \mu^{\text{dis}}_{\mathcal{L}_\infty}(C_{\mathfrak{K}_i}, C_{\mathfrak{K}}) = 0$ or $\lim_{i \to \infty} \mu^{\text{dis}}_{\mathcal{H}}(C_{\mathfrak{K}_i}, C_{\mathfrak{K}}) = 0$, but not necessary if $\lim_{i \to \infty} \mu^{\text{dis}}_{\mathcal{H}}(C_{\mathfrak{K}_i}, C_{\mathfrak{K}}) = 0$.

$\mathbf{P}^{\text{icst}}_e$ *Monotonicity.* Let $\hat{\omega}$ be a best candidate of $C_{\mathfrak{K}_1}$. Let $\omega \in \Omega$ be a probability distribution. Thus, $C_{\mathfrak{K}_1}(\hat{\omega}) \geq C_{\mathfrak{K}_1}(\omega)$. Since $1 \geq C_{\mathfrak{K}_2}(\omega)$, we have

$C_{\mathfrak{K}_1}(\hat{\omega}) \geq C_{\mathfrak{K}_1}(\omega) * C_{\mathfrak{K}_2}(\omega) = (C_{\mathfrak{K}_1} \Cup C_{\mathfrak{K}_2})(\omega) = (C_{\mathfrak{K}_1 \cup \mathfrak{K}_2})(\omega)$. Let $\hat{\omega}_{12}$ be a best candidate of $C_{\mathfrak{K}_1 \cup \mathfrak{K}_2}$. Thus, $\mu^{\text{icst}}(\mathfrak{K}_1) = 1 - C_{\mathfrak{K}_1}(\hat{\omega}) \leq 1 - C_{\mathfrak{K}_1 \cup \mathfrak{K}_2}(\hat{\omega}_{12}) = \mu^{\text{icst}}(\mathfrak{K}_1 \cup \mathfrak{K}_2)$.

$\mathbf{P}_f^{\text{icst}}$ *Strict monotonicity.* Let $\hat{\omega}$ be a best candidate of $C_{\mathfrak{K}_1}$. Let $\omega \in \Omega$ be a probability distribution. Thus, $C_{\mathfrak{K}_1}(\hat{\omega}) \geq C_{\mathfrak{K}_1}(\omega)$. if $0 < \mu^{\text{icst}}(\mathfrak{K}_2)$ then $1 > C_{\mathfrak{K}_2}(\omega)$, and we have $C_{\mathfrak{K}_1}(\hat{\omega}) > C_{\mathfrak{K}_1}(\omega) * C_{\mathfrak{K}_2}(\omega) = (C_{\mathfrak{K}_1 \cup \mathfrak{K}_2})(\omega)$. Hence $\mu^{\text{icst}}(\mathfrak{K}_1) < \mu^{\text{icst}}(\mathfrak{K}_1 \cup \mathfrak{K}_2)$.

$\mathbf{P}_g^{\text{icst}}$ *Minimality.* By definition of $\mathfrak{K} \models \mathfrak{k}$, we have $\mathfrak{k} \supseteq \hat{\Omega}_{C_{\mathfrak{K}}}$. Let $\hat{\omega}$ be a best candidate of $C_{\mathfrak{K}}$; hence $\hat{\omega} \in \mathfrak{k}$. Since the blur function satisfies *unanimity* (see principle $\mathbf{P}_8^{\mathbb{C}}$ on page 10), we have $C_{\mathfrak{k}}(\hat{\omega}) = 1$. Hence, $\mu_{\mathfrak{K}}^{\text{culp}}(\mathfrak{k}) = 1 - C_{\mathfrak{k}}(\hat{\omega}) = 0$.

$\mathbf{P}_h^{\text{icst}}$ *Consequence invariance.* By definition of $\mathfrak{K} \models \mathfrak{k}$, we have $\mathfrak{k} \supseteq \hat{\Omega}_{C_{\mathfrak{K}}}$. Let $\hat{\omega}_1$ be a best candidate of $C_{\mathfrak{K}}$; hence $\hat{\omega}_1 \in \mathfrak{k}$. Let $\hat{\omega}_2$ be a best candidate of $C_{\mathfrak{K} \cup \{\mathfrak{k}\}}$. By *unanimity* (see principle $\mathbf{P}_8^{\mathbb{C}}$), we have $C_{\mathfrak{k}}(\hat{\omega}_1) = 1$. Hence, $C_{\mathfrak{K}}(\hat{\omega}_2) \leq C_{\mathfrak{K}}(\hat{\omega}_1) * 1 = C_{\mathfrak{K}}(\hat{\omega}_1) * C_{\mathfrak{k}}(\hat{\omega}_1) = C_{\mathfrak{K} \cup \{\mathfrak{k}\}}(\hat{\omega}_1) \leq C_{\mathfrak{K} \cup \{\mathfrak{k}\}}(\hat{\omega}_2) = C_{\mathfrak{K}}(\hat{\omega}_2) * C_{\mathfrak{k}}(\hat{\omega}_2)$. Therefore, $C_{\mathfrak{K}}(\hat{\omega}_2) \leq C_{\mathfrak{K}}(\hat{\omega}_2) * C_{\mathfrak{k}}(\hat{\omega}_2)$ then $C_{\mathfrak{k}}(\hat{\omega}_2) = 1$. Finally, we deduce from $C_{\mathfrak{K}}(\hat{\omega}_1) = C_{\mathfrak{K} \cup \{\mathfrak{k}\}}(\hat{\omega}_2)$ that $\mu^{\text{icst}}(\mathfrak{K}) = \mu^{\text{icst}}(\mathfrak{K} \cup \{\mathfrak{k}\})$.

$\mathbf{P}_i^{\text{icst}}$ *Dominance.* Let $\omega \in \Omega$ be a probability distribution. Firstly, suppose that ω is in $\mathfrak{k}_1$ (hence in $\mathfrak{k}_2$); thus, $C_{\mathfrak{k}_1}(\omega) = 1 = C_{\mathfrak{k}_1}(\omega)$ by *unanimity* (see principle $\mathbf{P}_8^{\mathbb{C}}$). Secondly, suppose that $\omega \notin \mathfrak{k}_1$ and $\omega \in \mathfrak{k}_2$; thus, $C_{\mathfrak{k}_1}(\omega) < 1 = C_{\mathfrak{k}_2}(\omega)$ by $\mathbf{P}_8^{\mathbb{C}}$. Thirdly, suppose that $\omega \notin \mathfrak{k}_1$ and $\omega \notin \mathfrak{k}_2$; thus, $C_{\mathfrak{k}_1}(\omega) < 1$ and $C_{\mathfrak{k}_2}(\omega) < 1$ by $\mathbf{P}_8^{\mathbb{C}}$, and by *proximity* (see principle $\mathbf{P}_7^{\mathbb{C}}$ on page 10), $\mathfrak{k}_1 \models \mathfrak{k}_2$ implies $\mathfrak{k}_2 \supseteq \mathfrak{k}_1$ hence ω is nearer to $\mathfrak{k}_2$ than to $\mathfrak{k}_1$: $C_{\mathfrak{k}_1}(\omega) \leq C_{\mathfrak{k}_2}(\omega) < 1$. Thus, for any probability distribution $\omega \in \Omega$, we have $C_{\mathfrak{k}_1}(\omega) \leq C_{\mathfrak{k}_2}(\omega)$. Hence, $C_{\mathfrak{K} \cup \{\mathfrak{k}_1\}} = C_{\mathfrak{K}} * C_{\{\mathfrak{k}_1\}} \leq C_{\mathfrak{K}} * C_{\{\mathfrak{k}_2\}} = C_{\mathfrak{K} \cup \{\mathfrak{k}_2\}}$. Therefore, $\mu^{\text{icst}}(\mathfrak{K} \cup \{\mathfrak{k}_1\}) \geq \mu^{\text{icst}}(\mathfrak{K} \cup \{\mathfrak{k}_2\})$.

$\mathbf{P}_j^{\text{icst}}$ *Equitable distribution.* Suppose that $\mu_{\mathfrak{K}}^{\text{culp}}(\mathfrak{k})$ is defined for every knowledge item $\mathfrak{k}$ of $\mathfrak{K}$ (see Def. 28). Let $f(\vec{x}) \stackrel{\text{def}}{=} 1 - \prod(\vec{1} - \vec{x})$, where we denote by $\vec{x}$ the vector of culpability measures of the knowledge items of $\mathfrak{K}$. Let $\widetilde{\omega} \in \widetilde{\Omega}_{C_{\mathfrak{K}}}$. Thus, $\prod(\vec{1} - \vec{x})$ equals $\prod_{\mathfrak{k} \in \mathfrak{K}} 1 - \mu_{\mathfrak{K}}^{\text{culp}}(\mathfrak{k})$ equals $\sqrt[|\widetilde{\Omega}_0| + p]{\prod_{\widetilde{\omega} \in \widetilde{\Omega}_0} \prod_{\mathfrak{k} \in \mathfrak{K}} C_{\mathfrak{k}}(\widetilde{\omega}) * \prod_{i=1}^{p} \prod_{\mathfrak{k} \in \mathfrak{K}} \vartheta(\widetilde{\Omega}_i, \mathfrak{k})}$ where $\prod_{\mathfrak{k} \in \mathfrak{K}} \vartheta(\widetilde{\Omega}_i, \mathfrak{k}) = \exp\left(\frac{\int_{\widetilde{\Omega}_i} \ln(\prod_{\mathfrak{k} \in \mathfrak{K}} C_{\mathfrak{k}}(\widetilde{\omega})) \mathrm{d}\widetilde{\omega}}{\int_{\widetilde{\Omega}_i} 1 \mathrm{d}\widetilde{\omega}}\right)$. Since $C_{\mathfrak{K}}(\widetilde{\omega}) = \prod_{\mathfrak{k} \in \mathfrak{K}} C_{\mathfrak{k}}(\widetilde{\omega})$ is constant, $\prod_{\mathfrak{k} \in \mathfrak{K}} \vartheta(\widetilde{\Omega}_i, \mathfrak{k}) = \exp\left(\frac{\int_{\widetilde{\Omega}_i} \ln(C_{\mathfrak{K}}(\widetilde{\omega})) \mathrm{d}\widetilde{\omega}}{\int_{\widetilde{\Omega}_i} 1 \mathrm{d}\widetilde{\omega}}\right) = \exp\left(\ln(C_{\mathfrak{K}}(\widetilde{\omega})) * \frac{\int_{\widetilde{\Omega}_i} 1 \mathrm{d}\widetilde{\omega}}{\int_{\widetilde{\Omega}_i} 1 \mathrm{d}\widetilde{\omega}}\right) = C_{\mathfrak{K}}(\widetilde{\omega})$, and $\prod(\vec{1} - \vec{x})$ equals $\sqrt[|\widetilde{\Omega}_0| + p]{\prod_{\widetilde{\omega} \in \widetilde{\Omega}_0} C_{\mathfrak{K}}(\widetilde{\omega}) * \prod_{i=1}^{p} C_{\mathfrak{K}}(\widetilde{\omega})}$, which equals $C_{\mathfrak{K}}(\widetilde{\omega})$. Since $\widetilde{\omega}$ is a best candidate of $C_{\mathfrak{K}}$, we conclude that $f(\vec{x}) = 1 - C_{\mathfrak{K}}(\widetilde{\omega}) = \mu^{\text{icst}}(\mathfrak{K})$. □

Shapley Inconsistency Value

Besides, another culpability measure could be the *Shapley Inconsistency Value*, SIV for short, which appears in [18, Def. 9] for set of propositions. Let I be an inconsistency measure for knowledge contents, like μ^{icst}. Thus the SIV-based culpability measure of a knowledge item $\mathfrak{k}$ belonging to a knowledge content $\mathfrak{K}$ is defined as follows (where $\subseteq$ operates on multisets):

$$\mu_{I,\mathfrak{K}}^{\text{SIV}}(\mathfrak{k}) \stackrel{\text{def}}{=} \sum_{\mathfrak{C} \subseteq \mathfrak{K}} \frac{(|\mathfrak{C}|-1)!(|\mathfrak{K}|-|\mathfrak{C}|)!}{|\mathfrak{K}|!} (I(\mathfrak{C}) - I(\mathfrak{C} \setminus \{\mathfrak{k}\}))$$

Moreover, an SIV-based culpability measure induces the following SIV-based inconsistency measure (see [18, Def. 10]):

$$\mu_I^{\text{SIV}}(\mathfrak{K}) \stackrel{\text{def}}{=} \max_{\mathfrak{k} \in \mathfrak{K}} \mu_{I,\mathfrak{K}}^{\text{SIV}}(\mathfrak{k})$$

Further studies might define $I(\mathfrak{K})$ as the number of minimal inconsistent subsets of $\mathfrak{K}$, like in [18, Def. 11]:

$$I_{\text{MI}}(\mathfrak{K}) \stackrel{\text{def}}{=} |\{\, \mathfrak{C} \subseteq \mathfrak{K} \mid \Omega_{\mathfrak{C}} = \emptyset \text{ and } \forall \mathfrak{B} \subset \mathfrak{C}, \Omega_{\mathfrak{B}} \neq \emptyset \,\}|$$

While $\mu_{I_{\text{MI}}}^{\text{SIV}}$ satisfies significant principles like *language invariance*, it is not *continuous*, wrt neither $\mu_{\mathcal{L}_\infty}^{\text{dis}}$, $\mu_{\mathcal{H}}^{\text{dis}}$, nor $\mu_{\mathcal{H}}^{\text{dis}}$.

Towards a Σ-culpability measure

Our inconsistency measure (see Def. 26) together with its culpability measure (see Def. 28) does not satisfy *distribution* property proposed in [18, Def. 12]. This property expresses that the culpability measures of the items of a knowledge content $\mathfrak{K}$ should sum up to the inconsistency measure of $\mathfrak{K}$. By comparing this property with *equitable distribution* (see principle $\mathbf{P}_j^{\text{icst}}$), we remark that the former is an instance of the latter: *distribution* property forces the symmetric function f in $\mathbf{P}_j^{\text{icst}}$ to be $f(\vec{x}) \stackrel{\text{def}}{=} \sum \vec{x}$. The following inconsistency measure together with its Σ-culpability measure satisfies *distribution* property: $\mu^{\Sigma\text{icst}}(\mathfrak{K}) \stackrel{\text{def}}{=} -\ln(1 - \mu^{\text{icst}}(\mathfrak{K}))$ and $\mu_{\mathfrak{K}}^{\Sigma\text{culp}}(\mathfrak{k}) \stackrel{\text{def}}{=} -\ln(1 - \mu_{\mathfrak{K}}^{\text{culp}}(\mathfrak{k}))$. Notice that $\mu^{\Sigma\text{icst}}$ satisfies principles $\mathbf{P}_a^{\text{icst}}$ to $\mathbf{P}_j^{\text{icst}}$ since $-\ln(1-x)$ is not only continuously and strictly increasing when $x \in [0{:}1]$, but also equal to 0 when $x = 0$.

3.3.4 Conclusions

Our main contribution is twofold: firstly, we establish several principles to be satisfied by an inconsistency measure when applied to a knowledge content. Secondly, we define an inconsistency measure, together

with its culpability measure, that satisfies these principles. To our knowledge, neither principles nor inconsistency measures for knowledge bases in $\mathbb{K}$ exist in the literature.

Proposition 13 not only shows that our inconsistency measure is robust against slight fluctuations in the knowledge base, since it satisfies *continuity* (see principle $\mathbf{P}_d^{\text{icst}}$), but also that our inconsistency measure can be utilised to compare knowledge bases about different topics, since it satisfies *language invariance* (see principle $\mathbf{P}_a^{\text{icst}}$).

In future research, our inconsistency measure should be compared to the one defined in [18] for propositional knowledge bases, which are a special case of $\mathbb{K}^L$, and especially to the one defined in [42] for conditional probabilistic knowledge bases, which is again a subset of $\mathbb{K}$.

3.4 Incoherence measure μ^{icoh}: to quantify the consensus gap

An incoherence measure computes how far two candidacy functions C_1 and C_2 are from reaching a consensus, where these candidacy functions are underlain by a common propositional language.

3.4.1 Principles

$\mathbf{P}_{\text{I}}^{\text{icoh}}$ *Language invariance.* An incoherence measure is invariant by language enrichment.

$$\mu^{\text{icoh}}(C_1, C_2) = \mu^{\text{icoh}}(C_1 \oplus v, C_2 \oplus v)$$

$\mathbf{P}_{\text{II}}^{\text{icoh}}$ *Separation.* Candidacy functions are not incoherent iff they nominate at least one same candidate.

$$\mu^{\text{icoh}}(C_1, C_2) = 0 \text{ iff } \hat{\Omega}_{C_1} \cap \hat{\Omega}_{C_2} \neq \emptyset$$

$\mathbf{P}_{\text{III}}^{\text{icoh}}$ *Equivalence.* Equivalent pairs of candidacy functions have equal incoherence measure.

$$\begin{array}{l}\text{if } C_1 \equiv C_3 \text{ and } C_2 \equiv C_4\\ \text{then } \mu^{\text{icoh}}(C_1, C_2) = \mu^{\text{icoh}}(C_3, C_4)\end{array}$$

$\mathbf{P}_{\text{IV}}^{\text{icoh}}$ *Continuity.* When two candidacy functions change continuously, so their incoherence measure does, wrt a certain notion of convergence (see Def. 25 on page 20). This principle (together with *symmetry*, see $\mathbf{P}_{\text{V}}^{\text{icoh}}$) ensures a certain robustness for μ^{icoh} in the face of minor fluctuations.

$$\begin{array}{l}\text{if } \lim_{i\to\infty} \mu^{\text{dis}}(C_i, C) = 0 \text{ and } X \in \mathbb{C}\\ \text{then } \lim_{i\to\infty} \mu^{\text{icoh}}(C_i, X) = \mu^{\text{icoh}}(C, X)\end{array}$$

$\mathbf{P}_{\text{V}}^{\text{icoh}}$ *Symmetry.* An incoherence measure is commutative.

$$\mu^{\text{icoh}}(C_1, C_2) = \mu^{\text{icoh}}(C_2, C_1)$$

$\mathbf{P}_{\text{VI}}^{\text{icoh}}$ *Consequence invariance.* The incoherence measure between C_1 and C_2 equals the one between C_2 and the mergence of C_1 with one of its consequences C.

$$\text{if } C_1 \models C \text{ then } \mu^{\text{icoh}}(C_1, C_2) = \mu^{\text{icoh}}(C_1 \Cup C, C_2)$$

3.4.2 Vertical incoherence measure μ_V^{icoh}: potential versus real consistency degrees

Definition 29 (Vertical incoherence measure)**.** *The vertical incoherence measure of two candidacy functions C_1 and C_2 is the difference between the potential maximal "consistency degree", ie $C_1(\hat{\omega}_1) * C_2(\hat{\omega}_2)$, and the real "consistency degree", ie $(C_1 \Cup C_2)(\hat{\omega}_{12})$, where $\hat{\omega}_1 \overset{\text{def}}{\in} \hat{\Omega}_{C_1}$, $\hat{\omega}_2 \overset{\text{def}}{\in} \hat{\Omega}_{C_2}$, and $\hat{\omega}_{12} \overset{\text{def}}{\in} \hat{\Omega}_{C_1 \Cup C_2}$:*

$$\mu_V^{\text{icoh}}(C_1, C_2) \overset{\text{def}}{=} C_1(\hat{\omega}_1) * C_2(\hat{\omega}_2) - (C_1 \Cup C_2)(\hat{\omega}_{12})$$

Proposition 14. *μ_V^{icoh} satisfies principles $\mathbf{P}_{\text{I}}^{\text{icoh}}$, $\mathbf{P}_{\text{II}}^{\text{icoh}}$, $\mathbf{P}_{\text{III}}^{\text{icoh}}$ wrt $\overset{\text{i}}{\equiv}$ but not wrt $\overset{\text{e}}{\equiv}$, $\mathbf{P}_{\text{V}}^{\text{icoh}}$, but not $\mathbf{P}_{\text{IV}}^{\text{icoh}}$; μ_V^{icoh} satisfies nevertheless a weaker form of* continuity*.*

Proof. $\mathbf{P}_{\text{I}}^{\text{icoh}}$ *Language invariance.* From the definition of $\oplus$, we have $\forall \hat{\omega} \in \hat{\Omega}_C, \forall \hat{\omega}' \in \hat{\omega} \oplus v, C(\hat{\omega}) = (C \oplus v)(\hat{\omega}')$. Let $\omega_1' \in \omega_1 \oplus v$, $\omega_2' \in \omega_2 \oplus v$, and $\omega_{12}' \in \omega_{12} \oplus v$. Therefore, $\mu_V^{\text{icoh}}(C_1, C_2) = C_1(\hat{\omega}_1) * C_2(\hat{\omega}_2) - (C_1 \Cup C_2)(\hat{\omega}_{12}) = (C_1 \oplus v)(\hat{\omega}_1') * (C_2 \oplus v)(\hat{\omega}_2') - (C_1 \oplus v \Cup C_2 \oplus v)(\hat{\omega}_{12}') = \mu_V^{\text{icoh}}(C_1 \oplus v, C_2 \oplus v)$.

$\mathbf{P}_{\text{II}}^{\text{icoh}}$ *Separation.* $\mu_V^{\text{icoh}}(C_1, C_2) = 0$ iff $C_1(\hat{\omega}_1) * C_2(\hat{\omega}_2) = (C_1 \Cup C_2)(\hat{\omega}_{12})$ iff $C_1(\hat{\omega}_1) * C_2(\hat{\omega}_2) = C_1(\hat{\omega}_{12}) * C_2(\hat{\omega}_{12})$ iff $C_1(\hat{\omega}_1) = C_1(\hat{\omega}_{12})$ and $C_2(\hat{\omega}_2) = C_2(\hat{\omega}_{12})$ because C_1 is maximal for $\hat{\omega}_1$, and C_2 is maximal for $\hat{\omega}_2$. Therefore, $\mu_V^{\text{icoh}}(C_1, C_2) = 0$ iff $\hat{\omega}_{12} \in \hat{\Omega}_{C_1} \cap \hat{\Omega}_{C_2}$ iff $\hat{\Omega}_{C_1 \Cup C_2} \subseteq \hat{\Omega}_{C_1} \cap \hat{\Omega}_{C_2}$. Notice furthermore that $\hat{\Omega}_{C_1} \cap \hat{\Omega}_{C_2} \subseteq \hat{\Omega}_{C_1 \Cup C_2}$ always holds because $\hat{\omega} \in \hat{\Omega}_{C_1} \cap \hat{\Omega}_{C_2}, \forall \omega \in \Omega, (C_1 \Cup C_2)(\omega) = C_1(\omega) * C_2(\omega) \leq C_1(\hat{\omega}) * C_2(\hat{\omega}) = (C_1 \Cup C_2)(\hat{\omega})$. We thus conclude that $\hat{\Omega}_{C_1} \cap \hat{\Omega}_{C_2} \neq \emptyset$ iff $\hat{\Omega}_{C_1} \cap \hat{\Omega}_{C_2} = \hat{\Omega}_{C_1 \Cup C_2}$ iff $\mu_V^{\text{icoh}}(C_1, C_2) = 0$.

$\mathbf{P}_{\text{III}}^{\text{icoh}}$ *Equivalence.* In case $\Theta(C_1) = \Theta(C_2) = \Theta(C_3) = \Theta(C_4)$, ($C_1 \overset{\text{i}}{\equiv} C_3$ and $C_2 \overset{\text{i}}{\equiv} C_4$) iff ($C_1 = C_3$ and $C_2 = C_4$). Therefore, $\mu_V^{\text{icoh}}(C_1, C_2) = \mu_V^{\text{icoh}}(C_3, C_4)$. The following counter-example shows that μ_V^{icoh} does not satisfy $\mathbf{P}_{\text{III}}^{\text{icoh}}$ wrt $\overset{\text{e}}{\equiv}$. Recall that ($C_1 \overset{\text{e}}{\equiv} C_3$ and $C_2 \overset{\text{e}}{\equiv} C_4$) iff ($\hat{\Omega}_{C_1} = \hat{\Omega}_{C_3}$ and $\hat{\Omega}_{C_2} = \hat{\Omega}_{C_4}$). Suppose that $C_1 = C_3$, $\hat{\Omega}_{C_2} = \hat{\Omega}_{C_4}$, $\hat{\Omega}_{C_1} \cap \hat{\Omega}_{C_2} = \emptyset$, $\forall \hat{\omega} \in \hat{\Omega}_{C_4}, C_2(\hat{\omega}) = C_4(\hat{\omega})$, $\forall \omega \notin \hat{\Omega}_{C_4}, C_2(\omega) = \frac{C_4(\omega)}{2} > 0$ where $\Omega \setminus \hat{\Omega}_{C_2} \neq \emptyset$, and $\hat{\Omega}_{C_3 \Cup C_4} \cap \left(\hat{\Omega}_{C_3} \cup \hat{\Omega}_{C_4}\right)$. We thus have $C_1(\hat{\omega}_1) * C_2(\hat{\omega}_2) = C_3(\hat{\omega}_3) * C_4(\hat{\omega}_4)$, and $(C_1 \Cup C_2)(\hat{\omega}_{12}) = (C_3 \Cup C_2)(\hat{\omega}_{34})$, where $\hat{\omega}_i \in \hat{\Omega}_{C_i}$ and

$\hat{\omega}_{ij} \in \hat{\Omega}_{C_i \Cup C_j}$. However, $(C_3 \Cup C_2)(\hat{\omega}_{34}) = C_3(\hat{\omega}_{34}) * C_2(\hat{\omega}_{34}) = C_3(\hat{\omega}_{34}) * \frac{C_4(\hat{\omega}_{34})}{2} = \frac{(C_3 \Cup C_4)(\hat{\omega}_{34})}{2}$. Therefore, $\mu_V^{\text{icoh}}(C_1, C_2) = C_1(\hat{\omega}_1) * C_2(\hat{\omega}_2) - (C_1 \Cup C_2)(\hat{\omega}_{12}) = C_3(\hat{\omega}_3) * C_4(\hat{\omega}_4) - \frac{(C_3 \Cup C_4)(\hat{\omega}_{34})}{2} > \mu_V^{\text{icoh}}(C_3, C_4)$ whereas ($C_1 \overset{\text{e}}{\equiv} C_3$ and $C_2 \overset{\text{e}}{\equiv} C_4$).

$\mathbf{P}_{\text{IV}}^{\text{icoh}}$ *Continuity.* We now show that μ_V^{icoh} satisfies a weaker form of continuity, but not $\mathbf{P}_{\text{IV}}^{\text{icoh}}$. Let $\hat{\omega}_i \in \hat{\Omega}_{C_i}$, $\hat{\omega}_C \in \hat{\Omega}_C$, $\hat{\omega}_X \in \hat{\Omega}_X$, $\hat{\omega}_{iX} \in \hat{\Omega}_{C_i \Cup X}$, and $\hat{\omega}_{CX} \in \hat{\Omega}_{C \Cup X}$ be five best candidates. Thus we have $\lim_{i\to\infty} \mu_V^{\text{icoh}}(C_i, X) = \mu_V^{\text{icoh}}(C, X)$ iff $\lim_{i\to\infty} C_i(\hat{\omega}_i) * X(\hat{\omega}_X) - (C_i \Cup X)(\hat{\omega}_{iX}) = C(\hat{\omega}_C) * X(\hat{\omega}_X) - (C \Cup X)(\hat{\omega}_{CX})$, if not only $\lim_{i\to\infty} \mu_{\mathcal{L}_\infty}^{\text{dis}}(C_i, C) = 0$ but also both $\lim_{i\to\infty} \mu_{\mathcal{H}}^{\text{dis}}(C_i, C) = 0$ and $\lim_{i\to\infty} \mu_{\mathcal{H}}^{\text{dis}}(C_i \Cup X, C \Cup X) = 0$. Remark that the proof of Prop. 12 exhibit a counter-example where $\lim_{i\to\infty} \mu_{\mathcal{L}_\infty}^{\text{dis}}(C_i, C) = 0$ does not imply $\lim_{i\to\infty} \mu_{\mathcal{H}}^{\text{dis}}(C_i, C) = 0$.

$\mathbf{P}_{\text{V}}^{\text{icoh}}$ *Symmetry.* Let $\hat{\omega}_i \in \hat{\Omega}_{C_i}$ and $\hat{\omega}_{ij} \in \hat{\Omega}_{C_i \Cup C_j}$, where $i \in \{1, 2\}$. Thus, from the commutativity of $*$ and $\Cup$, it follows that $\hat{\Omega}_{C_2 \Cup C_1} = \hat{\Omega}_{C_1 \Cup C_2}$ and $\mu_V^{\text{icoh}}(C_1, C_2) = C_1(\hat{\omega}_1) * C_2(\hat{\omega}_2) - (C_1 \Cup C_2)(\hat{\omega}_{12}) = C_2(\hat{\omega}_2) * C_1(\hat{\omega}_1) - (C_2 \Cup C_1)(\hat{\omega}_{21}) = \mu_V^{\text{icoh}}(C_2, C_1)$.

□

3.4.3 Gap-based incoherence measure $\mu_{\mathcal{G}}^{\text{icoh}}$: the gap between the best candidates

Definition 30. *A gap-based incoherence measure computes the gap between the best candidates of two candidacy functions C_1 and C_2:*

$$\mu_{\mathcal{G}}^{\text{icoh}}(C_1, C_2) \overset{\text{def}}{=} \sqrt{2^n} * \mathcal{G}(\hat{\Omega}_{C_1}, \hat{\Omega}_{C_2})$$

Proposition 15. *$\mu_{\mathcal{G}}^{\text{icoh}}$ satisfies principles $\mathbf{P}_{\text{II}}^{\text{icoh}}$, $\mathbf{P}_{\text{III}}^{\text{icoh}}$ wrt $\overset{\text{e}}{\equiv}$ but not wrt $\overset{\text{i}}{\equiv}$, $\mathbf{P}_{\text{IV}}^{\text{icoh}}$ wrt $\mu_{\mathcal{H}}^{\text{dis}}$ (or $\mu_{\hat{\mathcal{H}}}^{\text{dis}}$) but not wrt $\mu_{\mathcal{L}_\infty}^{\text{dis}}$, $\mathbf{P}_{\text{V}}^{\text{icoh}}$, and $\mathbf{P}_{\text{VI}}^{\text{icoh}}$.*

Proof. $\mathbf{P}_{\text{II}}^{\text{icoh}}$ *Separation.* $\mu_V^{\text{icoh}}(C_1, C_2) = 0$ iff $\mathcal{G}(\hat{\Omega}_{C_1}, \hat{\Omega}_{C_2}) = 0$ iff $\hat{\Omega}_{C_1} \cap \hat{\Omega}_{C_2} \neq \emptyset$.

$\mathbf{P}_{\text{III}}^{\text{icoh}}$ *Equivalence.* Since $\overset{\text{i}}{\equiv}$ implies $\overset{\text{e}}{\equiv}$, it suffice to prove that $\mu_{\mathcal{G}}^{\text{icoh}}$ satisfies $\mathbf{P}_{\text{III}}^{\text{icoh}}$ wrt $\overset{\text{e}}{\equiv}$. If ($C_1 \overset{\text{e}}{\equiv} C_3$ and $C_2 \overset{\text{e}}{\equiv} C_4$) when a common underlying language is assumed, then ($\hat{\Omega}_{C_1} = \hat{\Omega}_{C_3}$ and $\hat{\Omega}_{C_2} = \hat{\Omega}_{C_4}$). Therefore, $\mathcal{G}(\hat{\Omega}_{C_1}, \hat{\Omega}_{C_2}) = \mathcal{G}(\hat{\Omega}_{C_3}, \hat{\Omega}_{C_4})$, as required.

$\mathbf{P}_{\text{IV}}^{\text{icoh}}$ *Continuity.* $\mu_{\mathcal{G}}^{\text{icoh}}$ is continuous wrt $\mu_{\mathcal{H}}^{\text{dis}}$ if $\lim_{i\to\infty} \mathcal{H}(\hat{\Omega}_{C_i}, \hat{\Omega}_C) = 0$ iff $\lim_{i\to\infty} \mu_{\mathcal{H}}^{\text{dis}}(C_i, C) = 0$. However, $\mu_{\mathcal{G}}^{\text{icoh}}$ is not continuous wrt $\mu_{\mathcal{L}_\infty}^{\text{dis}}$ for the same reason as $\lim_{i\to\infty} \mu_{\mathcal{L}_\infty}^{\text{dis}}(C_i, C) = 0$ does not imply $\lim_{i\to\infty} \mu_{\mathcal{H}}^{\text{dis}}(C_i, C) = 0$ (see Prop. 12).

$\mathbf{P}_{\text{V}}^{\text{icoh}}$ *Symmetry.* Since $\Cup$ and $\mathcal{G}$ are symmetric, $\mu_{\mathcal{G}}^{\text{icoh}}$ is symmetric.

$\mathbf{P}_{\text{VI}}^{\text{icoh}}$ *Consequence invariance.* If $C_1 \models C$ then $\forall \hat{\omega} \in \hat{\Omega}_{C_1}, C(\hat{\omega}) = 1$, hence $\hat{\Omega}_{C_1 \Cup C} = \hat{\Omega}_{C_1}$ Therefore, $\mu_{\mathcal{G}}^{\text{icoh}}(C_1, C_2) = \sqrt{2^n} * \mathcal{G}(\hat{\Omega}_{C_1}, \hat{\Omega}_{C_2}) = \sqrt{2^n} * \mathcal{G}(\hat{\Omega}_{C_1 \Cup C}, \hat{\Omega}_{C_2}) = \mu_{\mathcal{G}}^{\text{icoh}}(C_1 \Cup C, C_2)$.

□

We furthermore stress that $\mu_{\mathcal{G}}^{\text{icoh}}$ is σ-invariant, since Prop. 9 on page 14 states that the best candidates are σ-invariant. Besides, we conjecture that $\mu_{\mathcal{G}}^{\text{icoh}}$ satisfies *language invariance* (see principle $\mathbf{P}_{\text{I}}^{\text{icoh}}$).

On measuring the surprise If we deem the surprise to be the incoherence between what common sense would dictate us to believe wrt our beforehand knowledge C_{prior}, and what someone tells us now C_{now}, then we might define our first surprise measure as follows:

$$\mu_1^{\text{surp}}(C_{\text{now}}, C_{\text{prior}}) \overset{\text{def}}{=} \mu^{\text{icoh}}(C_{\text{now}}, C_{\mathcal{I}(C_{\text{prior}})})$$

where $\mathcal{I}$ is a commonsensical inference process (see chapter 4) returning the set of elected probability distributions for representing the real world; this set is interpreted as a knowledge item $\mathfrak{k}$, and μ_1^{surp} is the incoherence between C_{now} and $C_{\mathfrak{k}}$. Differently, and if we suppose that $\mathcal{I}$ to be an inference process returning a *unique* probability distribution, we then might define our surprise measure as the Kullback–Leibler divergence (or relative entropy) from $\{\omega_{\text{prior}}\} \overset{\text{def}}{=} \mathcal{I}(C_{\text{prior}})$ to $\{\omega_{\text{now}}\} \overset{\text{def}}{=} \mathcal{I}(C_{\text{now}})$:

$$\mu_{\text{KL}}^{\text{surp}}(C_{\text{now}}, C_{\text{prior}}) \overset{\text{def}}{=} \sum_{\alpha \in \alpha_\Theta} \omega_{\text{now}}(\alpha) * \ln\left(\frac{\omega_{\text{now}}(\alpha)}{\omega_{\text{prior}}(\alpha)}\right)$$

$\mu_{\text{KL}}^{\text{surp}}$ is related to the formal Bayesian theory of surprise appearing in [19].

On considering candidacy degrees We also propose an incoherence measure $\mu_{\tilde{\mathcal{G}}}^{\text{icoh}}$ which is similar to $\mu_{\mathcal{G}}^{\text{icoh}}$ but takes into account the candidacy degree of the best candidates. Thus, it satisfies a stronger version of *separation* (see principle $\mathbf{P}_{\text{II}}^{\text{icoh}}$): "candidacy functions are not incoherent iff they nominate at least one same candidate with the same candidacy degree".

$$\mu_{\tilde{\mathcal{G}}}^{\text{icoh}}(C_1, C_2) \overset{\text{def}}{=} \sqrt{2^n} * \mathcal{G}\left(\left\{ [\hat{\omega}; C_1(\hat{\omega})] \;\middle|\; \hat{\omega} \in \hat{\Omega}_{C_1} \right\}, \left\{ [\hat{\omega}; C_2(\hat{\omega})] \;\middle|\; \hat{\omega} \in \hat{\Omega}_{C_2} \right\}\right)$$

3.5 Precision measure μ^{pre}

Our precision measure counts the number of best candidates $\hat{\Omega}_C$ for representing the real world, wrt a candidacy function C. These best candidates are graphically represented by manifolds of different dimensions

in a Euclidean space. From the hypervolumes of these heterogeneous manifolds, our measure computes a real number representing the precision of C. Thus, $\mu^{\text{pre}}(C)$ quantifies the range of choice[4] an inference process (see chapter 4) have for electing one probability distribution from the best candidates: the more numerous the best candidates, the less precise the candidacy function.

3.5.1 Introduction

In a multisensor context, it may be useful to rank each sensor (or sensor group) from the most precise to the least in order to reconfigure the latter. This rank is defined by a *precision measure* that orders the knowledge (represented as candidacy functions) provided by the sensors. Again, to our knowledge, no precision measure exists for linear knowledge bases. We thus strive to exhibit the first principled precision measure for $\mathbb{C}$.

Introductory example Suppose we temporarily note $C^{[a:b]} \stackrel{\text{def}}{=} C_{\{\text{“}b\geq\omega(v)\text{”};\text{“}-a\geq-\omega(v)\text{”}\}}$ the candidacy function corresponding to the following knowledge: “the probability of v, ie getting a tail after tossing a coin, is between a and b”. If we have no information about this coin, then our initial knowledge is empty, ie $C^{[0\%:100\%]}$ is tautological. If someone tells us that this coin is perfectly designed such that the probability of getting a tail after tossing the coin is exactly 50%, then our updated knowledge is $C^{[50\%:50\%]}$. We claim that $C^{[0\%:100\%]}$ is strictly less precise than $C^{[50\%:50\%]}$, ie $C^{[0\%:100\%]} \prec^{\text{pre}} C^{[50\%:50\%]}$ where $\prec^{\text{pre}}$ is the strict order relation induced by a precision ordering $\preceq^{\text{pre}}$. We shall also attempt at defining a precision measure $\mu^{\text{pre}} : \mathbb{C} \mapsto \mathbb{R}$ such that $C_1 \preceq^{\text{pre}} C_2$ iff $\mu^{\text{pre}}(C_1) \leq \mu^{\text{pre}}(C_2)$. We furthermore claim that $C^{[0\%:100\%]} \prec^{\text{pre}} C^{[80\%:90\%]} \prec^{\text{pre}} C^{[0\%:50\%]}$.

The intuition behind $\preceq^{\text{pre}}$ (hence μ^{pre}) is that the smaller $[a{:}b]$, the more precise $C^{[a:b]}$. Word *smaller* refers to a notion of cardinality of the best candidates of $C^{[a:b]}$, which is roughly $b-a$ here (in fact, we shall show in the proof of Prop. 20 that $\hat{\Omega}_{C^{[a:b]}}$ does not necessary contain as many best candidates as $\hat{\Omega}_{C^{[a-\delta:b-\delta]}}$, where $0 < \delta \leq a \leq b$). We could define $C_1 \preceq^{\text{pre}} C_2$ as $\text{Cardinality}(\hat{\Omega}_{C_1}) \leq \text{Cardinality}(\hat{\Omega}_{C_2})$ if the set of best candidates would be finite (like $\hat{\Omega}_{C^{[50\%:50\%]}}$), but it can be infinite (like $\hat{\Omega}_{C^{[80\%:90\%]}}$). We shall therefore introduce a function $\mathcal{V}$ that returns the volume occupied by the best candidates in a Euclidean space: thus, $C_1 \preceq^{\text{pre}} C_2$ iff $\mathcal{V}(\hat{\Omega}_{C_1}) \leq^{\mathcal{V}} \mathcal{V}(\hat{\Omega}_{C_2})$. A similar idea of *concentration of possibilities* appears in [10] as the peakedness of probability distributions; in this thesis, we attempt to measure the peakedness of candidacy functions at the external level of knowledge content (see § 2.3.6 on page 15). Besides, remark that $C^{[a:b]}$ is as informative as $C^{[50\%:50\%]}$ according to the well known *entropy* information measure, for any a and b such that $0\% \leq a \leq 50\% \leq b \leq 100\%$. This entropy-based information measure is used in [23] to quantify the “information” of *one* probability distribution representative of a set of propositional sentences.

In the reminder of this section, we establish several principles for precision measures. We then intuitively and naively define our first precision measure, which satisfies most of the principles, yet not the *language invariance* (see principle $\mathbf{P}_A^{\text{preci}}$) stating that a measure should not change when the knowledge domain is dynamically enriched with new topics (ie new propositional variables); this occurs when a new kind of sensor is plugged into the multisensor system. Thus, this principle must be satisfied in such a system. After investigating the reason for our first measure to not being language invariant, we define a second precision measure holding this property, yet releasing other principles.

3.5.2 Principles

In this section, we state several principles to be satisfied by a precision measure μ^{pre} returning a positive real number when applied to a candidacy function. Most of these principles extend those listed in [3, page 224] that relates the works of Lozinskii [28], Konieczny [24], and Knight [23] about information measures for sets of propositional sentences.

$\mathbf{P}_A^{\text{preci}}$ *Language invariance.* A precision measure is invariant by language enrichment. This principle extends property 10 of [3, page 224].

$$\mu^{\text{pre}}(C) = \mu^{\text{pre}}(C \oplus v)$$

$\mathbf{P}_B^{\text{preci}}$ *Lower bound.* A candidacy function C nominates each probability distribution to be a best candidate iff C is not precise. This principle extends properties 1 and 8 of [3, page 225], which state that tautologies (eg $1_{\mathbb{C}}$) and contradictions (eg $0_{\mathbb{C}}$) have null information value.

$$\hat{\Omega}_C = \Omega \text{ iff } \mu^{\text{pre}}(C) = 0$$

$\mathbf{P}_C^{\text{preci}}$ *Singleton bound.* The most precise candidacy functions underlain by a propositional language Θ, denoted by $\mathcal{M}(\Theta)$, nominate a unique best candidate $\hat{\omega}$ such that $\hat{\omega}(\alpha) = 1$, where α is a minterm of Θ.

[4] Notice our definition for a precision measure is *not* the quantification of the difficulty (instead of the range of choice) for an inference process to elect one probability distribution from the best candidates. For example, if we consider inference process $\mathcal{I}^E_{\text{ME}}$, then it is easier to select one probability distribution from a convex set of best candidates than from the best candidates of a symmetric multimodal candidacy function centred in Ω.

This principle extends property 6 of [3, page 225].

$$\mathcal{M}(\Theta) \stackrel{\text{def}}{=} \left\{ C \in \mathbb{C} \;\middle|\; \begin{array}{l} \Theta(C) = \Theta \text{ and} \\ \hat{\Omega}_C = \{\hat{\omega}\} \text{ and} \\ \exists \alpha \in \alpha_\Theta, \hat{\omega}(\alpha) = 1 \end{array} \right\}$$

if $C_1 \notin \mathcal{M}(\Theta)$ and $C_2 \in \mathcal{M}(\Theta)$ and $\Theta(C_1) = \Theta(C_2)$ then $\mu^{\text{pre}}(C_1) < \mu^{\text{pre}}(C_2)$

$\mathbf{P}_D^{\text{preci}}$ *Language bound.* The most precise candidacy functions underlain by a given propositional language Θ are less precise than the most precise candidacy functions underlain by $\Theta \oplus v$, which is more expressive than Θ. This principle extends property $9'$ of [3, page 226]:

$$\begin{array}{l} \text{if } C_1 \in \mathcal{M}(\Theta) \text{ and } C_2 \in \mathcal{M}(\Theta \oplus v) \\ \text{then } \mu^{\text{pre}}(C_1) < \mu^{\text{pre}}(C_2) \end{array}$$

$\mathbf{P}_E^{\text{preci}}$ *Bounds.* A precision measure is bounded by constants 0 and $f(\Theta(C))$, where f is a strictly increasing function over propositional languages, ie $\Theta_1 \subset \Theta_2$ iff $f(\Theta_1) < f(\Theta_2)$. The lower bound 0 means that a candidacy function is not precise (see $\mathbf{P}_B^{\text{preci}}$). The upper bound $f(\Theta(C))$ means that the more expressive the language is, the more precise the candidacy function can be (see $\mathbf{P}_D^{\text{preci}}$). This principle extends property 7 of [3, page 225].

$$0 \leq \mu^{\text{pre}}(C) \leq f(\Theta(C))$$

$\mathbf{P}_F^{\text{preci}}$ *Strict monotonicity.* If the best candidates of a candidacy function C_1 form a strict superset of the best candidates of a candidacy function C_2, then C_1 is strictly less precise than C_2. This principle generalises the following property, which extends property 4 of [3, page 224]: if a knowledge item $\mathfrak{k}$ is not a consequence, wrt $\models_{\text{ff}}$, of a knowledge content $\mathfrak{K}$ such that $\Omega \cap \mathfrak{k} \cap \bigcap_{\mathfrak{t} \in \mathfrak{K}} \mathfrak{t} \neq \emptyset$, then $C_{\mathfrak{K}}$ is less precise than $C_{\mathfrak{K} \cup \{\mathfrak{k}\}}$.

$$\text{if } \hat{\Omega}_{C_1} \supset \hat{\Omega}_{C_2} \text{ then } \mu^{\text{pre}}(C_1) < \mu^{\text{pre}}(C_2)$$

$\mathbf{P}_G^{\text{preci}}$ *Equivalence.* Equivalent candidacy functions are equally precise. This principle extends properties 3 and 9 of [3, page 225].

$$\text{if } C_1 \equiv C_2 \text{ then } \mu^{\text{pre}}(C_1) = \mu^{\text{pre}}(C_2)$$

$\mathbf{P}_H^{\text{preci}}$ *Continuity.* When a candidacy function changes continuously, so its precision measure does. This principle ensures a certain robustness against slight fluctuations in the candidacy function. However, we do not require μ^{pre} of satisfying this principle if μ^{pre} is only used for ranking purpose.

$$\begin{array}{l} \text{if } \lim_{i \to +\infty} \mu^{\text{dis}}(C_i, C) = 0 \\ \text{then } \lim_{i \to +\infty} \mu^{\text{pre}}(C_i) = \mu^{\text{pre}}(C) \end{array}$$

Negation of a candidacy function Since we do not define the concept of negation for a candidacy function, we do not generalise property 2 and 5 of [3, page 224], which states that *the introduction of a contradiction decreases the amount of information.* Property 5 is formalised as follows: a consistent set Δ of propositional sentences is more informative than its mergence with $\neg\theta$ if Δ entails θ, ie if the minterms of $\wedge_{\phi \in \Delta}$ are also minterms of θ: $\alpha_{\wedge_{\phi \in \Delta} \phi} \subseteq \alpha_\theta$. We agree with property 5 in terms of sets of propositional sentences. However, in terms of knowledge bases, the mergence of the constraint $c \stackrel{\text{def}}{=}$ "$0 \geq \omega(\alpha_1)$" with the knowledge base $K \stackrel{\text{def}}{=} \{\text{"}\omega(\alpha_1) \geq \epsilon\text{"}\}$ is more precise than K if $\epsilon > 0$ tends towards 0 (which means that K tends to be tautological hence K tends to be void of information), even though K is consistent and $K \cup \{c\}$ is inconsistent.

3.5.3 Complete precision measure $\mu^{\text{pre}}_{\nearrow}$: the number of best candidates

In this section, we only consider candidacy functions of which their set of best candidates can be partitioned as $\{\hat{\Omega}_0, \hat{\Omega}_1, \ldots, \hat{\Omega}_P\}$ with $P \in \mathbb{N}$, where $\hat{\Omega}_0$ is a (possibly empty) finite set of points, and where each part $\hat{\Omega}_{p \geq 1}$ is a closed solo-dimensional manifold (see Def. 21 on page 18). We denote by $\mathbb{C}^{\int} \subset \mathbb{C}$ the set of such candidacy functions. For example, if K is a linear knowledge base, then $\hat{\Omega}_{C_K}$ is a closed convex set, hence $C_K \in \mathbb{C}^{\int}$.

Volume of the best candidates $\mathcal{V}(\hat{\Omega}_C)$

An i-hypervolume is the Lebesgue measure of a manifold in a Euclidean space of dimension i. An i-manifold is a manifold of dimension i, ie with a non-null i-hypervolume. A 0-manifold is a finite set of points and its 0-hypervolume is the cardinality of this finite set. Let $\mathcal{N} : \mathbb{R}^+ \mapsto [0{:}1]$ be a strictly increasing bijection, which will serve as normalising function; we arbitrarily choose $\mathcal{N}(x) \stackrel{\text{def}}{=} \frac{x}{1+x}$. Let $C \in \mathbb{C}^{\int}$ be a candidacy function underlain a propositional language with n variables. Let $\{\hat{\Omega}_0, \hat{\Omega}_1, \ldots, \hat{\Omega}_P\}$ be the partition for $\hat{\Omega}_C$. We now define $\mathcal{V}(\hat{\Omega}_C)$ as the vector $\vec{v} \in \mathbb{R}^{2^n+1}$ such that $\vec{v}_0 \stackrel{\text{def}}{=} \mathcal{N}(|\hat{\Omega}_0|)$ and its i^{th} element $\vec{v}_i$ is the Lebesgue measure in dimension i of $\hat{\Omega}_C$, denoted by ${}^i\!\int_{\hat{\Omega}_C}$: $\vec{v}_i \stackrel{\text{def}}{=} \mathcal{N}(\sum_{p=1}^{P} {}^i\!\int_{\hat{\Omega}_p} 1 \mathrm{d}\hat{\omega})$. Furthermore, $\vec{v}_i \stackrel{\text{def}}{=} 0$ for $i \geq 2^n + 1$.

For example, if a set of best candidates $\hat{\Omega}_C \subset \mathbb{R}^{2^n}$, with $n = 2$, is made of three isolated probability distributions and one tetrahedron having a 3-hypervolume equal to 0.2, then $\mathcal{V}(\hat{\Omega}_C)$ returns $[\mathcal{N}(3); \mathcal{N}(0); \mathcal{N}(0); \mathcal{N}(0.2); \mathcal{N}(0)]$, which we simply denote by $\mathcal{N}[3; 0; 0; 0.2; 0]$. If K is a linear knowledge base, then $\hat{\Omega}_{C_K}$ is a convex polytope of which the vol-

ume $\mathcal{V}(\hat{\Omega}_{C_K})$ is a vector full of zeros except for its i^{th} element that is the i-hypervolume of the polytope. An algorithm finding the exact hypervolume of a polytope is given in [32].

Proposition 16. *If Ω is underlain by a propositional language with n variables, then its 2^n-hypervolume is $\frac{\sqrt{2^n}}{(2^n-1)!}$.*

Proof. Let C be a candidacy function underlain by a propositional language with n variables and $J \stackrel{\text{def}}{=} 2^n$ minterms. Let V_J be the J-hypervolume of the polytope defined by the vertices $[\vec{0}, I_J]$, where each column is a vertex, $\vec{0}$ is the origin, and I_J is the identity matrix $J \times J$. Formally, V_J is inductively defined by $V_1 \stackrel{\text{def}}{=} 1$ and $V_J \stackrel{\text{def}}{=} \int_0^1 V_{J-1} * x^{J-1}\mathrm{d}x = V_{J-1} * \int_0^1 x^{J-1}\mathrm{d}x = V_{J-1} * \left(\frac{1^J}{J} - \frac{0^J}{J}\right) = V_{J-1} * \frac{1}{J} = \frac{1}{J!}$. Let S_J be the hypervolume of Ω, which is the polytope defined by the vertices represented by I_J, where each column denotes a vertex. Thus, S_J is a hypervolume in dimension $J-1$, which corresponds to the hypervolume of a main diagonal of the unit hypercube of the Euclidean space of dimension J. Let h_J be the altitude from the diagonal I_J to the origin $\vec{0}$. Let $\vec{x}$ be the foot of the altitude. Since point $\vec{x}$ belongs to the diagonal I_J, we have $1 = \sum_{j=1}^{J} \vec{x}_j$. By symmetry, each element $\vec{x}_j$ has the same value, hence $1 = \sum_{j=1}^{J} \vec{x}_j$ implies that $1 = J * \vec{x}_j$, then $\frac{1}{J} = \vec{x}_j$. The length of h_J is the distance between $\vec{x}$ and $\vec{0}$: $h_J = \|\vec{x} - \vec{0}\| = \sqrt{\sum_{j=1}^{J} \vec{\omega}_j^2} = \sqrt{\sum_{j=1}^{J} \frac{1}{J^2}} = \sqrt{J * \frac{1}{J^2}} = \frac{1}{\sqrt{J}}$. Since we have $V_J = \int_0^{h_J} S_J * \left(\frac{x}{h_J}\right)^{J-1} \mathrm{d}x = \frac{S_J}{h_J^{J-1}} * \int_0^{h_J} x^{J-1}\mathrm{d}x = \frac{S_J}{h_J^{J-1}} * \left(\frac{h_J^J}{J} - \frac{0^J}{J}\right) = \frac{S_J}{h_J^{J-1}} * \frac{h_J^J}{J} = S_J * \frac{h_J}{J}$, we also have $S_J = V_J * \frac{J}{h_J}$. Therefore, $S_J = \frac{1}{J!} * \frac{J}{\frac{1}{\sqrt{J}}} = \frac{\sqrt{J}}{(J-1)!}$. □

Precision ordering relation $\preceq^{\text{pre}}$

The precision ordering relation $\preceq^{\text{pre}}$ ranks two candidacy functions in $\mathbb{C}^{\int}$ as follows.

Definition 31. *C_1 is strictly less precise than C_2, denoted by $C_1 \prec^{\text{pre}} C_2$, iff $\vec{v} \stackrel{\text{def}}{=} \mathcal{V}(\hat{\Omega}_{C_1 \oplus vars(C_2)})$ is greater than $\vec{w} \stackrel{\text{def}}{=} \mathcal{V}(\hat{\Omega}_{C_2 \oplus vars(C_1)})$ wrt the following lexicographic order: $\exists i \in \mathbb{N}, \forall j > i, (\vec{v}_j = \vec{w}_j)$ and $(\vec{v}_i < \vec{w}_i)$. C_1 is (equally or) less precise than C_2, denoted by $C_1 \preceq^{\text{pre}} C_2$, iff $\vec{v} = \vec{w}$ or $C_1 \prec^{\text{pre}} C_2$.*

Remark that this order expresses that an i-hypervolume is lower than a j-hypervolume if $i < j$, ie, any finite set of points fits into any curve, any curve of finite length fits into any 2D-surface, and any 2D-surface of finite area fits into any 3D-volume.

Proposition 17. *$\mathcal{P}_\varepsilon(\vec{v}) < \mathcal{P}_\varepsilon(\vec{w})$ iff $\vec{v}$ is strictly greater than $\vec{w}$ wrt the lexicographic order ($\exists i \in \mathbb{N}, \forall j > i, (\vec{v}_j = \vec{w}_j)$ and $(\vec{v}_i < \vec{w}_i)$), where $\vec{v}, \vec{w} \in \mathbb{R}^{J+1}$, $J \in \mathbb{N}$, and $\varepsilon \in \mathbb{R}^+$ such that the smallest distinguishable difference between two j-hypervolumes $|\vec{v}_j - \vec{w}_j|$ is greater that $\varepsilon > 0$ for all $j \in \{1, 2, \ldots, J+1\}$, and where $\mathcal{P}_\varepsilon(\vec{v})$ is defined as follows:*

$$\mathcal{P}_\varepsilon(\vec{v}) \stackrel{\text{def}}{=} \sum_{j=0}^{|\vec{v}|} \vec{v}_j * \left(\frac{2}{\varepsilon} + 2\right)^j$$

Proof. In order to define our precision measure from a volume, ie a real number from a vector, we define an injective function $\mathcal{P}_\varepsilon : \mathbb{R}^{2^n+1} \mapsto \mathbb{R}$ that preserves the total order: $\vec{v} \leq^{\mathcal{V}} \vec{w}$ iff $\mathcal{P}_\varepsilon(\vec{v}) \leq \mathcal{P}_\varepsilon(\vec{w})$, where $\vec{v}$ and $\vec{w}$ are two volumes. We recall that a volume is a vector, $\vec{v}$ say, in $]-1{:}1]^{2^n} \times]0{:}1]$, and that $\vec{v}_j \stackrel{\text{def}}{=} 0$ for $j > 2^n + 1$. A univariate polynomial $P_{\vec{v}} \colon \mathbb{R} \mapsto \mathbb{R}$ is defined by $P_{\vec{v}}(x) \stackrel{\text{def}}{=} \sum_{j=0}^{+\infty} \vec{v}_j * x^j$, where $\vec{v}$ is the vector of its coefficients. We know that, for any two volumes $\vec{v}$ and $\vec{w}$, $\exists \hat{x} \in \mathbb{R}^+, \forall x > \hat{x} > 1, \left(\vec{v} <^{\mathcal{V}} \vec{w}\right) \iff P_{\vec{v}}(x) < P_{\vec{w}}(x)$. Such an $\hat{x}$ depends on the k^{th} element of any vectors $\vec{v}$ and $\vec{w}$, where k is such that $\forall j > k, \vec{v}_j = \vec{w}_j$ and $\vec{v}_k < \vec{w}_k$. Thus, such an $\hat{x}$ satisfies

$$P_{\vec{v}}(\hat{x}) < P_{\vec{w}}(\hat{x}) \tag{3.1}$$

iff $0 < P_{\vec{w}}(\hat{x}) - P_{\vec{v}}(\hat{x})$ iff $0 < \sum_{j=0}^{+\infty} \vec{w}_j * \hat{x}^j - \sum_{j=0}^{+\infty} \vec{v}_j * \hat{x}^j$ iff $0 < \sum_{j=0}^{+\infty} (\vec{w}_j - \vec{v}_j) * \hat{x}^j$ iff $0 < \left(\sum_{j=k+1}^{+\infty} (\vec{w}_j - \vec{v}_j) * \hat{x}^j\right) + \left((\vec{w}_k - \vec{v}_k) * \hat{x}^k\right) + \left(\sum_{j=0}^{k-1} (\vec{w}_j - \vec{v}_j) * \hat{x}^j\right)$ iff

$$0 < (\vec{w}_k - \vec{v}_k) * \hat{x}^k + \sum_{j=0}^{k-1} (\vec{w}_j - \vec{v}_j) * \hat{x}^j \tag{3.2}$$

Since $\vec{w}_j, \vec{v}_j \in [-1{:}1]$, we have $\vec{w}_j - \vec{v}_j \in [-2{:}2]$. The worst case for inequality (3.2) to be satisfied by $\hat{x}$ is reached when $\vec{w}_j - \vec{v}_j = -2$. Let $\epsilon \stackrel{\text{def}}{=} \vec{w}_k - \vec{v}_k \in \mathbb{R}^+$. Thus, $\hat{x}$ satisfies inequality (3.2) if $0 < \epsilon * \hat{x}^k + \sum_{j=0}^{k-1} -2 * \hat{x}^j$ iff $0 < \epsilon * \hat{x}^k - 2 * \frac{\hat{x}^k - \hat{x}^0}{\hat{x}-1}$ iff $\frac{\hat{x}^k - 1}{\hat{x}-1} < \frac{\epsilon}{2} * \hat{x}^k$ iff $\hat{x}^k - 1 < \frac{\epsilon}{2} * \hat{x}^k * (\hat{x} - 1)$ iff $1 - \hat{x}^{-k} < \frac{\epsilon}{2} * 1 * (\hat{x} - 1)$ iff $\hat{x}^{-k} > 1 + \frac{\epsilon}{2} * (1 - \hat{x})$ iff

$$\hat{x}^{-k} + \frac{\epsilon}{2} * \hat{x} > 1 + \frac{\epsilon}{2} \tag{3.3}$$

According to inequality (3.3), if $\hat{x}^{-k} > 1 + \frac{\epsilon}{2}$ or $\frac{\epsilon}{2} * \hat{x} > 1 + \frac{\epsilon}{2}$, then $\hat{x}$ satisfies inequality (3.1). Since condition $\frac{\epsilon}{2} * \hat{x} > 1 + \frac{\epsilon}{2}$ does not explicitly depend on k, we choose to make $\hat{x}$ satisfy it. Thus, $\hat{x}$ satisfies inequality (3.1) if $\frac{\epsilon}{2} * \hat{x} > 1 + \frac{\epsilon}{2}$ iff $\hat{x} > (1 + \frac{\epsilon}{2}) * \frac{2}{\epsilon}$ iff $\hat{x} > \frac{2}{\epsilon} + 1$, and finally, if $\hat{x} > \frac{2}{\vec{w}_k - \vec{v}_k} + 1$. Let $\varepsilon > 0$ be the smallest distinguishable difference between the j-hypervolumes

of two manifolds of dimension j, ie, for any two volumes $\vec{v}$ and $\vec{w}$, $\forall j \in \mathbb{N}, \varepsilon \leq |\vec{w}_j - \vec{v}_j|$. Therefore, for k as previously defined, ε is always smaller than $\vec{w}_k - \vec{v}_k$. Hence, if $\hat{x} \stackrel{\text{def}}{=} \frac{2}{\varepsilon} + 2$, then $\hat{x} > \frac{2}{\varepsilon} + 1$ and $\vec{v} <^{\mathcal{V}} \vec{w} \iff P_{\vec{v}}(\hat{x}) < P_{\vec{w}}(\hat{x})$. We thus conclude that all univariate polynomials are totally ordered at $\hat{x} \stackrel{\text{def}}{=} \frac{2}{\varepsilon} + 2$. □

Complete precision measure $\mu^{\text{pre}}_{\nearrow}$

Definition 32. *(Complete precision measure) The complete precision measure of a candidacy function $C \in \mathbb{C}^{\int}$ is founded upon the volume of its best candidates, where $\varepsilon > 0$ is the smallest distinguishable difference between two j-hypervolumes.*

$$\mu^{\text{pre}}_{\nearrow}(C) \stackrel{\text{def}}{=} \mathcal{P}_\varepsilon(\mathcal{V}(\Omega)) - \mathcal{P}_\varepsilon(\mathcal{V}(\hat{\Omega}_C))$$

Proposition 18. *$C_1 \preceq^{\text{pre}} C_2$ iff $\mu^{\text{pre}}_{\nearrow}(C_1) \leq \mu^{\text{pre}}_{\nearrow}(C_2)$, when $\Theta(C_1) = \Theta(C_2)$ and $\varepsilon > 0$ is the smallest distinguishable difference between two j-hypervolumes.*

Proof. This proposition follows from Prop. 17. □

Proposition 19. *If we only consider candidacy functions in $\mathbb{C}^{\int}$ and a tolerance $\varepsilon > 0$ representing the smallest distinguishable difference between two j-hypervolumes, then $\mu^{\text{pre}}_{\nearrow}$ satisfies principles $\mathbf{P}^{\text{preci}}_B$, $\mathbf{P}^{\text{preci}}_D$, $\mathbf{P}^{\text{preci}}_E$, $\mathbf{P}^{\text{preci}}_F$, and $\mathbf{P}^{\text{preci}}_G$. However, $\mu^{\text{pre}}_{\nearrow}$ does not satisfy principles $\mathbf{P}^{\text{preci}}_A$, $\mathbf{P}^{\text{preci}}_C$, and $\mathbf{P}^{\text{preci}}_H$.*

Proof. $\mathbf{P}^{\text{preci}}_A$ *Language invariance.* The counter-example in Prop. 20 shows that $\mu^{\text{pre}}_{\nearrow}$ does not satisfy principle $\mathbf{P}^{\text{preci}}_A$.

$\mathbf{P}^{\text{preci}}_B$ *Lower bound.* $\hat{\Omega}_C = \Omega$ iff $\mathcal{P}_\varepsilon(\mathcal{V}(\hat{\Omega}_C)) = \mathcal{P}_\varepsilon(\mathcal{V}(\Omega))$ iff $0 = \mathcal{P}_\varepsilon(\mathcal{V}(\hat{\Omega}_C)) - \mathcal{P}_\varepsilon(\mathcal{V}(\Omega)) = \mu^{\text{pre}}_{\nearrow}$.

$\mathbf{P}^{\text{preci}}_C$ *Singleton bound.* The following counter-example shows that $\mu^{\text{pre}}_{\nearrow}$ does not satisfy principle $\mathbf{P}^{\text{preci}}_C$. Let C_1 be a candidacy function such that its single best candidates $\hat{\omega}_1$ is such that $\forall \alpha \in \alpha_{\Theta(C_1)}, \hat{\omega}_1(\alpha) \neq 1$; hence, $C_1 \notin \mathcal{M}(\Theta)$. Let $C_2 \in \mathcal{M}(\Theta)$ and $\{\hat{\omega}_2\} \stackrel{\text{def}}{=} \hat{\Omega}_{C_2}$. Thus, $\mathcal{V}(\hat{\Omega}_{C_1}) = \mathcal{V}(\{\hat{\omega}_1\}) = \mathcal{N}[1;0;\ldots;0] = \mathcal{V}(\{\hat{\omega}_2\}) = \mathcal{V}(\hat{\Omega}_{C_2})$. Therefore, $\mu^{\text{pre}}_{\nearrow}(C_1) = \mu^{\text{pre}}_{\nearrow}(C_2)$. Besides, it seems that $\mu^{\text{pre}}_v(C) \stackrel{\text{def}}{=} \mathcal{P}_\varepsilon(\mathcal{V}(\Omega \oplus v)) - \mathcal{P}_\varepsilon(\mathcal{V}(\hat{\Omega}_{C \oplus v}))$ would satisfy $\mathbf{P}^{\text{preci}}_C$ if $v \notin vars(C)$.

$\mathbf{P}^{\text{preci}}_D$ *Language bound.* Let Θ be a language not containing a propositional variable v. Let $C_1 \in \mathcal{M}(\Theta)$ and $\{\hat{\omega}_1\} \stackrel{\text{def}}{=} \hat{\Omega}_{C_1}$. Let $C_2 \in \mathcal{M}(\Theta \oplus v)$ and $\{\hat{\omega}_2\} \stackrel{\text{def}}{=} \hat{\Omega}_{C_2}$. Thus, $\mathcal{P}_\varepsilon(\mathcal{V}(\hat{\Omega}_{C_1})) = \mathcal{P}_\varepsilon(\mathcal{N}[0,\ldots,0,1]) = \mathcal{P}_\varepsilon(\mathcal{V}(\hat{\Omega}_{C_2}))$, and $\mathcal{P}_\varepsilon(\mathcal{V}(\Omega)) < \mathcal{P}_\varepsilon(\mathcal{V}(\Omega \oplus v))$ since the dimension of $\Omega \oplus v$ is twice greater than the dimension of Ω. Therefore, $\mathcal{P}_\varepsilon(\mathcal{V}(\Omega)) < \mathcal{P}_\varepsilon(\mathcal{V}(\Omega \oplus v))$ iff $\mathcal{P}_\varepsilon(\mathcal{V}(\Omega)) - \mathcal{P}_\varepsilon(\mathcal{V}(\hat{\Omega}_{C_1})) < \mathcal{P}_\varepsilon(\mathcal{V}(\Omega \oplus v)) - \mathcal{P}_\varepsilon(\mathcal{V}(\hat{\Omega}_{C_2}))$ iff $\mu^{\text{pre}}_{\nearrow}(C_1) < \mu^{\text{pre}}_{\nearrow}(C_2)$.

$\mathbf{P}^{\text{preci}}_E$ *Bounds.* For any $C \in \mathbb{C}^{\int}$, and any probability distribution ω, $\mathcal{V}(\Omega) \geq \mathcal{V}(\hat{\Omega}_C) \geq \mathcal{V}(\{\omega\})$, hence $0 \leq \mu^{\text{pre}}_{\nearrow}(C) \leq \mathcal{P}_\varepsilon(\mathcal{V}(\Omega)) - \mathcal{P}_\varepsilon(\mathcal{V}(\{\omega\}))$. Therefore, $0 \leq \mu^{\text{pre}}_{\nearrow}(C) < f(\Theta(C))$ where $f(\Theta) \stackrel{\text{def}}{=} \mathcal{P}_\varepsilon(\mathcal{V}(\Omega))$.

$\mathbf{P}^{\text{preci}}_F$ *Strict monotonicity.* For any $C \in \mathbb{C}^{\int}$, $\hat{\Omega}_C$ is solodimensional, hence is a closed set. Thus, $\hat{\Omega}_{C_1} \supset \hat{\Omega}_{C_2}$ iff $\mathcal{V}(\hat{\Omega}_{C_1}) > \mathcal{V}(\hat{\Omega}_{C_2})$ because $\hat{\Omega}_{C_1}$ contains more isolated singleton than $\hat{\Omega}_{C_2}$ or there exists a Lebesgue measurable difference in some dimension(s) between $\hat{\Omega}_{C_1}$ and $\hat{\Omega}_{C_2}$. Therefore, $\mathcal{P}_\varepsilon(\mathcal{V}(\hat{\Omega}_{C_1})) > \mathcal{P}_\varepsilon(\mathcal{V}(\hat{\Omega}_{C_2}))$ and $\mu^{\text{pre}}_{\nearrow}(C_1) < \mu^{\text{pre}}_{\nearrow}(C_2)$.

$\mathbf{P}^{\text{preci}}_G$ *Equivalence.* Equivalent candidacy functions (underlain by a same language) have equal best candidates (wrt $\stackrel{\text{i}}{\equiv}$ and $\stackrel{\text{e}}{\equiv}$). Hence, the respective volumes of these sets of best candidates are indistinguishable. Therefore, equivalent candidacy functions are considered equally precise by $\mu^{\text{pre}}_{\nearrow}$.

$\mathbf{P}^{\text{preci}}_H$ *Continuity* The following counter-example shows that $\mu^{\text{pre}}_{\nearrow}$ does not satisfy principle $\mathbf{P}^{\text{preci}}_H$. Let $C \in \mathbb{C}^{\int}$ be such that $\hat{\Omega}_C$ is a square. Let $C_i \in \mathbb{C}^{\int}$ be a sequence of candidacy functions such that $\hat{\Omega}_{C_i}$ are cubes where an edge of the cube continuously decreases until obtaining the square $\hat{\Omega}_C$. Thus, we have $\lim_{i \to +\infty} \mu^{\text{dis}}(C, C_i) = 0$. Let both the length of the edges of the cube and the square equal to 0.1, and let η be the length of the decreasing edge of the cube. Thus, $\mathcal{V}(\hat{\Omega}_{C_i}) = \mathcal{N}[0;0;0;\eta * 0.1 * 0.1;\ldots;0]$ and $\mathcal{V}(\hat{\Omega}_C) = \mathcal{N}[0;0;0.1 * 0.1;0;\ldots;0]$. Let $\hat{x} \in \mathbb{R}^+$ be a strictly positive real number, namely $\frac{2}{\varepsilon} + 2$ (see Prop. 17). Thus, $\lim_{i \to +\infty} \mathcal{P}_\varepsilon(\mathcal{V}(\Omega_{C_i})) = \lim_{\eta \to 0} \mathcal{N}(\eta * 0.1 * 0.1) * \hat{x}^3 = 0$, which differs from $\mathcal{P}_\varepsilon(\mathcal{V}(\Omega_C)) = \mathcal{N}(0.1 * 0.1) * \hat{x}^2$. Hence $\lim_{i \to +\infty} \mu^{\text{pre}}_{\nearrow}(C_i) \neq \mu^{\text{pre}}_{\nearrow}(C)$.

□

Proposition 20. *$\mu^{\text{pre}}_{\nearrow}$ is not* language invariant *(see $\mathbf{P}^{\text{preci}}_A$).*

Proof. The following counter-example shows that $\mu^{\text{pre}}_{\nearrow}$ does not satisfy principle $\mathbf{P}^{\text{preci}}_A$. Let $K \stackrel{\text{def}}{=} \{\, \omega(v) = a \,\}$, with $a \in [0{:}1]$, be a linear knowledge base underlain by a language with one propositional variable v. Thus, each probability distribution in Ω is a point satisfying $\left\{ \begin{array}{c} \vec{0} \leq [\omega(v); \omega(\neg v)] \leq \vec{1} \\ \omega(v) + \omega(\neg v) = 1 \end{array} \right\}$. We now show that $\mu^{\text{pre}}_{\nearrow}(C_K) \neq \mu^{\text{pre}}_{\nearrow}(C_K \oplus v')$. Let $\omega' \in \Omega \oplus v'$ where v' is another propositional variable. Let $x' \stackrel{\text{def}}{=} \omega'(\neg v \wedge \neg v')$,

$y' \stackrel{\text{def}}{=} \omega'(\neg v \wedge v')$, $z' \stackrel{\text{def}}{=} \omega'(v \wedge \neg v')$, and $t' \stackrel{\text{def}}{=} \omega'(v \wedge v')$; notice that $\omega(v) = x' + y'$. Thus, $K' \stackrel{\text{def}}{=} K \oplus v' = \{ x' + y' = a \}$ and each probability distribution in $\Omega \oplus v'$ is a point satisfying $\left\{ \begin{array}{c} \vec{0} \leq [x';y';z';t'] \leq \vec{1} \\ x' + y' + z' + t' = 1 \end{array} \right\}$. Notice that K and K' are consistent knowledge bases. In order to compare $\mu^{\text{pre}}_{\nearrow}(C_K)$ and $\mu^{\text{pre}}_{\nearrow}(C_{K'})$, we need to compute the volume of the models of K and K'. The models of K' are the probability distribution satisfying $\left\{ \begin{array}{c} x' + y' = a \\ \vec{0} \leq [x';y';z';t'] \\ z' + t' = 1 - a \end{array} \right\}$. Notice that $\Omega_{K'}$ forms the simplex of which we want to compute the volume. Let $H \stackrel{\text{def}}{=} \begin{bmatrix} 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$ be a Hadamard matrix. Let $[x;y;z;t] \stackrel{\text{def}}{=} \frac{1}{2} * (H * [x';y';z';t'])$ be an orthonormal transformation, which does not change the volume of $\Omega_{K'}$. We thus express this simplex as $\left\{ \begin{array}{c} -y + t = a \\ \vec{0} \leq [x';y';z';t'] \\ y + t = 1 - a \end{array} \right\}$. We set $y \stackrel{\text{def}}{=} \frac{1}{2} - a$ and $t \stackrel{\text{def}}{=} \frac{1}{2}$ since $(-y + t) + (y + t) = (1 - a) + a$. Then, we rewrite this simplex as follows: $\left\{ \vec{0} \leq [x';y';z';t'] \right\}$ and we replace the variables by the transformed ones in order to obtain $\left\{ \begin{array}{l} 0 \leq \frac{1}{2} * (+x - y - z + t) \\ 0 \leq \frac{1}{2} * (-x - y + z + t) \\ 0 \leq \frac{1}{2} * (-x + y - z + t) \\ 0 \leq \frac{1}{2} * (+x + y + z + t) \end{array} \right\}$. After, we replace y and t by their definition and we have $\left\{ \begin{array}{l} 0 \leq +x - (\frac{1}{2} - a) - z + \frac{1}{2} \\ 0 \leq -x - (\frac{1}{2} - a) + z + \frac{1}{2} \\ 0 \leq -x + (\frac{1}{2} - a) - z + \frac{1}{2} \\ 0 \leq +x + (\frac{1}{2} - a) + z + \frac{1}{2} \end{array} \right\}$. After having isolated variable z, we finally we rewrite the simplex formed by $\Omega_{K'}$ as follows: $\left\{ \begin{array}{lcccl} x - a & \leq & z & \leq & x + a \\ -x + (a - 1) & \leq & z & \leq & -x - (a - 1) \end{array} \right\}$. This simplex is a rectangle (see the five colour-filled rectangles on Fig. 3.1, for $a = 0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1$) defined by the points $p_1 \stackrel{\text{def}}{=} [\frac{1}{2}; \frac{1}{2} - a; \frac{1}{2} - a; \frac{1}{2}]$, $p_2 \stackrel{\text{def}}{=} [-\frac{1}{2} + a; \frac{1}{2} - a; -\frac{1}{2}; \frac{1}{2}]$, $p_3 \stackrel{\text{def}}{=} [-\frac{1}{2}; \frac{1}{2} - a; -\frac{1}{2} + a; \frac{1}{2}]$, and $p_4 \stackrel{\text{def}}{=} [\frac{1}{2} - a; \frac{1}{2} - a; \frac{1}{2} + a; \frac{1}{2}]$, where p_1 is the intersection of $z = x - a$ with $z = -x - (a - 1)$, p_2 is the intersection of $z = x - a$ with $z = -x + (a - 1)$, p_3 is the intersection of $z = x + a$ with $z = -x + (a - 1)$, and p_4 is the intersection of $z = x + a$ with $z = -x - (a - 1)$. This rectangle can be a line when $a = 0$ or $a = 1$. In order to easily work on probability distributions denoted by this rectangle, we change again the coordinate system such that p_2 becomes the new origin, and such that vectors $\frac{\overrightarrow{p_2 p_1}}{||\overrightarrow{p_2 p_1}||} = [\frac{\sqrt{2}}{2}; 0; \frac{\sqrt{2}}{2}; 0]$ and $\frac{\overrightarrow{p_2 p_3}}{||\overrightarrow{p_2 p_3}||} = [\frac{\sqrt{2}}{-2}; 0; \frac{\sqrt{2}}{2}; 0]$ becomes the unit vectors for the new axes x_r and z_r, respectively. Let $R \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} & 0 \\ 0 & 1 & 0 & 0 \\ \frac{\sqrt{2}}{-2} & 0 & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ be a rotation matrix corresponding to a rotation by $\frac{\pi}{4}$ around axes y and t. Let $[x'';y'';z'';t''] \stackrel{\text{def}}{=} R * ([x;y;z;t] - p_2)$ be an orthonormal transformation (ie, a translation by $-p_2$ followed by rotation R), which does not change the volume of $\Omega_{K'}$. Let $p'' \stackrel{\text{def}}{=} [x'';0;z'';0]$ be a point in the rectangle, ie $x'' \in [0; p''_1.x]$ and $z'' \in [0; p''_3.z]$, where $p''_3.z \stackrel{\text{def}}{=} \sqrt{2} * a$ and $p''_1.x \stackrel{\text{def}}{=} \sqrt{2} * (1 - a)$. Let $p \stackrel{\text{def}}{=} (R^{-1} * p'') + p_2$ and $p' \stackrel{\text{def}}{=} \frac{(H * p)}{2}$; notice that p' denotes a probability distribution in the original Euclidean space Ω. The volume of $\Omega_{K'}$ is the surface of a rectangle $\sqrt{2} * (1 - a) \times \sqrt{2} * a$ if $0 < a < 1$, otherwise, this volume is the length of a segment line, namely $\sqrt{2}$. The volume of $\Omega \oplus v'$ is $\int_0^1 2a(1 - a)da = \frac{1}{3}$, as stated by the formula of proposition 16 for $n = 2$ propositional variables: $\frac{\sqrt{2^n}}{(2^n - 1)!}$. The volume of the simplex of Ω is $\frac{\sqrt{2^1}}{(2^1 - 1)!} = \sqrt{2}$. Thus, when $a = 0$, we obtain $\mathcal{V}(\hat{\Omega}_{C_K}) = \mathcal{V}(\Omega_K) = \mathcal{N}[1;0;0]$, $\mathcal{V}(\hat{\Omega}_{C_{K'}}) = \mathcal{V}(\Omega_{K'}) = \mathcal{N}[0;\sqrt{2};0;0]$ $\mathcal{V}(\Omega) = \mathcal{N}[0;\sqrt{2};0]$, and $\mathcal{V}(\Omega \oplus v') = \mathcal{N}[0;0;0;\frac{1}{3};0]$. Let $\varepsilon = \mathcal{N}(\frac{1}{3})$ be the smallest distinguishable difference between the involved hypervolumes; thus, $\left(\frac{2}{\varepsilon} + 2\right) = 9$. Therefore, $\mathcal{P}_\varepsilon(\mathcal{V}(\hat{\Omega}_{C_K})) = \mathcal{N}(1) * 9^0 = 1$, $\mathcal{P}_\varepsilon(\mathcal{V}(\hat{\Omega}_{C_{K'}})) = \mathcal{N}(\sqrt{2}) * 9^1 \approx 5.272$,

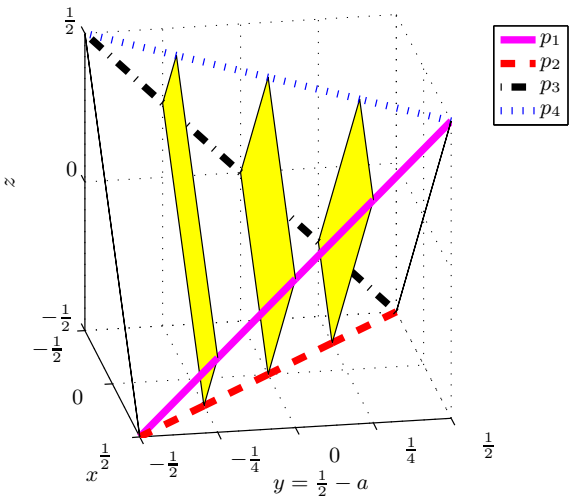


Figure 3.1: Set of probability distributions over a propositional language with two variables. All probability distributions within a same colour-filled rectangle denote the same probability distribution over a propositional language with one variable.

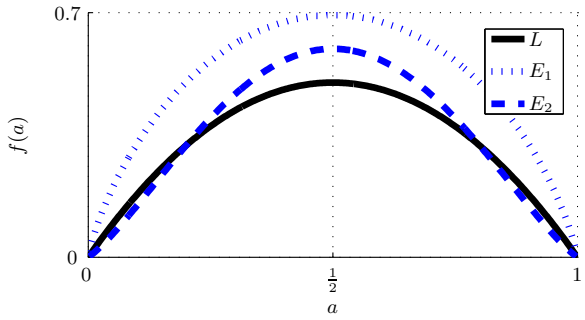


Figure 3.2: Curve L is the Lebesgue measure of the hypervolume of a colour-filled rectangle in Fig 3.1; thus, $f(a) \stackrel{\text{def}}{=} 2*a*(1-a)$. Curves E_1 and E_2 are the entropy functions for probability distributions over a language with respectively 1 and 2 propositional variables; thus $f(a) \stackrel{\text{def}}{=} E_1([a;1-a])$ for E_1 and $f(a) \stackrel{\text{def}}{=} 2*a*(1-a)*(E_1([a;1-a]) + \frac{1}{2})$ for E_2.

$\mathcal{P}_\varepsilon(\mathcal{V}(\Omega)) = \mathcal{N}(\sqrt{2}) * 9^1 \approx 5.272$, and $\mathcal{P}_\varepsilon(\mathcal{V}(\Omega \oplus v')) = \mathcal{N}(\frac{1}{3}) * 9^3 = 182.25$. We thus conclude that $\mu^{\text{pre}}_{\nearrow}(C_K) \approx 5.272 - 1$ differs from $\mu^{\text{pre}}_{\nearrow}(C_{K'}) \approx 182.25 - 5.272$. □

Concluding remarks and perspectives

In this paragraph, we sketch an explanation for the failure of our precision measure to satisfy *language invariance* (see principle $\mathbf{P}^{\text{preci}}_A$).

Let $K^a_1 \stackrel{\text{def}}{=} \{\, \omega(v_1) = a \,\}$ be knowledge base underlain by a propositional language Ω_1 with one variable v_1. Let $K^a_n \stackrel{\text{def}}{=} K^a_{n-1} \oplus v_n$, with $n \in \mathbb{N} \geq 2$, be a knowledge base equivalent to K^a_1 but underlain by $\Omega_n \stackrel{\text{def}}{=} \Omega_{n-1} \oplus v_n$. In the proof of Prop. 20, where K^a_1 is denoted by K and K^a_2 is denoted by K', we ascertain that *one* probability distribution in Ω_1, $[a;1-a]$ say, corresponds to a *set* of probability distributions in Ω_2, $\Omega_{K^a_2}$ say, which corresponds to a colour-filled rectangle in Fig. 3.1. Furthermore, the probability distributions in $\Omega_{K^a_2}$ correspond to a larger set of probability distributions $\Omega_{K^a_3}$, of which its probability distributions correspond to an even larger set $\Omega_{K^a_4}$, etc. By *equivalence* (see principle $\mathbf{P}^{\text{preci}}_G$), every knowledge base K^a_n should have the same precision measure, namely $\mu^{\text{pre}}_{\nearrow}(K^a_1)$, which only depends on a. The main idea behind our complete precision measure is to count the models of K^a_n. However, from Def 32, we remark that our precision measure applied to K^a_1 computes the volume of $\Omega_{K^a_1}$, whereas it should count all the probability distributions contained in $\bigcup_{n=1}^{+\infty} \Omega_{K^a_n}$. The hypervolumes of $\Omega_{K^a_1}$ are computed via the Lebesgue measure that counts indifferently all the probability distributions in $\Omega_{K^a_1}$. But we know that the probability distributions near $[\frac{1}{2};\frac{1}{2}]$ correspond to more probability distributions than those near $[0;1]$ or $[1;0]$; curve L illustrates this fact in Fig. 3.2. Therefore, we could redefine our hypervolume computation so that it counts the probability distributions in $\Omega_{K^a_1}$ while weighting them via a function f_1, which counts the probability distributions in $\Omega_{K^a_2}$ weighted by a function f_2, which counts the probability distributions in $\Omega_{K^a_3}$ weighted by a function f_3, etc. If such functions f_n exist, then they should satisfy the following property:

$$\forall \omega_n \in \Omega_{K^a_n}, f_n(\omega_n) = \int_{\Omega_{K^a_{n+1}}} f_{n+1}(\omega_{n+1}) \mathrm{d}\omega_{n+1}$$

where ω_n denotes a probability distribution over a propositional language with n variables. Hence, f_1 and f_2 should satisfy

$$f_1([a;(1-a)]) = \int_{\Omega_{K^a_2}} f_2(\omega) \mathrm{d}\omega \tag{3.4}$$

The hypervolume computation explained at section 3.5.3 defines f_1 and f_2 as being constant 1, which means that all the probability distributions get the same weight. If we look at the proof of Prop. 20, where $\Omega_{K^a_1}$ is $\{\, [a;1-a] \,\}$, where $\Omega_{K^a_2}$ denotes a colour-filled rectangle, and where $p' \stackrel{\text{def}}{=} [-\frac{z''+a}{\sqrt{2}}; \frac{z''}{\sqrt{2}}; -\frac{x''+1-a}{\sqrt{2}}; \frac{x''}{\sqrt{2}}]$, then we remark that $f_1([a;(1-a)]) = 1$, whereas we have $\int_0^{\sqrt{2}*a} \left(\int_0^{\sqrt{2}*(1-a)} f_2(p') \,\mathrm{d}x'' \right) \mathrm{d}z'' = 2*a*(1-a)$, since $f_2 = 1$. Thus, equation (3.4) is not satisfied; this partly explain why our measure fails to satisfy principle $\mathbf{P}^{\text{preci}}_A$.

Let $E_1([a;1-a]) \stackrel{\text{def}}{=} -(1-a)*\ln(1-a) - a*\ln(a)$ and $E_2([x';y';z';t']) \stackrel{\text{def}}{=} -x'*\ln(x') - y'*\ln(y') - z'*\ln(z') - t'*\ln(t')$ be the entropy functions for probability distributions over a language with respectively 1 and 2 propositional variables. By examining Fig. 3.2, we see that curve L has the same shape as curve E_1, namely they are both strictly concave and reach their maximum at $[\frac{1}{2};\frac{1}{2}]$. Moreover, the entropy functions are known to quantify the information on what they are applied to. Thus, we might hope that defining f_1 by E_1 and f_2 by E_2 leads to the satisfaction of equation (3.4). Unfortunately, $E_1([a;1-a])$ differs from $\int_0^{p''_3.z} \left(\int_0^{p''_1.x} E_2(p') \mathrm{d}x'' \right) \mathrm{d}z''$, as shown by the following equation:

$$\int_{\Omega_{K^a_2}} E_2(\omega) \mathrm{d}\omega = 2*a*(1-a)*\left(E_1([a;1-a]) + \frac{1}{2} \right) \tag{3.5}$$

In equation (3.5), we not only recognise the hypervolume of a colour-filled rectangle, ie $2*a*(1-a)$, but also the formula that we looked for, ie $E_1([a;1-a])$;

we are currently investigating an explanation for such a result that may guide us towards the definitions of f_n satisfying equation (3.4).

Maximum entropy as precision measure Finally, we conclude by exemplifying why the maximum entropy of probability distributions in $\hat{\Omega}_C$, ie $\max\left(E(\omega) \;\middle|\; \omega \in \hat{\Omega}_C \right)$, does not quantify the precision of a candidacy function C. We restate the example given in §3.5.1. Let v be a propositional variable meaning that we get a tail after tossing a coin. If we have no information about this coin, then our initial candidacy function, denoted by $C_{\{\omega(v)\in[0:1]\}}$, is empty (or is a tautology). If someone tells us that this coin is perfectly designed such that the probability to get a tail after tossing the coin is exactly $\frac{1}{2}$, then we update our candidacy function and we note it $C_{\{\omega(v)=\frac{1}{2}\}}$. Thus, as required by *strict monotonicity* (see principle $\mathbf{P}_F^{\text{preci}}$), we expect from a precision measure μ^{pre} to express that $C_{\{\omega(v)\in[0:1]\}}$ is less precise than $C_{\{\omega(v)=\frac{1}{2}\}}$. However, the maximum entropy expresses that the two candidacy functions do not contain any information, hence are equally imprecise: $\max(E(\omega) \mid \omega \in \hat{\Omega}_{C_{\{\omega(v)\in[0:1]\}}}) = E([\frac{1}{2};\frac{1}{2}]) = \max(E(\omega) \mid \omega \in \hat{\Omega}_{C_{\{\omega(v)=\frac{1}{2}\}}})$.

3.5.4 Language invariant precision measure $\mu^{\text{pre}}_{\ominus}$: the shadow lengths of the best candidates

Proposition 1 on page 7 states that a probability distribution ω' in $\Omega \oplus v$ is a unique probability distribution ω in Ω. We then define the language impoverishment operator for a probability distribution $\omega' \in \Omega \oplus v$ as $\omega' \ominus v \stackrel{\text{def}}{=} \omega$, where $\forall \alpha \in \alpha_{\Theta(\Omega)}, \omega(\alpha) = \omega'(\alpha)$, and for a set of probability distributions $W' \subseteq \Omega \oplus v$ as $W' \ominus v \stackrel{\text{def}}{=} \{\, \omega' \ominus v \mid \omega' \in W' \,\}$. Finally, we recursively define the subtraction of a set of variables *vars* from an object $\diamond$ as follows:

$$\diamond \ominus \mathit{vars} \stackrel{\text{def}}{=} \begin{cases} \diamond & \text{if } \mathit{vars} = \emptyset \text{ or } \mathit{vars} \supset \mathit{vars}(\diamond), \\ \diamond \ominus (\mathit{vars} \setminus \{v\}) & \text{else if } v \notin \Theta(\diamond), \\ (\diamond \ominus v) \ominus (\mathit{vars} \setminus \{v\}) & \text{else if } v \in \Theta(\diamond). \end{cases}$$

Notice that $(\diamond \oplus v) \ominus v = \diamond$ holds, whereas $(\diamond \ominus v) \oplus v = \diamond$ does not necessary hold.

An interesting property to design a language invariant precision measure (see $\mathbf{P}_A^{\text{preci}}$) is that the 1-hypervolume, ie the length, of $\Omega_K \ominus (\mathit{vars}(C) \setminus v)$ equals $\sqrt{2}$ when v does not appear in the constraints of K. We thus base our second precision measure $\mu^{\text{pre}}_{\ominus}(C)$ upon the sum of the Lebesgue measures in dimension 1, ie the lengths, of $\hat{\Omega}_C \ominus (\mathit{vars}(C) \setminus v)$ for each variable $v \in \mathit{vars}(C)$; $\hat{\Omega}_C \ominus (\mathit{vars}(C) \setminus v)$ is the set of best candidates projected on the segment line $[0{:}\sqrt{2}]$ representing the set of probability distributions underlain by one propositional variable v. By deeming a projected candidate to be the shadow of that candidate, we deem $\int_{\hat{\Omega}_C \ominus (\mathit{vars}(C)\setminus v)} 1\mathrm{d}x$ to be the shadow length, wrt v, of the best candidates of C.

Definition 33. *The language invariant precision measure of a candidacy function $C \in \mathbb{C}$ is founded upon the shadow lengths of its best candidates.*

$$\mu^{\text{pre}}_{\ominus}(C) \stackrel{\text{def}}{=} \sqrt{2} * |\mathit{vars}(C)| - \sum_{v \in \mathit{vars}(C)} \int_{\hat{\Omega}_C \ominus (\mathit{vars}(C)\setminus v)} 1\mathrm{d}x$$

On measuring the confidence Our first measure $\mu^{\text{pre}}_{\nearrow}$, which satisfies *strict monotonicity*, is able to reflect smaller variations of the precision of a candidacy function than $\mu^{\text{pre}}_{\ominus}$, which satisfies only a *non-strict* monotonicity. Nevertheless, when $\mu^{\text{pre}}_{\ominus}$ gives the same precision measure to two candidacy functions $C_1, C_2 \in \mathbb{C}$, we might still rank them as follows: C_1 is almost less precise than C_2 iff $\mu^{\text{pre}}_{\ominus}(C_1) = \mu^{\text{pre}}_{\ominus}(C_2)$ and $\mu^{\text{conf}}(C_1) < \mu^{\text{conf}}(C_2)$, where $\mu^{\text{conf}}(C) \stackrel{\text{def}}{=} \sup_{\omega\in\Omega} C(\omega) - \inf_{\omega\in\Omega} C(\omega)$ is the *confidence measure* of a candidacy function C. Roughly, a confidence measure quantifies a the flatness of a candidacy function. Notice that μ^{conf} is *language invariant* because it is "vertical" (see Fig. 3.3 on page 35).

In addition, notice that making a decision is answering a Yes/No question. Let $W \subseteq \Omega$ be the set of probability distributions where Yes should be answered to a given question, and let $\Omega \setminus W$ be the probability distributions where No should be answered. A decision is then identified with W. A (language invariant) measure of the confidence in a decision W wrt a knowledge $C \in \mathbb{C}$ is thus defined as follows: $\mu^{\text{conf}}(C, W) \stackrel{\text{def}}{=} \sup_{\omega\in W} C(\omega) - \sup_{\omega\in\Omega\setminus W} C(\omega)$, where $\sup_{\omega\in\emptyset} C(\omega) \stackrel{\text{def}}{=} 0$; roughly, $\sup_{\omega\in W} C(\omega)$ is the best support for Yes, and $\sup_{\omega\in\Omega\setminus W} C(\omega)$ is the best support for No, because each probability function ω is seen as an argument of strength $C(\omega)$ that supports either Yes if $\omega \in W$, or No if $\omega \notin W$. This measure is strictly positive if the answer should be Yes, strictly negative if the answer should be No, and neutral otherwise (for other notions of decision quality, see [14, 46]). If the best candidates of a candidacy function C are all in W, then the more peaked the candidacy function, the higher is the confidence measure. When the confidence measure is neutral, or when its absolute value is lower than a given threshold, it would be reasonable to postpone the decision until enough information is gathered; this idea appears in [13] for deciding under ignorance. In case postponing is impossible, adhering to common-

sensical principles like those stated in chapter 4 may guide us towards the best decision to make.

Concluding remarks In a multisensor system, if new sensors can be added to the system, they may bring new kinds of data, ie constraints involving new variables. Hence, the underlying language of the candidacy function may grow. Thus, this second precision measure satisfying *language invariance* is more convenient than our first measure $\mu^{\text{pre}}_{\nearrow}$. Furthermore, $\mu^{\text{pre}}_{\ominus}$ is defined on $\mathbb{C}$ whereas $\mu^{\text{pre}}_{\nearrow}$ is only defined on $\mathbb{C}^{\int} \subset \mathbb{C}$. Besides, our first precision measure, which satisfies *strict monotonicity*, is able to reflect smaller variations of a candidacy function than our second measure satisfying only a non-strict monotonicity. Nevertheless, a language invariant confidence measure is defined to rank two candidacy functions $C_1, C_2 \in \mathbb{C}$ having the same precision measure wrt $\mu^{\text{pre}}_{\ominus}$: C_1 is said to be almost less precise than C_2 iff $\mu^{\text{pre}}_{\ominus}(C_1) = \mu^{\text{pre}}_{\ominus}(C_2)$ and $\mu^{\text{conf}}(C_1) < \mu^{\text{conf}}(C_2)$.

3.5.5 Conclusions

Our main contribution is twofold: firstly, we establish eight principles to be satisfied by a precision measure when applied to a candidacy function. Secondly, we define two principled precision measures. They quantify the range of choice an inference process (see chapter 4) have for electing one probability distribution from the best candidates. To our knowledge, precision measure for probabilistic knowledge does not exist in the literature.

Our two precision measures count the best candidates of a given candidacy function (eg, number of probability distributions satisfying a consistent knowledge base); the less there exist such probability distributions, the more the candidacy function is precise. However, our counting function (see $\mathcal{V}(\hat{\Omega}_C)$ in §3.5.3) suffers from two problems.

The first problem is intrinsic to probability distributions (see proposition 1). Probability distributions contain different quantity of information (as shown in Fig. 3.1). Briefly, an infinite set of probability distributions is hidden behind each probability distribution ω, and the *cardinality* of this hidden set gives the quantity of information of ω. Hence, in this section, we strive to compute this recursively defined cardinality of infinite set (see concluding remarks at §3.5.3). This first problem hinders our first measure to satisfy *language invariance* (see principle $\mathbf{P}^{\text{preci}}_{A}$). However, we managed to make our second measure satisfy this principle by projecting all the probability distributions onto several 1-dimensional spaces (one for each underlying propositional variable). This projection over-approximates the cardinality of the set constraining our second measure to satisfy only a weaker form of *strict monotonicity* (see principle $\mathbf{P}^{\text{preci}}_{F}$). Thus, our second precision measure is less "complete" than our first one.

The second problem is intrinsic to Lebesgue measure. A set of probability distributions over a propositional language with n variables can be represented by manifolds in a 2^n-dimensional Euclidean space. Thus, we use Lebesgue measure to compute the hypervolume of this set. However, these manifolds (points, segment lines, 2D-surfaces,..., 2^n-manifolds) may have different dimensions, so their respective hypervolume (cardinality, metre, square metre, ..., metre$^{2^n}$). We provide a method (see Prop. 17) to merge all these heterogeneous Lebesgue measures into one real number. This real number is intended to represent the quantity of best candidates, which is taken as the precision of a given candidacy. However, this merging suffers from discontinuity, eg, when the geometrical representation of a set of best candidates goes continuously from a cube to a square, our precision measures decrease continuously until the set representation becomes the square; then our precision measure do a jump discontinuity. This second problem hinders our two measures to satisfy *continuity* (see principle $\mathbf{P}^{\text{preci}}_{H}$). Nevertheless, this is not problematic when we only need to order the candidacy functions by precision.

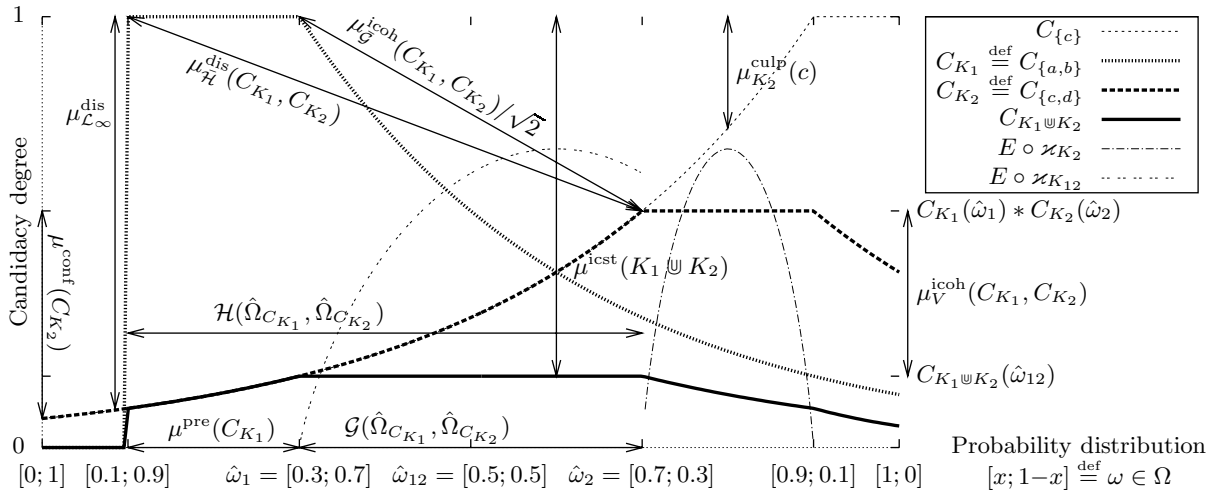
We demonstrate that our precision measures satisfy different subsets of principles. These subsets guide us towards the precision measure that best fits our specifications. For example, in our application (see section 5.1 in chapter 5), new sensors may be added to the multisensor system. Thus, these sensors may bring new kinds of data, ie, new variables may be inserted into the underlying propositional language. Therefore, we need a precision measure satisfying *language invariance* (see principle $\mathbf{P}^{\text{preci}}_{A}$)

Several aspects of the relation between sets of probability distributions of different dimensions (ie, the relation between Ω and $\Omega \oplus \mathit{vars}$) remain obscure, yet this study gives grounds to discussions and provides clues (like equation (3.5)) for future investigations.

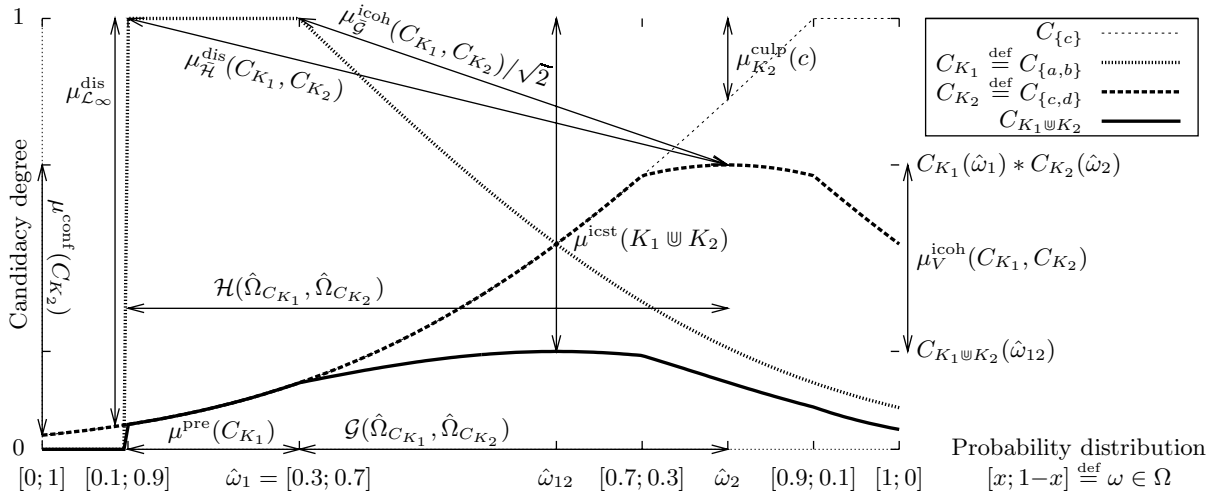
3.6 Conclusions and perspectives

In this chapter, we show how convenient (see Fig. 3.3) the candidacy functions are for formalising in a single probabilistic framework several notions like dissimilarity, surprise, inconsistency, incoherence, confidence, or precision.

The principles and measures introduced in this chapter are indubitably preliminary to further research. Thus, the next step is to collect from the literature then unify such different notions and associated desiderata,



(a) Exponential blur ($\sigma = 0.95$)



(b) Half-Gaussian blur ($\sigma = 0.90$)

Figure 3.3: Measures for paraconsistent probabilistic reasoning, where $vars(\Omega) = \{v\}$, $x = \omega(v)$, $a \stackrel{\text{def}}{=}$ "$-0.1 \geq -x$", $b \stackrel{\text{def}}{=}$ "$0.3 \geq x$", $c \stackrel{\text{def}}{=}$ "$-0.9 \geq -x$", and $d \stackrel{\text{def}}{=}$ "$0.7 \geq x$" are four constraints, which compose three knowledge bases $K_1 \stackrel{\text{def}}{=} \{a, b\}$, $K_2 \stackrel{\text{def}}{=} \{c, d\}$, and $K_1 \Cup K_2 = \{a, b, c, d\}$. The candidacy function of each knowledge base is founded upon the exponential blur (see Fig. 3.3(a) and Def. 15 on page 14) or upon the half-Gaussian blur (see Fig. 3.3(b) and Def. 16 on page 14). The reliability level corresponding to constraints b, c, and d is 0.95 on Fig. 3.3(a) and is 0.9 on Fig. 3.3(b), whereas that of a is 1; hence, $C_{\{a\}}$, C_{K_1}, and $C_{K_1 \Cup K_2}$ are partially absorbing: $\forall x \in [0{:}0.1], C_{\{a\}}([x; 1-x]) = 0$. Thus, wrt C_{K_1}, the non-candidate probability distributions are $\{[x; 1-x] \mid x \in [0{:}0.1]\}$, the best candidates (which are also the models of K_1 since K_1 is consistent) are $\hat{\Omega}_{C_{K_1}} = \{[x; 1-x] \mid x \in [0.1{:}0.3]\} = \Omega_{K_1}$, and the candidates (which may become best candidates, like $\hat{\omega}_{12}$, after merging K_1 with another base) are $\{[x; 1-x] \mid x \in [0.1{:}1]\}$. We denote by $\hat{\omega}_1$ any best candidates of C_{K_1}, by $\hat{\omega}_2$ a best candidate of C_{K_2} such that $E(\hat{\omega}_2) \geq E(\hat{\omega}), \forall \hat{\omega} \in \hat{\Omega}_{C_{K_2}}$, and by $\hat{\omega}_{12}$ a best candidate of $C_{K_1 \Cup K_2}$ such that $E(\hat{\omega}_{12}) \geq E(\hat{\omega}), \forall \hat{\omega} \in \hat{\Omega}_{C_{K_1 \Cup K_2}}$, where E is the entropy of a distribution.

then to compare them; the final step being the characterisation of these notions by a small set of intuitive principles.

Chapter 4

Inference processes to elect the least-biased most-probable worlds

> Ne rien trouver ridicule est
> le signe de l'intelligence complète.
>
> Valéry Larbaud, in [25]

Chapter 2 introduces a knowledge representation called *candidacy functions*. A candidacy function C totally orders the set of probability distributions Ω such that the best probability distributions $\hat{\Omega}_C$ are those that best represent the real world. However, $\hat{\Omega}_C$ may not be a singleton, which happens when C is tautological.

In chapter 4, we thus theoretically address the problem of electing a unique probability distribution that *best* represents the real world, wrt a certain knowledge. A solution to such a problem is called an inference process[1]. After adapting several principles (stated in [33, 38]) to candidacy functions, we present in §4.2.1 an inference process $\mathcal{I}^E_{\text{ME}}$ that returns the best candidates having the maximal entropy, wrt a candidacy function.

4.1 Principles

In this section, we state several principles to be satisfied by an inference process $\mathcal{I}$ electing a non-empty set of probability distributions when applied to a candidacy function $C \in \mathbb{C}$. The ideas underlying these principles proceed from [33, chapter 7], which deals with consistent knowledge bases. Since we desire an inference process to elect a unique probability distribution, we suppose that $\mathcal{I}$ satisfies *uniqueness* (see $\mathbf{P}^{\mathcal{I}}_{\alpha}$) when applied to a candidacy function corresponding to a linear knowledge base (like in principles $\mathbf{P}^{\mathcal{I}}_{\vartheta}$ and $\mathbf{P}^{\mathcal{I}}_{\iota}$).

Notice that these principles *abstractly* employ $\hat{\Omega}_C$: its definition (see Def. 10) does not matter as long as the set of best candidates $\hat{\Omega}_{C_K}$ coincides with the set of models Ω_K when K is a consistent knowledge base; this is required by *unanimity* (see principle $\mathbf{P}^{\mathbb{C}}_{8}$).

$\mathbf{P}^{\mathcal{I}}_{\alpha}$ *Uniqueness.* An inference process should deterministically elect a unique probability distribution.

$$\forall \omega_1, \omega_2 \in \mathcal{I}(C), \omega_1 = \omega_2$$

$\mathbf{P}^{\mathcal{I}}_{\beta}$ *Irrelevant information* (extends [33, page 87]). Entirely irrelevant information should be ignored by an inference process. Let $C_1, C_2 \in \mathbb{C}$ be two candidacy functions such that $\mathit{vars}(C_1) \cap \mathit{vars}(C_2) = \emptyset$.

$$\mathcal{I}(C_1) = \mathcal{I}(C_1 \Cup C_2) \ominus \mathit{vars}(C_2)$$

If *uniqueness* is satisfied, then *irrelevant information* can be defined without the language impoverishment operator $\ominus$ (see § 3.5.4 on page 33) as follows. Let $\theta_1 \in \Theta(C_1)$ be a propositional sentence. Since C_2 is entirely irrelevant to C_1 and θ_1, $\mathcal{I}$ should satisfy $(\mathcal{I}(C_1))(\theta_1) = (\mathcal{I}(C_1 \Cup C_2))(\theta_1)$.

$\mathbf{P}^{\mathcal{I}}_{\gamma}$ *Equivalence* (extends [33, page 82]). Equal information should be inferred from equivalent candidacy functions, wrt a certain equivalence relation (see $\overset{\text{i}}{\equiv}$ at Def. 11 on page 9 and $\overset{\text{e}}{\equiv}$ at Def. 17 on page 15).

$$\text{if } C_1 \equiv C_2 \text{ then } \mathcal{I}(C_1) = \mathcal{I}(C_2)$$

$\mathbf{P}^{\mathcal{I}}_{\delta}$ *Renaming* (extends [33, page 95]). An inference process should be insensitive to a renaming of the propositional variables, hence of the minterms. Let π be a bijection over the minterms $\alpha_{\Theta(C)}$, where $C \in \mathbb{C}$. If ω is a probability distribution, then $(\pi(\omega))(\alpha) \overset{\text{def}}{=} \omega(\pi(\alpha))$. Also, if W is a set of probability distributions, then $\pi(W) \overset{\text{def}}{=} \{\, \pi(\omega) \mid \omega \in W \,\}$. Finally, if C a candidacy function, then $\forall \omega \in \Omega, (\pi(C))(\omega) \overset{\text{def}}{=} C(\pi(\omega))$.

$$\mathcal{I}(\pi(C)) = \pi(\mathcal{I}(C))$$

[1] We interpret Valéry Larbaud's statement *Never pouring ridicule is a sign of complete intelligence* as *A completely intelligent inference process considers all the knowledge items as useful, even those that are contradicting.*

$\mathbf{P}^{\mathcal{I}}_{\varepsilon}$ *Obstinacy* (extends [33, page 90]). Additional support for what is already known should be ignored by an inference process.

$$\text{if } \mathcal{I}(C_1) \cap \hat{\Omega}_{C_2} \neq \emptyset \text{ then } \mathcal{I}(C_1) \cap \hat{\Omega}_{C_2} = \mathcal{I}(C_1 \Cup C_2)$$

$\mathbf{P}^{\mathcal{I}}_{\zeta}$ *Continuity* (extends [33, page 89]). Microscopic changes in the knowledge should not cause macroscopic changes in the inferred information. This principle ensures a certain robustness in face of minor fluctuations in the candidacy function. Formally, when a candidacy function changes continuously, so the inferred information does.

$$\begin{array}{l}\text{if } \lim_{i\to\infty} \mu^{\text{dis}}(C_i, C) = 0 \\ \text{then } \lim_{i\to\infty} \mathcal{H}(\mathcal{I}(C_i), \mathcal{I}(C)) = 0\end{array}$$

$\mathbf{P}^{\mathcal{I}}_{\eta}$ *Open-mindedness* (extends [33, page 95]). An inference process should give the benefit of the doubt; this principle is a kind of precautionary principle. Formally, if some best candidates for representing the real world, wrt $C \in \mathbb{C}$, consider $\theta \in \Theta(C)$ as probable, then the elected probability distribution should not consider θ as improbable.

$$\begin{array}{l}\text{if } \exists \hat{\omega} \in \hat{\Omega}_C, \hat{\omega}(\theta) > 0 \text{ and } \hat{\Omega}_C \text{ is convex} \\ \text{then } \forall \hat{\omega} \in \mathcal{I}(C), \hat{\omega}(\theta) > 0\end{array}$$

If $\hat{\Omega}_C$ is not convex, it is questionable to require this property. For example, suppose a knowledge base K being such that $\Omega_K = \{\hat{\omega}_1, \hat{\omega}_2\}$ where $\hat{\omega}_1 \stackrel{\text{def}}{=} [0; 0.2; 0.4; 0.4]$, $\hat{\omega}_2 \stackrel{\text{def}}{=} [0.2; 0; 0.4; 0.4]$, and $\Theta(K)$ has two propositional variables. We would expect from an inference process $\mathcal{I}$ to elect $\hat{\omega}_1$ or $\hat{\omega}_2$; by doing so, $\mathcal{I}$ would not be open-minded because if $\mathcal{I}$ elects $\hat{\omega}_1$, which considers α_1 as improbable, then $\mathcal{I}$ should have elected $\hat{\omega}_2$ instead of $\hat{\omega}_1$ since $\hat{\omega}_2(\alpha_1) > 0$, but if $\mathcal{I}$ elects $\hat{\omega}_2$, which considers α_2 as improbable, then $\mathcal{I}$ should have elected $\hat{\omega}_1$ instead of $\hat{\omega}_2$ since $\hat{\omega}_1(\alpha_2) > 0$.

$\mathbf{P}^{\mathcal{I}}_{\vartheta}$ *Independence* (due to [33, page 101]). The absence of any information linking two events should be identified with the conditional independence; justifications for this principle are given in [37]. Let K be a knowledge base such that $vars(K) = \{v_1, v_2, v_3\}$. Let v_1, v_2, and v_3 be three propositional variables such that $\{v_1, v_2, v_3\} = vars(K)$, where K is the following (non-normalised) knowledge base:

$$K \stackrel{\text{def}}{=} \left\{ \begin{array}{r} \omega(v_1) = a \\ \omega(v_2 \mid v_1) = b \\ \omega(v_3 \mid v_1) = c \end{array} \right\} \text{ with } a, b \in [0{:}1]$$

Independence states that v_2 and v_3 should be treated as conditionally independent given v_1:

$$(\mathcal{I}(C_K))(v_2 \wedge v_3 \mid v_1) = b * c$$

$\mathbf{P}^{\mathcal{I}}_{\iota}$ *Relativisation* (due to [33, page 100]). The probabilities an inference process would give if an event φ occurred should only depend on the knowledge conditioned by the occurrence of event φ. Let K, K_1, and K_2 be three knowledge bases defined in a non-normalised form as follows, where $a_{ij}, b_i, a'_{ij}, b'_i, c \in \mathbb{R}$, $k, k', l_i, l'_i \in \mathbb{N}$, $\theta, \theta_i, \theta'_i, \varphi \in \Theta$:

$$K \stackrel{\text{def}}{=} \{\ \text{“}c = \omega(\varphi)\text{”}\ \} \text{ with } 0 < c < 1$$

$$K_1 \stackrel{\text{def}}{=} \left\{ \text{“}b_i = \sum_{j=1}^{l_i} a_{ij} * \omega(\theta_i \mid \varphi)\text{”} \;\middle|\; i = 1, \ldots, k \right\}$$

$$K_2 \stackrel{\text{def}}{=} \left\{ \text{“}b'_i = \sum_{j=1}^{l'_i} a'_{ij} * \omega(\theta'_i \mid \neg\varphi)\text{”} \;\middle|\; i = 1, \ldots, k' \right\}$$

Notice that K_1 expresses knowledge relative to the occurrence of φ, whereas K_2 expresses knowledge relative to the non-occurrence of φ. Then, *relativisation* states that the probability of θ given φ should only depend on $K \Cup K_1$, when $K \Cup K_1 \Cup K_2$ is consistent:

$$\begin{array}{l}\text{if } \Omega_{K \Cup K_1 \Cup K_2} \neq \emptyset \\ \text{then } (\mathcal{I}(C_{K \Cup K_1}))(\theta \mid \varphi) = (\mathcal{I}(C_{K \Cup K_1 \Cup K_2}))(\theta \mid \varphi)\end{array}$$

$\mathbf{P}^{\mathcal{I}}_{\kappa}$ *Best candidates.* An inference process should elect best candidates only.

$$\mathcal{I}(C) \subseteq \hat{\Omega}_C$$

4.2 Entropy-based inference processes $\mathcal{I}^E$

An entropy-based inference process elects probability distributions with high entropy in order to minimise the risk of being surprised, ie the risk of having elected distributions that poorly represent the real world.

4.2.1 Paraconsistent Maximum Entropy inference process $\mathcal{I}^E_{\text{ME}}$

Let $K \in \mathbb{K}$ be a consistent knowledge base. The Maximum Entropy inference process ME returns the models of K that has a maximal entropy:

$$E(\omega) \stackrel{\text{def}}{=} -\sum_{j=1}^{J} \omega_j * \ln(\omega_j) \qquad \text{ME}(K) \stackrel{\text{def}}{=} \arg\max_{\omega \in \Omega_K} E(\omega)$$

The demonstration of the following characterisation theorem appears in [36], and its generalisation to consistent polynomial knowledge bases appears in [38, theorem 18].

Theorem 1 (See [33, theorem 7.9])**.** *When dealing with a consistent knowledge base $K \in \mathbb{K}^{=}$,* ME *is the unique inference process satisfying the principles of* irrelevant information*,* equivalence*,* renaming*,* obstinacy*,* independence*,* continuity*,* open-mindedness*, and* relativisation*.*

Definition 34. *The Paraconsistent Maximum Entropy inference process $\mathcal{I}^E_{\text{ME}}$ elects the probability distributions having the maximum entropy from the best candidates for representing the real world, wrt C:*

$$\mathcal{I}^E_{\text{ME}}(C) \stackrel{\text{def}}{=} \arg\max_{\hat{\omega}\in\hat{\Omega}_C} E(\hat{\omega})$$

Proposition 21. *$\mathcal{I}^E_{\text{ME}}$ satisfies principles $\mathbf{P}^{\mathcal{I}}_{\beta}$, $\mathbf{P}^{\mathcal{I}}_{\gamma}$, $\mathbf{P}^{\mathcal{I}}_{\delta}$, $\mathbf{P}^{\mathcal{I}}_{\varepsilon}$, $\mathbf{P}^{\mathcal{I}}_{\zeta}$ wrt $\mu^{\text{dis}}_{\mathcal{H}}$ or $\mu^{\text{dis}}_{\mathcal{H}}$, $\mathbf{P}^{\mathcal{I}}_{\eta}$, $\mathbf{P}^{\mathcal{I}}_{\vartheta}$, $\mathbf{P}^{\mathcal{I}}_{\iota}$, and $\mathbf{P}^{\mathcal{I}}_{\kappa}$. If $\hat{\Omega}_C$ is convex, then $\mathcal{I}^E_{\text{ME}}$ satisfies also $\mathbf{P}^{\mathcal{I}}_{\alpha}$.*

Proof. By *unanimity* (see principle $\mathbf{P}^{\mathbb{C}}_{\text{s}}$), $\hat{\Omega}_{C_K} = \Omega_K$ when $K \in \mathbb{K}$ is a consistent knowledge base. Therefore, $\mathcal{I}^E_{\text{ME}}$ generalises ME, which has been proved to satisfy $\mathbf{P}^{\mathcal{I}}_{\beta}$ to $\mathbf{P}^{\mathcal{I}}_{\iota}$.

$\mathbf{P}^{\mathcal{I}}_{\alpha}$ *Uniqueness.* If $\hat{\Omega}_C$ is convex, then maximising the strictly concave function E over $\hat{\Omega}_C$ returns a unique argument.

$\mathbf{P}^{\mathcal{I}}_{\beta}$ *Irrelevant information.* In the sequel of this proof, which follows the one in [38, page 18], superscript numbers are used for naming instead of exponentiation, except for powers of 2. Let C^1 and C^2 be two candidacy functions underlain by disjoint propositional languages, ie $\mathit{vars}(C^1) \cap \mathit{vars}(C^2) = \emptyset$. Let α^1_i and α^2_j be the respective minterms of $\Theta(C^1)$ and $\Theta(C^2)$, with $i = 1, \ldots, I$ and $j = 1, \ldots, J$, where $I \stackrel{\text{def}}{=} 2^m$ and $J \stackrel{\text{def}}{=} 2^n$, and where n and m are the number of variables of $\Theta(C^1)$ and $\Theta(C^2)$. Let $C \stackrel{\text{def}}{=} C^1 \Cup C^2$. Let $\omega^1 \in \Omega(C^1)$, $\omega^2 \in \Omega(C^2)$, and $\omega \in \Omega(C)$. We define six probability distributions as follows:

$$\begin{array}{ll}
\tau^1 \stackrel{\text{def}}{\in} \mathcal{I}^E_{\text{ME}}(C^1) & \nu^1(\alpha^1_i) \stackrel{\text{def}}{=} \nu(\alpha^1_i) \\
\tau^2 \stackrel{\text{def}}{\in} \mathcal{I}^E_{\text{ME}}(C^2) & \nu^2(\alpha^2_j) \stackrel{\text{def}}{=} \nu(\alpha^2_j) \\
\tau(\alpha^1_i \wedge \alpha^2_j) \stackrel{\text{def}}{=} \tau^1(\alpha^1_i) * \tau^2(\alpha^2_j) & \nu \stackrel{\text{def}}{\in} \mathcal{I}^E_{\text{ME}}(C)
\end{array}$$

Let $\diamond$ be either τ or ν, but not mix of τ and ν. Let $\diamond^1_i \stackrel{\text{def}}{=} \diamond^1(\alpha^1_i)$, $\diamond^2_j \stackrel{\text{def}}{=} \diamond^2(\alpha^2_j)$, and $\diamond_{ij} \stackrel{\text{def}}{=} \diamond(\alpha^1_i \wedge \alpha^2_j)$.

• Firstly, we show several direct consequences of these definitions.
$1 = \sum_{i=1}^I \tau^1_i = \sum_{j=1}^J \tau^2_j = \sum_{i,j=1}^{I,J} \nu_{ij}$ since τ^1, τ^2, and ν are probability distributions over $\Theta(C^1)$, $\Theta(C^2)$, and $\Theta(C^1 \Cup C^2)$.
τ is a probability distribution over $\Theta(C)$ since $\tau^1_i * \tau^2_j \in [0{:}1]$ and $\sum_{i,j=1}^{I,J} \tau(\alpha^1_i \wedge \alpha^2_j) = \sum_{i,j=1}^{I,J} \tau^1_i * \tau^2_j = \sum_{i=1}^I \left(\tau^1_i * \sum_{j=1}^J \tau^2_j\right) = \sum_{i=1}^I \tau^1_i * 1 = 1$.
τ characterises the probability distributions τ^1 and τ^2 over respectively $\Theta(C^1)$ and $\Theta(C^2)$ since $\tau(\alpha^1_i) = \tau(\alpha^1_i \wedge (\alpha^2_1 \vee \cdots \vee \alpha^2_J)) = \sum_{j=1}^J \tau(\alpha^1_i \wedge \alpha^2_j) = \sum_{j=1}^J \tau^1_i * \tau^2_j = \tau^1_i * \sum_{j=1}^J \tau^2_j = \tau^1_i$, and similarly for τ^2.
ν^1 and ν^2 are two probability distributions over respectively $\Theta(C^1)$ and $\Theta(C^2)$ since $\nu^1_i = \nu(\alpha^1_i) \in [0{:}1]$ and $\sum_{i=1}^I \nu(\alpha^1_i) = \sum_{i=1}^I \nu(\alpha^1_i \wedge (\alpha^2_1 \vee \cdots \vee \alpha^2_J)) = \sum_{i,j=1}^{I,J} \nu(\alpha^1_i \wedge \alpha^2_j) = 1$, and similarly for ν^2.

• Secondly, we prove that $\nu^1 \in \hat{\Omega}_{C^1}$.
$\nu \in \hat{\Omega}_C$ iff $C(\omega) \leq C(\nu)$ iff $(C^1 \Cup C^2)(\omega) \leq (C^1 \Cup C^2)(\nu)$ iff $(C^1 \oplus \mathit{vars}(C^2))(\omega) * (C^2 \oplus \mathit{vars}(C^1))(\omega) \leq (C^1 \oplus \mathit{vars}(C^2))(\nu) * (C^2 \oplus \mathit{vars}(C^1))(\nu)$ iff $C^1(\omega^1) * C^2(\omega^2) \leq C^1(\nu^1) * C^2(\nu^2)$ iff $C^1(\tau^1) * C^2(\tau^2) \leq C^1(\nu^1) * C^2(\nu^2)$. Since $\tau^1 \in \hat{\Omega}_{C^1}$, $\tau^2 \in \hat{\Omega}_{C^2}$, we have $C^1(\tau^1) * C^2(\tau^2) = C^1(\nu^1) * C^2(\nu^2)$. Suppose $C^1(\nu^1) < C^1(\tau^1)$; then we obtain this contradiction: $C^1(\nu^1) * C^2(\tau^2) < C^1(\tau^1) * C^2(\tau^2) = C^1(\nu^1) * C^2(\nu^2) \leq C^1(\nu^1) * C^2(\tau^2)$. Therefore, we have $C^1(\nu^1) \geq C^1(\tau^1)$, hence $\nu^1 \in \hat{\Omega}_{C^1}$.

• Thirdly, we prove that $\tau \in \hat{\Omega}_C$.
Remember that τ characterises τ^1 and τ^2, hence $\tau \in \tau^1 \oplus \mathit{vars}(C^2) \cap \tau^2 \oplus \mathit{vars}(C^1)$. Since $\tau^1 \in \hat{\Omega}_{C^1}$ and $\tau^2 \in \hat{\Omega}_{C^2}$, we have $C^1(\omega^1) \leq C^1(\tau^1)$ and $C^2(\omega^2) \leq C^2(\tau^2)$. Thus, by definition of $\oplus$, $(C^1 \oplus \mathit{vars}(C^2))(\omega) \leq (C^1 \oplus \mathit{vars}(C^2))(\tau)$ and $(C^2 \oplus \mathit{vars}(C^1))(\omega) \leq (C^2 \oplus \mathit{vars}(C^1))(\tau)$. By multiplying these two inequalities, and by applying the definition of $\Cup$, we obtain $C(\omega) \leq C(\tau)$, hence $\tau \in \hat{\Omega}_C$.

• Fourthly, recall that $E(\tau^1) = E(\nu^1)$ and $E(\tau) = E(\nu)$, according to [38, theorem 6].

• Finally, we show that $\mathcal{I}^E_{\text{ME}}(C^1) = \mathcal{I}^E_{\text{ME}}(C^1 \Cup C^2) \ominus \mathit{vars}(C^2)$.
We have $\nu^1 \in \mathcal{I}^E_{\text{ME}}(C^1)$ since $\nu^1 \in \hat{\Omega}_{C^1}$ and $E(\tau^1) = E(\nu^1)$. Hence, $\forall \tau \in \mathcal{I}^E_{\text{ME}}(C^1 \Cup C^2), \exists \tau^1 \in \mathcal{I}^E_{\text{ME}}(C^1), \forall \alpha^1 \in \alpha_{\Theta(C^1)}, \nu^1(\alpha^1) = \nu(\alpha^1)$. Besides, we have $\tau \in \mathcal{I}^E_{\text{ME}}(C)$ since $\tau \in \hat{\Omega}_C$ and $E(\tau) = E(\nu)$. Hence $\forall \nu^1 \in \mathcal{I}^E_{\text{ME}}(C^1), \exists \nu \in \mathcal{I}^E_{\text{ME}}(C^1 \Cup C^2), \forall \alpha^1 \in \alpha_{\Theta(C^1)}, \nu^1(\alpha^1) = \nu(\alpha^1)$. Therefore, $\mathcal{I}^E_{\text{ME}}(C^1) = \mathcal{I}^E_{\text{ME}}(C^1 \Cup C^2) \ominus \mathit{vars}(C^2)$.

$\mathbf{P}^{\mathcal{I}}_{\gamma}$ *Equivalence.* If $C_1 \stackrel{\text{i}}{\equiv} C_2$ then $C_1 \stackrel{\text{e}}{\equiv} C_2$ then $\hat{\Omega}_{C_1} = \hat{\Omega}_{C_2}$ hence $\mathcal{I}^E_{\text{ME}}(C_1) = \mathcal{I}^E_{\text{ME}}(C_2)$.

$\mathbf{P}^{\mathcal{I}}_{\delta}$ *Renaming.* Since $\hat{\Omega}_{\pi(C)} = \arg\max_{\omega\in\Omega} \pi(C)(\omega) = \arg\max_{\omega\in\Omega} C(\pi(\omega)) = \arg\max_{\omega\in\pi(\Omega)} C(\omega) = \pi(\arg\max_{\omega\in\Omega} C(\omega)) = \pi(\hat{\Omega}_C)$, we have $\mathcal{I}^E_{\text{ME}}(\pi(C)) = \arg\max_{\hat{\omega}\in\hat{\Omega}_{\pi(C)}} E(\hat{\omega}) = \arg\max_{\hat{\omega}\in\pi(\hat{\Omega}_C)} E(\hat{\omega}) = \pi\left(\arg\max_{\hat{\omega}\in\hat{\Omega}_C} E(\hat{\omega})\right) = \pi(\mathcal{I}^E_{\text{ME}}(C))$.

$\mathbf{P}^{\mathcal{I}}_{\varepsilon}$ *Obstinacy.* Let $\tau \stackrel{\text{def}}{\in} \mathcal{I}^E_{\text{ME}}(C_1) \cap \hat{\Omega}_{C_2}$. Thus, $\hat{\Omega}_{C_1} \cap \hat{\Omega}_{C_2}$ contains at least τ. Let $\omega \in \Omega$. Thus,

$C_1(\tau) \geq C_1(\omega)$ and $C_2(\tau) \geq C_2(\omega)$, hence $C_1(\tau) * C_2(\tau) \geq C_1(\omega) * C_2(\omega)$ iff $(C_1 \Cup C_2)(\tau) \geq (C_1 \Cup C_2)(\omega)$ iff $\tau \in \hat{\Omega}_{C_1 \Cup C_2}$. Let $\eta \overset{\text{def}}{\in} \mathcal{I}^E_{\text{ME}}(C_1 \Cup C_2)$. Since $\tau, \eta \in \hat{\Omega}_{C_1 \Cup C_2}$, we have $(C_1 \Cup C_2)(\tau) = (C_1 \Cup C_2)(\eta)$, then $C_1(\tau) * C_2(\tau) = C_1(\eta) * C_2(\eta)$. Suppose $\eta \notin \hat{\Omega}_{C_1}$; thus, $C_1(\tau) > C_1(\eta)$ since $\tau \in \hat{\Omega}_{C_1}$. Therefore, we should have $C_2(\eta) > C_2(\tau)$ in order to satisfy $C_1(\tau) * C_2(\tau) = C_1(\eta) * C_2(\eta)$. However, we know that $C_2(\tau) \geq C_2(\eta)$ since $\tau \in \hat{\Omega}_{C_2}$. Therefore, our assumption is wrong because we obtain the following contradiction: $C_2(\eta) > C_2(\tau) \geq C_2(\eta)$. Hence, $\eta \in \hat{\Omega}_{C_1}$. Similarly, we can show that $\eta \in \hat{\Omega}_{C_2}$.

Since $E(\tau)$ is maximal for $\hat{\Omega}_{C_1}$ and $\eta \in \hat{\Omega}_{C_1}$, $E(\tau) \geq E(\eta)$. Since $E(\eta)$ is maximal for $\hat{\Omega}_{C_1 \Cup C_2}$ and $\tau \in \hat{\Omega}_{C_1 \Cup C_2}$, $E(\eta) \geq E(\tau)$. Therefore, $E(\tau) = E(\eta)$.

Since $\tau \in \hat{\Omega}_{C_1 \Cup C_2}$, $E(\tau)$ is maximal for $\hat{\Omega}_{C_1 \Cup C_2}$, hence $\tau \in \mathcal{I}^E_{\text{ME}}(C_1 \Cup C_2)$. Since $\eta \in \hat{\Omega}_{C_1}$, $E(\eta)$ is maximal for $\hat{\Omega}_{C_1}$, hence $\eta \in \mathcal{I}^E_{\text{ME}}(C_1)$.

We thus conclude that $\forall \tau \in \mathcal{I}^E_{\text{ME}}(C_1) \cap \hat{\Omega}_{C_2}, \tau \in \mathcal{I}^E_{\text{ME}}(C_1 \Cup C_2)$ and that $\forall \eta \in \mathcal{I}^E_{\text{ME}}(C_1 \Cup C_2), \eta \in \mathcal{I}^E_{\text{ME}}(C_1) \cap \hat{\Omega}_{C_2}$, as required to conclude $\mathcal{I}^E_{\text{ME}}(C_1) \cap \hat{\Omega}_{C_2} = \mathcal{I}^E_{\text{ME}}(C_1 \Cup C_2)$.

$\mathbf{P}^{\mathcal{I}}_{\zeta}$ *Continuity.* If $\lim_{i \to \infty} \mu^{\text{dis}}_{\mathcal{H}}(C_i, C) = 0$, then $\lim_{i \to \infty} \mu^{\text{dis}}_{\mathcal{H}}(C_i, C) = 0$, and then $\lim_{i \to \infty} \mathcal{H}(\hat{\Omega}_{C_i}, \hat{\Omega}_C) = 0$. By the continuity of E on Ω, we conclude $\lim_{i \to \infty} \mathcal{H}(\mathcal{I}^E_{\text{ME}}(C_i), \mathcal{I}^E_{\text{ME}}(C)) = 0$.

$\mathbf{P}^{\mathcal{I}}_{\eta}$ *Open-mindedness.* Since $\hat{\Omega}_C$ is supposed to be convex, the proof given in [33, page 95] holds for $\mathcal{I}^E_{\text{ME}}$.

$\mathbf{P}^{\mathcal{I}}_{\vartheta}$, $\mathbf{P}^{\mathcal{I}}_{\iota}$ *Independence*, *Relativisation.* Since the knowledge bases involved in these principles are linear and consistent, $\mathcal{I}^E_{\text{ME}}$ coincides with ME, which satisfies these principles.

$\mathbf{P}^{\mathcal{I}}_{\kappa}$ *Best candidates.* This principle is satisfied since $\mathcal{I}^E_{\text{ME}}$ returns the argument of a maximisation over $\hat{\Omega}_C$.

□

We furthermore stress that $\mathcal{I}^E_{\text{ME}}$ is σ-invariant, since Prop. 9 on page 14 states that the best candidates are σ-invariant.

Notice that E could be substituted by another "electing" function $f_n : \Omega \mapsto \mathbb{R}$, where $|vars(\Omega)| = n$, which elects some probability distributions among the best candidates of a candidacy function. This would yield to another kind of inference process that would inherit from the good properties of electing best candidates. For example, such an inference process would be *continuous* wrt $\mu^{\text{dis}}_{\mathcal{H}}$ or $\mu^{\text{dis}}_{\mathcal{H}}$ if f_n was continuous, and it would satisfy *irrelevant information* if f_n satisfied $f_I(\tau^1) = f_I(\nu^1)$ and $f_{I*J}(\tau) = f_{I*J}(\nu)$ (see the proof of the satisfaction of principle $\mathbf{P}^{\mathcal{I}}_{\beta}$ in Prop. 21).

Another consequence of founding an inference process upon the best candidates is that a probability distribution ω considered as a non-candidate wrt an absorbing candidacy function $0_{\mathbb{C}}$ (ie $\forall \omega, 0_{\mathbb{C}}(\omega) = 0$) can nevertheless be drafted as candidate then be elected. For example, let ω be the probability distribution having the maximum entropy, namely $\omega \overset{\text{def}}{=} [\frac{1}{2^n}; \ldots; \frac{1}{2^n}]$ where n is the number of propositional variables underlying $0_{\mathbb{C}}$; thus ω is drafted by $\hat{\Omega}_{0_{\mathbb{C}}}$ then elected by $\mathcal{I}^E_{\text{ME}}(0_{\mathbb{C}})$.

4.2.2 Internal entropy-based inference process $\mathcal{I}^E_i$

Definition 35. *The internal entropy-based inference process $\mathcal{I}^E_i$ elects the probability distributions with a high entropy while being (nearly) a best candidate for representing the real world, wrt C:*

$$\mathcal{I}^E_i(C) \overset{\text{def}}{=} \arg\max_{\omega \in \Omega} E(\omega) * C(\omega)$$

$\mathcal{I}^E_i$ satisfies *equivalence* wrt $\overset{\text{i}}{\equiv}$. It is *continuous* wrt $\mu^{\text{dis}}_{\mathcal{L}^{\vec{}}_{\infty}}$ if C is continuous and unimodal (which is the case for any candidacy function C_K corresponding to a linear knowledge base K having reliability degree $\sigma \in]0{:}1[$). Therefore, $\mathcal{I}^E_i$ tends to coincide with ME when K is consistent and tends to be reliable, ie $\sigma \to 1$, but $\mathcal{I}^E_i$ does not satisfy *best candidates*. Besides, $\mathcal{I}^E_i$ is rather "internal", whereas $\mathcal{I}^E_{\text{ME}}$ is rather "external" since $\mathcal{I}^E_{\text{ME}}$ satisfies *equivalence* wrt $\overset{\text{e}}{\equiv}$ and *continuity* wrt $\mu^{\text{dis}}_{\mathcal{H}}$.

4.3 Conclusions and perspectives

The question we address in this chapter is *which probability distributions best correspond to the real world, according to a given (possibly inconsistent) knowledge base (seen as a candidacy function) and some common sense?* After having extended already discussed principles in [33, 37, 38], we define a new inference process: the Paraconsistent Maximum Entropy inference process $\mathcal{I}^E_{\text{ME}}$ (see Def. 34). To our knowledge, $\mathcal{I}^E_{\text{ME}}$ is the first to both tolerate inconsistencies and always coincide with the Maximum Entropy inference process ME when applied to a consistent knowledge base. Besides, [1, 35] extended ME to unary predicate languages (see [40] for further details). A perspective is thus to extend $\mathcal{I}^E_{\text{ME}}$ from propositional probabilistic logic to purely unary predicate probabilistic logic.

Chapter 5

Potential applications: persuade the indecisive, reconcile the *schizo*phrenic

In this chapter, we present several potential applications of our measures and inference process. The first section is about autonomy and robustness in decision making for spacecraft. We notably show that consensus decision making might be tractable. In the second section, we consider the merging of knowledge bases in a multiagent context, where each agent shares its knowledge with others, and takes different attitudes towards the other agents' knowledge; this attitude ranges from scepticism to credulity.

5.1 Autonomous and robust decision making aboard spacecrafts

Future space science missions will involve unmanned spacecrafts performing in hazardous environment at far distance from Earth. We therefore theoretically address the problem of making autonomous and robust decisions wrt inconsistent and uncertain information. To achieve such decisions, we suggest to equip a spacecraft with paraconsistent probabilistic reasoning, intended to define *common sense*. We also suggest to program the spacecraft behaviours in a synchronous language, which is utilised to develop, verify, and certify safety-critical embedded systems. By injecting some common sense into decision systems, we hope to make them more trustworthy.

5.1.1 Behaviour-based programming

The success of future space missions will rely on the spacecraft aptitude for making reliable decisions. In this section, we thus propose a methodology for onboard decision making, focused on spacecraft autonomy. This theoretical methodology is twofold. On earth, space engineers specify the deterministic spacecraft behaviours. Aboard, these uploaded behaviours conduct activities according to sensory data and some *common sense*. We depict this methodology in Fig. 5.1.

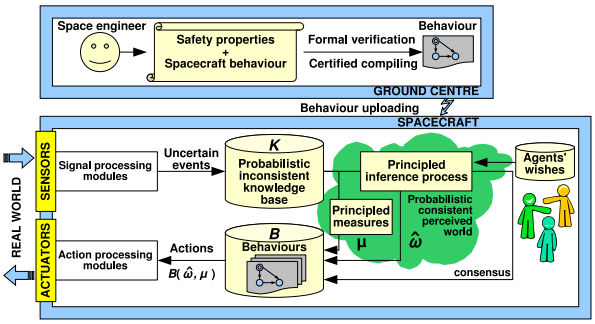


Figure 5.1: Methodology for robust decision making.

After sketching the spacecraft behaviours programming performed by engineers on Earth, we introduce the spacecraft decision process that manages these behaviours aboard. We use $\mathbb{K}^*$ as a knowledge representation formalising the sensory data. We then provide an example of consensus decision making: this problem can only be solved by a paraconsistent probabilistic inference process, like $\mathcal{I}^E_{\mathrm{ME}}$. Finally, we argue for using principled measures and inference process to design autonomous and robust decision making systems.

Spacecraft behaviours design

Firstly, on earth, space engineers specify the deterministic spacecraft behaviours. Behaviours represent tasks to realise wrt the current situation. In the following example, behaviour b_3 executes subbehaviours b_1 and b_2 conditionally to c_1, which depends on the probability that event e_1 occurs in the current situation:
e_1: "*camera-1 detects life on Mars*"
b_1: "*inform ground centre about the probability of* e_1"
b_2: "*focus camera-2 on camera-1's target*"
c_1: "*probability of* e_1 *is higher than 80%*"
b_3: "*if* (c_1) *then (suspend low priority behaviours and*

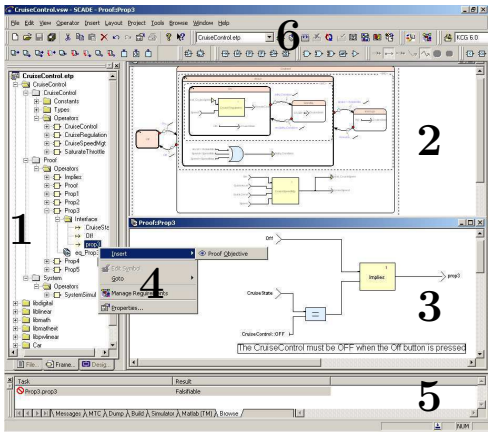


Figure 5.2: SCADE Suite screenshot showing the design, the verification, and the compilation of a behaviour.

execute simultaneously b_1 and b_2; when b_1 and b_2 terminate, resume low priority behaviours)"

Notice that behaviour b_3 is deterministic iff c_1 is either **t**rue or **f**alse, ie iff the probability of e_1 is computable and is a *unique* value (see principle $\mathbf{P}^{\mathcal{I}}_{\alpha}$).

A behaviour, together with its set of safety properties, is written in a synchronous programming language. Such "languages have been designed to allow the unambiguous description of reactive, embedded real-time systems. The common foundation for these languages is the synchrony hypothesis, which considers computations to not take any time. This abstraction allows to separate the concerns functionality and real-time characteristics, and thus facilitates the design of complex embedded systems"[1]. In this methodology, we propose to use SCADE Suite[2], which is an Integrated Development Environment to design, verify, then generate certified[3] code. It provides graphical and textual formal languages, both with data-flow and control-flow synchronous programming styles. These languages comprise instructions to modularly parallelise, sequentialise, suspend, resume, and abort behaviours. The SCADE Suite screenshot in Fig. 5.2 shows **1**) the list of nodes, where a node represents a behaviour or a property, **2**) a behaviour, written in both data-flow (yellow blocks) and control-flow (blue and pink blocks), **3**) a property that a behaviour should satisfy, **4**) the model checking of the property, **5**) the result of the verification, and **6**) the behaviour compilation. The forbidden sign at the bottom left corner indicates that the behaviour does not satisfy the property. In which case, a scenario leading to the violation of the property is generated, helping thus engineers to debug the behaviour. Finally, the behaviour is uploaded aboard the spacecraft into a repository called B.

Behaviours driven by common sense

Once onboard, behaviours B determine the spacecraft decisions, wrt the current situation depicted by sensors. Because of the hazardous spacecraft environment, sensory data are tainted with uncertainty; eg, the processing of the *camera-1* images could lead to uncertain events, eg "*probability of e_1 is lower than 30%*"; such events may be imprecise due to missing or partial sensory data resulting from sensor failure or power loss. Uncertain events are stored into a knowledge base $K \in \mathbb{K}^*$, which tends to be inconsistent due to the multisensor context. Thus, the spacecraft must act wrt an imprecise, possibly inconsistent, probabilistic knowledge base. In chapter 4, we propose a process, called $\mathcal{I}^E_{\mathrm{ME}}$, that infers from K one precise[4] (hence probabilistically consistent) world model $\hat{\omega}$. In addition to $\mathcal{I}^E_{\mathrm{ME}}$, we define in chapter 3 several principled measures μ for knowledge bases. These measures enable engineers to specify behaviours such as "*if ($\mu^{\mathrm{pre}}(K_{\text{camera-1}}) \leq 20\%$) then (execute b_2)*", which commands to the camera-2 to focus on camera-1's target when camera-1 provides the spacecraft with too imprecise data. Thus, the spacecraft actions are computed by evaluating behaviours B wrt $\hat{\omega} \stackrel{\text{def}}{=} \mathcal{I}^E_{\mathrm{ME}}(C_K)$ and the measures: $B(\hat{\omega}, \mu)$. The key for making autonomous and robust decision resides in the *common sense* underlying $\mathcal{I}^E_{\mathrm{ME}}$ and μ.

5.1.2 A treatment for a *schizo*phrenic rover

Voting theory is a theory of electing a societal preference from individual preferences. In the following example, a rover will have to achieve a consensus about resource allocation from the possibly conflicting preferences of its embedded agents; whence we qualify this rover as *schizo*phrenic, ie having *multiple* personalities.

Suppose a rover is scouting a surface for soil sampling. This rover embeds several scientific agents, ie computer programs, that together decide on the amount of each soils to carry back to the main station where further analysis will be performed. During its journey, the rover stows the soils in a storage box having sliding walls (see Fig. 5.3): this allows to adjust the

[1] This description is excerpted from the SYNCHRON'2009 workshop website: `http://www.dagstuhl.de/09481`

[2] SCADE Suite is a trademark of Esterel Technologies SA. All rights reserved. See `http://www.esterel-technologies.com/`

[3] Code generation qualifiable for DO-178B up to Level A, certifiable for IEC 61508 certified at SIL 3 and EN 50128 certified at SIL 3/4.

[4] Throughout section 5.1, we suppose satisfied *uniqueness* (see principle $\mathbf{P}^{\mathcal{I}}_{\alpha}$), ie we suppose the knowledge bases are such that applying an inference process like $\mathcal{I}^E_{\mathrm{ME}}$ on them elects a single probability distribution: $K \in \mathbb{K}^*$ can thus be any linear knowledge base.

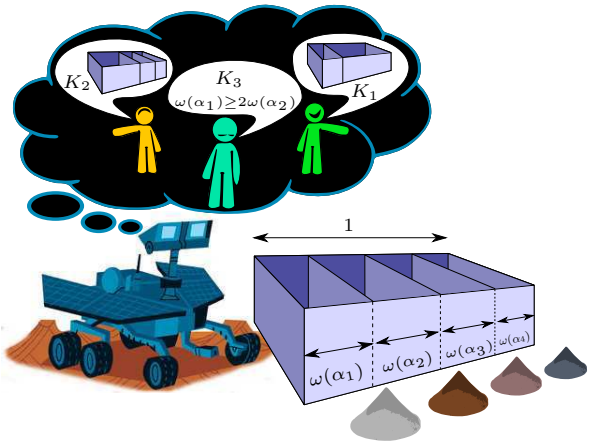


Figure 5.3: A *schizo*phrenic rover and its storage box with sliding walls; the capacity of each compartment is adjusted to the amount of carried soils.

capacity of each compartment to the amount of a collected soil. The agents have different interests, eg, one agent focuses on organic chemistry, whereas another agent focuses on rare soils. Due to the finite capacity of the storage box, these interests may be conflicting, eg, the latter agent may want to carry back a maximum amount of a rare soil, even if this soil is much less inspiring from the organic standpoint than an abundant soil. Notwithstanding the possibly conflicting agents' interests, the rover must achieve a consensus about the capacity of each compartment of the storage box; we formally state this problem as follows.

A rover embedding $I \in \mathbb{N}$ agents stows soil samples in $J \in \mathbb{N}$ compartments $\{\alpha_1, \alpha_2, \ldots, \alpha_J\}$ of a box with sliding walls. A space distribution ω is a function that maps each compartment to its capacity and satisfies these two assumptions: (A1) the box volume is 1 cubic decimetre, ie $1 = \sum_{j=1}^{J} \omega(\alpha_j)$, and (A2) each compartment capacity is positive, ie $\forall j \in \{1, 2, \ldots, J\}, \omega(\alpha_j) \geq 0$.

Each agent i independently expresses a set $K_i \in \mathbb{K}^*$ of wishes for the space distribution ω. For example, i may wish to allocate at least twice more space to soil α_1 than to soil α_2, ie $\omega(\alpha_1) \geq 2{*}\omega(\alpha_2)$, and may wish that the total amount allocated to α_1 and α_2 be within 0.2 and 0.3 cubic decimetre, ie $0.2 \leq \omega(\alpha_1) + \omega(\alpha_2) \leq 0.3$. Besides, the rover affixes to each agent i a reliability level $\sigma_i \in]0{:}1[$, which tends towards 1 as the rover deems i more reliable; eg, i will be labelled as reliable if, in case the rover had fulfilled i's wishes without considering other agents' wishes, its wishes would have enabled a high science return.

Thus, the rover must implement a voting system $\mathcal{I}$ yielding the space distribution $\hat{\omega}$ that *best* conciliates the wishes K_i of each agent i, according to their reliability σ_i and some *common sense*; formally, $\hat{\omega} \stackrel{\text{def}}{=} \mathcal{I}(\cup_{i=1}^{I} K_i^{\sigma_i})$, where $\mathcal{I}$ must satisfy several principles intended to define common sense. By interpreting assumptions (A1) and (A2) as Kolmogorov's axioms for probability, space distributions can be identified with *probability* distributions (see Def. 2 on page 4). We therefore take the *probabilistic* standpoint to define $\mathcal{I}$ as a principled inference process, like $\mathcal{I}^E_{\text{ME}}$ (see Def. 34 on page 39).

5.1.3 Significance of the principles for making autonomous decisions

$\mathcal{I}^E_{\text{ME}}$ satisfies principles $\mathbf{P}^{\mathcal{I}}_{\alpha}$ to $\mathbf{P}^{\mathcal{I}}_{\iota}$ ensuring:

- autonomy, ie decisions are taken without recourse to humans: *uniqueness* (see $\mathbf{P}^{\mathcal{I}}_{\alpha}$) ensures that an evaluation of a condition in a behaviour (see condition c_1 in the example at § 5.1.1 on page 41) always returns either **t**rue or **f**alse;
- determinism, ie decisions are explainable: *determinism* (see $\mathbf{P}^{\mathcal{I}}_{\alpha}$) ensures that $\mathcal{I}^E_{\text{ME}}$ does not use any random function, hence the evaluation of the conditions in the behaviours are deterministic, therefore, the whole behaviour is deterministic;
- robustness, ie decisions are robust against slight fluctuations of sensory data: *continuity* (see $\mathbf{P}^{\mathcal{I}}_{\zeta}$) ensures the continuity of $\mathcal{I}^E_{\text{ME}}$, although a value of a condition in a behaviour can wobble. If this effect is undesirable, an engineer could design more sophisticated behaviours which compute spacecraft actions by applying some continuous functions to $\mathcal{I}^E_{\text{ME}}(C_K)$ (so that the spacecraft actions continuously depend on the sensory data), but in which case, the model checker may not be able to formally verify the behaviour;
- fairness, ie $\mathcal{I}^E_{\text{ME}}$ equally trusts, or fairly conciliates, each uncertain event in K: this property is ensured by the knowledge formalisation, because K is a multiset, and by *proximity* (see $\mathbf{P}^{\subset}_{7}$), which avoids knowledge items to be ignored, even when they are inconsistent or incoherent;
- backwards compatibility, ie the spacecraft decisions are not influenced by the addition of new sensors if these sensors provide data on new topics (hence a decision taken before the spacecraft upgrade is still valid): this is ensured by *irrelevant information* (see $\mathbf{P}^{\mathcal{I}}_{\beta}$).
- semantic analysis, ie decisions depend on the meaning of K, not on the syntax: the knowledge normalisation and the other principles are intended

to make an inference process syntax invariant (eg, see $\mathbf{P}^{\mathcal{I}}_{\gamma}$ and $\mathbf{P}^{\mathcal{I}}_{\delta}$).

In addition to $\mathcal{I}^E_{\mathrm{ME}}$, we propose in chapter 3 several principled measures for K. These measures allow engineers to define behaviours that establish strategies for:

- mission planning, by measuring the *incoherence* between the current situation and the mission target, both described in terms of knowledge bases. A behaviour could be "if the mission target is too incoherent from the current situation depicted by the sensors, then the spacecraft should select a more achievable target".
- tackling unexpected events, ie the known unknowns, by measuring the *dissimilarity* between the current situation and an expected one, both described in terms of knowledge bases;
- self-healing, by measuring the *culpability* of each sensor for making K inconsistent: spacecraft may decide to check then repair such a sensor.
- sensors recalibration, by measuring the *redundancy* of sensory data. For example, an exploring spacecraft may decide to widen its sensor coverage by decreasing the overlap of each sensor coverage, ie by increasing the dissimilarity between sensory data. However, when the spacecraft detects an interesting event, it may decide to focus its sensors on this event by increasing the overlap of each sensor coverage.
- postponing a decision until enough information is gathered, by measuring the *confidence* in such a decision wrt the sensory data.

5.1.4 Computational complexities

In the sequel, we denote by m the number of inequalities in a knowledge base $K \in \mathbb{K}$, and by n the number of propositional variables. The space complexity of our knowledge representation is exponential wrt n. Besides, the time complexity of our inference process $\mathcal{I}^E_{\mathrm{ME}}$ depends on the space complexity. Thus, in order to make $\mathcal{I}^E_{\mathrm{ME}}$ tractable, we are investigating techniques that exponentially reduce the space complexity, like those in [20, 44].

Space and time complexities The naive space complexity of our knowledge representation is $\mathcal{O}(m * 2^n)$. The following partitioning technique exponentially reduces this complexity. A knowledge base can be partitioned into p subbases of inequalities such that each subbase does not share any propositional variable with the other subbases. Notice that the knowledge in a partition is irrelevant to the knowledge in another partition. This partitioning technique is legitimate for any inference process satisfying *irrelevant information* like $\mathcal{I}^E_{\mathrm{ME}}$ (see principle $\mathbf{P}^{\mathcal{I}}_{\beta}$ on page 37). Hence, the space complexity of the partitioned knowledge is only $\sum_{i=1}^{p} \mathcal{O}(m_i * 2^{n_i})$, with $n = \sum_{i=1}^{p} n_i$ and $i = 1, \ldots, p$ where m_i and n_i are respectively the number of inequalities and propositional variables of the i^{th} partition. Due to the partitioning, $\mathcal{I}^E_{\mathrm{ME}}$ applied to a knowledge base computes $p * 2$ optimisations over 2^{n_i} variables within [0:1] instead of two optimisations over 2^n variables. If p is large then $2^{n_i} \ll 2^n$, and $\sum_{i=1}^{p} \mathcal{O}(m_i * 2^{n_i})$ might become a tractable space complexity.

The time complexity of $\mathcal{I}^E_{\mathrm{ME}}(C_K)$ relies on the time complexity for maximising p times the function E over $\hat{\Omega}_{C_{K_i}}$, which is a maximisation of C_{K_i} over 2^{n_i} variables with $i = 1, \ldots, p$, where K_i is a partition of K. We know that C_{K_i} is not only continuous and log-concave but also non-smooth (see the non-smoothness of $C^{h_2}_K$ in Fig. 2.2). Thus, the time complexity of $\hat{\Omega}_{C_{K_i}}$ is the same as maximising a concave non-smooth function over the convex set $[0{:}1]^{2^{n_i}}$ constrained by the linear equality $1 = \sum_{j=1}^{2^{n_i}} \omega_j$ (see [45]).

Bounding and approximating techniques In addition, there exist techniques to smooth out a log-concave function (see [27]) enabling us to not only use faster optimisation algorithms (see [26]), but to also compute a hat function (see [16]) that allows arbitrarily precise approximation of $\hat{\Omega}_{C_{K_i}}$. Furthermore, an easier-to-compute entropy function is proposed in [29], which accelerates each function evaluation during the optimisation process.

Tractable consensus decision making In §5.1.2, we propose to use $\mathcal{I}^E_{\mathrm{ME}}$ for computing a consensus among the agents about the capacity of the J compartments, where $J \in \mathbb{N}$ must be a power of two[5]. In which situation, the space complexity of a knowledge base $K \in \mathbb{K}$ containing m agents' wishes is $\mathcal{O}(m * J)$. Furthermore, if $J = 2^1$ and $K \in \mathbb{K}^{=}$ then we conjecture that $\mathcal{I}^E_{\mathrm{ME}}(C_K)$ simply selects the median space distribution that is the nearest to $[\frac{1}{2}; \frac{1}{2}]$; the time complexity would then be $\mathcal{O}(m * \ln(m))$. Thus, $\mathcal{I}^E_{\mathrm{ME}}(C_K)$ may be tractable.

5.1.5 Conclusions

Paraconsistent probabilistic reasoning is the solution to a certain kind of consensus decision making (see §5.1.2),

[5] If there is only $J' < J = 2^n$ compartments, then it suffices to merge the agents' wishes K with the following knowledge base, of which the reliability level σ equals 1: $\{\, \omega(\alpha_j) = 0 \mid j \in \mathbb{N}, J' < j \leq J \,\}$.

and is a theoretical solution to autonomous and robust decision making (see §5.1.1). The significance of the principles for measures and inference processes (see chapters 3 and 4) is exhibited in §5.1.3. We stress the importance of *uniqueness*, *continuity*, and *irrelevant information* to make autonomous and robust decisions. Furthermore, the satisfaction of the latter principle is necessary to exponentially reduce the computational complexities of our knowledge representations, hence of our measures and inference processes (see §5.1.4).

Tractable approximations of our knowledge representations, our measures, and our inference processes are needed to make viable our methodology for onboard decision making. Nevertheless, the problem of consensus decision making may be tractable due to the quadratic space complexity of the knowledge representation.

5.2 A continuum of mergences

Consider a multiagent context, where each agent owns one knowledge base and has access to other agents' knowledge bases. When reasoning, each agent may adopt different strategies to merge all these bases; eg, a sceptical agent may trust more its own knowledge than others' knowledge.

Suppose each agent trusts each agent's knowledge to a certain degree; these degrees of trust are represented by a trust matrix σ of dimension $n \times n$ where $n \in \mathbb{N}$ is the number of agents. Element $\sigma_{ij} \in]0{:}1[$ denotes the trust of agent i in the knowledge $K_j \in \mathbb{K}$ of agent j; agent i is sceptical of j if σ_{ij} tends to 0, credulous if σ_{ij} tends to 1. Notice that K_j may be inconsistent hence cannot be fully trustworthy ($\sigma_{ij} \neq 0$). Also, we want the agents to be aware of the whole available knowledge without ignoring some agents' knowledge ($\sigma_{ij} \neq 1$). For a given agent i, let $C_i \stackrel{\text{def}}{=} \mathbb{U}_{j=1}^{n} C_{K_j}^{\sigma_{ij}}$ be the mergence of all the agents' knowledge from the viewpoint of i (this is our *continuum of mergences*). Agent i can then take decisions wrt $\mathcal{I}^E_{\mathrm{ME}}(C_i)$.

Chapter 6

Conclusion

> La théorie des probabilités n'est, au fond,
> que le bon sens réduit au calcul.
>
> Pierre-Simon Laplace, in [30, page 275]

Inconsistency is essentially a form of uncertainty, which should not hinder us from reasoning. In this thesis, we thus define the *paraconsistent probabilistic reasoning* as a set of theoretical tools (measures and inference processes satisfying commonsensical principles) designed to tolerate inconsistency while performing on a probabilistic knowledge representation. Our main contributions are summarised as follows.

Chapter 2. Knowledge representations. We introduce a new knowledge representation, named candidacy functions, which resolves the concept of (in)consistency present in probabilistic propositional knowledge bases.

Following the approach in J.B. Paris's book [33], we define a knowledge base as a set of constraints on a probability distribution (see §2.2.2), where such a distribution returns the probability that a proposition is **t**rue. Knowledge bases generalise sets of propositions and conditional probabilistic knowledge bases. Besides, we define a candidacy function (see §2.2.4) as returning the degree to which each probability distribution is candidate for representing the real world. The probability distributions maximising a candidacy function are called the best candidates. We then establish principles (see §2.3.5) guiding us towards the construction of a candidacy function from a knowledge base. Such a candidacy function expresses the degree to which each probability distribution satisfies all the constraints of a knowledge base, even when this knowledge base is inconsistent. Moreover, reliability levels can be given to the constraints.

Having bridged knowledge bases and candidacy functions, we design the following tools only for candidacy functions.

Chapter 3. Measures We propose several new principled formalisations of the following four notions.

Section 3.2. Dissimilarity measure. We endow the set of candidacy functions $\mathbb{C}$ with two principled metrics: the *internal* dissimilarity measure, denoted by $\mu^{\text{dis}}_{\mathcal{L}_\infty}$ and defined as the uniform norm of two candidacy functions, and the *external* one, denoted by $\mu^{\text{dis}}_{\mathcal{H}}$ and defined as the Hausdorff distance between the respective best candidates of two candidacy functions. Our metrics extends those discussed in [33, page 89–91]. Each metric induces a notion of convergence in $\mathbb{C}$, hence a notion of *continuity* for the tools performing on $\mathbb{C}$.

Section 3.3. Inconsistency measure. The reliability of a source of information might be partially determined by how far its information is to be consistent: this distance is given by our inconsistency measure μ^{icst}. We furthermore quantify the *culpability* of each item of information in making the whole inconsistent. μ^{icst} satisfies several principles extending those stated by A. Hunter and S. Konieczny in [18] for sets of propositions.

Section 3.4. Incoherence measure. The reliability of a source of information might be partially determined by how non-consensual its information and our beforehand knowledge are. We formalise this quantity by two different incoherence measures, one rather internal, denoted by μ^{icoh}_{V}, and one external, denoted by $\mu^{\text{icoh}}_{\mathcal{G}}$ and defined as the gap between the respective best candidates of two candidacy functions. We also propose a tentative definition of the *surprise* as the incoherence between a new information and our beforehand knowledge.

Section 3.5. Precision measure. The more numerous the best candidates, the less precise the candidacy function. We show that a naive formalisation of the previous statement, ie a precision measure defined as a kind of volume of the best candidates, leads to two issues: one intrinsic to Lebesgue measure (the Lebesgue measure does not count frontier points), and one intrinsic to probabilities (different probability distributions may have different volumes). We exhibit a workaround for the latter issue, and we consider the first issue as insignificant when only ordering candidacy functions by

precision. We also extend the principles collected by A. Hunter and S. Konieczny in [3, page 222–224]. Besides, we suggest a definition for the *confidence* one may have in making a particular decision wrt a given knowledge.

Chapter 4. Inference processes. An inference process elects the probability distributions that *best* represent the real world, wrt a candidacy function. J.B. Paris and A. Vencovská stated in [34, 36] several principles characterising one inference process operating on knowledge bases. We extend to candidacy functions these principles and the inference process, which elects the best candidates having the maximum entropy.

Chapter 5. Potential applications. We show that paraconsistent probabilistic reasoning is the only solution to a problem from voting theory, where a group of agents has to consensually elect a (probability) distribution (see §5.1.2). Notably, this solution may be tractable due to the quadratic space complexity of the knowledge representation. However, paraconsistent probabilistic reasoning is intractable when applied to scenarios recognition (see §5.1.1): approximate inference processes are still needed.

Bibliography

[1] O.W. Barnett and J.B. Paris. Maximum entropy inference with quantified knowledge. *Logic Journal of the IGPL*, 2008.

[2] T. Bergstrom and M. Bagnoli. Log-concave probability and its applications. *Economic Theory*, 10.1007:445–469, 2005.

[3] Leopoldo Bertossi, Anthony Hunter, and Torsten Schaub. *Inconsistency Tolerance*, volume 3300 of *LNCS*. Springer, 2005.

[4] Denis Bouyssou, Didier Dubois, and Marc Pirlot. *Concepts and Methods of Decision-Making*. ISTE Ltd and John Wiley & Sons Inc, 2009.

[5] Luc Bovens and Stephan Hartmann. An impossibility result for coherence rankings. *Philosophical Studies*, 128(1), 2006.

[6] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[7] Tom DeMarco. *Controlling software projects: management, measurement & estimation*. Yourdon Press, New York, NY, 1982.

[8] René Descartes. *Discours de la méthode*. Project Gutenberg, ∼1600.

[9] Igor Doven and Wouter Meijs. Measuring coherence. *Synthese*, 156:405–425, 2007.

[10] Didier Dubois and Eyke Hüllermeier. Comparing probability measures using possibility theory: A notion of relative peakedness. *International Journal of Approximate Reasoning*, 45:364–385, July 2007.

[11] Faramarz Faghihi. Generalization of the dissimilarity measure of fuzzy sets. *International Mathematical Forum, 2, 2007, no. 68, 3395 - 3400*, 2007.

[12] David H. Glass. Coherence measures and their relation to fuzzy similarity and inconsistency in knowledge bases. *Artif. Intell. Rev.*, 26(3):227–249, 2006.

[13] Rolf Haenni. Towards a unifying theory of logical and probabilistic reasoning. In *ISIPTA*, pages 193–202, 2005.

[14] Rolf Haenni. Probabilistic argumentation. *Journal of Applied Logic*, 2009.

[15] Thomas Hobbes. *Leviathan*. The Harvard Classics, 1588–1679.

[16] W. Hormann and J. Leydold. Automatic random variate generation for simulation input. In *The 2000 Winter Simulation Conference*, pages 675–682. IEEE Press, 2000.

[17] A. Hunter and S. Konieczny. Shapley inconsistency values. In *KR'06*, pages 249–259, 2006.

[18] A. Hunter and S. Konieczny. Measuring inconsistency through minimal inconsistent sets. In *KR'08*, pages 358–366, 2008.

[19] L. Itti and P. F. Baldi. Bayesian surprise attracts human attention. In *Advances in Neural Information Processing Systems, Vol. 19 (NIPS*2005)*, pages 547–554, Cambridge, MA, 2006. MIT Press.

[20] Gabriele Kern-Isberner and Thomas Lukasiewicz. Combining probabilistic logic programming with the power of maximum entropy. *Artif. Intell.*, 157(1-2):139–202, 2004.

[21] E.P. Klement, R. Mesiar, and E. Pap. *Triangular norms*. Springer Netherlands, 2000.

[22] G.J. Klir and B. Yuan. *Fuzzy sets and fuzzy logic: theory and applications*. Prentice Hall Upper Saddle River, NJ, 1995.

[23] Kevin M. Knight. Two information measures for inconsistent sets. *J. of Logic, Lang. and Inf.*, 12(2):227–248, 2003.

[24] S. Konieczny, J. Lang, and P. Marquis. Quantifying information and contradiction in propositional logic through epistemic tests. *IJCAI'03*, pages 106–111, 2003.

[25] Valéry Larbaud. *A. O. Barnabooth – Son journal intime*. Gallimard, Paris, 1922.

[26] László Lovász and Santosh Vempala. Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. In *FOCS*. IEEE Press, 2006.

[27] László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Struct. Algorithms*, 30(3):307–358, 2007.

[28] E Lozinskii. Information and evidence in logic systems. *J. of Experimental and Theoretical Artificial Intelligence*, 6:163–193, 1994.

[29] Cheng Ma, Chao Liu, Shaoxian Ma, and Chengshun Jiang. The application of a new entropy function and mutative scale chaos optimization strategy in two-dimensional entropic image segmentation. In *Computational Intelligence and Security*, volume 2, pages 1647–1652. IEEE Press, Nov. 2006.

[30] Pierre-Simon (marquis de) Laplace. *Essai philosophique sur les probabilités*. Bachelier, 1825. Cinquième édition.

[31] Enrique Miranda. A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 48(2):628–658, 2008.

[32] H.L. Ong, H.C. Huang, and W.M. Huin. Finding the exact volume of a polyhedron. *Advances in Engineering Software*, 34:351–356(6), 2003.

[33] J.B. Paris. *The Uncertain Reasoner's Companion: A Mathematical Perspective*, volume 39 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 1994.

[34] J.B. Paris. Common sense and maximum entropy. *Synthese*, 117:75–93, 2000.

[35] J.B. Paris and S.R. Rad. Inference processes for quantified predicate knowledge. In *WoLLIC '08: Proceedings of the 15th international workshop on Logic, Language, Information and Computation*, pages 249–259, Berlin, Heidelberg, 2008. Springer-Verlag.

[36] J.B. Paris and A. Vencovská. A note on the inevitability of maximum entropy. *Int. J. Approx. Reasoning*, 4(3):183–223, 1990.

[37] J.B. Paris and A. Vencovská. In defence of the maximum entropy inference process. *International Journal of Approximate Reasoning*, 17:17–103, 1997.

[38] J.B. Paris and A. Vencovská. Common sense and stochastic independence. In Jon Williamson David Corfield, editor, *Fondation of Bayesianism*, number 24 in Applied Logic Series, chapter Logic, Mathematics and Bayesianism, pages 203–240. Kluwer, 2001.

[39] David Picado Muiño. *Deriving Information from Inconsistent Knowledge Bases: A Probabilistic Approach*. PhD thesis, The University of Manchester, 2008.

[40] Soroush Rafiee Rad. *Inference Processes for Probabilistic First Order Languages*. PhD thesis, The University of Manchester, 2009.

[41] Philippe Smets. Data fusion in the transferable belief model. In *Information Fusion, 2000. FUSION 2000. Proceedings of the Third International Conference on*, volume 1, pages PS21–PS33 vol.1, July 2000.

[42] Matthias Thimm. Measuring inconsistency in probabilistic knowledge bases. In Jeff Bilmes and Andrew Ng, editors, *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI'09)*, Montreal, Canada, June 2009.

[43] Peter Walley. Coherent upper and lower previsions. The Imprecise Probabilities Project, 1998.

[44] Jon Williamson. Objective bayesian nets. In *We Will Show Them! Essays in Honour of Dov Gabbay*, pages 713–730. College Publications, 2005.

[45] Jin Yu, S. V. N. Vishwanathan, Simon Günter, and Nicol N. Schraudolph. A quasi-newton approach to non-smooth convex optimization. In *ICML*, pages 1216–1223, 2008.

[46] Anbu Yue, Weiru Liu, and Anthony Hunter. Measuring the ignorance and degree of satisfaction for answering queries in imprecise probabilistic logic programs. In *SUM*, pages 386–400, 2008.

Index

Symbols

C

E

F

Définition d'une logique probabiliste tolérante à l'inconsistance

– appliquée à la reconnaissance de scénarios & à la théorie du vote –

Résumé : Les humains raisonnent souvent en présence d'informations contradictoires. Dans cette thèse, j'ébauche une axiomatisation du *sens commun* sous-jacent à ce raisonnement dit paraconsistant. L'implémentation de cette axiomatisation dans les ordinateurs autonomes sera essentielle si nous envisageons de leur déléguer des décisions critiques ; il faudra également vérifier formellement que leurs réactions soient sans risque en toute situation, même incertaine.

Une situation incertaine est ici modélisée par une base de connaissances probabilistes éventuellement inconsistante ; c'est un multi-ensemble de contraintes éventuellement insatisfiable sur une distribution de probabilité de phrases d'un langage propositionnel, où un niveau de confiance peut être attribué à chaque contrainte. Le principal problème abordé est l'inférence de la distribution de probabilité qui représente au mieux le monde réel, d'après une base de connaissances donnée. Les réactions de l'ordinateur, préalablement programmées puis vérifiées, seront déterminées par cette distribution, modèle probabiliste du monde réel.

J.B. Paris *et al* ont énoncé un ensemble de sept principes, dit de sens commun, qui caractérise l'inférence dans les bases de connaissances probabilistes consistantes. Poursuivant leurs travaux de définition du sens commun, je suggère l'adhésion à de nouveaux principes régissant le raisonnement dans les bases inconsistantes.

Ainsi, je définis les premiers outils théoriques fondés sur des principes pour raisonner de manière probabiliste en tolérant l'inconsistance. Cet ensemble d'outils comprend non seulement des mesures de dissimilarité, d'inconsistance, d'incohérence et de précision, mais aussi un processus d'inférence coïncidant avec celui de J.B. Paris dans le cas consistant. Ce processus d'inférence résout un problème de la théorie du vote, c'est-à-dire l'obtention d'un consensus parmi des opinions contradictoires à propos d'une distribution de probabilité telle que la répartition d'un investissement financier.

Finalement, l'inconsistance n'est qu'une forme d'incertitude qui ne doit pas entraver notre raisonnement, ni celui des ordinateurs : peut-être qu'une plus grande confiance leur sera accordée s'ils fondent leurs décisions sur notre sens commun.

Mots clés : logique, probabilité, inconsistance, base de connaissances, raisonnement, mesure

Paraconsistent probabilistic reasoning

– applied to scenario recognition & voting theory –

Abstract: If we envisage delegating critical decisions to an autonomous computer, we should not only endow it with *common sense*, but also formally verify that such a machine is programmed to *safely* react in every situation, notably when the situation is depicted with uncertainty.

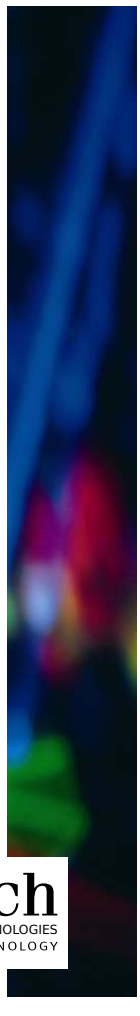
In this thesis, I deem an uncertain situation to be a possibly inconsistent probabilistic propositional knowledge base, which is a possibly unsatisfiable multiset of constraints on a probability distribution over a propositional language, where each constraint can be given a reliability level. The main problem is to infer one probabilistic distribution that best represents the real world, with respect to a given knowledge base. The reactions of the computer, previously programmed then verified, will be determined by that distribution, which is the probabilistic model of the real world.

J.B. Paris *et al* stated a set of seven commonsensical principles that characterises the inference from consistent knowledge bases. Following their approach, I suggest adhering to further principles intended to define common sense when reasoning from an inconsistent knowledge base.

My contribution is thus the first principled framework of paraconsistent probabilistic reasoning that comprises not only an inference process, which coincides with J.B. Paris's one when dealing with consistent knowledge bases, but also several measures of dissimilarity, inconsistency, incoherence, and precision. Besides, I show that such an inference process is a solution to a problem originating from voting theory, namely reaching a consensus among conflicting opinions about a probability distribution; such a distribution can also represent a distribution of a financial investment.

To conclude, this study enhances our understanding of common sense when dealing with inconsistencies; injecting common sense into decision systems should make them more trustworthy.

Keywords: logic, probability, inconsistency tolerance, knowledge base, reasonning, measure



MINES ParisTech

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY